Wen Siang Lew
Gerard Joseph Lim
Putu Andhita Dananjaya *Editors*

Emerging Non-volatile Memory Technologies

Physics, Engineering, and Applications

Emerging Non-volatile Memory Technologies

Wen Siang Lew · Gerard Joseph Lim ·
Putu Andhita Dananjaya
Editors

Emerging Non-volatile Memory Technologies

Physics, Engineering, and Applications

Editors
Wen Siang Lew
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore, Singapore

Gerard Joseph Lim
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore, Singapore

Putu Andhita Dananjaya
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore, Singapore

ISBN 978-981-15-6910-4   ISBN 978-981-15-6912-8 (eBook)
https://doi.org/10.1007/978-981-15-6912-8

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021, corrected publication 2021
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

A common driving force in the development of emerging memory technologies is the urgency to overcome challenges faced by transistor-based memory and logic. Semiconductor devices have to keep pace with the increasing demands for even faster speeds, smaller footprint, and greater efficiency, while reducing the total energy consumption. For integrated circuits to keep up with increasing transistor density over the next few decades, novel approaches such as new transistor architectures, 3D stacking, advanced lithography techniques, and new materials are imperative. While these concerns are being tackled, it is also worth looking into emerging, beyond-CMOS technologies. Even the International Roadmap for Devices and Systems (IRDS) sees beyond-CMOS as a milestone in the evolution of computing. Two very promising emerging memory technologies currently in high industry demand are spintronic-based devices such as magnetic random-access memory (MRAM), and resistive switching devices also known as resistive random-access memory (RRAM).

In spintronic-based devices, the spin property of electrons is exploited alongside its conventional charge property. The electronic spin that manifests as magnetization can be used as memory elements for non-volatile information storage, such as in commercially available MRAM, using magnetic tunnel junctions (MTJ). On the other hand, RRAM devices utilize the intrinsic defects modulation of metal–insulator–metal structures under an external electric field. The change in the structural defects configuration results in significant alteration in conductance. RRAM devices have been implemented as both memory unit and computing element due to their unique device dynamics and highly scalable architecture.

The dynamic response of spintronic and resistive switching devices continue to be thoroughly investigated. With advanced material and device engineering, the characteristics of these devices can be rigorously tuned to achieve functional dynamics such as stochastic and deterministic behaviours. This aspect of the device has an especially crucial role in the development of non-conventional computing and neuromorphic engineering applications.

This book is a timely resource that discusses the physics and applications of spintronic and RRAM devices. The first part of the book mainly discusses spintronic devices such as nano-oscillators, domain wall memory, nano-magnetic and programmable logic, circuit design, as well as their applications in neuromorphic engineering, while the second part of the book focuses on RRAM device physics and array architectures, various device characterizations and modelling, its implementation as a synaptic device, and specific application on automatic sound classification. The content of the book covers the interdisciplinary area of device physics, material science and engineering, circuit design, and neuromorphic engineering. This book provides fundamental insight into emerging memory technologies, as well as comprehensive research findings, challenges, and prospects in these areas. Thus, it can be used as a reference for scientists and engineers of various backgrounds, who wish to gain understanding from the basics to the more advanced aspect of the corresponding fields.

Singapore

Wen Siang Lew
Gerard Joseph Lim
Putu Andhita Dananjaya

Contents

Spintronic-based Devices

Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions 3
P. K. Muduli, Raghav Sharma, Dhananjay Tiwari, Naveen Sisodia, Afshin Houshang, O. G. Heinonen, and Johan Åkerman

Spin Transfer Torque Magnetoresistive Random Access Memory 45
Wai Cheung Law and Shawn De Wei Wong

Current-Driven Domain Wall Dynamics in Magnetic Heterostructures for Memory Applications 103
Sachin Krishnia and Wen Siang Lew

Electric-Field-Controlled MRAM: Physics and Applications 149
James Lourembam and Jiancheng Huang

Chiral Magnetic Domain Wall and Skyrmion Memory Devices 175
Kyujoon Lee, Dong-Soo Han, and Mathias Kläui

Circuit Design for Non-volatile Magnetic Memory 203
Tony Tae-Hyoung Kim

Domain Wall Programmable Magnetic Logic 225
Sarjoosing Goolaup, Chandrasekhar Murapaka, and Wen Siang Lew

3D Nanomagnetic Logic 259
Markus Becherer

Spintronics for Neuromorphic Engineering 297
Gerard Joseph Lim, Calvin Ching Ian Ang, and Wen Siang Lew

Resistive Random-Access Memory

Resistive Random Access Memory Device Physics and Array Architectures 319
Victor Yiqian Zhuo, Zhixian Chen, and King Jien Chui

RRAM Device Characterizations and Modelling 345
Peng Huang, Bin Gao, and Jinfeng Kang

RRAM-Based Neuromorphic Computing Systems 383
Putu Andhita Dananjaya, Roshan Gopalakrishnan, and Wen Siang Lew

An Automatic Sound Classification Framework with Non-volatile Memory 415
Jibin Wu, Yansong Chua, Malu Zhang, Haizhou Li, and Kay Chen Tan

Correction to: Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions C1
P. K. Muduli, Raghav Sharma, Dhananjay Tiwari, Naveen Sisodia, Afshin Houshang, O. G. Heinonen, and Johan Åkerman

Spintronic-based Devices

Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions

P. K. Muduli, Raghav Sharma, Dhananjay Tiwari, Naveen Sisodia, Afshin Houshang, O. G. Heinonen, and Johan Åkerman

Abstract A nanopillar of a magnetic tunnel junction (MTJ) consisting of two magnetic layers (free and fixed) separated by an oxide barrier can be used as a microwave generator, and also as a microwave detector, by using the concept of spin-transfer torque. When the MTJ nanopillar is used as a microwave generator, the device is known as a spin-torque nano-oscillator (STNO) and offers advantages of high output power over metallic-based STNOs. We first discuss the bias dependence of spin torque and, in particular, the field-like torque present in these devices. We then discuss the mechanism governing linewidth broadening and show how multimode excitation and mode-hopping can affect the phase noise in these STNOs. The potential of these oscillators for communication applications is discussed, with the focus on the synchronization and modulation of the MTJ-STNOs with an external signal. Finally the application of these devices as microwave detectors is discussed, and we concentrate on approaches for enhancing their microwave sensitivity.

The original version of this chapter was revised: This chapter was revised with the missed out corrections. The correction to this chapter is available at https://doi.org/10.1007/978-981-15-6912-8_14

P. K. Muduli (✉) · R. Sharma · D. Tiwari · N. Sisodia
Department of Physics, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India
e-mail: muduli@physics.iitd.ac.in

O. G. Heinonen
Materials Science Division, Argonne National Laboratory, Lemont, IL 60439, USA
e-mail: heinonen@anl.gov

A. Houshang · J. Åkerman
Department of Physics, University of Gothenburg, Gothenburg 412 96, Sweden
e-mail: johan.akerman@physics.gu.se

NanOsc AB, Electrum 205, Kista 164 40, Sweden

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021, corrected publication 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_1

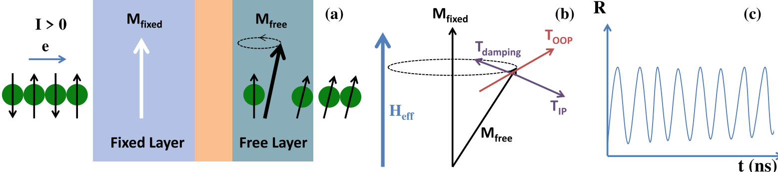


Fig. 1 **a** Schematic of the operation of a spin-torque nano-oscillator in a system consisting of a ferromagnet/spacer/ferromagnet trilayer. **b** Schematic presentation of damping torque, in-plane spin torque, T_{IP}, and out-of-plane torque, T_{OPP} on the local magnetization. **c** Schematic presentation of resistance oscillations arising due to the spin-torque-induced magnetization precession, as a result of the GMR or TMR effect

1 Introduction

The theoretical prediction of the spin-transfer torque (STT) effect in the year 1996 constituted a remarkable breakthrough in the development of spintronics. This prediction made independently by Slonczewski [1] and Berger [2] after the 1988 discovery of giant magnetoresistance (GMR) [3, 4] and the first report of high tunneling magnetoresistance (TMR) in room-temperature magnetic tunnel junctions (MTJ) in 1995 [5, 6]. STT made possible new methods of manipulating magnetization dynamics using spin-polarized currents instead of magnetic fields, opening several opportunities for the development of novel spintronics devices. The two most important types of STT devices are (i) magnetic memories, such as magnetoresistive random-access memories (MRAM) [7, 8], and (ii) nanoscale microwave signal generators/detectors. In this chapter, we will focus on the second application of STT. The reader may also refer to reviews of STT-based MRAM and STNOs in Ref. [9–15].

The basic principle of operation of an STNO is illustrated in Fig. 1. When electrons pass through the thick fixed magnetic layer, their spins become aligned with the magnetization of this layer. When these spin-polarized electrons enter the free layer, after passing through the non-magnetic/insulating barrier, they exert a torque on the local magnetization of the free layer if current is sufficiently high, i.e. more than some threshold value. It can lead to magnetization switching if the torque is sufficiently large. Smaller torque values result in magnetization precession around the effective magnetic field. In the general case, an STNO consists of a nanopillar with a giant magnetoresistance (GMR) or a magnetic tunnel junction (MTJ) structure. The oscillation of the magnetization is detected as a microwave or radio frequency (RF) voltage signal by virtue of either the GMR or TMR effect (Fig. 1c).

STNOs are particularly useful in wireless telecommunications, where they offer several advantages over commercially used oscillators, such as yttrium iron garnet (YIG) tuned oscillators (YTOs), dielectric resonator oscillators (DROs) based on YIG, and semiconductor-based voltage-controlled oscillators (VCOs). These advantages include a very wide frequency tuning range [16–19], high modulation rates [20–26], submicron footprints [27], and compatibility with standard complementary metal oxide semiconductor processes [7, 8]. However, low output power and

high linewidth are the main limitations for commercialization of STNOs. MTJ-based STNOs offer significantly greater output power than metallic GMR-based STNOs, and have recently reached 10 μW [28], in vortex based MTJ oscillators, which meets the requirements of commercial applications. However, the frequency of operation of such vortex STNOs is around 300 MHz, and hence they do not take full advantage of the wide-frequency tunability of STNOs, which can usually be tuned in the 1–65 GHz range [17]. STNOs have the capability to go beyond 200 GHz [29] with conventional materials in high fields, and can even reach the terahertz range [30] with the use of antiferromagnets.

MTJ nanopillars can also be used to rectify radio frequency (RF) signals through spin-transfer torque [31–36]. With an RF current injected into an MTJ nanopillar, a dc voltage develops over the device when the frequency of the RF current is close to the eigenfrequency of the nanomagnet. Thus, an MTJ nanopillar can act as a microwave detector, also known as spin-torque diode (STD). For applications, a high sensitivity—defined as the ratio of the voltage detected to the input RF power—is required. Preliminary calculations predicted that optimized MTJs should reach sensitivities exceeding 10,000 mV/mW [37]. With advances in the design of the MTJs in terms of tunneling magnetoresistance (TMR) effect and voltage-controlled magnetocrystalline anisotropy (VCMA) and with improvements in the understanding of strong nonlinear effects such as stochastic resonance [38] and non-linear resonance [35, 36], the sensitivity performance of STD has reached values as high as 10^5 mV/mW, exceeding that of semiconductor Schottky diode detectors at room temperature. These results open a path for the design of a new generation of high-sensitivity microwave detectors for wireless power transfer, which can rectify microwave power from different energy sources, such as satellite, sound, television, and Wi-Fi signals. The resulting dc voltage could be used to operate low-powered nanodevices [39].

2 Linear Spin Wave Modes in MTJ-Based Nanopillars

Spin waves can be excited by thermal excitations in MTJ nanopillars in the presence of an external field. A frequency domain spectrum of the time-varying magnetization in the MTJ will reveal that magnetic moments can oscillate at several different frequencies. Each of these coexisting frequencies is called a frequency eigenmode. The study of these modes gives an insight into the coherence of spin motion inside the ferromagnet. Because the excited spin waves are contained mostly in the ferromagnetic material and die out fast outside it, different spin wave modes are excited only from the magnetic part of the nanopillars—i.e., the free layer and the fixed layer. Moreover, as the fixed layer is pinned strongly in one particular direction due to exchange bias, a very high field is usually required to excite spin waves in it. Figure 2 shows the field variation of the excited spin wave modes in an MTJ nanopillar as reported in Ref. [40]. The MTJ investigated in [40] had a circular cross-section with a diameter 240 nm consisting of multilayers of IrMn (5)/CoFe (2.1)/Ru (0.81)/CoFe

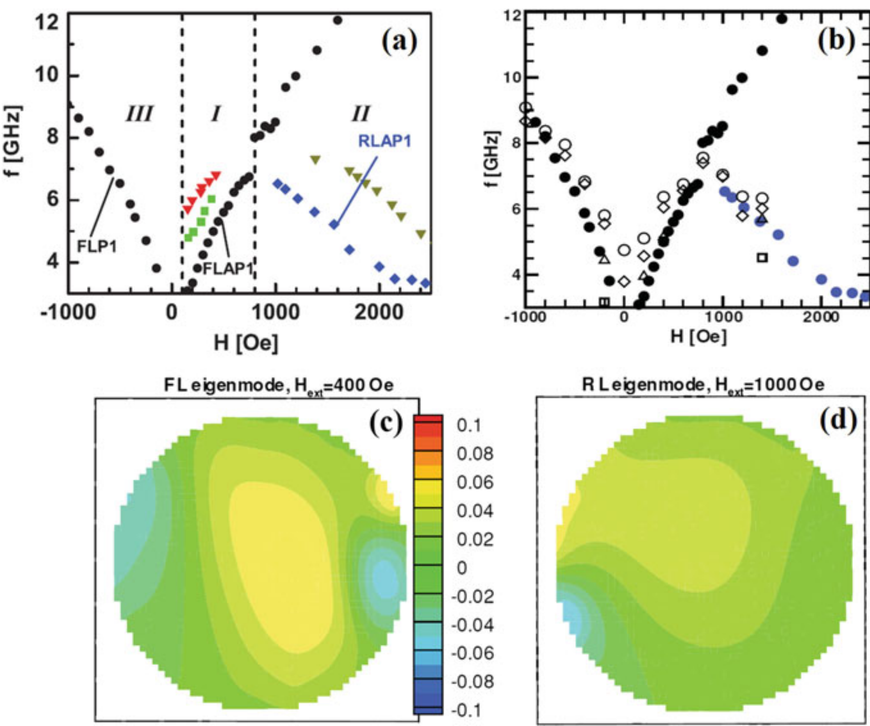


Fig. 2 Field-excited spin wave modes in an MTJ nanopillar. **a** Three different modes originating from free and fixed layers at different applied external fields. Depending on the direction of the applied field, the free and fixed layers can be aligned parallel (P state) or antiparallel (AP state) to each other. Here, FLP1 and FLAP1 represent the free layer modes in the P and AP configurations, respectively. RLAP1 represents the reference layer mode in the P state. **b** shows similar modes obtained using simulations (open squares, triangles, diamonds, and circles). The crossover from increasing frequency modes ($<$800 Oe) to decreasing frequency modes ($>$800 Oe) is qualitatively and quantitatively reproduced in the simulations. **c** and **d** show the eigenmode distribution (the amplitude of the magnetization oscillations) of the lowest frequency mode at applied fields of 400 Oe and 1000 Oe. A dominant bulk-like mode in FL (RL) can be seen at 400 Oe (1000 Oe). Reprinted with permission from Muduli et al. [40] copyright (2011) by the American Physical society

(1)/CoFeB (1.5)/MgO(1)/CoFeB (3.5) (thicknesses in nm). The bottom CoFeB layer is the pinned layer (PL), the composite CoFe/CoFeB represents the reference Layer (RL), and the top CoFeB layer is the free layer (FL). All the layers are magnetized with the magnetization in-plane. The antiferromagnetic IrMn layer provides exchange bias on the CoFe pinned layer (PL), which tends to keep its magnetization direction in a fixed direction. The PL is strongly coupled antiferromagnetically through 0.81 nm thick Ru layer to the composite CoFe/CoFeB fixed layer, or reference layer (RL). This CoFe/Ru/CoFe/CoFeB structure is referred to as a synthetic antiferromagnet (SAF), as its net magnetic moment is close to zero. In Fig. 2, we can distinguish three regions in which three different behaviors of the eigenmodes are

observed. The first region (I) corresponds to the positive field, $100 < H_{\rm ext} \leq 800$ Oe; here the FL and RL magnetizations are in an antiparallel (AP) configuration. In this region, the frequencies of the modes increase with an external field. These modes correspond to free layer excitations, labeled FLAP. The second region (II) corresponds to higher positive fields, $H_{\rm ext} \geq 800$ Oe, where additional lower-frequency modes appear with frequencies that decrease with $H_{\rm ext}$. These new modes are RL modes. Finally, the third region (III) corresponds to the case of $H_{\rm ext} < 100$ Oe, with the magnetization of the FL and RL in parallel (P) configuration, which is also an FL mode, labeled FLP1.

By resolving these modes spatially, the region of the MTJ nanopillar that is excited during oscillations at any particular mode can be extracted. It is observed that broadly two kinds of spatial modes can be excited: (i) bulk modes and (ii) edge modes. With bulk modes, the magnetization in the bulk (central region) of the ferromagnetic material moves in phase with each other and most of the output power (in the frequency spectra) comes from this central region. The internal field is mostly homogeneous in this region of the MTJ. In edge modes, the amplitude of magnetization oscillations is significant mostly only around the edges, where the internal field is not homogeneous. Since spatial resolution of the frequencies is experimentally difficult, micromagnetic simulations utilizing the Landau–Lifshitz–Gilbert–Slonczewski (LLGS) equation have been increasingly used for this purpose. The micromagnetic simulation results shown in Fig. 2b reproduces the experimental spin wave modes in all the three regions. In Fig. 2c, d, spatially resolved plots of bulk spin wave modes from free (at 400 Oe) and fixed (at 1000 Oe) layers are shown using micromagnetic simulations. Several factors, such as the lateral size of MTJ and the current-induced Oersted field, have also been reported to influence spin wave generation [41–43].

3 Bias Dependence of In-Plane and Field-Like Torques

Zhang et al. [44] have pointed out that spin torque not only has the in-plane component predicted by Slonczewski [1] and Berger [2], but also possesses a perpendicular component. The in-plane torque (T_{IP}) is also referred to simply as the spin-transfer torque (STT), $\tau_{\parallel}$, while the perpendicular or out-of-plane torque (T_{OOP}) is also referred to as the field-like torque (FLT), $\tau_{\perp}$. The in-plane torque $\tau_{\parallel}$ lies in the plane defined by the magnetization of the fixed layer M_{fixed} and the magnetization of the free layer M_{free}, while $\tau_{\perp}$ points out of the plane defined by M_{fixed} and M_{free}, as shown in Fig. 1. The FLT $\tau_{\perp}$ is small in metallic systems such as spin valves [45–47], but much larger ($\sim 40\%$ of $\tau_{\parallel}$) in MTJs [48], where it significantly affects the magnetization dynamics. In order to control the magnetization using spin-torque effects, the FLT $\tau_{\perp}$ needs to be known and understood. However, the magnitude of this component is both difficult to calculate [49–51] and to measure.

The magnetization dynamics of the free layer, including both these torques, can be described by the Landau–Lifshitz–Gilbert–Slonczewski (LLGS) equation:

$$\frac{d\hat{m}}{dt} = -\gamma[\hat{m} \times \hat{H}_{eff}] + \alpha\left[\hat{m} \times \frac{d\hat{m}}{dt}\right] + \frac{\gamma}{\mu_0 M_S}[a_j(\hat{m} \times (\hat{m} \times \hat{p})) + b_j(\hat{m} \times \hat{p})] \tag{1}$$

where $\hat{m}$ is the normalized magnetic moment ($|\hat{m}| = M_{free}/M_{sat}$), γ is the gyromagnetic ratio of the electron, M_{sat} is the saturation magnetization, $\hat{H}_{eff}$ is the effective external field, and $\hat{p}$ is the direction of the fixed layer with ϕ as the angle between free and fixed layer. The coefficients a_j and b_j represent the strengths of $\tau_{\|}$ and $\tau_{\perp}$ respectively. When current is injected into the free layer, depending on the current direction, $\tau_{\|}$ will either support the damping or act opposite as an anti-damping source. However, $\tau_{\perp}$ can simulate the action of the field on M_{free}, which can also modify the energy variations in the free layer magnetization.

Measuring the spin-torque magnitude and direction in MTJ-STNOs, especially as a function of bias voltage, is crucial for both fundamental understanding and for applications. However, this is not a trivial task, since these measurements are often subject to experimental artifacts. Despite significant progress, there is controversy about the bias dependence of $\tau_{\|}$ and $\tau_{\perp}$ in MTJ-STNOs, with results so far disagreeing both qualitatively and quantitatively [32, 33, 52–55]. Theoretical works are also far from conclusive, as both quadratic [49, 51] and linear dependence [56] of the $\tau_{\perp}$ on bias voltage are predicted.

Sankey et al. [32] measured the perpendicular spin-torque component by fitting the line shape of spin-torque ferromagnetic resonance (ST-FMR) to elliptical MgO MTJs. ST-FMR is a common technique for measuring spin torque by applying an RF current to the MTJs. They showed that the perpendicular spin torque is quadratic in applied voltage, in good agreement with the first-principle calculations by Heileger and Stiles [51]. Similar measurements have also been carried out on MgO MTJs

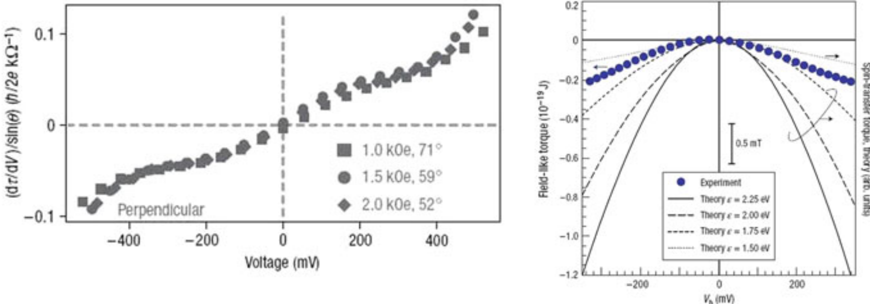


Fig. 3 Measured bias dependence of perpendicular torque showing quadratic behavior. **a** Magnitudes of perpendicular torquance ($d\tau_{\perp}/dV$) determined from room temperature ST-FMR signals for three different values of applied magnetic field. Adapted by permission from Macmillan Publishers Ltd: (Nature Physics) Sankey et al. [32], copyright (2008). **b** The bias dependence of perpendicular torque showing experimental and calculated results. Adapted by permission from Macmillan Publishers Ltd: (Nature Physics) Kubota et al. [33], copyright (2008)

by Kubota et al. [57]. The results of Sankey et al. [32] and Kubota et al. [57] are summarized in Fig. 3.

On the other hand, Petit et al. [52] measured the change in resonance frequency as a function of bias current in circular Al_2O_3-based MTJs and concluded from the observed changes that the $\tau_\perp$ is linearly proportional to the biasing current density, and thus changes sign with bias voltage. This result was in qualitative agreement with the measurement of Li et al. [50], who extracted the perpendicular torque by measuring the switching current of MgO MTJs, and by carefully accounting for heating effects due to the rather large current densities required for switching. However, since the measurements by Li et al. [50] were based on switching currents, the voltage range used was necessarily higher, approximately 0.5 V–1.0 V, than those of Sankey et al. [32] and Kubota et al. [57]. Measurements of the perpendicular spin-torque effect in asymmetric MgO MTJs (where the fixed and free layers were not identical) were carried out by Oh et al. [54]. The perpendicular spin torque $\tau_\perp$ was obtained by fitting plots of the FL magnetic configuration relative to that of the fixed layer as function of applied external field and bias voltage, and showed that there is a linear term in the dependence of perpendicular spin-torque effective field on bias voltage, in addition to a quadratic one, in agreement with theoretical predictions [58].

In Fig. 4, we summarize two recent experiments [40, 55] in which $\tau_\perp$ was obtained from the measurement of resonance frequency. The resonance frequency depends on $\tau_\perp$ as follows [52]:

$$\frac{d\omega_r}{dj} \simeq -2\alpha a_j/j + b_j/j \tag{2}$$

Since $\alpha \ll 1$, the FLT term b_j has a pronounced effect on the resonance frequency and shows a strong linear dependence on the bias, as shown in Fig. 4a, b. Both these studies show that the perpendicular torque varies linearly with bias voltage (for a

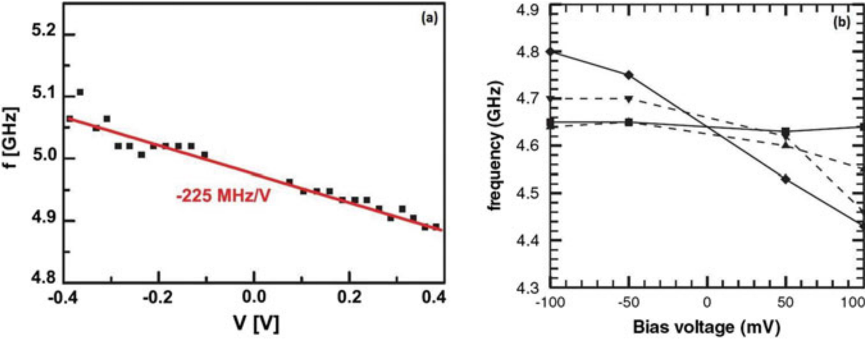


Fig. 4 Measured resonance frequencies versus bias voltage in MgO-based MTJs from **a** Reprinted with permission from Muduli et al.. [40] copyright (2011) by the American Physical society and **b** Reprinted with permission from Heinonen et al. [55] Copyright (2010) by the American Physical society. Both behaviors can be explained by a linear dependence of the perpendicular spin torque on bias voltage

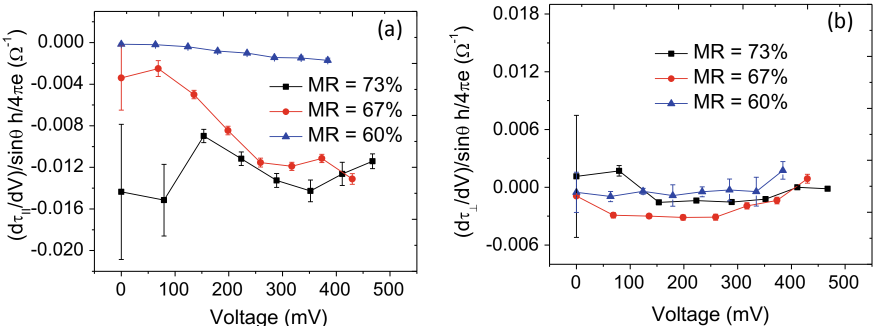


Fig. 5 Bias dependence of the (**a**) IP torkance, and (**b**) OOP torkance determined from the room temperature FM-STFMR signals, for the three devices having different MR ratios. Reprinted from Tiwari et al. [59], with the permission of AIP Publishing

certain range of bias voltages) even for nominally symmetric MTJs, and that any quadratic term is negligible.

In a recent work by Tiwari et al. [59], bias dependence of spin torques in MgO-based magnetic tunnel junction using a field-modulated spin torque ferromagnetic resonance measurement technique is studied for three devices with tunneling magnetoresistance (MR) of 60, 67, and 73%. The devices with lower MR ratio showed the presence of multiple modes, which can be ascribed to asymmetric modes originating from inhomogeneous magnetization precession in the FL due to non-uniform currents (arising, e.g., from pinholes), while the device with higher MR device (73%), showed a single resonance mode. It is found that the quality of the barrier certainly has an effect on the in-plane torkance, and larger MR means larger efficiency in spin-symmetric injection, which should lead to larger in-plane spin transfer torque.

The bias dependence of in-plane (IP) and out-of-plane (OOP) torkance are shown in Figs. 5a, b, respectively. Figure 5a shows that IP torkance depends strongly on the bias voltage. The magnitude of STT is higher for higher MR device. However, the OOP torkance is almost constant for all the devices, which implies a linear dependence of OOP torque on the bias voltage, which is consistent with studies performed in Ref. [40, 50, 52, 55] but inconsistent with other works [32, 57]. Also, the magnitude of OOP torkance is two orders of magnitude below the magnitude of IP torkance, in contrast from previous works [48, 60]. The OOP torkance also shows a change of its sign at higher bias. They also observe this behavior in other low MR devices (<65%) and it is because of the sign change of the antisymmetric amplitude (V_a) at higher bias. The origin of this behavior was attributed to second–order term in the field-like torque which appears as a linear term in torkance. These results suggest that bias dependence of torques depends on the MR ratio and this could be the reason for reported different behaviors of torques as functions of bias voltage, especially the field-like torque.

4 Auto-oscillations in MTJ Nanopillar

The term "auto-oscillatory phenomenon" refers to the generation of periodic signals when the input signal is not periodic [61]. In case of STNOs, input in the form of dc bias leads to a nonideal sinusoidal signal that is periodic in nature and which shows some phase deviation. Here, the auto-oscillations in STNOs are because of the STT, which provides sufficient energy for the local magnetization of the free layers and acts opposite to the natural damping force. Slavin and Tiberkevich [14] provided the general theoretical model of the auto-oscillations in STNO devices, which is a generalized nonlinear auto-oscillator model:

$$\frac{dc}{dt} + i\omega\,(p)c + \Gamma_{+}(p)c - \Gamma_{-}(p, I)c = f\,(t) \tag{3}$$

where $c(t)$ defines the complex amplitude of the auto-oscillations characterized by the resonant frequency, $\omega(p)$, power, $p = |c|^2$, and phase, $\Phi = arg(c)$. The third term, $\Gamma_{+}(p)$, defines the natural damping that in general describes energy dissipation. The fourth term, $\Gamma_{-}(p)$, defines the source of energy pumped to the system that can counteract the natural damping and cause the system to generate auto-oscillations. As this is opposite to the natural damping, $\Gamma_{-}(p)$ is usually known as a negative damping term. The nonlinearity in the auto-oscillation comes from the dependence of the resonant frequency and of both damping terms on the auto-oscillation power [14, 62–69]. The term on the right hand side, $f\,(t)$, describes the interaction of the STNO with other sources or perturbations, such as thermal fluctuations. The resonant frequency is power-dependent and under nonlinear dynamics can be described by: $\omega(p) = \omega_0 + Np$. Here, ω_0 is the ferromagnetic resonance frequency and N is the nonlinear frequency shift due to amplitude–phase coupling in the STNO, resulting in a nonlinear contribution to the phase noise. The precessional motion in the STNOs, especially the tuning of resonant frequency $\omega(p)$, depends on the applied current and effective magnetic field. The applied current can directly influence the oscillation power p and thus $\omega(p)$. The power dependence of $\omega(p)$ mainly comes from the demagnetizing field, which is included in the H_{eff} term of Eq. (1). However, the main factor that can tune the $\omega(p)$ in a wide range is the externally applied magnetic field H_{ext} in the H_{eff} term. An important feature of the auto-oscillatory regime is the threshold nature of the auto-oscillation. According to Ref. [14], this threshold current is theoretically given by

$$I_{th} = \Gamma_g/\sigma. \tag{4}$$

where Γ_g represents the half-linewidth of the linear ferromagnetic resonance given by $\Gamma_g = \omega_0\alpha_G$, where ω_0 is the ferromagnetic resonance frequency and α_G is the intrinsic Gilbert damping constant. The coefficient σ is given by:

$$\sigma = \frac{\varepsilon\, g\mu_B}{2eM_0LS}, \tag{5}$$

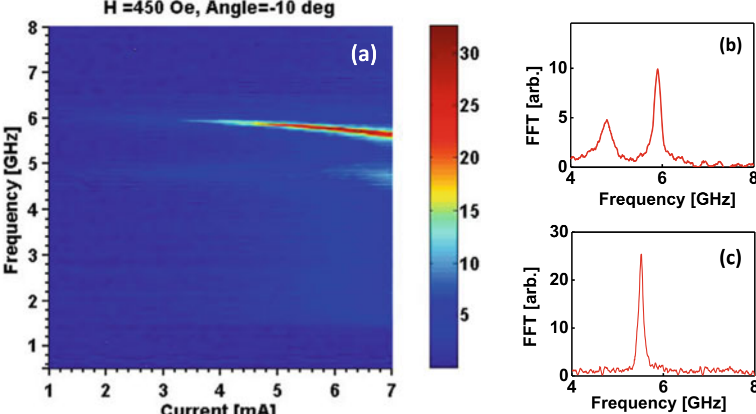


Fig. 6 **a** Map of STNO power as a function of STNO frequency and dc current. **b** and **c** show the spectra measured by the spectrum analyzer at dc currents of 4 mA (below threshold) and 7 mA (above threshold), respectively

Here, ε is a dimensionless quantity that represents a spin-polarized efficiency, g is the spectroscopic Landé factor, μ_B is the Bohr magneton, e is the electronic charge, L is the thickness of the free layer, and S is the area of the current-carrying region of the STNO. Equation (4) shows that threshold current required for the microwave generation by the STT depends on the damping constant and the coefficient σ. These factors depend on the material properties as well as the applied field. The free-running auto-oscillation frequency ω_p can be tuned by using both magnetic field and dc current. Furthermore, the applied dc current can also control the STNO frequency directly through the Oersted field in the H_{eff} by altering the spatial variation [70, 71].

Figure 6 shows the example of spectra measured from an MTJ-based STNO. Auto-oscillation is achieved at a current of 6.4 mA. At lower currents, low-power thermally excited ferromagnetic resonance (TE-FMR) signals are seen due to the high magnetoresistance (MR) of MTJ devices [53]. These signals are often not observed for metallic-based devices. At low currents, the spectra can often contain multiple modes [72], as shown in Fig. 6b for a current of 4 mA, which is below the auto-oscillation threshold. Figure 6c shows a sharp peak due to auto-oscillations at a current of 7 mA, which is above the threshold current of 6.4 mA.

5 Auto-oscillations in MTJ Hybrid Nanocontacts

MTJs are traditionally shaped into nanopillars in order to confine the current and force it to go through the insulating barrier. Although nanopillar MTJs can generate much higher microwave power, they suffer from larger linewidths compared to nanocontact structures [73–76]. The reason for that is attributed to the smaller effective volume

of the nanopillar structure and also the in-homogeneous magnetization that arises from the serrated edges of the nanopillars, which can lead to incoherent precession of magnetization. To combine the low linewidths of nanocontact structures and high output power of MTJs, a hybrid nanocontact structure has been proposed [77, 78]. A schematic of this type of hybrid device is shown in Fig. 7a, b. In these type of devices, two modifications were implemented in order to force the current to go through the insulating barrier. First, the MTJ cap layer is thinned down to minimize the lateral current shunting as can be seen in Fig. 7b. Second, a MgO layer with a low-resistance area (RA) product of 1.5 $\Omega\mu m^2$ was used to promote the tunneling through the insulating barrier. As will be seen in the following, these devices show a rich magneto-static and magneto-dynamic behaviour.

5.1 Magnetostatic Properties of Nanocontact MTJs

The complete layer sequence of the MTJ stack used for the study is: Ta(3)/CuN(30)/ Ta(5)/PtMn(20)/$CoFe_{30}$(2)/Ru(0.85)/$CoFe_{40}B_{20}$(2)/$CoFe_{30}$(0.5)/MgO/$CoFe_{30}$(0.5)/ $CoFe_{40}B_{20}$(1.5)/Ta(3)/Ru(7), with thicknesses in nm. This stack has an in-plane easy axis. Figure 7c shows the hysteresis loop of the above-mentioned stack with the magnetic field applied along the in-plane easy axis while the corresponding resistance of an MTJ-STNO with a nominal diameter of $d_{NC} = 150$ nm, is shown in Fig. 7d. The MTJ-STNO device displays a 36% magnetoresistance and the field dependence of the MTJ-STNO device and the unpatterned film stack shows a very good agreement. This indicates that during the processing of the MTJ stack there was minimal damage to the film and more importantly, the free layer (FL) has remained intact. In Fig. 7c the magnetic state of some of the key points in the hysteresis loop is demonstrated. Decreasing the magnetic field from the fully saturated state (1), causes the reference layer (RL) to gradually become antiparallel to FL. This is manifested with a substantial increase in the resistance (R) of the MTJ-STNO, shown in Fig. 7d. Decreasing the field even more from state 2 $\rightarrow$ 3, the FL and RL become parallel again and as a result, R decreases as well. Further decreasing the field from 3 $\rightarrow$ 4, the pinned layer (PL) slowly switches to be parallel to the RL, working against a strong antiferromagnetic coupling (AFC). Going from 5 $\rightarrow$ 6 $\rightarrow$ 1, The FL switches at low field and RL switches at high fields to align with the applied field.

5.2 Magnetodynamic Properties of Nanocontact MTJs

To study the magnetodynamical behaviour of MTJ-STNOs, their power spectral density was extracted as a function of magnetic field strength applied at $\theta_{ex} = 85°$ for different drive currents. The results of such measurement are shown in Fig. 8. At low currents ($I_{dc} = -5$ & -6 mA), the strongest mode that can be detected is the spin wave bullet [16, 76, 80], with a much lower frequency than the ferromagnetic

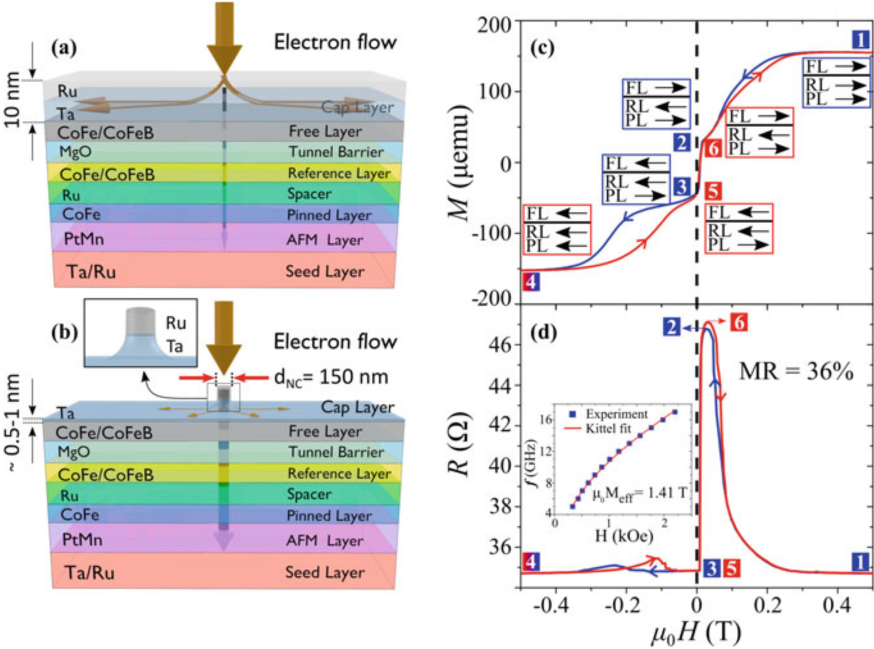


Fig. 7 Schematic of the layer sequence and the current distribution in (**a**) a normal nanocontact and (**b**) a hybrid nanocontact structure in which the cap layer is thinned down. **c** Hystersis loop of the unpatterned MTJ stack with the field being applied along the in-plane easy axis. **d** Resistance of a MTJ-STNO device with the field being applied along the in-plane easy axis yielding a magnetoresistance of 36%. Inset in (**d**) in the uniform FMR mode frequency (blue dots and the Kittel fit (red line) to extract an effective magnetization of 1.41 T. Reprinted from Houshang et al. [79] Copyright (2018) by Nature Communications

resonance frequency (f_{FMR}, red dashed line). This mode eventually dies out at $H_{\mathrm{ex}} =$ 1.35 T where the self-localization of the bullet is no longer possible. We calculate the internal angle of magnetization at this field from the following boundary conditions:

$$H_{\mathrm{ex}} \cos\theta_{\mathrm{ex}} = H_{\mathrm{int}} \cos\theta_{\mathrm{int}}, \tag{6}$$

$$H_{\mathrm{ex}} \sin\theta_{\mathrm{ex}} = (H_{\mathrm{int}} + M_{\mathrm{s}}) \sin\theta_{\mathrm{int}}. \tag{7}$$

where, H_{int} and θ_{int} are the internal magnetic field and out-of-plane angle, respectively. The internal angle is found to be $\theta_{\mathrm{int}}^{\mathrm{crit}} = 60°$ and it is in a good agreement with the theoretical predictions $\theta_{\mathrm{int}}^{\mathrm{crit}} = 55°$ [81] Above the critical field, a weaker mode can be seen having a frequency only slightly higher than f_{FMR} which is consistent with the ordinary Slonczewski propagating spin wave mode. Increasing the drive current to $I_{\mathrm{dc}} = -7$ mA, a strong mode with a frequency well above f_{FMR} starts to appear. Further increasing the amount of current, a sharp jump to a mode with even higher frequency is observed at about 1.6 T (Fig. 8d). At I_{dc} = -9 mA and -10 mA,

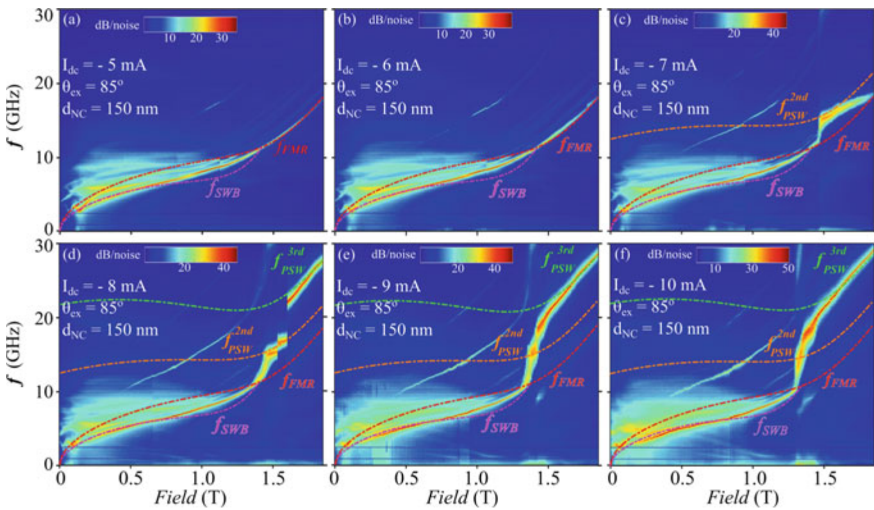


Fig. 8 Power spectral density of auto-oscillations as a function of applied magnetic field strength for **a** $I_{dc} = -5$ mA, **b** $I_{dc} = -6$ mA, **c** $I_{dc} = -7$ mA, **d** $I_{dc} = -8$ mA, **e** $I_{dc} = -9$ mA, and **f** $I_{dc} = -10$ mA. The FMR frequency (f_{FMR}) and the SW bullet frequency (f_{SWB}), are shown by red and pinck dashed lines, and the second- (f^{2nd}_{PSW}), and third-order (f^{3rd}_{PSW}) Slonczewski modes are shown by brown and green dashed lines, respectively. Reprinted from Houshang et al. [79] Copyright (2018) by Nature Communications

the new mode dominates the high field region. These new high frequency modes can be perfectly analyzed by the higher-order propagating SW modes put forward by Slonczewski [82]. Slonczewski predicted that propagating SWs can assume a discrete set of wave vectors $r_{NC}k \simeq 1.2, 4.7, 7.7 \ldots$. Using the free layer (CoFe/CoFeB) exchange stiffness $A_{ex} = 23 \times 10^{-12}$ J/m, and allowing for a reasonable lateral current spread, one cat fit the high frequency modes accurately, by taking higher order wave vectors into account. The experimental results together with the calculated SW mode frequencies are shown in Fig. 8c–f. Furthermore, the need for higher currents to excite these higher order modes are also in-line with Slonczewsiki's expectations [82]. The details of calculating the frequency of higher order modes, dependence of M_s on I_{dc}, and the role of azimuthal modes in addition to the radial ones considered by Slonczewski, can be found in Ref. [79].

5.3 Mutual Synchronization of Nanocontact MTJs

Mutual synchronization of STNOs has been extensively studied in the literature [71, 83–85]. Here, the mutual synchronization of all three Slonczewski modes are experimentally shown (Fig. 9), further demonstrating their propagating nature. Field and current dependence of MTJ-STNOs with two NCs having 150 nm nominal diameter

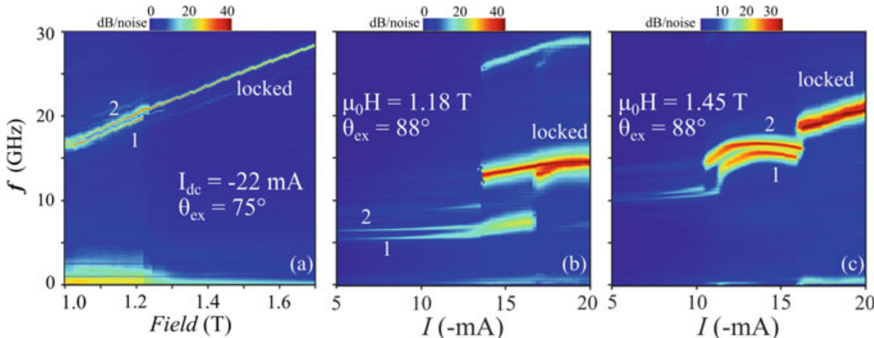


Fig. 9 Mutual synchronization on the **a** first-, **b** second-, and **c** third-order Slonczewski modes for a MTJ-STNO with two NCs having a nominal diameter of $d_{NC} = 150$ nm and a center to center spacing of 300 nm. Reprinted from Houshang et al. [79] Copyright (2018) by Nature Communications

separated 300 nm are shown in Fig. 9a–c. The behavior of the device as a function of field at a constant drive current, is shown in Fig. 9a. At fields lower than 1.22 T, two distinct modes can be observed at high frequencies with a substantial noise at low frequencies, indicative of a substantial interaction between the two modes. By increasing the field, these two modes merge into one and at the same time, the low-frequency noise disappears which is consistent with mutual synchronization. A set of new devices with a thicker bottom electrode to sustain more current, were used to study the synchronization behaviour as a function of drive current. As can be seen in Fig. 9b, at 1.18 T, two fist order Slonczewski modes can be detected at currents $I_{\rm dc} < -13$ mA. Increasing the amount of drive current, first one of the NCs jump to second-order at about $I_{\rm dc} = -13$ mA and then the second NC at about $I_{\rm dc} = -17$ mA after which the two modes merge and synchronize at about -18 mA. Increasing the applied field to 1.45 T, the second order modes do not synchronize at all while the third-order modes are synchronized. Therefore, it is possible to tune the applied field and drive current so that synchronization is achieved for each of the three Slonczewski modes.

Being able to generate high-order Slonczewski modes means that one can generate modes with much shorter wavelengths, in this 120 nm and 74 nm for the second- and third-order modes. Having access to larger wavevectors also indicates that the group velocity of these waves are also higher since the group velocity increases with wavevector $v_{\rm gr} \simeq 4\gamma A_{\rm ex} k / M_{\rm s}$. This means not only SWs can travel much faster but much further before they get damped out which has significant implications for practical applications of magnonic devices.

6 Generation Linewidth of MTJ-Based STNOs

The understanding of the origin of linewidth in STNOs is crucial for applications. In the following, we will discuss the linewidth of single modes and multimode STNOs in separate subsections.

6.1 Single-Mode Excitation

The linewidth, Δf of the single mode in STNO devices can be described by the single-mode theory developed by Kim et al. [14, 67, 68].

According to this theory, Δf of a nonlinear oscillator is given by:

$$\Delta f = \Gamma_g (1 - \tfrac{I_{dc}}{I_{th}}), \ \text{for}\ I << I_{th} \tag{8}$$

$$= \Delta f_L (1 + \nu^2), \ \text{for}\ I >> I_{th} \tag{9}$$

where I_{dc} is the dc bias current, and the nonlinear linewidth amplification is $(1 + \nu^2) = 1 + \left(\frac{p_0 N}{\Gamma_p}\right)^2$, where $N = \frac{d\omega}{dp}$ is the nonlinear frequency shift, and Γ_p is the power restoration rate (Γ_p^{-1} is the correlation time of the power fluctuations); $\Delta f_L = \Gamma_g \frac{kT}{E(p_0)}$ is the intrinsic thermal linewidth—i.e., the linewidth of a linear ($\nu = 0$) oscillator. Here, $E(p_0)$ is the total energy of the oscillator. Above threshold ($I_{dc} \gg I_{th}$), the nonlinear amplification of the linewidth is controlled by the ratio of the nonlinear frequency shift N to the power restoration rate Γ_p.

George et al. [69] successfully applied the single-mode theory to MTJ-based STNOs. They showed that the nonlinear coefficient ν can be calculated from the equation:

$$A_{NL} = 1 + \left(\frac{I_{dc}}{\Gamma_g}\frac{df}{dI_{dc}}\right)^2. \tag{10}$$

Here, A_{NL} is the nonlinear amplification equal to $(1 + \nu^2)$. Figure 10 shows the experimental observation of the linewidth below and above the threshold current of 1 mA. Above the threshold current, the linewidth is amplified by a factor of $(1 + \nu^2)$ due to nonlinear oscillations maintained by the STT.

The high power of MTJ-based STNOs allows direct measurement of the signal in the time-domain, which is not possible in low-power metallic-based STNOs. Bianchini et al. [86] have shown that it is possible to directly measure the nonlinear coefficient ν, as well as the linear linewidth, from time-domain data using the following relation of the phase variance [66, 86]:

$$\delta\phi_R^2 = 2\Delta\omega_0 \left[(1 + \nu^2)t - \nu^2 \frac{1 - e^{-2\Gamma_p}}{2\Gamma_p}\right]. \tag{11}$$

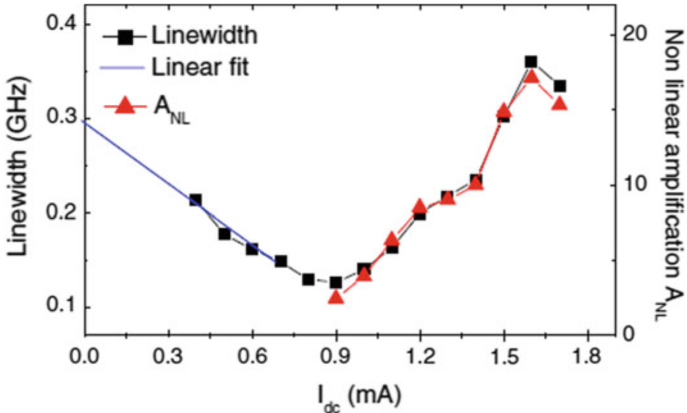


Fig. 10 Linewidth as a function of I_{dc} for $H_{ext} = 110$ Oe and $T = 300$ K. The right axis shows the experimentally determined nonlinear amplification factor A_{NL} with I_{dc} as fitted by Eq. (10). Reprinted from Georges et al. [69] Copyright (2009) by the American Physical Society

They derived Γ_p from the auto-correlation function of power fluctuations using the following expression [87]:

$$\kappa_P(\tau) = \langle \delta p(\tau)\delta p(0)\rangle = A(p_0, \Gamma_p)e^{-2\Gamma_p(\tau)}. \tag{12}$$

Here, δp is the power fluctuations around the mean value p_0 and $A(p_0, \Gamma_p)$ is the amplitude. The other way of measuring Γ_p is given by Refs. [88, 89] from the noise spectra. In particular, Ref. [88] estimates the nonlinear amplitude relaxation frequency, $f_p = \Gamma_p/\pi$, from the crossover of the linear and nonlinear contributions in the phase noise plot.

6.1.1 Temperature Dependence of Linewidth

The linewidth of the single mode in STNO devices, as discussed above, can be well described by white noise generated primarily due to thermal noise. White noise is a random noise whose intensity is independent of frequency. The thermal fluctuations add a Gaussian noise to the STNO spectra, leading to a Lorentzian line shape. This Gaussian white frequency noise exists in all electronic devices. The effect of thermal fluctuations on the emission linewidth has been studied in MTJ-STNOs in different experimental studies [69, 90, 91]. Reference [69] showed that intrinsic noise is not dominated by thermal fluctuations but rather by chaotic dynamics of the magnetic system induced by the spin-transfer torque. Sierra et al. [91] showed a linear dependency of linewidth in nanopillar MTJ-STNOs in contrast to the result shown by Ref. [69]. Sierra et al. showed that the emission linewidth (Δf) strongly depends on nonlinear parameters like ν and Γ_p. For the nonlinear oscillator, where

$\nu > 0$, emission linewidth could be dependent on the temperature in different ways. For the case of $2\pi\Delta f << \Gamma_p$, emission linewidth shows a linear dependency on the temperature, whereas for the case of $2\pi\Delta f > \Gamma_p$, the emission linewidth increases nonlinearly with the temperature in good agreement with theoretical predictions [14, 66].

6.2 Multimode Excitation and Mode-Hopping

In the previous section, the analytic theory of linewidth generation was based on the single auto-oscillatory mode. However, various studies have shown the presence of multiple modes in STNO devices [43, 92, 93]. Figure 11 shows so-called Wigner plots of instantaneous frequency of MTJ-based STNO for a duration of 20 ns. Clear mode-hopping can be seen at $\varphi = 220°$, where φ represents the angle between the free and reference layer. The right-hand side column shows the autocorrelation function of the power fluctuation. The decay of the autocorrelation function $\Gamma(\tau)$ is faster in the case of $\varphi = 220°$ because of mode hopping events and decoherence. For the case of $\varphi = 196°$, a single mode is observed and the decay of the autocorrelation function $\Gamma(\tau)$ is slower. Detailed investigation into the bias current and field angle shows that mode-hopping is minimized near the current threshold for the antiparallel alignment of the free layer with the reference layer magnetization. Away from the antiparallel alignment, mode hopping limits oscillator coherence. Furthermore, evidence of mode hopping [40, 72, 94], mode transitions [16, 95], and coexistence [70] has been reported in STNOs.

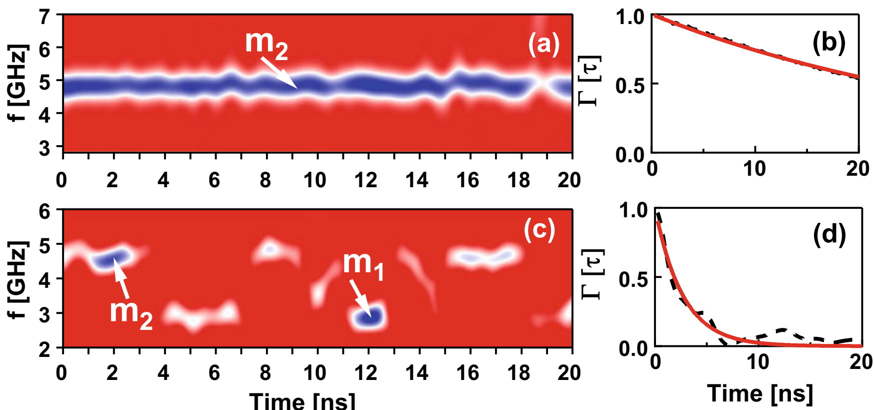


Fig. 11 Wigner distribution (first column) and normalized autocorrelation envelope of the dominant mode (second column) at φ= (**a**), (**b**) 196°, and (**c**), (**d**) 220° for $I_{dc} = 8$ mA and $H_{ext} = 450$ Oe. Red lines are exponential fits to the experimental autocorrelation envelope from the Eq. (12). Dashed lines are the experimental data. Reprinted from Muduli et al. [72] Copyright (2012) by the American Physical Society

There have been some recent works to extend the single-mode theory to multiple modes [72, 96–99], which is necessary in order to explain experimental observations of multi-mode system as well as of the generation linewidth of such systems. These works draw on an analogy between equations describing STNO multimode generation with equations similar to those describing multimode ring lasers [100, 101] in laser technology. The theory presented in Ref. [99] is a generalization of the single-mode theory by Slavin and Tiberkevich [14] that includes contributions to nonlinear frequency shift as well as to nonlinear damping and pumping from mode-interactions mediated by exchange and magnetostatic coupling. An important part of the theory in Ref. [99] is that coupling through a bath of thermally excited magnons leads to an effective linear mode-coupling term. This is the term that leads to mode-hopping in the presence of a stochastic force, and it also leads to temperature-driven effects, such as temperature-induced switching from one resonant mode to another, as observed by Muduli, Heinonen, and Åkerman [96]. If we consider a system with an in-plane MTJ-STNO and a total current I perpendicular to the plane of the magnetic layers and with two dominant modes 1 and 2 between which the system exhibits mode-hopping while the other modes are thermally populated, one can derive equations for the slow time-evolution of amplitudes $A_1(t)$ and $A_2(t)$. Here, "slow" means slow compared to the typical nano-second or sub-nanosecond timescales of resonant modes of such a system. As was shown in Ref. [99], the effective equations for the amplitudes of the coupled modes are:

$$
\begin{aligned}
\frac{dA_1(t)}{dt} &= -i\left[\eta_{1,1}|A_1|^2 + \eta_{1,2}|A_2|^2\right]A_1 - \Gamma_{G,1}\left[1 + P_{1,1}|A_1|^2 + P_{1,2}|A_2|^2\right]A_1 \\
&\quad + \sigma_0 I \cos\beta\left[1 - Q_{1,1}|A_1|^2 - Q_{1,2}|A_2|^2\right] + R_{1,2}(T)A_2 \\
\frac{dA_2(t)}{dt} &= -i\left[\eta_{2,1}|A_1|^2 + \eta_{2,2}|A_2|^2\right]A_2 - \Gamma_{G,2}\left[1 + P_{2,1}|A_1|^2 + P_{2,2}|A_2|^2\right]A_2 \\
&\quad + \sigma_0 I \cos\beta\left[1 - Q_{2,1}|A_1|^2 - Q_{2,2}|A_2|^2\right] + R_{2,1}(T)A_2.
\end{aligned}
\tag{13}
$$

Here, $\sigma_0 I$ is the effective field that arises because of STT, β is the angle between the reference layer magnetization and an in-plane external magnetic field, and $\Gamma_{G,i} = \alpha\omega_i$, $i = 1, 2$, with ω_i the linearized undamped frequency of eigenmode i. The coefficients $\eta_{i,i}$ are the usual non-linear frequency shift due to the power of mode i, but now there are also off-diagonal non-linear frequency shifts $\eta_{i,j}$, $i \neq j$, of mode i due to the power of mode j. Similarly, the non-linear damping and pumping now have off-diagonal contributions $Q_{i,j}$ and $P_{i,j}$, $i \neq j$, to the dynamics of mode i because of the power in mode j, in addition to the diagonal contributions $Q_{i,i}$ and $P_{i,i}$. Note that while $Q_{1,2} = Q_{2,1}$ it is *not* the case in general that $P_{1,2} = P_{1,2}$. Finally, the complex coefficients $R_{i,j}(T)$, with T the temperature, couple the amplitude of mode j to the time-evolution of mode i. These are the couplings that arise from interactions with the thermal bath of magnons. If the coefficients $R_{i,j}$ are zero, the system may exhibit single motion according to modes 1 or 2, or may exhibit mode coexistence. Only if $R_{i,j}$ are non-zero [100, 101], and in the presence of a stochastic force such as thermal agitation, will the system exhibit mode-hopping.

6.2.1 Impact of 1/f Frequency Noise on Linewidth

Keller et al. [75] showed an additional mechanism of linewidth broadening that arises from the so-called *1/f frequency noise* or color noise. They obtained a frequency noise spectrum spanning more than five decades of Fourier frequencies and showed that the frequency noise spectrum is white at large Fourier frequencies and varies as *1/f* at small Fourier frequencies. They showed that *1/f frequency noise* can causes both broadening and a change in shape of the oscillator's spectral line as the measurement time increases. The *1/f frequency noise* was observed in other types of STNOs [75, 102, 103] as well as in MTJ-based STNOs [88, 104].

In general, the STNO waveform is a nonideal sinusoidal waveform where instantaneous frequency/phase and amplitude varies as a function of time. Keller et al. [75, 102] measured the phase noise using the zero crossover method. The phase noise is calculated from the power spectral density of the phase deviations, which is calculated using the equation, $\phi_{tot} = 2\pi f_0 t + \phi(t) = n\pi$. Here, ϕ_{tot} is the total phase, f_0 is the nominal frequency, and $\phi(t)$ is the phase deviation. Following this method, the frequency noise can simply be calculated using the Fourier relation $S_\nu^2 = f^2 S_\phi^2$ [88, 102]. Here, S_ν is the frequency noise (power spectral density of instantaneous frequency deviation) and S_ϕ is the phase noise (power spectral density of instantaneous phase deviation). Figure 12 shows the frequency noise plots as a function of field angle (φ) between the free and fixed layers and the dc bias current (I_{dc}) in an MTJ-based STNO. Note that the free and fixed layers are aligned antiparallel—that is, $\varphi = 180°$. The threshold current of the device is 6 mA for the chosen magnetic field of 400 Oe. A strong single mode was observed for the case of $I_{dc} = 7$ mA and $\varphi = 196°$, whereas for the other case, $I_{dc} = 7$ mA and $\varphi = 220°$, as many as 5 clear modes were observed [72]. To quantify the $1/f$ frequency noise, a parameter known as transition frequency (f_t) [104] is defined as the point on the frequency noise plot at which $1/f$ frequency noise transforms into white frequency noise. It can be seen from the above threshold condition (Fig. 12a) that under mode-hopping conditions of $\varphi = 220°$, f_t increases to 440 kHz. In the sub-threshold region (i.e., with $I_{dc} = 3$ mA), where thermal noise dominates, multiple modes have been observed and hence f_t is higher. In this case also, f_t increases in the case of $\varphi = 220°$, where again significant mode hopping is observed (Fig. 12b). A significant increase in the linewidth was observed

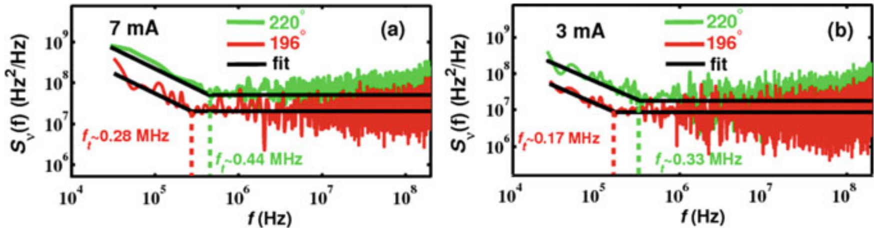


Fig. 12 Comparison of frequency noise at $\varphi = 196°$ and $\varphi = 220°$ for **a** $I_{dc} = 7$ mA and **b** $I_{dc} = 3$ mA. The dashed lines show the transition frequency where $1/f$ frequency noise begins. The data is similar to Ref. [104]

for the case where $1/f$ frequency noise is high. Furthermore, it was shown that the $1/f$ frequency noise is significantly high due to mode-hopping. The line-shape fitting shows a better fit with a Voigt line-shape compared to a Lorentzian line-shape. This verifies both previous observations [75, 98] of change in the lineshape for $1/f$ frequency noise. But, most importantly, it is established that mode-hopping can lead to $1/f$ frequency noise.

6.2.2 Mode-Hopping Contributions to the Nonlinear STNO Generation Linewidth

A single-mode STNO will exhibit a Lorentzian lineshape in the frequency domain. If there is mode-hopping present, one would expect it to contribute to the linewidth as mode-hopping events generally reduce the coherence of the motion. Based on the mode-coupling theory in Refs. [72, 96, 97, 99], Iacocca et al. [98] investigated the effect of mode-hopping on the nonlinear STNO generation linewidth. As one might expect, mode-hopping events do indeed increase the generation linewidth. However, multi-mode generation first of all fundamentally changes the generation line shape to a Voigt line shape in general, but under certain circumstances, the line shape may be reduced to a Lorentzian. Mode-hopping events can be described as a Poisson process which provides a purely Lorentzian contribution to the line shape. At high mode-hopping rates, mode-hopping contributions to the line width will dominate and the line shape will then regain a Lorentzian character. The mode-hopping rate can, in this case, be described by an Arrhenius distribution, and measurements of the linewidth can provide direct estimates of the energy barrier that has to be overcome by thermal fluctuations during mode-hopping events.

The analysis starts by directly calculating for a two-mode system the autocorrelation function $\mathcal{K}(t-t')$ of the system, which is defined as

$$\mathcal{K}(t-t') = \langle [A_1(t) + A_2(t)] \left[A_1^*(t') + A_2^*(t')\right] \rangle, \tag{14}$$

where $A_i(t)$ is the slow amplitude of mode i. By recognizing that individual mode-hopping events occur on time-scales very much smaller than the time-scale of evolution of $A_i(t)$, mode-hopping events can then be added as a Poisson process with a rate λ. After some lengthy calculations one can then show that the autocorrelation function $\mathcal{K}(\tau)$, where $\tau = t - t'$, in general contains three kinds of behavior in τ. The first two are factors that are exponentials in $-|\tau|$ and $-\tau^2$, respectively, and arise from the coupled mode equations. The third factor is an exponential in $-\lambda|\tau|$, and it arises from the mode-hopping. In general, the line shape is then given by a convolution of exponentials in $|\tau|$ with a Gaussian in τ, which is a Voigt line shape. This was also confirmed by Iacocca et al. [98] by directly integrating the coupled mode-equations in the presence of a thermal stochastic force, calculating the auto-correlation function, and Fourier transforming it.

Close to a mode transition, the rate mode hopping events are observed to increase dramatically, and the line width is then dominated by mode-hopping events. Measure-

ments of the line width can then provide a direct measurement of the mode-hopping rate λ. Under the assumption that mode-hopping events obey an Arrhenius distribution, the energy barrier ΔE that must be overcome by thermal agitation in order to produce a mode-hopping event can then be obtained from

$$\Delta E(I_{dc}) = k_B T \log \frac{f_a}{\lambda}. \tag{15}$$

Here, we have indicated that the energy barrier $\delta E(I_{dc})$ depends sensitively on the dc current I_{dc}; the frequency f_a is an attempt frequency. k_B is the Boltzmann constant and T is the temperature. Using a simple assumption that the attempted frequency is twice the relaxation frequency and be extracting $\Delta E(I_{dc})$ from a single measurement, Iacocca et al. [98] obtained good quantitative and qualitative agreement between theory and experiment [96] for the linewidth as a function of temperature.

7 Parametric Synchronization

Injection locking is a phenomenon in oscillators where one oscillator can synchronize with the other when operating at nearby frequencies. In case of STNOs, injection locking [105–109] is useful in understanding the synchronization behavior of an STNO to an external RF source. Furthermore, injection locking is useful for understanding the mutual synchronization [71, 83–85, 110–112] of multiple STNO, which is essential for improving the power of STNOs. In addition to injection locking, an external RF source operating at a fraction or harmonic of the STNO eigenfrequency can also synchronize the STNO. In particular, parametric synchronization [113–118], in which the external RF source frequency (f_e) is twice the STNO operating frequency (f_0), allows the synchronization of an STNO without interference from the external RF signal.

Parametric excitation is a related phenomena that was first reported by Urazhdin et al. [113] in a nanocontact-based STNO at the very low temperature of 5 K. Urazhdin et al. [113] applied a microwave field via a separately fabricated strip line on top of the STNO. They applied a frequency f_e, equal to twice the resonance frequency f_0 of the STNO, and parametrically induced oscillations at a DC current below the oscillation threshold. The first room-temperature measurement of parametric excitation was performed by Bortolotti et al. [90] in vortex-based MTJ-STNOs using RF currents instead of an RF magnetic field. In this work, the Oersted or orthoradial field due to the RF current was used to parametrically excite the vortex gyration.

Both parametric excitation and parametric synchronization have been demonstrated in MTJ-based STNOs in Ref. [117], as shown in Fig. 13. At a current of $I_{dc} = 4$ mA , well below the threshold current $I_{th} = 6.3$ mA, parametric excitation was observed, while at a high current of $I_{dc} = 7$ mA, parametric synchronization was observed. In the case of parametric excitation, the total integrated power around the STNO single mode increases, unlike in the case of injection locking, where the

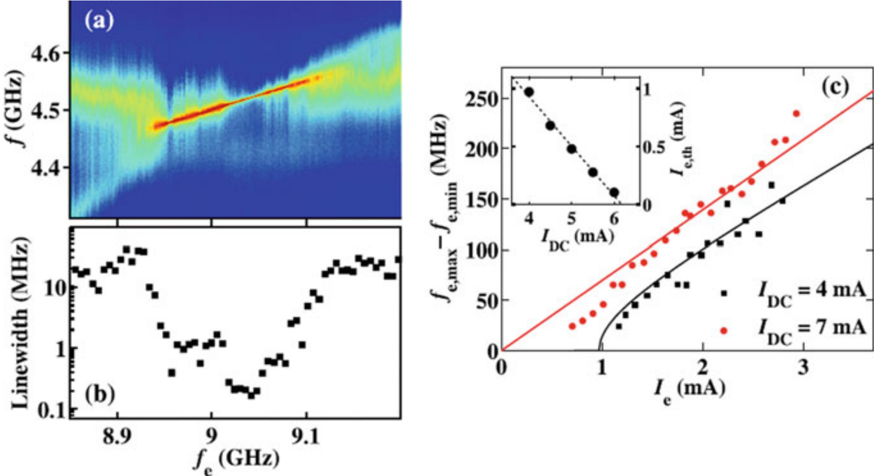


Fig. 13 Parametric synchronization and excitation of the MTJ-based STNO. **a** Spectrum of the STNO frequency as a function of injected RF signal with a frequency f_e for an injected RF current of $I_e = 2.6$ mA and $I_{dc} = 7$ mA showing parametric synchronization. **b** Variation of linewidth as a function of I_e. **c** Excitation bandwidth as a function of the external signal strength I_e at $I_{dc} = 4$ mA (parametric excitation) and $I_{dc} = 7$ mA (parametric synchronization). The corresponding solid lines for 4 mA and 7 mA are fits to Eqs. (16) and (17), respectively. Inset: Excitation threshold $I_{e,th}$ as a function of dc bias current. Reprinted from Dürrenfeld et al. [117], with the permission of AIP Publishing

integrated power remain constant. This shows that thermally activated subthreshold oscillations can become coherent through parametric excitation. According to Ref. [113, 117], the excitation bandwidth is given by:

$$\Delta\omega_e = 4\sqrt{V^2 I_e^2 - \Gamma_I^2} \tag{16}$$

Here, V is the coupling between the external source and the STNO and Γ_I is the linear damping parameter in the subthreshold bias current: $\Gamma_I = \Gamma_o(1 - I_{dc}/I_{\text{th}})$. Figure 13 shows the excitation bandwidth calculated for the case of $I_{dc} = 4$ mA. The threshold RF current required to see any excitation bandwidth is $I_e = 0.99$ mA.

For the case of parametric synchronization at $I_{dc} = 7$ mA, the excitation bandwidth is higher than in the case of $I_{dc} = 4$ mA. The excitation bandwidth for the parametric synchronization is given by [113, 117]:

$$\Delta\omega_e = 4\sqrt{1 + \nu^2} V I_e \tag{17}$$

where, ν is the dimensionless nonlinearity coefficient. The above equation implies that there is no threshold microwave current requirement for parametric synchronization. The solid red line in Fig. 13 is a fit of the above equation for $I_{dc} = 7$ mA. The good fit indicates a nearly zero threshold within the uncertainties of experimental

data. The inset of Fig. 13 shows that the excitation threshold decreases in the subthreshold region and reaches zero at $I_{dc} = 7$ mA. In Ref. [117], it was also shown that parametric synchronization favors single-mode oscillations in the case of multimode excitation.

7.1 Time-Domain Stability of Parametric Synchronization

The thermal stability of any electronic device is a crucial issue. Although the synthetic antiferromagnetic structure gives good thermal stability to STNO devices, the magnetization dynamics of the free layer are in general affected by thermal agitation, which can also affect parametric synchronization. The time-averaged spectrum analyzer data cannot reveal the real-time dynamics of the signal generated by the free layer precession. For this purpose, time-domain study is required. A study of the effect of thermal noise is required to understand the coupling of STNOs in arrays. George et al. [107] was the first to discuss and demonstrate the effects of thermal Gaussian noise in injection-locked spin-valve-based STNO devices. According to this work, thermal noise can weaken injection locking by affecting the range of locking bandwidth. On the other hand, in an another study based on micromagnetic simulation [119], injection-locked spin-valve-based STNOs were found to be quite stable against thermal Gaussian noise. Recently, Lebrun *et al.* [120] described an experimental time-domain study of fractional synchronization under different conditions: $f_e \sim f_0/2$, $f_e \sim f_0$, and $f_e \sim 2f_0$, revealing that the power required for phase noise squeezing (where $1/f^2$ phase noise is completely suppressed) is different for different fractional synchronization conditions. In MTJ-STNO, a study of parametric synchronization [117] revealed that the frequency domain STNO spectra show a random unlocking at low microwave power, which can be attributed to thermal fluctuation.

A recent time-domain study of parametric synchronization [121] by the authors of this chapter revealed random unlocking at low RF power due to thermally driven frequency stability, while the frequency-domain data show the expected synchronization of any STNO device. The instantaneous frequency calculated from the Hilbert transform method reveals that the instantaneous frequency is not always inside the locking bandwidth (32 MHz, shown by blue dotted line), as shown in Fig. 14a at $P_{rf} = -11$ dBm. At the high RF power value of -5 dBm, the instantaneous frequency is inside the locking bandwidth (107 MHz, shown by red dotted line) for the total time of 5 μsec, as shown in Fig. 14a. However, even at -5 dBm, frequency fluctuations are present, resulting in a phase noise that exceeds the injected signal. We only observe phase slips of π, which, according to Ref. [120], is a signature of phase locking with an injected signal at $f_{rf} = 2f_0$. These π phase slips still leads to $1/f^2$ phase noise. To understand the random unlocking at low RF power, macrospin simulations were performed by adding the thermal field to H_{eff} by following Brown's approximation [122, 123]. The macrospin simulations reproduce the experimental results and reveal that the random unlocking during synchronization is driven by thermal fluctuations, as shown in Fig. 14b. Macrospin simulation further suggests

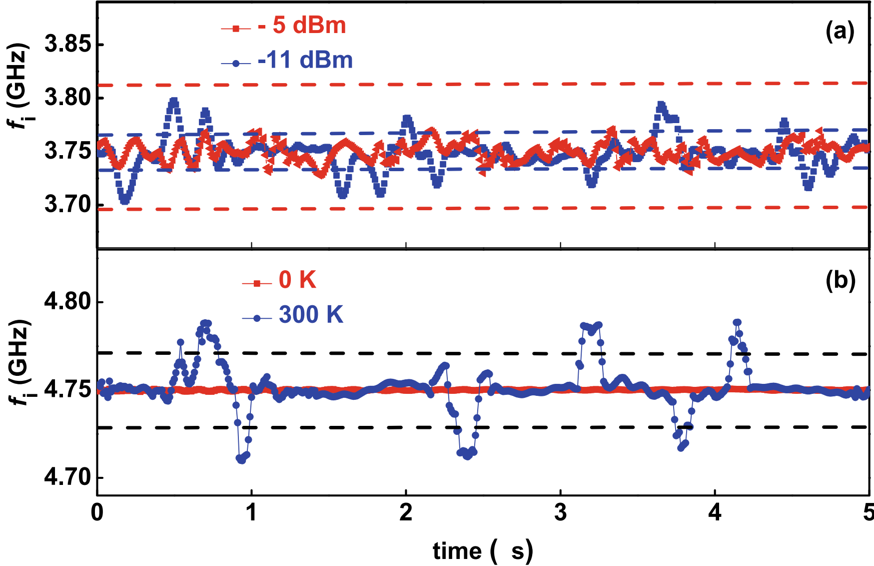


Fig. 14 Experimentally measured instantaneous frequency f_i of the synchronized signal at $P_{rf} = -11$ dBm (blue) and $P_{rf} = -5$ dBm (red) at $f_{rf} = 7.5$ GHz. The blue and red dotted lines show the locking bandwidths corresponding to $P_{rf} = -11$ dBm and $P_{rf} = -5$ dBm, respectively. **b** Simulated instantaneous frequency for the locked signal at $T = 0$ K and 300 K at $f_{rf} = 9.5$ GHz. The black dotted line show the locking bandwidth

that increasing RF power can lead to a complete suppression of $1/f^2$ phase noise, where not even π phase slips exists and results in a constant phase (i.e., STNO phase follows an injected signal phase: a condition of perfect phase synchronization) with a complete suppression of phase noise. Macrospin simulations suggest that a lower temperature and a higher positive field-like torque both reduce the threshold RF power required for phase noise squeezing under parametric synchronization.

8 Modulation of MTJ-Based STNOs

For communication applications, standard modulation techniques will have to be applied to STNOs to test their suitability for communicating information. Different modulation schemes, such as frequency modulation (FM) [20, 22, 26, 124], amplitude shift keying (ASK) modulation, on–off keying (OOK) modulation [125–127], and frequency shift keying (FSK) modulation [128], have been demonstrated in STNOs for communication applications.

The most explored modulation technique is *frequency modulation*. In frequency modulation, the high frequency carrier signal (STNO frequency) is modified by a low frequency information signal. The frequency and amplitude of the STNO depend

nonlinearly on the dc bias current. Due to this nonlinearity, the fundamental models describing linear frequency modulation fail to describe the modulation behavior. In 2005, nonlinear frequency modulation (NFM) was reported by Pufall et al. [20] for a nanocontact-based STNO. To fit their experimental power of carrier and sidebands, the NFM method was implemented. In 2010, Muduli et al. [22] showed that the nonlinear frequency and amplitude modulation (NFAM) [22–24, 129, 130] theory is the most appropriate model for explaining the modulation of STNOs. The NFAM model was required to explain two main experimental observations: (1) unequal sideband amplitudes around the carrier spectrum and (2) shifts in the frequency of the carrier spectrum with the increase in the modulation current. Pogoryelov et al. [25] showed the validity of the NFAM model on a pair of synchronized nanocontact STNOs.

With regard to the modulation of MTJ-based STNOs, Martin et al. [124] first reported modulation of MTJ-vortex-based STNO devices. They also applied the NFM model to their experimental data and showed that the characteristic time of the vortex dynamics must be taken into account to explain the modulation data. This is because the modulation period approaches the transient time.

8.1 Single Sideband Modulation of MTJ-Based STNO

Recently, Sharma et al. [131] have demonstrated the phenomenon of single sideband (SSB) modulation in MTJ-based STNOs, as shown in Fig. 15. Figure 15a, b shows the experimental observation of the SSB over a broad range of modulation frequencies, f_m = 100–500 MHz at modulation current, I_m = 1 mA, and over a broad range of modulation currents of I_m = 0–1.5 mA at a modulation frequency of f_m = 500 MHz,

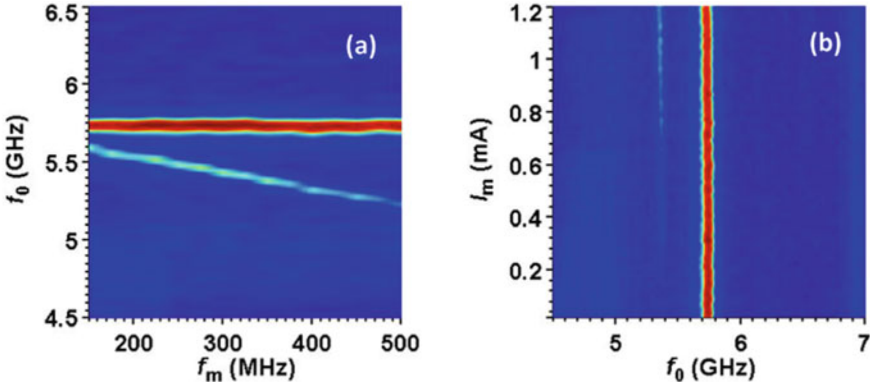


Fig. 15 Single sideband modulation of MTJ-based STNOs. **a** Map of power versus frequency and modulation frequency at I_m = 1.5 mA and I_{dc} = 4.6 mA and **b** map of power versus frequency and modulation current at f_m = 500 MHz at I_{dc} = 4.6 mA

respectively. The figure shows only the presence of a lower sideband; the upper sideband is completely suppressed. This phenomenon of a carrier with one sideband is known as single sideband (SSB) modulation and is a result of the strong nonlinear variation of frequency and amplitude with bias current in MTJ-STNOs. The performance of the STNOs as an SSB generator can be tuned with operating parameters such as in-plane field angle [132] as well as with the field-like torque term present in MTJ-based devices [131]. The in-plane field angle changes the frequency tunability with bias current and thus the SSB onset frequency and carrier power for the SSB transmission [132]. The field-like term also affects the frequency tunability with the bias current [40, 52, 55]. The observation of lower single side band (LSSB) in the experimental results of Ref. [131] indicates the presence of smaller field-like torque, which is consistent with other works on similar devices [40, 133].

Transmission through SSB generation is particularly advantageous for wireless communication since: (1) less transmission power is required as only one sideband, carrying the same information as the other one, will be transmitted; (2) there is less utilization of the available spectrum in comparison with double sideband (DSB) transmission; (3) there is less selective fading or interference with other signals. Furthermore, SSB generation through MTJ-based STNO devices has an edge over the existing technology for SSB generation, due to its compact size and fast modulation rate. The traditional method of generating SSB using, for example, a Hartley modulator [134] or a Weaver modulator [135], requires a number of components for SSB generation that makes the overall set-up a little bulky. However, STNO uses a single device for the carrier and sideband generation due to its nonlinearity and is suitable for chip technology, due to its compatibility with commercialized CMOS technology.

8.2 Modulation of Parametric Synchronized MTJ-STNOs

Most often, parametric synchronization and modulation are studied in MTJ-STNOs as a separate study to understand the effect of the external perturbation of the RF signal on the magnetization dynamics. Recently, an effort has been made to understand the modulation in a parametrically synchronized STNO in Ref. [136], as shown in Fig. 16. The major motivation for this study was to take advantage of the high power and low linewidth due to parametric synchronization and then to determine the data rate of the synchronized state.

For modulation of the parametrically synchronized STNO, in addition to dc current, two additional RF signals were superimposed. The first RF signal was injected at twice the frequency of the STNO and serves as the signal for parametric synchronization, while the second RF signal with a frequency range of $f_{mod} = 50$ MHz–500 MHz was used for modulation. The synchronization was found to break below $f_{unlock} \sim 170$ MHz, as shown in Fig. 16a. Figure 16b shows that macrospin simulation performed under similar bias conditions also exhibit similar unlocking behavior. The value of f_{unlock} increases with the increase in amplitude of the modulation cur-

rent, I_{mod}. The phenomena is described as a non-resonant unlocking mechanism. The mechanism responsible behind this phenomena can be explained by an analytic model that shows the phase dynamics of the STNO under the effect of two competing processes—namely, parametric synchronization and modulation. The modulation source changes the synchronized phase ($\Delta\phi_0$), shifting it to a new value of $\Delta\phi_1$. At the maximum modulating amplitude, $\Delta\phi_0$ reaches $\Delta\phi_{max}$, at which the STNO cannot lock anymore as $\Delta\phi_0 > \pi$. A linear relationship thus exists between f_{unlock} and the modulation amplitude. The maximum phase difference can be calculated by [136, 137]:

$$\Delta f_{max} = 2\nu\Gamma_p\Gamma_-(p_0)\frac{I_{mod}}{I_{DC}}\frac{1}{\sqrt[2]{4\Gamma_p^2 + f_{mod}^2}} \tag{18}$$

At the unlocking condition—that is, at $f_{mod} = f_{unlock}$, the experimental data for f_{unlock} versus modulation strength in Fig. 16c can be fitted with the above Eq. (18) by assuming Δf_{max} to be equal to the excitation bandwidth. Thus the unlocking observed in Ref. [136] can be identified as the nonresonant mechanism [138].

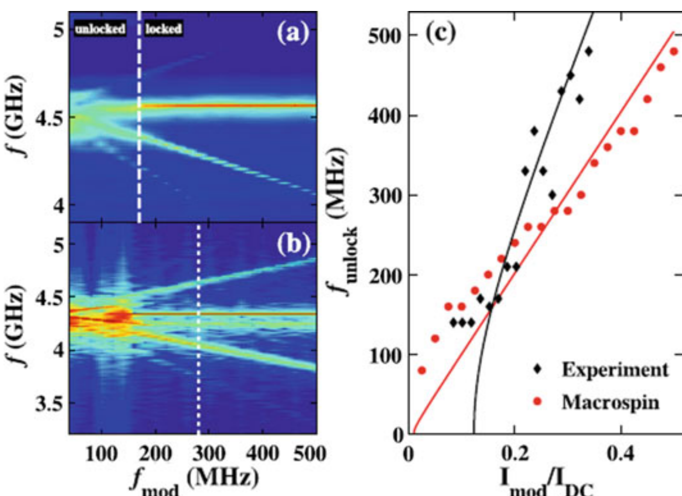


Fig. 16 Modulation of a locked STNO. **a** Experimental and **b** simulated spectra of the synchronized STNO versus f_{mod} at a modulation current of $I_{mod} = 0.99$ mA. **c** The relation between f_{unlock} and I_{mod}/I_{DC} shows an increase with I_{mod} for both experiments and numerical macrospin simulations. The solid lines are fits to the data according to Eq. (18). Reprinted from Dürrenfeld et al. [136], with the permission of AIP Publishing

9 MTJ-Based Microwave Detector

Microwave or radio frequency (RF) detectors can be realized when the dc input is replaced by a microwave current in an MTJ-based STNO. The microwave detection in an MTJ nanopillar is a result of the giant TMR and magnetization dynamics induced by the spin-transfer torque. In the case of a RF current $I_{rf} = I\sin(2\pi f_e t)$ with frequency f_e close to the resonance frequency of the STNO, a large dc voltage is produced by the microwave current mixing with the signal generated by the dynamic response of the STNO in a phenomenon called the spin-torque diode effect [31]. The experimental set-up of the spin-torque diode effect used by Tulapurkar et. al. [31] is shown in Fig. 17a. The rectified voltage generated by the device in response to the alternating current is shown in Fig. 17b. The microwave detector sensitivity of the tunnel junction, which is defined as the ratio of the detected voltage to the input RF power, is important for applications. In the very first work by Tulapurkar et. al. [31] a sensitivity of 1.4 mV/mW was obtained, which is too small for applications. Later, Wang et al. [37] reported higher sensitivity of 54 mV/mW, and predicted even higher values of 10,000 mV/mW.

A report by Gui et al. [139] shows that even in the absence of any external applied magnetic field, an MTJ can produce a significant direct voltage under microwave radiation at frequencies far from the ferromagnetic resonance condition. However, we will not focus on this type of detector; interested reader may refer to the review article by Hemour et al. [140].

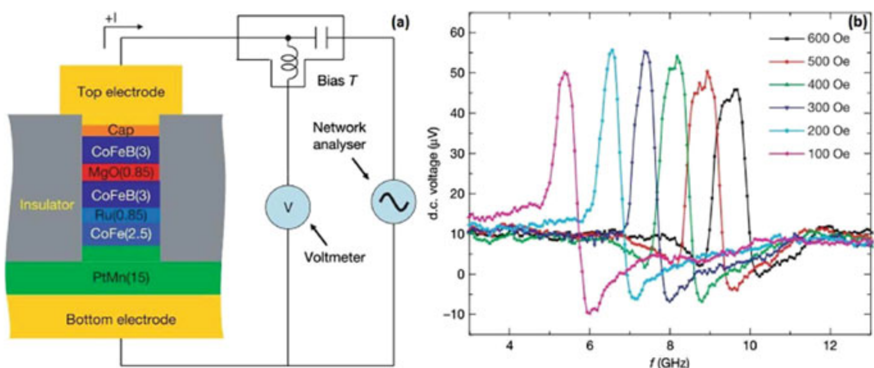


Fig. 17 Experimental demonstration of the spin-torque diode effect by Tulapurkar et al. [31] **a** Schematic diagram of the experimental set-up and cross-sectional view of the magnetic tunnel junction (MTJ) device. The thicknesses of various layers of the device in nanometers are given in brackets. A microwave current and a direct current are applied simultaneously through a bias-tee to an MTJ, which excites the free-layer magnetization and causes resistance oscillations at the driving frequency of the microwave current. **b** The dc voltage generated by the device in response to the alternating current. The dc voltage is plotted as a function of the frequency. Adapted by permission from Macmillan Publishers Ltd: (Nature) Tulapurkar et al. [31], copyright (2005)

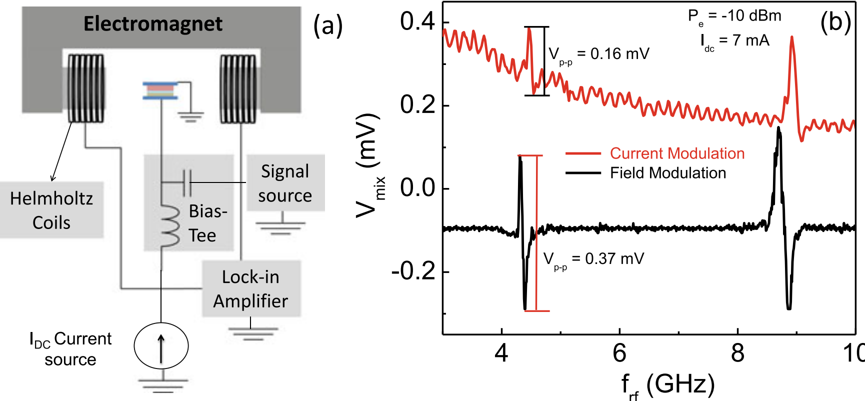


Fig. 18 **a** Schematic of the setup used for the field modulation spin-torque ferromagnetic resonance technique **b** Comparison of current and field-modulated spectra measured via the spin-torque diode effect using spin-torque ferromagnetic resonance at $H_{ext} = 400$ Oe

From an experimental point of view, two modulation techniques—namely, RF current modulation [31–33] and field modulation [133, 141]—are used for the detection of rectified voltage. In the RF current modulation scheme, a modulating microwave current is supplied to the STNO and the rectified voltage is obtained across a lock-in amplifier. The detected resonant spectra may, however, be dominated by frequency-dependent nonmagnetic background oscillations. These oscillations are due to the frequency-dependent variation of I_{rf} in the transmission line, which arise from impedance mismatch. To suppress these oscillations, the field modulation technique is useful [133]. In this technique, the static magnetic field is modulated with a small ac field (~5 to 10 Oe) produced by a pair of Helmholtz coils. These coils are supplied with an ac current at a reference frequency from a lock-in amplifier, which also measures the spin-torque diode voltage. An RF current I_{rf} and a direct current I_{dc} are applied simultaneously through a bias-tee to the STNO, which excites the free layer magnetization and causes resistance oscillations. The schematic for field modulation technique is shown in Fig. 18a. The oscillating frequencies and processional modes in both modulation schemes (current and field) are close, but vary in peak to peak voltages V_{pp} with the applied external magnetic field. The field modulation method offers better signal-to-noise ratio as well as a higher sensitivity (see Fig. 18b).

In the following, we will discuss approaches to improving the sensitivity of the MTJ based microwave detector.

9.1 *Approaches to Enhance the Sensitivity of Microwave Detection*

9.1.1 Enhancement of Sensitivity Using Parametric Synchronization

Parametric synchronization, as discussed in Sect. 7, can be used to efficiently enhance the signal sensitivity of an MTJ-based radio frequency STNO [133]. The synchronization increases the signal sensitivity at $2f_0$, where f_0 is the eigen frequency of the free-running STNO. This happens when the frequency of the external microwave signal (f_e) is close to twice the eigen frequency f_0 of the STNO [133]. Figure 19a shows the spin-torque diode voltage signal obtained at an injected frequency of f_e $\sim$ 9 GHz, which is twice the resonance frequency of the STNO, $f_0 \sim$ 4.5 GHz. The signal at $\sim$ 9 GHz increases with dc current I_{dc}, and exceeds the amplitude of the signal at $\sim$4.5 GHz for a current of 7 mA. The threshold current I_{th} for auto-oscillation is around 6.4 mA [40, 72, 96]. Hence, the sensitivity at $2f_0$ becomes higher than at f_0 only above the threshold of the auto-oscillations. Figure 19b, c shows the macrospin simulated behavior of the STNO below and above the threshold, respectively. Parametric synchronization happens above the threshold, as shown in Fig. 19c, and this enhances the microwave sensitivity. The advantage of this approach is that the frequency of detection is twice that of the eigen frequency of the STNO, and hence the technique also offers higher operation frequencies.

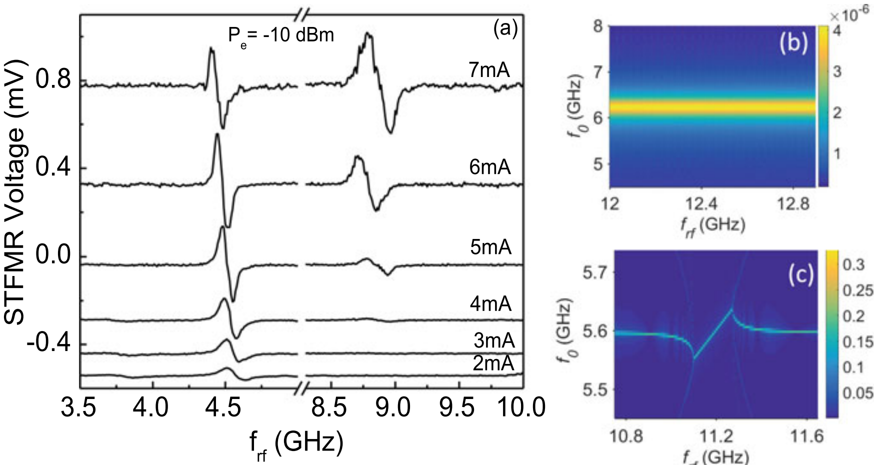


Fig. 19 Enhancement of sensitivity using parametric synchronization. **a** FM-STFMR spectra measured at varying dc biases at $H_{ext} = 400$ Oe and an injected RF power of $P_e = -10$ dBm. **b** and **c** Macrospin simulated spectra of the STNO as a function of f_{rf} at $H_{ext} = 400$ Oe for $I_{dc} = 2$ mA and 7 mA, respectively

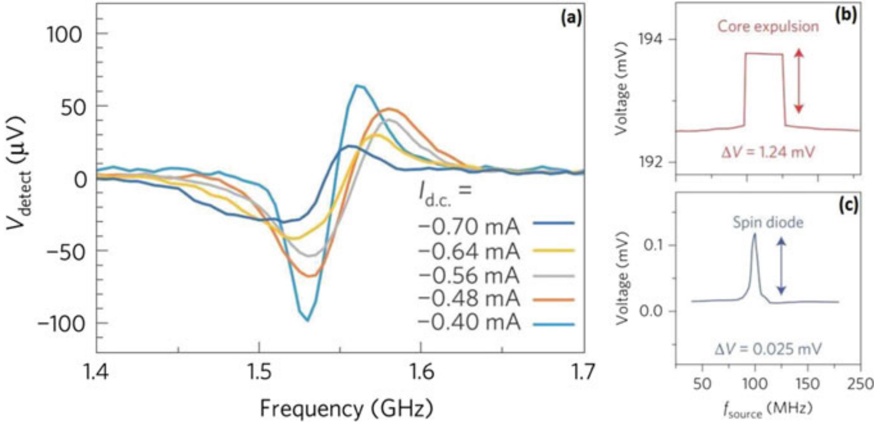


Fig. 20 **a** RF detection voltage (V_{detect}) as a function of the RF input frequency under various dc bias currents (I_{dc}). The radio frequency detection sensitivity of 12,000 mV/mW is achievable at room temperature. Adapted by permission from Macmillan Publishers Ltd: (Nature Materials) Miwa et al. [35], copyright (2013). **b**, **c** shows the rectified voltage observed for the core expulsion and spin-torque diode measurements taken at 6 and 0 mA, respectively, for I_{rf} = 0.2 mA. The maximum effective sensitivity observed is as large as 80,000 mV/mW [143]. Adapted by permission from Macmillan Publishers Ltd: (Nature Nanotechnology) Jekins et al. [144], copyright (2016)

9.1.2 Enhancement of Sensitivity Using Nonlinear FMR

Miwa et al. [35] showed a much greater microwave detection sensitivity of 12,000 mV/mW at room temperature, which exceeds that of the commercial semiconductor diode detectors (3800 mV/mW). The results of Miwa et al. [35] are shown in Fig. 20a. They showed that sensitivity increases significantly when dc bias currents are applied to an MTJ. This high sensitivity has been explained on the basis of nonlinear FMR [142]. In the nonlinear FMR process, the center of the precession orbit is energy-dependent because of the higher-order nonlinear terms. This causes a change in the relative angle between the free-layer magnetization and the pinned-layer magnetization and, consequently, a change in the dc resistance, leading to higher sensitivity. However, the frequency of operation was limited to 1–2 GHz.

9.1.3 Enhancement of Sensitivity Using Vortex Expulsion

Jenkins et al. [144] reported a significant enhancement in the sensitivity of a vortex-based MTJ-based STNOs using the phenomenon of resonant expulsion of the vortex core. This method is shown in Fig. 20b; it produces a large, sharp change in the resistance, associated with the difference in magnetoresistance between the vortex ground state and the final C-state magnetization configuration. Sensitivities as high as 40,000 mV/mW were reported by Tsunegi et al. [143], which were later improved to 80,000 mV/mW (Fig. 20c). A proof of concept for simultaneous detection of

multiple RF signals was also shown using these detectors [145]. The advantage of this technique is that the generated voltage signal is independent of the input RF power. However, the technique does not work for low RF power, and a threshold RF power is needed for the detection. The frequency of operation (<500 MHz) of the detector is also significantly lower than that of standard in-plane magnetized MTJ-based detectors.

9.1.4 Enhancement of Sensitivity Using Voltage Controlled Magnetic Anisotropy

MTJs with large perpendicular magnetic anisotropy (PMA) often show a change in the anisotropy value when a bias (or electric field) is applied across the device [146, 147]. This electric field induces certain interfacial magnetoelectric effects that affect the availability of electronic states for different spin orientations [148]. The resulting imbalance leads to a change in out-of-plane anisotropy. This effect, called voltage-controlled magnetic anisotropy (VCMA), has been shown to have uses in several applications, such as magnetization switching [12, 149–152] and phase noise reduction [153]. VCMA is also shown to excite voltage induced FMR dynamics at radio frequencies (GHz) [34, 36, 154–156].

In 2012, Nozaki et al. [154] reported the possibility to excite ferromagnetic resonance by using VCMA in FeCo monolayers. Zhu et al. [34] later reported that the inclusion of VCMA in exciting FMR dynamics leads to an extra antisymmetric term in the STFMR spectra, which increases the RF sensitivity. Figure 21a shows the calculated STFMR spectra of an MTJ including only VCMA for two different field directions. The calculation is done using macrospin simulations by neglecting the

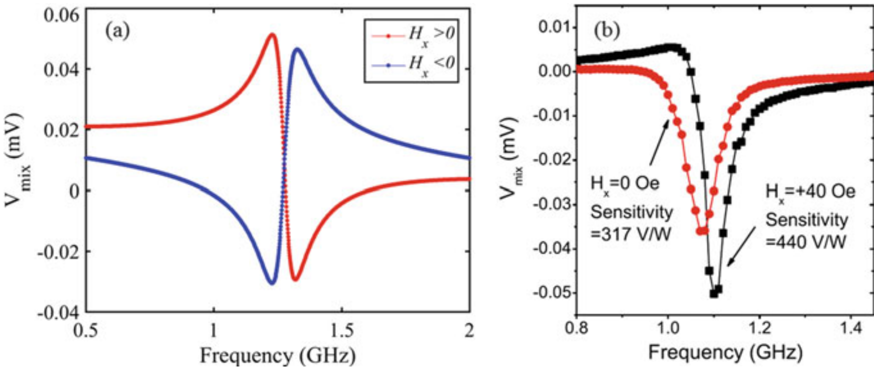


Fig. 21 **a** Calculated STFMR spectra generated by resistance oscillations due to the effect of VCMA. The opposite signs of the antisymmetric component at the 0° ($H_x > 0$) and 180° ($H_x < 0$) field directions (w.r.t. the fixed layer polarization) are a signature of VCMA. **b** Experimental result showing an increase in sensitivity of $\sim 39\%$ with the inclusion of VCMA effect. Reprinted from Zhu et al. [34] Copyright (2012) by the American Physical Society

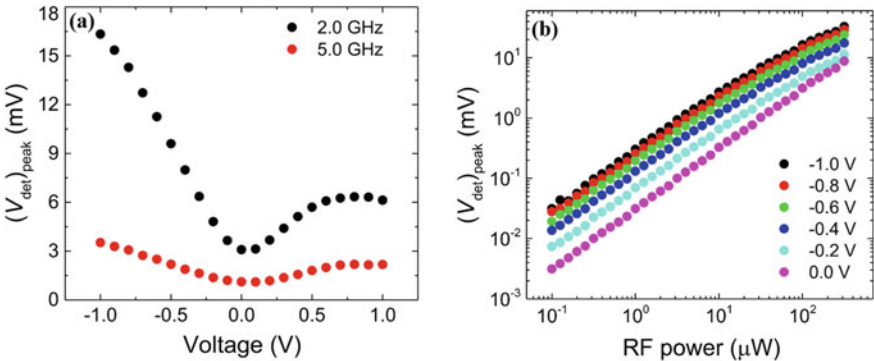


Fig. 22 **a** Bias dependence of peak-to-peak voltage in FMR spectra at 2 GHz and 5 GHz. An increase in the detected homodyne voltage can be seen at both polarities of the applied bias. **b** Shows a linear increase in the peak-to-peak homodyne voltage with increasing RF power under different bias conditions. A linear variation is maintained even for very high power values, which is not possible using spin-torque effects where the free layer is destabilized at high currents. Reprinted from Shiota et al. [157], with the permission of AIP Publishing

spin torque (both in-plane and field-like torque) in the LLGS equation, but including VCMA. The effects of STT on the spectra are neglected here, assuming a very high resistance area (RA) product, due to which only a small quantity of current flows through the device. According to Zhu et al. [34], a sensitivity of 440 mV/mW was achieved with a 40 Oe external field using VCMA, which is 39% more than the value obtained without using VCMA at zero field (Fig. 21b).

A voltage-induced FMR study was conducted by Shiota et al. [157] to analyze the effect of bias and input power on the RF sensitivity. This study highlighted the advantages of using the voltage-induced FMR technique, as opposed to the standard STFMR technique which is based on STT. The study reports an improvement in the peak-to-peak voltage with applied bias in both polarities (Fig. 22a). Applying higher RF power also increases the magnitude of the detected peak-to-peak voltage (Fig. 22b). The increase in peak-to-peak homodyne detected voltage was found to be linear, even at very high power values (300 μW). This result is in stark contrast to those of studies based on the STFMR technique, where the linear increase is seen only at low powers since, at higher input powers, the current destabilizes the free layer. Moreover, more than 30 mV peak-to-peak output voltage was obtained using this method—ten times the value reported by Miwa et al. [35] using the STFMR technique. The highest input current that could be injected was also 3 times higher. However, it should be noted that the maximum sensitivity achieved was 300 mV/mW, which is much less than the value obtained by Miwa et al. (12000 mV/mW) [35]. However, recent studies have [158, 159] shown that further optimization of sensitivity is possible by tuning the external field angle which changes the contribution of STT (symmetric) and VCMA (symmetric as well as anti-symmetric) terms in the STFMR spectrum.

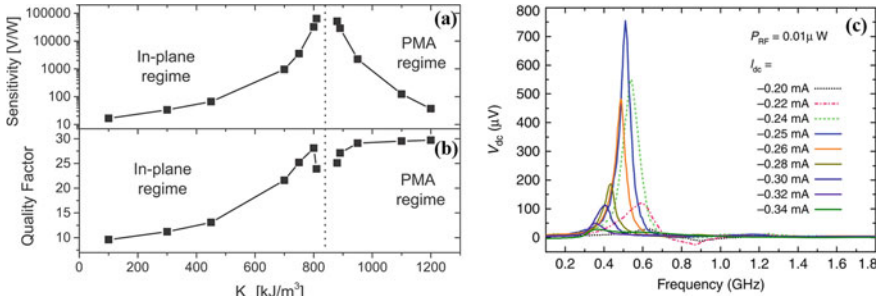


Fig. 23 **a** and **b** show the dependence of sensitivity and quality factor, respectively, on the uniaxial anisotropy of the free layer. Sensitivity varies by several orders of magnitude near the region where the in-plane to out-of-plane transition takes place (dotted line). This happens due to the large precession angle of the magnetic spins. The quality factor deteriorates in this region owing to an increase in effective damping and exchange energy. Reprinted from Frankowski et al. [160], with permission from Elsevier. **c** shows the enhancement in sensitivity using injection locking. The voltage signal generated as a function of the microwave frequency is shown under various dc bias currents (I_{dc}). Adapted by permission from Macmillan Publishers Ltd: (Nature Communication) Fang et al. [36], copyright (2016)

Frankowski et al. [160], in a study of VCMA-based MTJs, showed that sensitivity on the order of tens of thousands of mV/mW is attainable. Using micromagnetic simulations, the dependence of the sensitivity on the uniaxial anisotropy constant was studied. They showed that, near the region where the transition from in-plane configuration to out-of-plane configuration takes place, the sensitivity changes by several orders of magnitude (Fig. 23a). The reason for this behavior was attributed to the decrease in effective anisotropy near the transition region, which leads to large-angle magnetic precession. Resistance oscillations, which depend on the relative angle between the free and fixed layer, also become very large in this case. Such large resistance oscillations lead to an increase in RF sensitivity. Figure 23b shows the variation in the quality factor with the anisotropy energy. The quality factor is defined as the ratio of the eigenmode frequency (f) and the linewidth (f_0). A decrease in quality factor was seen near the transition region. This was explained by the increase in effective damping and the high exchange energy in the transition regime. In an another experiment on microwave detection using VCMA, Fang et al. [36] also managed to achieve a very large sensitivity of 75,400 mV/mW at zero applied field by exploiting the phenomena of injection locking (Fig. 23c). In this case, the driving AC signal locks to the resonant eigenmode of the device and produces a large rectified voltage. Note that the locking happens only over a narrow range of injected dc bias. Using the same method of injection locking, Zhang et al. [161] later improved these sensitivity values to over 2×10^5 mV/mW by optimizing various parameters in MTJ device fabrication.

It is also possible to achieve parametric synchronization as discussed in Sect. 9.1.1 in perpendicular MTJs having VCMA. Sisodia et al. [162] showed that by applying an in-plane external field and optimizing the DC bias, an enhancement in sensitivity can be achieved similar to injection locking. Figure 24a shows the STFMR spectra

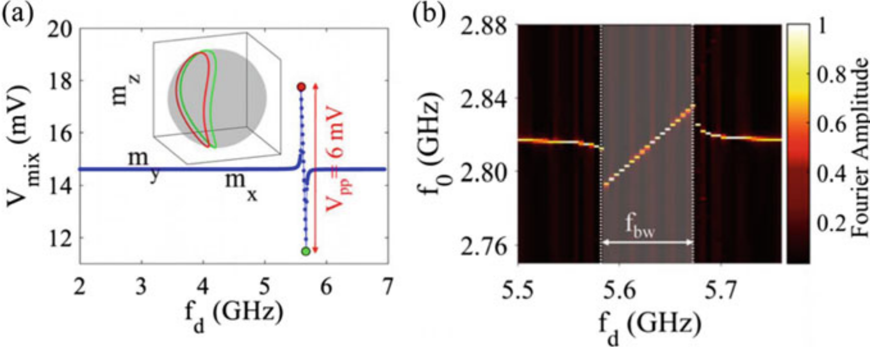


Fig. 24 **a** Simulated STFMR spectra accounting for the VCMA effect for an in-plane external field, $H = -1000$ Oe, $I_{dc} = 0.2$ mA and $P_{rf} = -42$ dBm. The obtained sensitivity is of the order of 10^5 mV/mW. The inset shows the clamshell trajectory at the two extreme points in the STFMR spectra. **b** shows the excited frequency of the MTJ as a function of the driving RF signal clearly showing parametric synchronization in the shaded region with a locking bandwidth, f_{bw}. Reprinted from Sisodia et al. [162], with permission from AIP Publishing

for a perpendicular MTJ with VCMA at an external in-plane field of $H = -1000$ Oe. The applied dc bias of $I_{dc} = 0.2$ mA is above the threshold dc bias ($I_{th} = 0.13$ mA) for auto-oscillations which is a necessary requirement for this method. The injected power is -42 dBm. Figure 24b shows the parametric synchronization, where the excited frequency is exactly half of the driving frequency in a certain bandwidth (f_{bw}). Using parametric synchronization, along with an increase in sensitivity, the detection frequency is also enhanced since the RF signal being detected is close to $2f_0$. Numerical simulations predict that the sensitivity values in this case will also be of the order of 10^5 mV/mW with comparably much higher detection frequencies (4–6 GHz).

The values of sensitivity and detection frequency obtained by the above discussed methods are summarized in Table 1. These results suggests that VCMA together with parametric synchronization offer best combination of high RF sensitivity as well as high detection frequency.

10 Spin Hall Nano-oscillators

In 1999, Hirsch revisited [164] a different phenomenon, the so-called spin Hall effect (SHE), first described by D'yakonov and Perel' in 1971 [165, 166]. The SHE can produce a pure spin current in a direction perpendicular to a charge current [167, 168], which can, in turn, exert substantial STT on an adjacent magnetic layer. It is thus possible to use the SHE in a nonmagnetic metal to achieve, in principle, all the STT-related functions-such as magnetic switching [169, 170] and driven magnetization precession [171, 172] that previously required a separate magnetic

Table 1 Summary of sensitivity and detection frequency values from literature

Study	Geometry	Technique	Sensitivity (mV/mW)	Frequency (GHz)
Tulapurkar et al. [31]	ip-MTJ	STT	0.4	5–10
Tiwari et al. [133]	ip-MTJ	STT+PS	16	8–9
Jenkins et al. [163]	ip-MTJ	Vortex Expulsion	40,000	$\sim$ 0.1–0.2
Tsunegi et al. [143]	ip-MTJ	Vortex Expulsion	80,000	$\sim$ 0.45
Zhu et al. [34]	p-MTJ	VCMA+STT	440	$\sim$1.1
Shiota et al. [157]	p-MTJ	VCMA+nl-FMR	300	2
Frankowski et al. [160]	p-MTJ	VCMA	$>10^4$	$\sim$1
Miwa et al. [35]	p-MTJ	VCMA+nl-FMR	12,000	1.5–2
Fang et al. [36]	p-MTJ	VCMA+IL	75,400	1–1.4
Zhang et al. [161]	p-MTJ	VCMA+IL	2.1×10^5	$\sim$1
Sisodia et al. [162]	p-MTJ	VCMA+PS	$> 10^5$	4–6

ip-MTJ in-plane magnetic tunnel junction; *p-MTJ* perpendicular magnetic tunnel junction; *STT* spin transfer torque; *PS* parametric synchronization; *VCMA* voltage controlled magnetic anisotropy; *nl-FMR* non-linear ferromagnetic resonance; *IL* injection locking

layer to spin-polarize the charge current. With the recent emergence of the spin Hall effect, a new type of spintronic oscillator, the SHNO, has been developed [172]. This provides a new route for the development of microwave and magnonic devices [173]. SHNOs exhibit several advantages over STNOs, including easier nano-fabrication [174–176], lower threshold current, direct optical access to the magnetodynamically active area [177–179], smaller and voltage tunable radiation losses [180, 181], and suppressed nonlinear damping process [173]. While the linewidth of single SHNOs is of the same order as that of STNOs, and their output power is currently lower than that of STNOs, they can be mutually synchronized in both long chains and large two-dimensional arrays, such that record high signal quality factors of 170,000 can be achieved [182, 183]. SHNOs hence offer an opportunity to implement novel nanoscale microwave sources and emitters for wireless communications, nonlinear frequency modulation, and magnonics applications [184–186]. For further details of SHNO developments, issues, challenges, and application, we refer the reader to a recent review article [15].

11 Summary and Outlook

In summary, MTJ-based STNOs are detectors continue to hold promise for applications, due to their high microwave power output and higher frequency tunability. In

this chapter, we discussed the bias dependence of spin torque, the mechanism governing linewidth broadening, synchronization, and modulation of the MTJ-STNOs. Approaches to enhance microwave sensitivity for microwave detector applications were also discussed. MTJ based STNOs have already reached the power needed for applications. Recently the power of MTJ-based STNOs and the sensitivity of MTJ based detectors have reached the level required for commercial applications. However, significant challenges remain to achieve the high power of MTJ-based STNOs and the high sensitivity of MTJ based detectors to-gather with high frequency tunability.

References

1. J.C. Slonczewski, J. Magn. Magn. Mater. **159**, L1 (1996)
2. L. Berger, Phys. Rev. B **54**, 9353 (1996)
3. M.N. Baibich, J.M. Broto, A. Fert, F.N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, J. Chazelas, Phys. Rev. Lett. **61**(21), 2472 (1988)
4. G. Binasch, P. Grünberg, F. Saurenbach, W. Zinn, Phys. Rev. B **39**(7), 4828 (1989)
5. T. Miyazaki, N. Tezuka, J. Magn. Magn. Mater. **139**, L231 (1995)
6. J.S. Moodera, L. Kinder, T. Wong, R. Meservey, Phys. Rev. Lett. **74**(16), 3273 (1995)
7. B.N. Engel, J. Åkerman, B. Butcher, R.W. Dave, M. Deherrera, M. Durlam, G. Grynkewich, J. Janesky, S.V. Pietambaram, N.D. Rizzo, J.M. Slaughter, K. Smith, J.J. Sun, S. Tehrani, IEEE Trans. Magn. **41**(1), 132 (2005)
8. J. Åkerman, Science **308**, 508 (2005)
9. J.A. Katine, E.E. Fullerton, J. Magn. Magn. Mater. **320**, 1217 (2008)
10. J.V. Kim, Solid State Phys. **63**, 217 (2012)
11. T.J. Silva, W.H. Rippard, J. Magn. Magn. Mater. **320**, 1260 (2008)
12. W.G. Wang, M. Li, S. Hageman, C. Chien, Nat. Mater. **11**(1), 64 (2012)
13. N. Locatelli, V. Cros, J. Grollier, Nature Mater. **13**(1), 11 (2013)
14. A. Slavin, V. Tiberkevich, IEEE Trans. Magn. **45**(4), 1875 (2009)
15. T. Chen, R.K. Dumas, A. Eklund, P.K. Muduli, A. Houshang, A.A. Awad, P. Dürrenfeld, B.G. Malm, A. Rusu, J. Åkerman, Proc. IEEE **104**(10), 1919 (2016)
16. S. Bonetti, V. Tiberkevich, G. Consolo, G. Finocchio, P. Muduli, F. Mancoff, A. Slavin, J. Åkerman, Phys. Rev. Lett. **105**(21), 217204 (2010)
17. S. Bonetti, P. Muduli, F. Mancoff, J. Åkerman, Appl. Phys. Lett. **94**(10), 102507 (2009)
18. W.H. Rippard, M.R. Pufall, S. Kaka, T.J. Silva, S.E. Russek, Phys. Rev. B **70**(10), 100406 (2004)
19. P.K. Muduli, O.G. Heinonen, J. Åkerman, J. Appl. Phys. **110**(7), 076102 (2011)
20. M.R. Pufall, W.H. Rippard, S. Kaka, T.J. Silva, S.E. Russek, Appl. Phys. Lett. **86**(8), 082506 (2005)
21. M. Manfrini, T. Devolder, J.V. Kim, P. Crozat, N. Zerounian, C. Chappert, W. van Roy, L. Lagae, G. Hrkac, T. Schrefl, Appl. Phys. Lett. **95**(19), 192507 (2009)
22. P.K. Muduli, Y. Pogoryelov, S. Bonetti, G. Consolo, F. Mancoff, J. Åkerman, Phys. Rev. B **81**(14), 140408 (2010)
23. P.K. Muduli, Y. Pogoryelov, Y. Zhou, F. Mancoff, J. Åkerman, Integr. Ferroelectr. **125**(1), 147 (2011)
24. P.K. Muduli, Y. Pogoryelov, F. Mancoff, J. Åkerman, IEEE Trans. Magn. **47**, 1575 (2011)
25. Y. Pogoryelov, P.K. Muduli, S. Bonetti, E. Iacocca, F. Mancoff, J. Åkerman, Appl. Phys. Lett. **98**(19), 192501 (2011)
26. Y. Pogoryelov, P.K. Muduli, S. Bonetti, F. Mancoff, J. Åkerman, Appl. Phys. Lett. **98**(19), 192506 (2011)

27. P. Villard, U. Ebels, D. Houssameddine, J. Katine, D. Mauri, B. Delaet, P. Vincent, M.C. Cyrille, B. Viala, J.P. Michel, J. Prouvee, F. Badets, IEEE J. Solid-State Circuits **45**(1), 214 (2010)
28. S. Tsunegi, K. Yakushiji, A. Fukushima, S. Yuasa, H. Kubota, Appl. Phys. Lett. **109**(25), 252402 (2016)
29. M. Hoefer, M. Ablowitz, B. Ilan, M. Pufall, T. Silva, Phys. Rev. Lett. **95**(26), 267206 (2005)
30. R. Cheng, D. Xiao, A. Brataas, Phys. Rev. Lett. **116**(20), 207603 (2016)
31. A.A. Tulapurkar, Y. Suzuki, A. Fukushima, H. Kubota, H. Maehara, K. Tsunekawa, D.D. Djayaprawira, N. Watanabe, S. Yuasa, Nature **438**, 339 (2005)
32. J.C. Sankey, Y.T. Cui, J.Z. Sun, J.C. Slonczewski, R.A. Buhrman, D.C. Ralph, Nat. Phys. **4**(1), 67 (2008)
33. H. Kubota, A. Fukushima, K. Yakushiji, T. Nagahama, S. Yuasa, K. Ando, H. Maehara, Y. Nagamine, K. Tsunekawa, D.D. Djayaprawira et al., Nat. Phys. **4**(1), 37 (2008)
34. J. Zhu, J.A. Katine, G.E. Rowlands, Y.J. Chen, Z. Duan, J.G. Alzate, P. Upadhyaya, J. Langer, P.K. Amiri, K.L. Wang, I.N. Krivorotov, Phys. Rev. Lett. **108**(19), 197203 (2012)
35. S. Miwa, S. Ishibashi, H. Tomita, T. Nozaki, E. Tamura, K. Ando, N. Mizuochi, T. Saruya, H. Kubota, K. Yakushiji, T. Taniguchi, H. Imamura, A. Fukushima, S. Yuasa, Y. Suzuki, Nat. Mater. **13**(1), 50 (2014)
36. B. Fang, M. Carpentieri, X. Hao, H. Jiang, J.A. Katine, I.N. Krivorotov, B. Ocker, J. Langer, K.L. Wang, B. Zhang et al., Nat. Commun. **7**, (2016)
37. Z.K. Wang, V.L. Zhang, H.S. Lim, S.C. Ng, M.H. Kuok, S. Jain, A.O. Adeyeye, Appl. Phys. Lett. **94**(8), 083112 (2009)
38. X. Cheng, C.T. Boone, J. Zhu, I.N. Krivorotov, Phys. Rev. Lett. **105**(4), 047202 (2010)
39. B. Fang, M. Carpentieri, S. Louis, V. Tiberkevich, A. Slavin, I.N. Krivorotov, R. Tomasello, A. Giordano, H. Jiang, J. Cai, Y. Fan, Z. Zhang, B. Zhang, J.A. Katine, K.L. Wang, P.K. Amiri, G. Finocchio, Z. Zeng, Phys. Rev. Appl. **11**, 014022 (2019)
40. P.K. Muduli, O.G. Heinonen, J. Åkerman, Phys. Rev. B **83**(18), 184410 (2011)
41. A. Helmer, S. Cornelissen, T. Devolder, J.V. Kim, W. Van Roy, L. Lagae, C. Chappert, Phys. Rev. B **81**(9), 094416 (2010)
42. M. Pauselli, G. Carlotti, J. Phys. D Appl. Phys. **48**(41), 415001 (2015)
43. Z. Zeng, K.H. Cheung, H.W. Jiang, I.N. Krivorotov, J.A. Katine, V. Tiberkevich, A. Slavin, Phys. Rev. B **82**(10), 100410 (2010)
44. S. Zhang, P.M. Levy, A. Fert, Phys. Rev. Lett. **88**(23), 236601 (2002)
45. K. Xia, P.J. Kelly, G.E. Bauer, A. Brataas, I. Turek, Phys. Rev. B **65**(22), 220401 (2002)
46. M.A. Zimmler, B. Özyilmaz, W. Chen, A.D. Kent, J.Z. Sun, M.J. Rooks, R.H. Koch, Phys. Rev. B **70**(18), 184438 (2004)
47. S. Urazhdin, N.O. Birge, W.P. Pratt, J. Bass, Phys. Rev. Lett. **91**(14), 146803 (2003)
48. R. Matsumoto, A. Chanthbouala, J. Grollier, V. Cros, A. Fert, K. Nishimura, Y. Nagamine, H. Maehara, K. Tsunekawa, A. Fukushima et al., Appl. Phys. Express **4**(6), 063001 (2011)
49. I. Theodonis, N. Kioussis, A. Kalitsov, M. Chshiev, W.H. Butler, Phys. Rev. Lett. **97**(23), 237205 (2006)
50. Z. Li, S. Zhang, Z. Diao, Y. Ding, X. Tang, D. Apalkov, Z. Yang, K. Kawabata, Y. Huai, Phys. Rev. Lett. **100**(24), 246602 (2008)
51. C. Heiliger, M. Stiles, Phys. Rev. Lett. **100**(18), 186805 (2008)
52. S. Petit, C. Baraduc, C. Thirion, U. Ebels, Y. Liu, M. Li, P. Wang, B. Dieny, Phys. Rev. Lett. **98**(7), 077203 (2007)
53. A.M. Deac, A. Fukushima, H. Kubota, H. Maehara, Y. Suzuki, S. Yuasa, Y. Nagamine, K. Tsunekawa, D.D. Djayaprawira, N. Watanabe, Nat. Phys. **4**, 803 (2008)
54. S. Oh, S. Park, A. Manchon, M. Chshiev, J. Han, H. Lee, J. Lee, K. Nam, Y. Jo, Y. Kong, B. Dieny, K. Lee, Nat. Phys. **5**, 898 (2009)
55. O. Heinonen, S. Stokes, J. Yi, Phys. Rev. Lett. **105**(6), 066602 (2010)
56. A. Manchon, R. Matsumoto, H. Jaffres, J. Grollier, Phys. Rev. B **86**(6), 060404 (2012)
57. H. Kubota, A. Fukushima, K. Yakushiji, T. Nagahama, S. Yuasa, A. Ando, H. Maehara, Y. Nagamine, K. Tsunekawa, D.D. Djayaprawira, N. Watanabe, Y. Suzuki, Nat. Phys. **4**, 37 (2008)

58. J. Xiao, G.E.W. Bauer, A. Brataas, Phys. Rev. B **77**(22), 224419 (2008)
59. D. Tiwari, R. Sharma, O. Heinonen, J. Åkerman, P. Muduli, Appl. Phys. Lett. **112**(2), 022406 (2018)
60. S. Wang, R. Ward, G. Du, X. Han, C. Wang, A. Kohn, Phys. Rev. B **78**(18), 180411 (2008)
61. J. Alejandro, Phys. Rep. **525**(2), 167 (2013)
62. A.N. Slavini, P. Kabos, IEEE Trans. Magn. **41**, 1264 (2005)
63. S.M. Rezende, F.M. de Aguiar, A. Azevedo, Phys. Rev. Lett. **94**(3), 037202 (2005)
64. J.V. Kim, Phys. Rev. B **73**(17), 174412 (2006)
65. V. Tiberkevich, A. Slavin, J.V. Kim, Appl. Phys. Lett. **91**(19), 192506 (2007)
66. V.S. Tiberkevich, A.N. Slavin, J.V. Kim, Phys. Rev. B **78**(9), 092401 (2008)
67. J.V. Kim, Q. Mistral, C. Chappert, V.S. Tiberkevich, A.N. Slavin, Phys. Rev. Lett. **100**(16), 167201 (2008)
68. J.V. Kim, V. Tiberkevich, A.N. Slavin, Phys. Rev. Lett. **100**(1), 017207 (2008)
69. B. Georges, J. Grollier, V. Cros, A. Fert, A. Fukushima, H. Kubota, K. Yakushijin, S. Yuasa, K. Ando, Phys. Rev. B **80**(6), 060404 (2009)
70. R.K. Dumas, E. Iacocca, S. Bonetti, S.R. Sani, S.M. Mohseni, A. Eklund, J. Persson, O. Heinonen, J. Åkerman, Phys. Rev. Lett. **110**, 257202 (2013)
71. A. Houshang, E. Iacocca, P. Dürrenfeld, S. Sani, J. Åkerman, R. Dumas, Nat. Nanotechnol. **11**(3), 280 (2016)
72. P.K. Muduli, O.G. Heinonen, J. Åkerman, Phys. Rev. Lett. **108**(20), 207203 (2012)
73. P.K. Muduli, O.G. Heinonen, J. Åkerman, Phys. Rev. B **86**, 174408 (2012)
74. T. Silva, M. Keller, IEEE Trans. Magn. **46**(9), 3555 (2010)
75. M. Keller, M. Pufall, W. Rippard, T. Silva, Phys. Rev. B **82**(5), 054416 (2010)
76. S. Bonetti, V. Puliafito, G. Consolo, V. Tiberkevich, A. Slavin, J. Åkerman, Phys. Rev. B **85**(17), 174427 (2012)
77. H. Maehara, H. Kubota, Y. Suzuku, T. Seki, K. Nishimura, Y. Nagamine, K. Tsunekawa, A. Fukushima, A.M. Deac, S. Yuasa, Appl. Phys. Express **6**(11), 113005 (2013)
78. H. Maehara, H. Kubota, Y. Suzuki, T. Seki, K. Nishimura, Y. Nagamine, K. Tsunekawa, A. Fukushima, H. Arai, T. Taniguchi, H. Imamura, K. Ando, S. Yuasa, Appl. Phys. Express **7**(2), 023003 (2014)
79. A. Houshang, R. Khymyn, H. Fulara, A. Gangwar, M. Haidar, S.R. Etesami, R. Ferreira, P.P. Freitas, M. Dvornik, R.K. Dumas, J. Åkerman, Nat. Commun. **9**(1), 4374 (2018)
80. A. Slavin, V. Tiberkevich, Phys. Rev. Lett. **95**(23), 237201 (2005)
81. G. Gerhart, E. Bankowski, G.A. Melkov, V.S. Tiberkevich, A.N. Slavin, Phys. Rev. B **76**(2), 024437 (2007)
82. J. Slonczewski, J. Magn. Magn. Mater. **195**(2), 261 (1999)
83. F.B. Mancoff, N.D. Rizzo, B.N. Engel, S. Tehrani, Nature **437**, 393 (2005)
84. S. Kaka, M.R. Pufall, W.H. Rippard, T.J. Silva, S.E. Russek, J.A. Katine, Nature **437**, 389 (2005)
85. S. Sani, J. Persson, S.M. Mohseni, Y. Pogoryelov, P. Muduli, A. Eklund, G. Malm, M. Käll, A. Dmitriev, J. Åkerman, Nat. Commun. **4**, (2013)
86. L. Bianchini, S. Cornelissen, J.V. Kim, T. Devolder, W. van Roy, L. Lagae, C. Chappert, Appl. Phys. Lett. **97**(3), 032502 (2010)
87. V. Tiberkevich, I. Krivorotov, G. Gerhart, A. Slavin, J. Magn. Magn. Mater. **321**, L53 (2009)
88. M. Quinsat, D. Gusakova, J.F. Sierra, J.P. Michel, D. Houssameddine, B. Delaet, M.C. Cyrille, U. Ebels, B. Dieny, L.D. Buda-Prejbeanu, J.A. Katine, D. Mauri, A. Zeltser, M. Prigent, J.C. Nallatamby, R. Sommet, Appl. Phys. Lett. **97**(18), 182507 (2010)
89. M. Quinsat, F. Garcia-Sanchez, A.S. Jenkins, V.S. Tiberkevich, A.N. Slavin, L.D. Buda-Prejbeanu, A. Zeltser, J.A. Katine, B. Dieny, M.C. Cyrille, U. Ebels, Appl. Phys. Lett. **105**(15), 152401 (2014)
90. P. Bortolotti, E. Grimaldi, A. Dussaux, J. Grollier, V. Cros, C. Serpico, K. Yakushiji, A. Fukushima, H. Kubota, R. Matsumoto, S. Yuasa, Phys. Rev. B **88**, 174417 (2013)
91. J. Sierra, M. Quinsat, F. Garcia-Sanchez, U. Ebels, I. Joumard, A. Jenkins, B. Dieny, M. Cyrille, A. Zeltser, J. Katine, Appl. Phys. Lett. **101**(6), 062407 (2012)

92. S.I. Kiselev, J.C. Sankey, I.N. Krivorotov, N.C. Emley, M. Rinkoski, C. Perez, R.A. Buhrman, D.C. Ralph, Phys. Rev. Lett. **93**(3), 036601 (2004)
93. J.C. Sankey, I.N. Krivorotov, S.I. Kiselev, P.M. Braganca, N.C. Emley, R.A. Buhrman, D.C. Ralph, Phys. Rev. B **72**(22), 224427 (2005)
94. I.N. Krivorotov, N.C. Emley, R.A. Buhrman, D.C. Ralph, Phys. Rev. B **77**(5), 054440 (2008)
95. S. Bonetti, V. Puliafito, G. Consolo, V.S. Tiberkevich, A.N. Slavin, J. Åkerman, Phys. Rev. B **85**(17), 174427 (2012)
96. P. Muduli, O. Heinonen, J. Åkerman, Phys. Rev. B **86**(17), 174408 (2012)
97. O. Heinonen, P. Muduli, E. Iacocca, J. Åkerman, IEEE Trans. Magn. **49**(7), 4398 (2013)
98. E. Iacocca, O. Heinonen, P.K. Muduli, J. Åkerman, Phys. Rev. B **89**(5), 054402 (2014)
99. S.S.L. Zhang, Y. Zhou, D. Li, O. Heinonen, J. Magn. Magn. Mater. **414**, 227 (2016)
100. S. Beri, L. Gelens, M. Mestre, G. Van der Sande, G. Verschaffelt, A. Scirè, G. Mezosi, M. Sorel, J. Danckaert, Phys. Rev. Lett. **101**(9), 093903 (2008)
101. G. Van der Sande, L. Gelens, P. Tassin, A. Scirè, J. Danckaert, J. Phys. B At. Mol. Opt. Phys. **41**(9), 095402 (2008)
102. M.W. Keller, A.B. Kos, T.J. Silva, W.H. Rippard, M.R. Pufall, Appl. Phys. Lett. **94**(19), 193105 (2009)
103. A. Eklund, S. Bonetti, S. Sani, S.M. Mohseni, J. Persson, S. Chung, S. Amir Hossein Banuazizi, E. Iacocca, M. Östling, J. Åkerman, Appl. Phys. Lett. **104**(9), 092405 (2014)
104. R. Sharma, P. Dürrenfeld, E. Iacocca, O. Heinonen, J. Åkerman, P.K. Muduli, Appl. Phys. Lett. **105**(13), 132404 (2014)
105. W.H. Rippard, M.R. Pufall, S. Kaka, T.J. Silva, S.E. Russek, J.A. Katine, Phys. Rev. Lett. **95**(6), 067203 (2005)
106. Y. Zhou, J. Persson, J. Åkerman, J. Appl. Phys. **101**(9), 09A510 (2007)
107. B. Georges, J. Grollier, M. Darques, V. Cros, C. Deranlot, B. Marcilhac, G. Faini, A. Fert, Phys. Rev. Lett. **101**(1), 017201 (2008)
108. Y. Zhou, J. Persson, S. Bonetti, J. Åkerman, Appl. Phys. Lett. **92**(9), 092505 (2008)
109. S. Urazhdin, P. Tabor, V. Tiberkevich, A. Slavin, Phys. Rev. Lett. **105**, 104101 (2010)
110. A. Ruotolo, V. Cros, B. Georges, A. Dussaux, J. Grollier, C. Deranlot, R. Guillemet, K. Bouzehouane, S. Fusil, A. Fert, Nat. Nanotechnol. **4**(8), 528 (2009)
111. T. Kendziorczyk, T. Kuhn, Phys. Rev. B **93**(13), 134413 (2016)
112. A. Awad, P. Dürrenfeld, A. Houshang, M. Dvornik, E. Iacocca, R. Dumas, J. Åkerman, Nat. Phys. (2016)
113. S. Urazhdin, V. Tiberkevich, A. Slavin, Phys. Rev. Lett. **105**, 237204 (2010)
114. S.Y. Martin, N. de Mestier, C. Thirion, C. Hoarau, Y. Conraux, C. Baraduc, B. Diény, Phys. Rev. B **84**(14), 144434 (2011)
115. A. Dussaux, A.V. Khvalkovskiy, J. Grollier, V. Cros, A. Fukushima, M. Konoto, H. Kubota, K. Yakushiji, S. Yuasa, K. Ando, A. Fert, Appl. Phys. Lett. **98**(13), 132506 (2011)
116. A. Hamadeh, N. Locatelli, V. Naletov, R. Lebrun, G. De Loubens, J. Grollier, O. Klein, V. Cros, Appl. Phys. Lett. **104**(2), 022408 (2014)
117. P. Dürrenfeld, E. Iacocca, J. Åkerman, P.K. Muduli, Appl. Phys. Lett. **104**(5), 052410 (2014)
118. M. Quinsat, J.F. Sierra, I. Firastrau, V. Tiberkevich, A. Slavin, D. Gusakova, L.D. Buda Prejbeanu, M. Zarudniev, J.P. Michel, U. Ebels, B. Dieny, M.C. Cyrille, J.A. Katine, D. Mauri, A. Zeltser, Appl. Phys. Lett. **98**(18), 182503 (2011)
119. M. d'Aquino, C. Serpico, R. Bonin, G. Bertotti, I.D. Mayergoyz, Phys. Rev. B **82**(6), 064415 (2010)
120. R. Lebrun, A. Jenkins, A. Dussauz, N. Locatell, S. Tsunegi, E. Grimaldi, H. Kubota, P. Bortolotti, K. Yakushiji, J. Grollier, A. Fukushima, S. Yuasa, V. Cros, Phys. Rev. Lett. **115**(11), 017201 (2015)
121. R. Sharma, N. Sisodia, P. Dürrenfeld, J. Åkerman, P.K. Muduli, Phys. Rev. B **96**(2), 024427 (2017)
122. W.F. Brown Jr., J. Appl. Phys. **34**(4), 1319 (1963)
123. J. Xiao, A. Zangwill, M.D. Stiles, Phys. Rev. B **72**(1), 014446 (2005)
124. S.Y. Martin, C. Thirion, C. Hoarau, C. Baraduc, B. Dieny, Phys. Rev. B **88**(2), 024421 (2013)

125. S.Y. Choi, H. S. andKang, S.J. Cho, I.Y. Oh, M. Shin, H. Park, C. Jang, B. Min, C. Min, S.I. Kimm, S.Y. Park, C.S. Park, Sci. Rep. **4**(18), 4586 (2014)
126. R. Sharma, P. Dürrenfeld, M. Ranjbar, R.K. Dumas, J. Åkerman, P.K. Muduli, IEEE Trans. Magn. **51**(11), 1401304 (2015)
127. I.Y. Oh, S.Y. Park, D.H. Kang, C.S. Park, IEEE Microwave Wireless Comp. Lett. **24**(18), 502 (2014)
128. M. Manfrini, T. Devolder, J.V. Kim, P. Crozat, C. Chappert, W. van Roy, L. Lagae, J. Appl. Phys. **109**(8), 083940 (2011)
129. G. Consolo, V. Puliafito, L. Lopez-Diaz, F. Nizzoli, L. Giovannini, G. Valenti, B. Azzerboni, IEEE Trans. Magn. **46**, 3629 (2010)
130. P.K. Muduli, Y. Pogoryelov, G. Consolo, F. Mancoff, J. Åkerman, A.I.P. Conf, Proc. **1347**(1), 318 (2011)
131. R. Sharma, N. Sisodia, E. Iacocca, A.A. Awad, J. Åkerman, P. Muduli, Sci. Rep. **7**(1), 13422 (2017)
132. R. Sharma, N. Sisodia, P. Muduli, *2018 IEEE International Magnetic Conference (INTERMAG)*, pp. 1–5 (2018)
133. D. Tiwari, N. Sisodia, R. Sharma, P. Dürrenfeld, J. Åkerman, P. Muduli, Appl. Phys. Lett. **108**(8), 082402 (2016)
134. R. Carson, John. Method and means for signaling with high-frequency waves (1923). US Patent 1,449,382
135. D.K. Weaver, Proc. IRE **44**(12), 1703 (1956)
136. P. Dürrenfeld, E. Iacocca, J. Åkerman, P.K. Muduli, Appl. Phys. Lett. **105**(25), 252404 (2014)
137. E. Iacocca, J. Åkerman, Phys. Rev. B **85**(18), 184420 (2012)
138. E. Iacocca, J. Åkerman, Phys. Rev. B **87**(21), 214428 (2013)
139. Y. Gui, Y. Xiao, L. Bai, S. Hemour, Y. Zhao, D. Houssameddine, K. Wu, H. Guo, C.M. Hu, Appl. Phys. Lett. **106**(15), 152403 (2015)
140. S. Hemour, K. Wu, Proc. IEEE **102**(11), 1667 (2014)
141. A. Gonçalves, I. Barsukov, Y.J. Chen, L. Yang, J. Katine, I. Krivorotov, Appl. Phys. Lett. **103**(17), 172406 (2013)
142. C. Wang, Y.T. Cui, J.Z. Sun, J.A. Katine, R.A. Buhrman, D.C. Ralph, Phys. Rev. B **79**(22), 224416 (2009)
143. S. Tsunegi, T. Taniguchi, K. Yakushiji, A. Fukushima, S. Yuasa, H. Kubota, Appl. Phys. Express **11**(5), 053001 (2018)
144. A. Jenkins, R. Lebrun, E. Grimaldi, S. Tsunegi, P. Bortolotti, H. Kubota, K. Yakushiji, A. Fukushima, G. de Loubens, O. Klein, et al., Nat. Nanotechnol. (2016)
145. S. Menshawy, A.S. Jenkins, K.J. Merazzo, L. Vila, R. Ferreira, M.C. Cyrille, U. Ebels, P. Bortolotti, J. Kermorvant, V. Cros, AIP Adv. **7**(5), 056608 (2017)
146. T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami et al., Nat. Nanotechnol. **4**(3), 158 (2009)
147. T. Nozaki, Y. Shiota, M. Shiraishi, T. Shinjo, Y. Suzuki, Appl. Phys. Lett. **96**(2), 022506 (2010)
148. C.G. Duan, J.P. Velev, R.F. Sabirianov, Z. Zhu, J. Chu, S.S. Jaswal, E.Y. Tsymbal, Phys. Rev. Lett. **101**(13), 137201 (2008)
149. S. Kanai, M. Yamanouchi, S. Ikeda, Y. Nakatani, F. Matsukura, H. Ohno, Appl. Phys. Lett. **101**(12), 122403 (2012)
150. N.A. Pertsev, Sci. Rep. **3**, 2757 (2013)
151. R. Tomasello, V. Puliafito, B. Azzerboni, G. Finocchio, IEEE Trans. Magn. **50**(7), 1 (2014)
152. T. Nozaki, H. Arai, K. Yakushiji, S. Tamaru, H. Kubota, H. Imamura, A. Fukushima, S. Yuasa, Appl. Phys. Express **7**(7), 073002 (2014)
153. P. Wisniowski, M. Dabek, W. Skowronski, T. Stobiecki, S. Cardoso, P.P. Freitas, Appl. Phys. Lett. **105**(8), 082404 (2014)
154. T. Nozaki, Y. Shiota, S. Miwa, S. Murakami, F. Bonell, S. Ishibashi, H. Kubota, K. Yakushiji, T. Saruya, A. Fukushima et al., Nat. Phys. **8**(6), 491 (2012)

155. S. Kanai, M. Gajek, D. Worledge, F. Matsukura, H. Ohno, Appl. Phys. Lett. **105**(24), 242409 (2014)
156. E. Hirayama, S. Kanai, J. Ohe, H. Sato, F. Matsukura, H. Ohno, Appl. Phys. Lett. **107**(13), 132404 (2015)
157. Y. Shiota, S. Miwa, S. Tamaru, T. Nozaki, H. Kubota, A. Fukushima, Y. Suzuki, S. Yuasa, Appl. Phys. Lett. **105**(19), 192408 (2014)
158. C. Gonzalez-Fuentes, C. Garcia, P. Landeros, R. Gallardo, Phys. Rev. B **96**(17), 174440 (2017)
159. N. Sisodia, P.K. Muduli, J. Magn. Magn. Mater. **515**, 167301 (2020)
160. M. Frankowski, J. Checiński, W. Skowroński, T. Stobiecki, J. Magn. Magn. Mater. **429**, 11 (2017)
161. L. Zhang, B. Fang, J. Cai, M. Carpentieri, V. Puliafito, F. Garescì, P.K. Amiri, G. Finocchio, Z. Zeng, Appl. Phys. Lett. **113**(10), 102401 (2018)
162. N. Sisodia, P.K. Muduli, Appl. Phys. Lett. **115**(10), 102401 (2019)
163. A.S. Jenkins, R. Lebrun, E. Grimaldi, S. Tsunegi, P. Bortolotti, H. Kubota, K. Yakushiji, A. Fukushima, G.d. Loubens, O. Klein, S. Yuasa, V. Cros, Nat. Nanotech. **11**, 360 (2016)
164. J. Hirsch, Phys. Rev. Lett. **83**(15), 1834 (1999)
165. M. D'yakonov, V. Perel', Sov. Phys. JETP Lett. **13**(11), 467 (1971)
166. M.I. Dyakonov, V.I. Perel, Phys. Lett. A **35**(6), 459 (1971)
167. Y.K. Kato, R. Myers, A. Gossard, D. Awschalom, Science **306**(5703), 1910 (2004)
168. J. Wunderlich, B. Kaestner, J. Sinova, T. Jungwirth, Phys. Rev. Lett. **94**(4), 047204 (2005)
169. I.M. Miron, K. Garello, G. Gaudin, P.J. Zermatten, M.V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, P. Gambardella, Nature **476**(7359), 189 (2011)
170. L. Liu, C.F. Pai, Y. Li, H.W. Tseng, D.C. Ralph, R.A. Buhrman, Science **336**(6081), 555 (2012)
171. L. Liu, C.F. Pai, D.C. Ralph, R.A. Buhrman, Phys. Rev. Lett. **109**(18), 186602 (2012)
172. V.E. Demidov, S. Urazhdin, H. Ulrichs, V. Tiberkevich, A. Slavin, D. Baither, G. Schmitz, S.O. Demokritov, Nat. Mater. **11**(12), 1028 (2012)
173. M. Ranjbar, P. Dürrenfeld, M. Haidar, E. Iacocca, M. Balinskiy, T.Q. Le, M. Fazlali, A. Houshang, A.A. Awad, R.K. Dumas, J. Åkerman, I.E.E.E. Magn, Lett. **5**, 3000504 (2014)
174. V.E. Demidov, S. Urazhdin, A. Zholud, A.V. Sadovnikov, S.O. Demokritov, Phys. Rev. Lett. **105**(17), 172410 (2014)
175. P. Durrenfeld, A.A. Awad, A. Houshang, R.K. Dumas, J. Akerman, Nanoscale **9**(3), 1285–1291 (2017)
176. M. Haidar, A.A. Awad, M. Dvornik, R. Khymyn, A. Houshang, J. Akerman, Nat. Commun. **10**(1), 1–6 (2019)
177. H. Mazraati, S. Chung, A. Houshang, M. Dvornik, L. Piazza, F. Qejvanaj, Appl. Phys. Lett. **109**(24), 242402 (2016)
178. M. Zahedinejad, H. Mazraati, H. Fulara, J. Yue, S. Jiang, A.A. Awad, Appl. Phys. Lett. **112**(13), 132404 (2018)
179. A.A. Awad, A. Houshang, M. Zahedinejad, R. Khymyn, J. Akerman, Appl. Phys. Lett. **116**(23), 232401 (2020)
180. H. Fulara, M. Zahedinejad, R. Khymyn, A. Awad, S. Muralidhar, M. Dvornik, Sci. Adv. **5**(9), eaax8467 (2019)
181. H. Fulara, M. Zahedinejad, R. Khymyn, M. Dvornik, S. Fukami, S. Kanai, Nat. Commun. **11**(1), 1–7 (2020)
182. A.A. Awad, P. Durrenfeld, A. Houshang, M. Dvornik, E. Iacocca, R.K. Dumas, Nat. Phys. **13**(3), 292–299 (2017)
183. M. Zahedinejad, A.A. Awad, S. Muralidhar, R. Khymyn, H. Fulara, H. Mazraati, Nat. Nanotechnol. **15**(1), 47–52 (2020)
184. V.E. Demidov, S. Urazhdin, A. Zholud, A.V. Sadovnikov, A.N. Slavin, S.O. Demokritov, Sci. Rep. **5**, 8578 (2015)
185. M. Zahedinejad, A.A. Awad, P. Dürrenfeld, A. Houshang, Y. Yin, P.K. Muduli, J. Åkerman, IEEE Magn. Lett. **8**, 1 (2017)
186. A. Giordano, M. Carpentieri, A. Laudani, G. Gubbiotti, B. Azzerboni, G. Finocchio, Appl. Phys. Lett. **105**(4), 042412 (2014)

Spin Transfer Torque Magnetoresistive Random Access Memory

Wai Cheung Law and Shawn De Wei Wong

Abstract Faced with an ever-increasing consumer demand for electronics and the need to remain cost-competitive, a variety of emerging non-volatile memory (NVM) technologies have been proposed to cater to the performance gaps within the memory hierarchy system. The magnetoresistive random access memory (MRAM), an emerging and promising NVM technology, will be the key focus of this chapter. At the core of MRAM is the magnetic tunnel junction (MTJ), a storage element embedded within the CMOS process, which has undergone several iterations since its conception. The novelty of the MTJ structure and the use of magnetic materials meant that the conventional semiconductor industry have to innovate solutions that can allow integration with CMOS technology to enable manufacturability. In this chapter, we will cover the development history of MTJ, the material stack designing and patterning process.

1 Introduction

Electronics has been an integral part of our daily lives since the age of technology, and will continue to ever increasingly place an important role especially with the advent of Artificial Intelligence, 5G and the Internet of Things. One key aspect of electronics is the memory system, which has been divided into sub-classes to satisfy the different demands, as shown in Fig. 1. The advantages and limitations of each class lead to the categorization based on its function to serve and their corresponding requirements, i.e. fast read/write speed or bit density. The CPU registers sits at the top of a typical computer memory system hierarchy, occupying between 512 Bytes to a few kilo Bytes (kB), which is mainly achieved through high speed static random access memory (SRAM). L0 to L4 caches would be supporting CPU registers in the form of a mixture of SRAM and dynamic random access memory (DRAM) systems. Random access memories (RAM) stored in the motherboard falls into the third level of the hierarchy system, which has a read/write requirement of 10–100 ns. The bulk

W. C. Law (✉) · S. De W. Wong
Nanyang Technological University, Singapore, Singapore
e-mail: wlaw002@e.ntu.edu.sg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_2

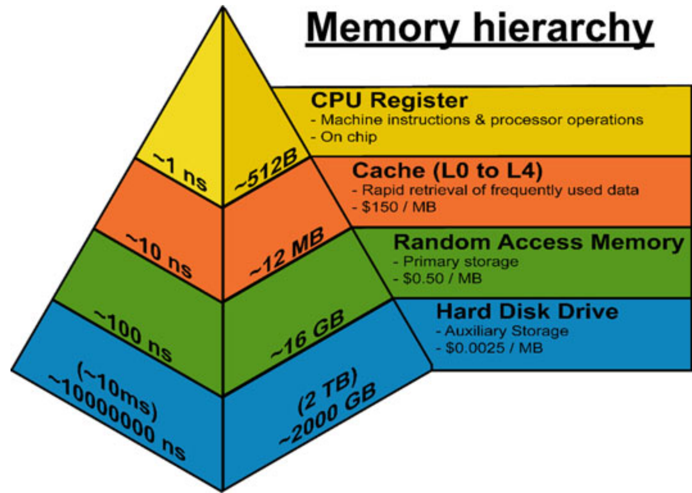


Fig. 1 Memory hierarchy of a computer system

of the memory can be stored in ultra-dense hard disk drives, which allows for Tb/inch of storage space at the expense of access speed. However, the difference in memory systems means that additional peripherals and integration schemes have to be set in place to ensure that all systems can communicate seamlessly [1]. Although such a hierarchy system has been adequate thus far, the need for scaling of electronics continues for greater computational power means that issues such as leakage current and high power consumption can no longer be ignored.

In 2014, the International Technology Roadmap for Semiconductors (ITRS) announced the reorganization into ITRS 2.0 to better address the challenges in the semiconductor industry and the sunset trend predicted by Moore's Law. Under the category of "Beyond CMOS (complementary metal-oxide-semiconductor)", spintronics devices became recognized as one of the emerging technologies to resolve the scaling issues of CMOS. Spintronics, an amalgam of the words "spin" and "electronics", is an active research field that studies the manipulation of both the spin and transport properties of electrons flow to control the localized moments of magnetic materials, and vice versa. One of the key research focuses in spintronics is the development of magnetoresistive random access memory (MRAM), which is considered as a viable option for the future of non-volatile memory system due to its high read/write speed, radiation hardness, scalability, low power consumption and seemingly infinite endurance [1, 2, 3, 4].

The advantages of MRAM are inherent to the ferromagnetic properties of its storage element, known as the magnetic tunnel junction (MTJ). The rudimentary form of an ideal perpendicular MTJ (pMTJ) stack consists of an insulating dielectric sandwiched between two ferromagnetic layers with their preferred magnetization direction pointing normal to the film plane, as illustrated in Fig. 2. The magnetization orientation of the hard layer is fixed in one direction, while the soft layer is able to

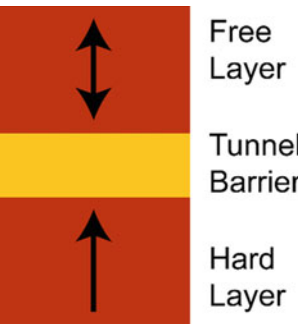


Fig. 2 Illustration of a rudimentary pMTJ stack

toggle between parallel and antiparallel configurations with respect to the hard layer. The hard layer serves as a reference for the soft layer in providing a signal readout, while the soft layer stores the binary information ("0" or "1") with respect to its magnetization direction (arrows "up" or "down"). The magnetization direction of the soft layer should remain unperturbed, until sufficiently large electric current or magnetic field is sent across the pMTJ to reverse its orientation, which is known as a write operation. For a read operation, the orientation of the free layer indicates the resistance state when a small sensing current is passed through the contact electrodes.

Since ferromagnetism is used to represent the binary bit, MRAM is classified as a non-volatile memory as its storage data does not get erased in the event of a power cutoff. An array of MTJs can be embedded within the CMOS back-end-of-line (BEOL) interconnects for specific applications, or used as a standalone memory system.

Emerging memory technologies have to demonstrate compelling advantages to drive the market towards its development and commercialization. Although often envisioned to be the "universal memory" that can potentially replace existing memory technologies [1, 4, 5, 6], MRAM is still considered to be in its infancy stage as it attempts to compete with cheaper and ultra-dense memory systems such as DRAM and NAND Flash. The establishment and maturity of CMOS technology raises the entry level requirement for emerging memory technologies, which will have to demonstrate the ease of its integration into the existing CMOS BEOL processes. A list of device performance for different technologies has been summarized in Table 1. More recently, the roadmaps of several key industry players have outlined the move towards MRAM as a last level cache [6, 7, 8, 9]. Thus, the current state-of-the-art design will have to continue undergoing iterations to improve or circumvent the problems associated with device performance and integration process.

Table 1 Overview of key performance indicators for existing stand-alone memory technologies. Key strengths are shaded in green, while limiting factors are shaded in orange or red [10, 11, 12]

Types / Features	Mature Memory Technologies			Emerging Memory Technologies			
	SRAM	DRAM	Flash	PCM	RRAM	FeRAM	MRAM
Non-volatile	No	No	Yes	Yes	Yes	Yes	Yes
Endurance (# of cycles)	10^{16}	10^{16}	10^{5}	10^{8}	10^{6}	10^{10}-10^{14}	>10^{15}
Read Speed (ns)	1 - 100	30	50	20 – 50	10	20 - 80	2 - 20
Write Speed (ns)	5 - 10	10	10^{4}	75	5 – 20	50	2 - 20
Write Power	Low	Low	Very High	Medium	Low	Low	Low to Medium
Cell Size (F^2)	50 - 120	6 - 10	4	6 – 12	6 – 12	15 - 34	6 - 12
Scalability	Limited	Limited	Yes	Yes	Yes	Under Research	Yes

2 History

The emergence of modern spintronic devices originates from the independent discovery of giant magnetoresistance (GMR) effect by Albert Fert and Peter Grünberg in 1988 [13, 14]. We will cover the magnetoresistance effect in greater detail in Sect. 3.1. For now, one may simply consider that the directions of the magnetization can result in a difference in spin-dependent scattering rate, thereby resulting in a change of resistance states, which is defined as GMR $= \frac{R_{AP}-R_P}{R_P}$, where R_{AP} and R_P refers to the resistance of the GMR device when the two magnetization states are in the antiparallel and parallel states, respectively.

The term GMR was coined due to enhanced effect as compared to anisotropic magnetoresistance (AMR) and has been extensively employed in a various applications, such as magnetic field sensors and hard disk drive read heads in the form of spin valve structures [15, 16, 17, 18, 19]. The spin valve stack is considered as the early concept of MTJ, consisting of two magnetic layers sandwiching a metallic layer (typically copper), in which an electron current can flow either along the electrodes (Current In-Plane/CIP) or perpendicular through the electrodes (Current Perpendicular to Plane/CPP).

Nonetheless, interest within the community was momentarily piqued, as the concept was still plagued with numerous challenges such as achieving observable effects at room temperature as well as low throughput since molecular beam epitaxy is used in sample fabrication. In 1990, Parkin et al. were able to demonstrate large GMR in sputtered multilayers, which greatly ease the manufacturing of GMR-based read sensor [20]. As such, spin valves have begun to take over as read head sensors in hard disk drive applications that require a low resistance-area (RA) product despite

its relative complexity to AMR sensors [16, 17, 18, 21, 22]. These devices have been used since the mid-1990s to detect the transitions between the magnetic domains, in which encode the information stored on hard drives. To enable further scaling of hard disk drives, CPP spin valves are be preferred as the size of a read sensor in a hard drive must be comparable to the bit size. It is easier to design the read/write head with a small read sensor if the flow of current is perpendicular to the spin valves [21, 22]. However, the signal-to-noise ratio (SNR) is still too low to be considered for RAM applications.

The search for a higher SNR led to the discovery of tunnel magnetoresistance (TMR) in Fe/Ge–O/Co-junctions by Michel Jullière in 1975 [23]. The TMR effect is attributed to quantum tunneling through an insulating barrier. The reported change in resistance was around 14% at 4.2 K, similar to the magnitude of GMR and did not attract much attention initially. Other tunnel barriers, such as NiO and Gd_2O_3, had also low TMR values at low temperatures [24, 25]. Eventually, room temperature TMR values of 11.8 to 18% were attained when amorphous AlOx tunnel barrier was used in conjunction with Fe, Co and CoFe electrodes in 1995 [26, 27], up till 70% when CoFeB was used in 2004 [28]. The second major breakthrough in achieving higher SNR occurs when bcc-MgO(100) was used as the tunnel barrier. The highest reported TMR till date is at 604% at room temperature, or 1144% at 5 K [29], and can theoretically reach beyond thousands of percent [30, 31, 32, 33, 34].

Although the major conceptual concerns have been addressed, a multitude of practical designs such as scalability, integration and read/write schemes of MRAM devices have to be considered too. The first generation of MRAM products available on the commercial market have the easy axes of the magnetization of the free and reference layers lying along the plane of film, otherwise known as magnetic materials with in-plane magnetic anisotropy (IMA) [35]. The MTJ devices were patterned into an ellipsoid shape to capitalize on shape anisotropy, inducing the magnetization to preferentially align along the long axis of the ellipsoid, as shown in Fig. 3b. This poses challenges to scalability of the bit density and the risk of potential anomalous switching due to unintended dipole field formation [36, 37]. In addition, the initial conception of MRAM requires write lines to generate a local Oersted field to drive the magnetization along a given orientation. While the MTJ devices can scale with relative ease, these current carrying wires face the challenges in scaling of current density and electromigration effects (~10^7 A/cm^2 for copper). Crosstalk issue may also arise during the scaling of MTJ bits due to the generated Oersted field unintentionally disturbing neighboring MTJ bits.

To overcome the above issue, toggling and heat-assisted switching have been proposed and utilized by several companies [22, 38, 39, 40, 41, 42]. Nonetheless, the added complexity of such switching mechanisms did not increase the overall attractiveness of MRAM. The discovery of current induced switching, termed spin-transfer-torque (STT), revolutionized the design of MRAM. It was discovered that a pure spin current polarized by the reference layer can impart its angular momentum to the soft layer, induce magnetization reversal when the minimum threshold is exceeded. With the read and write currents now passing through the same contact

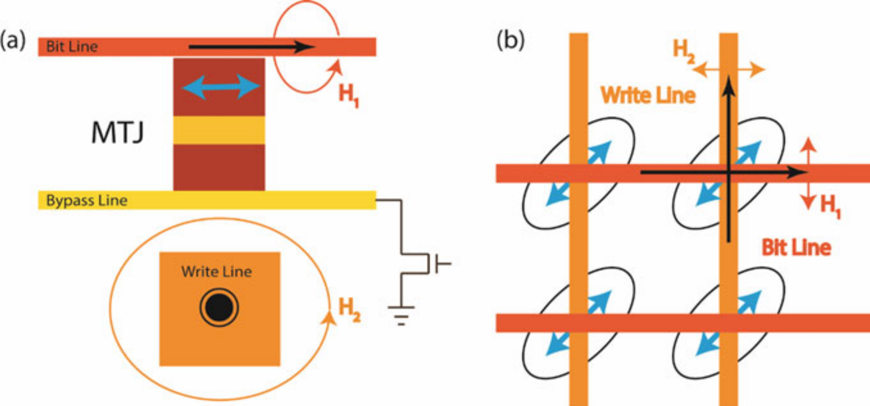


Fig. 3 Schematics of the first generation toggle MRAM embedded within CMOS BEOL, with cyan arrows depicting the easy axis of the soft layer. **a** Cross-sectional view of a single MTJ bit. Black arrows depicts the write currents flowing through the bit and write line at staggered timings, generating magnetic field H_1 and H_2 to induce magnetization reversal of the free layer. **b** Top-down view of the array of elliptical MTJs

electrodes, the need of write lines to create Oersted fields for magnetization reversal is effectively eliminated, hence resolving a significant challenge in scalability.

To improve the thermal stability and scalability while reducing write current density, MTJ designs utilizing materials with perpendicular magnetic anisotropy (PMA) were conceptualized in the 1980s, inspired in part by advances in the hard disk drive industry. We denote such stack designs as pMTJ, in contrast to early concepts of MTJs using IMA (denoted as iMTJ). In addition, pMTJs also have higher efficiencies in spin transfer switching due to the magnetization switching path trajectory.

Current bottlenecks include further downscaling of pMTJ, large writing current density, read and write margins, while ensuring thermal robustness as the integration with the CMOS BEOL processes requires stack stability even at 400 °C. We will discuss this in further details in Sect. 5.

3 Basic Concepts

3.1 Magnetoresistance

Magnetoresistance was first discovered in 1857 by William Thomson [43], where an increase in resistance was observed when a current flows along the magnetization of the ferromagnetic material. Known as anisotropic magnetoresistance (AMR), the effect is a consequence of spin–orbit interaction, where the 3d orbital cloud appears as a larger scattering cross section to electrons traversing in the direction of magneti-

zation. The resultant angular dependence of AMR can be expressed in the following form [44, 45]:

$$\rho(\theta) = \rho_{\perp} + (\rho_{||} - \rho_{\perp})\cos^2\theta, \tag{1}$$

where ρ is the resistivity, θ is the angle between M and the current and the subscripts $||$ and $\perp$ refers to the parallel and perpendicular component respectively. Therefore, the resistance is at its maximum when the electric current is aligned with the magnetization M, which can be controlled using an externally applied magnetic field H_{ext}. Here, the magnetoresistive coefficient $\frac{\rho_{||}-\rho_{\perp}}{\rho} = \frac{\Delta\rho}{\rho}$ is the key figure of merit to evaluate the AMR effect. The relatively large magnetoresistive coefficient and small magneostriction effect for $Ni_{81}Fe_{19}$ leads to a shift from inductive head technology towards MR-based read heads [46, 47, 48]. Note that in the absence of H_{ext}, the average resistivity of a demagnetized sample would be $\rho = \rho_{average} = \frac{1}{3}\rho_{||} + \frac{2}{3}\rho_{\perp}$.

For Giant Magnetoresistance (GMR) effect, we begin by considering the density of states (DOS) of a single ferromagnetic electrode. It is well known from quantum mechanics that the 3d-orbital bands in ferromagnetic materials (e.g. Fe, Co, Ni) are exchange-split, resulting in non-zero magnetization as the two bands are not filled equally at the Fermi energy level. We denote electrons with spin parallel to the overall magnetization as majority carriers, also known as spin-up electrons, while electrons with their spin anti-parallel to the overall magnetization as minority carriers, or spin-down electrons. Due to the heavier effective mass of 3d electrons that are more tightly bounded to the nucleus, electrical conductivity is mainly due to the 4s electrons. Scattering of electrons from 4s-3d states will result in larger resistivity, which will be less common in the majority channel due to the lack of available DOS. The two-current model proposed by Mott can then provide a qualitatively understanding of spin-dependent conduction in ferromagnetic materials, which consider a ferromagnetic material having two independent current channels parallel to each other [49]. Assuming no spin-flip processes, the resistivity ρ for a ferromagnetic measured at a temperature lower than the Curie temperature T_c can be expressed as:

$$\frac{1}{\rho} = \frac{1}{\rho^{\uparrow}} + \frac{1}{\rho^{\downarrow}}, \tag{2}$$

where the superscripts $\uparrow$ and $\downarrow$ refers to the majority and minority carriers, respectively. Due to the difference in resistivity experienced by the majority and minority carriers, a net balance of majority charge carriers will prevail, resulting in a spin polarized current.

If the model is extended to the scenario where an ultra-thin non-magnetic metal is sandwiched between two ferromagnetic electrodes, two resistance states can be obtained depending on the magnetization orientation of the two ferromagnetic electrodes as shown in Fig. 4. Since the electrons will spend half of the time on average within each ferromagnetic electrode, the resistance is split into two components, where r and R refers to the resistance encountered by the itinerant electrons in the

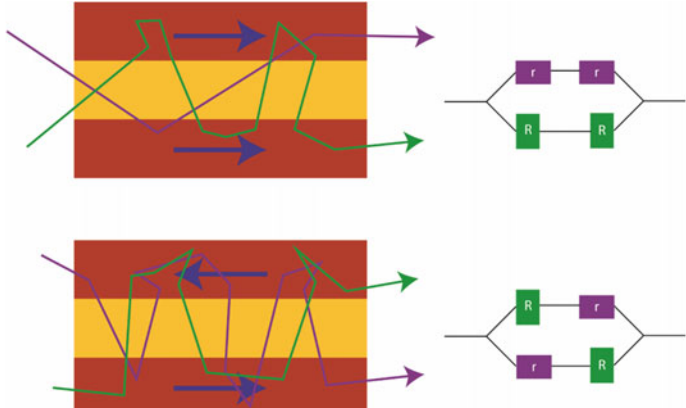


Fig. 4 Illustration of the scattering effect experienced by the majority (purple) and minority (green) charge carriers in a CIP configuration. Figure on the left corresponds to the circuit diagram of the resistance experienced according to the two-channel model

majority and minority spin channels, respectively. When the magnetization of the two ferromagnetic electrodes are parallel to each other, the resistance in the parallel configuration, R_P, would be

$$R_P = \left(\frac{1}{\frac{r}{2} + \frac{r}{2}} + \frac{1}{\frac{R}{2} + \frac{R}{2}} \right)^{-1} = \frac{rR}{r + R}. \tag{3}$$

When the magnetization of the ferromagnetic electrodes are antiparallel, both majority and minority carriers will appear at one point as parallel to one of the ferromagnets and anti-parallel to the other. The corresponding resistance, R_{AP}, would be:

$$R_{AP} = \left(\frac{1}{\frac{R}{2} + \frac{r}{2}} + \frac{1}{\frac{R}{2} + \frac{r}{2}} \right)^{-1} = \frac{r + R}{4}. \tag{4}$$

The GMR ratio can then be defined as:

$$\mathrm{GMR} = \frac{R_{AP} - R_P}{R_{AP}} = \frac{(R - r)^2}{4rR} = \frac{(1 - \alpha)^2}{4\alpha}, \tag{5}$$

where $\alpha = \frac{r}{R}$ is the defined as the scattering asymmetry.

In order to ensure that the spin-independent scattering events do not occur during current flow, the thickness of non-magnetic material is crucial to develop devices with high GMR, which in turn depends on the direction of the current flow. For CIP configuration, the resistance would be determined by the mean free paths for electron

scattering in the ferromagnetic and non-magnetic materials. On the other hand, the resistance in CPP configuration would depend on the spin diffusion lengths in both the ferromagnetic and non-magnetic materials.

The two current model can also be applied to the tunnel magnetoresistance (TMR) effect, wherein the non-magnetic metal is replaced with an insulating barrier. Therefore, the itinerant electrons will have to undergo additional tunneling effect as opposed to instead of ohmic transport in the non-magnetic metallic spacer. Jullière's model was useful to describe amorphous tunnel barriers such as AlO_x [23], where the TMR can be described as:

$$TMR = \frac{R_{AP} - R_P}{R_P} = \frac{G_P - G_{AP}}{G_{AP}} = \frac{2P_1P_2}{1 - P_1P_2}. \tag{6}$$

Here, P_1 and P_2 refers to the spin polarization of the ferromagnetic electrodes, which is further defined as:

$$P = \frac{D^{\uparrow} - D^{\downarrow}}{D^{\uparrow} + D^{\downarrow}}, \tag{7}$$

where $D^{\uparrow}$ and $D^{\downarrow}$ are the DOS of the majority and minority carriers at the Fermi level. Using the two-current model and assuming no spin-flipping occurs during tunneling, the conductance in each channel is dependent on the Fermi's golden rule and is proportional to the tunneling probability. Therefore, when the two ferromagnetic electrodes are parallel, the conductance G_P has the following expression:

$$G_P \propto D_1^{\uparrow} D_2^{\uparrow} + D_1^{\downarrow} D_2^{\downarrow}, \tag{8}$$

where the subscripts refer to the ferromagnetic materials 1 and 2 as depicted in Fig. 5. Likewise, the conductance for the antiparallel configuration G_{AP} can be written as:

$$G_P \propto D_1^{\uparrow} D_2^{\downarrow} + D_1^{\downarrow} D_2^{\uparrow}. \tag{9}$$

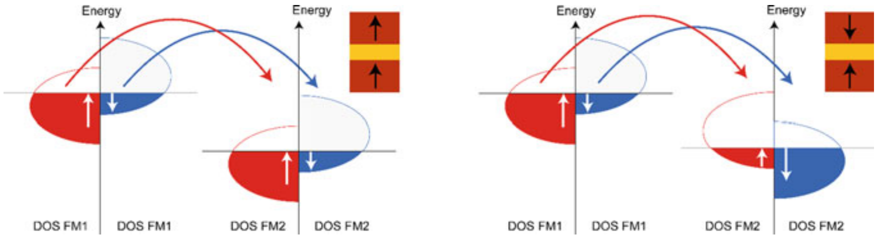


Fig. 5 3d band diagrams illustrating the spin-dependent tunneling process when the MTJ is in the **a** parallel state and **b** anti-parallel state

However, higher TMR ratio observed in bcc-CoFeB/MgO tunnel barrier could not be explained by Jullière's model, as it was observed that CoFeB has lower spin polarization upon annealing [50]. The findings suggest that tunneling of electron wave functions in crystalline MTJ structures do not occur on equal tunneling probabilities as assumed in the simplified picture above, but depends on the symmetry matching between the Bloch states and the evanescent wave functions within the tunnel barrier. For the sake of brevity, the quantum mechanics and tunneling mechanism will not be described in detail, in which readers may refer to reference [51]. Instead, we note that evanescent wave functions decay at different rates within the barrier, wherein Δ_1, Δ_2, $\Delta_{2'}$ and Δ_5 are wave function symmetries compatible with bcc-Fe grown epitaxially on MgO. Similar ab initio calculations can be applied to Co and CoFe electrodes, in good agreement with experimental results. The resultant conductance G_P and G_{AP} will depend on the decay rate in $D^{\uparrow}$ and $D^{\downarrow}$ as the electron tunnels through the ferromagnetic materials 1 and 2 as per Eqs. (8) and (9).

3.2 Spin Transfer Torque

The concept of STT was first predicted by Slonczewski and Berger independently, wherein a spin polarized current can transfer angular momentum to a ferromagnetic layer [52, 53]. Consider the first case as shown in Fig. 6a, where an electron current is spin polarized as it passes through the reference layer of a MTJ. If the polarized spins

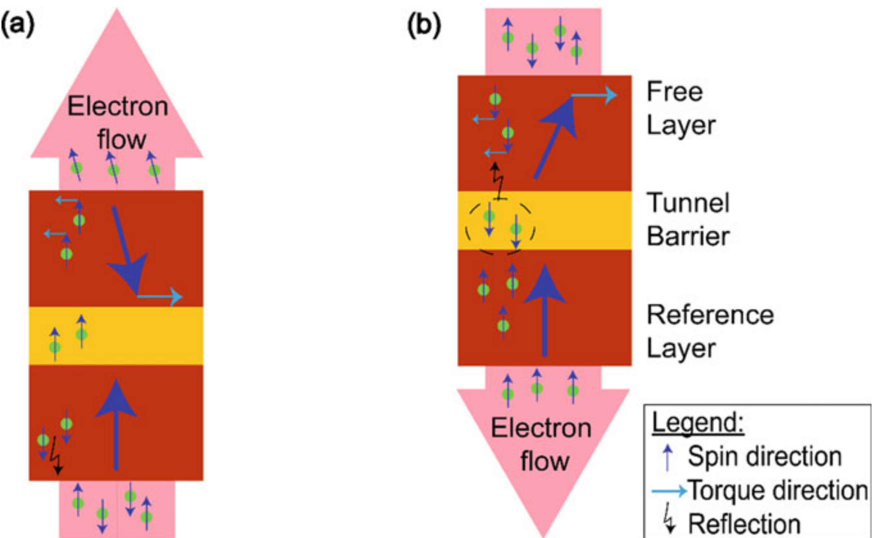


Fig. 6 2-D illustration of STT acting on a MTJ for the case of **a** anti-parallel to parallel state, **b** parallel to anti-parallel state

are not collinear with the free layer magnetization, each itinerant electron will align itself with the localized moment due to exchange interaction. As a consequence, the component of the moment s normal to the free layer magnetization M will be lost. However, the loss in s will be absorbed by the free layer in the form of a torque due to the conservation of angular momentum. With an adequate amount of spin polarized current, either steady state precession or magnetization reversal may be induced.

To switch from parallel to anti-parallel configuration, the electron current would have to be sent from the free layer downwards, as depicted in Fig. 6b. The up-spins, being aligned in the same direction as the reference layer, will transmit through the MTJ with ease, while the down-spins will get reflected by the reference layer. These reflected spins will then exert the torque in the same manner as previously described.

STT is a localized effect occurring at the interface between the tunnel barrier and the free layer. Therefore, by summing up the torque exerted by each spin polarized electron s onto the free layer with magnetization M and volume V, the rate of change of free layer magnetization due to the torque $\tau_{||}$ exerted along the same plane as M, can be expressed as:

$$\left(\frac{dM}{dt}\right)_{\tau_{||}} = \frac{1}{V} P \frac{dN}{dt} \frac{g\mu_B}{2M_s}(M \times (M \times s)) = \tau_{||}(M \times (M \times s)), \tag{10}$$

where P is the spin polarization efficiency, $\frac{dN}{dt} = \frac{I}{e}$ is the rate of unpolarized electron flow per unit time, g is the Landé g-factor, μ_B is the Bohr magneton, and M_s is the saturation magnetization of the free layer for normalization. Equation (10) is also known as the in-plane torque term, which will be useful in describe the magnetization dynamics (see Sect. 3.4). The amount of torque induced on the free layer magnetization is also dependent on the relative alignment between the spin polarized current and the free layer moment (assuming a macrospin approximation).

In addition, an additional torque $\tau_{\perp}$ perpendicular to the M may be induced in the case of a 3D model if the spin moment s has an out-of-plane component. The origin of $\tau_{\perp}$ remains under hot debate, but its effect is similar to an external magnetic field and can be described as:

$$\left(\frac{dM}{dt}\right)_{\tau_{\perp}} = \tau_{\perp}(M \times s). \tag{11}$$

3.3 *Magnetization Energies*

The origin of magnetism arises from quantum mechanism; wherein an electron, in addition to its intrinsic spin angular momentum, has an orbital angular momentum which gives rise to a magnetic moment. These electrons, each generating its own dipole moment (expressed in Bohr magnetons), fill in the lowest possible energy

states of the electron shells in accordance to Hund's rules. In the case where the outermost electronic shells are not to completely fill up, unpaired electrons would occupy the states with the same spin orientation. Such materials are known as paramagnetic, as they weakly respond to external magnetic field, giving rise to a net positive dipole moment until the external magnetic field is removed.

For the case of ferromagnetism, additional exchange interactions that occur between unpaired electrons are included, such as between neighboring atoms (direct exchange) or between orbitals within the same atom (intra-atomic exchange). According to the Pauli exclusion principle, no fermions can occupy the same state, resulting in an energy difference when two spins are aligned parallel or antiparallel with each other. Based on the Heisenberg model for crystalline structures with N atoms, the exchange energy from the sum of Hamiltonian for all S_i and S_j atoms can be expressed as;

$$E_{ex} = -\frac{1}{2}\sum_{i,j}^{N} J(S_i \cdot S_j), \tag{12}$$

where the 1/2 factor is to account for double counting when performing the summation. Ferromagnetism occurs when the sign of the exchange integral J is positive, leading to spontaneous net moment even in the absence of external magnetic field. As the overlapping of the electron wave functions is limited to nearest neighbors, the exchange interaction effect falls off rapidly (i.e. a short range effect) but its magnitude is sizeable (~10^{-2} eV).

Even so, naturally occurring ferromagnets are rarely observed, let alone materials with uniform magnetization. Instead, microscopy techniques such as magnetic force microscopy and magneto-optical Kerr effect microscopy reveal that it is common for ferromagnets to have small magnetic domains separated by domain walls. Since the aforementioned exchange interaction will tend to align all spins together, other longer range interaction forces play the role of further minimizing the energy configuration, stabilizing the energy configuration of the ferromagnet. To help understand the interplay between the magnetostatic interactions, the continuity model is commonly used to represent a distribution of individual spins within a region of interest in ferromagnetic materials as magnetization $M(r, t)$ as a function of time and space. For a ferromagnetic system having an uniform temperature undergoing an isothermal process, one would then able to determine the lowest energy state. In this section, we aim to briefly touch on key magnetostatic terms that influence the design of a functional MTJ stack.

3.3.1 Demagnetizing Energy

If each magnetic moment within the ferromagnet is considered as a magnetic dipole, these "tiny magnets" induces their own magnetic field which interacts with other

magnetic moments. Such magnetostatic interaction can have a long range effect, leading to the formation of domains in an effort to lower the energy cost in contrast to keeping spins aligned parallel as a single domain. As such, it earns its name as the demagnetization term since it competes with the short range exchange interaction. Therefore, at an intermediate range where the exchange interaction is no longer strong enough to hold spins in the parallel configuration, magnetic domains are formed with transitions between domains known as domain walls. Domain walls could be an alternative method to store information along with the STT effect as a writing mechanism to drive domain walls along a nanowire [54, 55, 56].

It is generally difficult to calculate the demagnetization term for a magnet that is either non-uniform or arbitrarily shaped. However, in the case of an ellipsoid, the demagnetization field can be expressed as [57, 58]:

$$H_{dem} = -\overline{\overline{N}}\, M = -\begin{pmatrix} N_x & 0 & 0 \\ 0 & N_y & 0 \\ 0 & 0 & N_z \end{pmatrix} M, \tag{13}$$

where $\overline{\overline{N}}$ is the dimensionless demagnetization tensor of rank 2, with its trace $N_x + N_y + N_z = 1$ for a coordinate system orientated along the principle axes of the ellipsoid. The energy of the demagnetizing energy is simply the integral over the volume of the magnet:

$$E_{dem} = -\frac{\mu_0}{2} \int_V M \cdot H_{dem} dV. \tag{14}$$

The demagnetization term can be utilized to induce shape anisotropy, which is crucial for the development of iMTJ. Shape anisotropy has also been proposed to create pMTJs smaller than 10 nm (see Sect. 5.6). In the case of infinitely extended thin films, the demagnetization factors can be set as $N_x = N_y = 0$, leaving $N_z = 1$.

3.3.2 Anisotropy Energy

In MRAM applications, magnetic anisotropy is an important effect as it results in an easy axis where magnetic moments tend to align towards to for energy minimization. The direction of the magnetic moment and subsequently the bit information is therefore confined to this axis, which can be used to achieve maximum TMR effect. In addition, the anisotropy energy would constitute as the energy barrier required for magnetic reversal, influencing the thermal stability of the bit information.

The total effective anisotropy energy is summed up due to the contribution of the bulk contribution K_v (e.g. magnetocrystalline anisotropy), the surface anisotropy term $\frac{2K_s}{t}$ and the demagnetization energy ($2\pi M_s^2$ for PMA), leads to the following expression [59, 60]:

$$K_{eff} = K_v + \frac{2K_s}{t} - 2\pi M_s^2. \quad (15)$$

If the resultant K_{eff} value is positive, then the easy axis is directed normal to the film plane (i.e. PMA).

Magnetocrystalline anisotropy arises from the interaction between the magnetic moment and the crystal lattice. This coupling is known as spin–orbit interaction, as the total angular momentum is also dependent on the orbital angular momentum of the electron, which is influenced by the crystal structure. Therefore, uniaxial or cubic anisotropy can be induced depending on the crystalline structure [61]. While relatively weak as compared to exchange interactions, it can go up to a few hundred Oersteds, sufficient to induce anisotropy.

Surface anisotropy can be induced due to electronic hybridization or when the symmetry at the interfaces of ultrathin films is broken. This can be significant when the films are ultrathin, as with the case for MTJ design. It is believed that the origin of PMA observed in CoFeB arises due to the surface anisotropy, as its PMA follows a thickness dependence.

3.3.3 Interlayer Exchange Coupling

As mentioned in Sect. 3.3, the origin of ferromagnetism is attributed to the exchange interaction, which can be due to direct exchange and intra-layer exchange coupling mechanisms. However, other forms of exchange coupling mechanisms such as Ruderman–Kittel–Kasuya–Yosida (RKKY), interlayer exchange coupling (IEC) or exchange bias may also give rise to interesting physical phenomena. IEC between Fe/Cr multilayers can be depicted as the polarization of the conduction electrons by a magnetic ion, which polarizes another magnetic ion within the vicinity [62]. The IEC phenomenon is similar to the RKKY exchange coupling effect which predicted such oscillatory behavior, except that its effect is mediated over a spacer layer in contrast to metallic impurities [13, 20, 62, 63].

An alternative explanation is the quantum interference model [64, 65, 66], which is attributed to the confinement of electrons in a quantum well with both ferromagnetic electrodes considered as potential barriers. An electron of wavevector k may undergo multiple reflections as it propagates through the two potential barriers, leading to an interference effect with a phase shift ϕ of an oscillatory nature dependent on the length scale of the quantum well (i.e. thickness of the spacer layer).

One of the main challenges behind the observation of IEC arises from the difficulty in the fabrication of ultrathin multilayer films without pinholes, as the direct exchange interaction between the two ferromagnetic layers through these pinholes will dominate the indirect exchange coupling effect. The oscillatory nature of the IEC is also found to be dependent on various factors, such the thickness of the ferromagnet, defects and/or roughness and the capping layer [52, 66, 67]. The magnetic coupling energy per unit area (erg/cm^2) is expressed as:

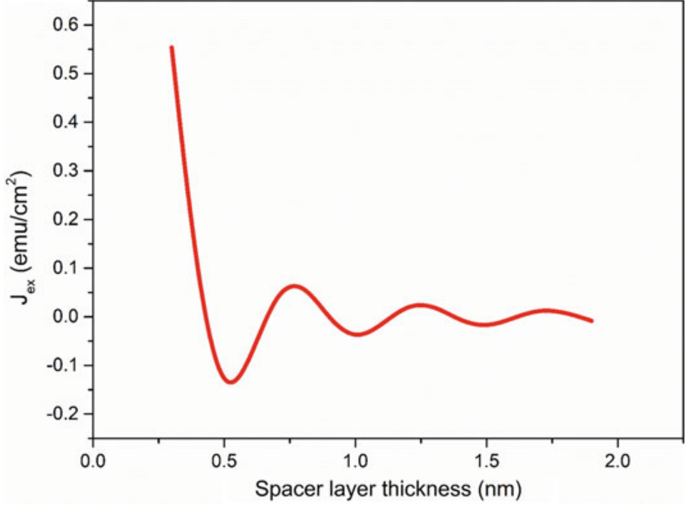


Fig. 7 Illustration of the oscillatory nature of J_{ex} as a function of spacer layer thickness

$$E(\theta) = -J_{ex}\cos(\theta), \tag{16}$$

where J_{ex} is the exchange coupling strength and θ is the angle between the two ferromagnetic layer. J_{ex} can be expressed as a function of spacer layer thickness t with an oscillation period Λ:

$$J_{ex} = \frac{J_0}{t^2}\sin(\frac{2\pi}{\Lambda}t + \phi_1 + \phi_2), \tag{17}$$

where J_0 is the oscillation amplitude (expressed in energy, or erg), t is the thickness of the spacer layer and $\phi_{1,2}$ correspond to the phases of the reflect wave within ferromagnetic electrodes 1 and 2 [66]. J_{ex} decays in a sinusoidal function with respect to the thickness, as shown in Fig. 7. This allows for flexibility to toggle between ferromagnetic and antiferromagnetic configuration, although the coupling strength decreases significantly with thickness. A positive (negative) J_{ex} value will result in the two ferromagnetic layers to preferentially couple in a ferromagnetic (antiferromagnetic) behavior in order to minimize its energy density.

3.3.4 Zeeman Energy

Zeeman energy is the result of interaction between the magnetic moment and an external magnetic field, in which the magnetization would align itself along the direction of the external field in an effort to minimize the energy. The Zeeman energy can be expressed as:

$$E_{Zeeman} = -\mu_0 \int_V M \cdot H_{ext} dV \tag{18}$$

3.4 Magnetization Dynamics

After the introduction on the magnetostatic interactions in the previous section, we will examine the magnetization dynamics which is responsible for the fast read/write response time for MRAM. According to quantum theory, the spin momentum $\boldsymbol{\mu}$ is related to the angular moment $\boldsymbol{L}$ of the electron by the gyromagnetic ratio γ by:

$$\boldsymbol{\mu} = -\gamma \mathbf{L}. \tag{19}$$

The change in angular momentum $\boldsymbol{L}$ can occurs when an externally applied magnetic field H_{ext} exerts a torque on the spin momentum $\boldsymbol{\mu}$ (Zeeman effect), which can be expressed as

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\mu} \times H_{ext}. \tag{20}$$

By substituting Eq. (19) into (20), Eq. (21) is obtained:

$$\frac{d\boldsymbol{\mu}}{dt} = -\gamma \boldsymbol{\mu} \times H_{ext}. \tag{21}$$

For a given magnetization volume under an effective field H_{eff} due to the combination of the external magnetic field, the demagnetizing field and additional anisotropic terms, (22) can be expressed as:

$$\frac{dM}{dt} = -\gamma M \times H_{eff}, \tag{22}$$

which is the basis of the model first proposed by Landau and Lifshitz. Since the magnetization cannot precess indefinitely, an additional phemonological dissipation term λ was added to account for the eventual relaxation of magnetization along the axis of H_{eff}:

$$\frac{\delta M}{\delta t} = -\gamma M \times H_{eff} - \lambda M \times \left(M \times H_{eff}\right). \tag{23}$$

One may observe that $\frac{\delta M}{\delta t} \to \infty$ as $\lambda \to \infty$, yielding unphysical results when the damping factor of ferromagnetic material is large. By modifying the damping term above with one that is dependent on the time-derivative of the magnetization, Gilbert

showed that the magnitude of the magnetization is conserved through the following expression [68]:

$$\frac{\delta M}{\delta t} = -\gamma M \times H_{eff} + \frac{\alpha}{M_s}\left(M \times \frac{\delta M}{\delta t}\right), \tag{24}$$

where α is the phenomenological gilbert damping parameter. Since $\frac{\delta M}{\delta t}$ occurs on both sides of Eq. (24) making it inconvenient to use, an alternative expression can be obtained by vector multiplying both sides by M to give:

$$M \times \frac{\delta M}{\delta t} = -\gamma M \times (M \times H_{eff}) + \frac{\alpha}{M_s} M \times \left(M \times \frac{\delta M}{\delta t}\right). \tag{25}$$

Using the vector identity $a \times (b \times c) = b(a \cdot c) - c(a \cdot b)$, $a \cdot b = |a||b| \cos\theta$ and that $M \cdot \frac{dM}{dt} = 0$ due to the conservation of magnitude of magnetization:

$$M \times \frac{\delta M}{\delta t} = -\gamma M \times (M \times H_{eff}) + 0 - \frac{\alpha}{M_s} M_s^2 \frac{\delta M}{\delta t}. \tag{26}$$

Equation (26) can now be substituted into (24) to give:

$$\frac{\delta M}{\delta t} = -\gamma M \times H_{eff} + \frac{\alpha}{M_s}(-\gamma M \times (M \times H_{eff}) - \alpha M_s \frac{\delta M}{\delta t}), \tag{27}$$

which can be simplified as:

$$\frac{\delta M}{\delta t} = -\frac{\gamma}{1+\alpha^2} M \times H_{eff} - \frac{\gamma\alpha}{(1+\alpha^2)M_s}(M \times (M \times H_{eff})). \tag{28}$$

One may find similarity between the recast form of LLG (28) with the LL Eq. (23) by substituting the following terms

$$\gamma' = \frac{\gamma}{1+\alpha^2} \text{ and } \lambda' = \frac{\gamma\alpha}{(1+\alpha^2)M_s} \text{ to give } \frac{\delta M}{\delta t} = -\gamma M \times H_{eff} - \lambda' M \times \left(M \times H_{eff}\right). \tag{29}$$

Slonczewski and Berger proposed that additional terms are to be included within the LLG model if a spin polarized current leads to the transfer of angular momentum of ferromagnetic layer with magnetization M [52, 53]. Known as the Spin Transfer Torque (STT) as described in Sect. 3.2, the spin current exerts an additional torque that can affect the magnetization dynamics, which can be expressed after the Landau-Lifshitz-Gilbert–Slonczewski (LLGS) equation as:

$$\frac{\delta M}{\delta t} = -\gamma M \times H_{eff} + \alpha M \times \frac{\delta M}{\delta t} + \tau_{||} M \times (M \times s) + \tau_{\perp}(M \times s), \tag{30}$$

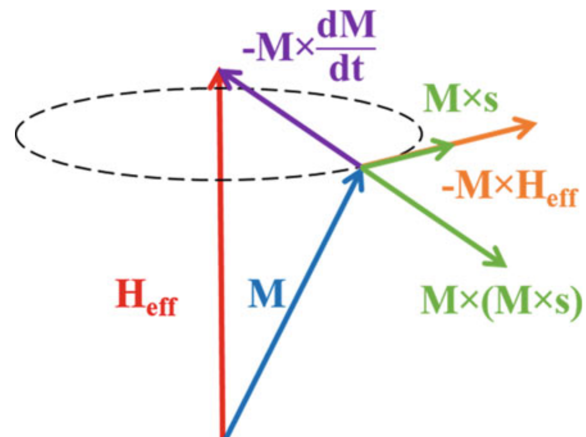


Fig. 8 Torque terms acting upon M due to the influence of an effective field and a spin-polarized current. Green arrows indicate the STT terms due to spin-down current transferring its angular momentum to M

where s refers to the unit directional vector of the spin current, while $\tau_{\parallel}$ and $\tau_{\perp}$ correspond to the in-plane and perpendicular torque terms, respectively. In Fig. 8, the vector representations of the four terms within the right hand side of LLGS equation are illustrated, in which STT terms in green depend on the direction of the spin current (against the damping term in this case). Therefore, a large enough spin polarized current (i.e. the third term greater than the second term) can induce magnetization reversal of the storage layer even in the absence of external magnetic field. This leads to much more compact design in the subsequent generation of MRAM as the current can be injected via the top or bottom electrodes instead of relying on external fields.

A variety of potential spintronics applications have since opened up due to a greater understanding of the magnetization dynamics over the years. For example, by injecting an appropriate level of DC current (which will subsequently be polarized either via STT or SOT), steady-state magnetization precession on the GHz scale can be achieved. As such, a new class of spintronics product known as the Spin-Torque Nano-Oscillator (STNO) can be developed to generate high-frequency microwaves [69, 70, 71]. Alternatively, incident microwaves may also be rectified as the precession of the free layer induces a high and low resistance state every half a cycle, leading to another class of spintronics application known as the spin diode that could potentially be used for energy harvesting technology [72, 73, 74, 75, 76, 77].

4 Performance Indicators in STT-MRAM

While MRAM has been claimed to potentially be the universal memories due to its overall superiority in every aspect required of a memory system, there remains a significant amount of effort needed to address further downscaling for future technology nodes. A pentalemna exists as the requirements of a functional memory

system are closely intertwined with each other [2]. For example, increasing thermal stability will correspondingly require a higher write current density to induce magnetization switching, while raising TMR via increasing the tunnel barrier thickness will also increasing the resistance, the read/write speed and time-dependent dielectric breakdown. As we have covered the basic concepts of the magnetic and electrical transport properties in the previous section, this section will be devoted towards the discussion on the key figure of merits one must consider in a STT-MRAM device.

4.1 Thermal Stability

In the previous sections, we have only considered switching of the free layer due to the STT effect. However, unintended magnetization reversal induced from thermal fluctuation can arise from self-heating (e.g. write operations) or from its immediate vicinity. Therefore, thermal stability Δ is an important key figure of merit particularly for MRAM applications, and is defined for a given temperature T as:

$$\Delta = \frac{E_b}{k_B T}, \tag{31}$$

where E_b is the energy barrier separating between the two binary states (up and down magnetization states of the free layer) and k_B is the Boltzmann constant. It is obtained from the Arrhenius model for a binary state system, in which the bit flip rate due to thermal agitation can be expressed as [42, 78, 79]:

$$N = N_0(1 - e^{-\frac{t}{\tau_0 \exp(\Delta)}}), \tag{32}$$

where N is the number of bits that flipped from an initial population of N_0 bits and τ_0 is the characteristic timescale in which a bit attempts to reverse (~1 ns). Therefore, in order to ensure that the error rate due to thermally induced bit flipping is less than 100 bits per million over a 10 year span set by industrial standards, the corresponding thermal stability would have to be greater than 70. To quantify the thermal stability at device level without having to wait for 10 years, E_b is estimated by measuring the coercivity of the free layer H_c and fitting it the Sharrock's equation [80];

$$H_c(t) = H_{eff}\left\{1 - \left[\frac{1}{\Delta}\ln(\frac{t}{\tau_0 \ln 2})\right]^n\right\}, \tag{33}$$

where n is a number indicating the randomness of the magnetization.

High thermal stability is one of the key strength of MRAM devices, which has seemingly unlimited cycling endurance [81]. In a MTJ device, the free layer E_b is defined as:

$$E_b = K_{eff} V = \frac{M_s H_{eff} V}{2}, \tag{34}$$

where K_{eff} is the effective anisotropy energy and V is the volume of the magnetic layer.

However, a high thermal stability would require an increase in switching current density J_c to induce magnetization reversal within the free layer. The lack of suitable magnetic material with low M_s and α for low J_c, makes it hard to achieve a free layer with high thermal stability. In addition, H_{eff} is observed to degrade at a much faster rate in comparison to M_s in the operating temperature range of STT-MRAM products [82]. A possible but inefficient strategy is to increase H_{eff} instead of M_s to increase Δ as H_{eff} is proportional to H_{eff} (T = 0 K). Furthermore, the thermal stability is proportional to the magnetic volume of the free layer, but will have a direct impact on the scalability of MRAM devices. While the thickness of the free layer can be increased for iMTJ to maintain a high thermal stability, the bulk anisotropy effect will start to dominate CoFeB-based free layer which uses interfacial PMA in pMTJ stack designs. Another potential concern to thermal stability would be the "magnetic dead layer" effect caused by energetic bombardment of atoms during sputtering or interlayer diffusion during annealing [59, 82, 83, 84, 85, 86]. Since the thermal stability is a function of temperature, thermal-assisted switching can be adopted [39, 87, 88, 89, 90]. This device requires the bit to be subjected to an elevated temperature such that the thermal stability will be lowered momentarily for a write current to induce magnetization reversal. Upon cooling, the remanence magnetization will be locked at the new magnetic state to be stored.

4.1.1 SNR (TMR)

As mentioned in Sect. 3.1, TMR is the resistance ratio between the relative orientations of the magnetization direction between the two ferromagnetic layers within the MTJ device. The minimum requirement for TMR is typically 100% in order to achieve clean readout after factoring in error correction code (ECC). ECC is used to correct soft errors arising from software and are not permanent in nature (i.e. read/write disturbance, thermal fluctuations or radiation effects that might perturb the bit [91, 92, 93]. Hard errors are due to device damage arising from sidewall redeposition during etching, pinholes in tunnel barrier during thin film deposition, or dielectric breakdown due to voltage overstress during operation.

TMR is heavily dependent on the crystalline state of MgO tunnel barrier and CoFeB in order to induce coherent tunneling [21, 29, 30, 32, 33, 94, 95, 96, 97, 98, 99]. This means that it is important to tune and optimize the sputtering conditions of MgO, such as the working gas pressure, annealing process and thickness of MgO.

Another motivation for stack optimization is to reduce the Néel coupling effect, which causes dipole formation between the reference layer and the free layer that can lower the TMR ratio. Néel coupling effect, otherwise known as the orange peel

coupling, can be understood as interlayer exchange coupling due to proximity effect that could be enhanced by the roughness (or waviness) between the layers. Assuming a sinusoidal roughness profile, the Néel coupling field H_N can be of the form [100, 101, 102]:

$$H_N = \frac{\pi^2}{\sqrt{2}} \frac{A^2}{\lambda t_F} M_s e^{(-2\pi\sqrt{2}t_B/\lambda)}, \tag{35}$$

where A and λ are the amplitude and wavelength of the roughness profile, respectively, and t_F and t_B are the thickness of the free layer and that of the barrier, respectively.

TMR is also dependent on bias voltage applied and temperature, in which a larger resistance drop can be observed when the two ferromagnetic electrodes are in the antiparallel configuration [99, 103]. This is largely attributed to magnon excitation and defects within the tunnel barrier creating trap states for electrons to co-tunnel, diluting the spin polarization P in the process [104, 105, 106, 107, 108].

At present, research is currently focusing on Heusler alloys exhibiting nearly 100% spin polarization, which is required to achieve high TMR [109]. Interested readers may wish to refer to [110] for a more comprehensive review on Heusler alloys.

4.1.2 Read/Write Current Density

Although MRAM does not require periodic refreshing as compared to DRAM, the large write current density required to perform magnetization reversal is still a significant challenge. The intrinsic write current density, J_{c0}, calculated for a macrospin model at zero temperature follows the equation:

$$J_{c0} = \frac{2\alpha e M_s t_F}{\hbar P} H, \tag{36}$$

where α is the gilbert damping, e is the electron charge, t_F is the thickness of the free layer, $\hbar$ is the reduced Planck's constant and P is the polarization efficiency. H is dependent on the switching trajectory, which acts in-plane and out-of-plane for pMTJ and iMTJ, respectively (see Eq. (10) and Fig. 8). As such, the STT switching component would have to overcome the out-of-plane demagnetizing energy in the case of iMTJ, resulting in the following expression [52, 111]:

$$H = H_{eff} + 2\pi M_s, \tag{37}$$

leading to an increase in J_c in accordance to Eq. (38) without an increase in Δ. On the other hand, H is simply H_{eff} for circular pMTJ nanopillars since the STT switching trajectory is along the same path as the demagnetization term. In reality, P can be

lower when switching from the parallel to anti-parallel configuration [112, 113, 114]. This can be explained by the free layer attempting to polarize the conduction electrons as they travel towards the reference layer, reducing the amount of the minority carriers that will backscatter at the reference layer.

If thermal fluctuation is to be included due to finite temperature, then the write current density J_c can be expressed as:

$$J_c(\tau) = J_{c0}\left[1 - \frac{1}{\Delta}\ln\frac{\tau}{\tau_0}\right], \tag{38}$$

where τ_0 is the characteristic flip attempt time (~1 ns) and τ is the time taken for magnetization reversal. From Eq. (40), one may obtain faster magnetization reversal ($\tau < 1$ ns) when overdriving the current, *i.e.* setting J_c to be greater than J_{c0}. It follows from Eq. (40) that a high thermal stability factor results in a higher current density required to overcome the energy barrier, a conundrum that has yet see a viable solution till date. Therefore, the current figure of merit is the switching efficiency is defined as $\eta_{STT} = \frac{\Delta}{I_{SW}}$.

To achieve pulse width τ of the order of <1 ns, the write current of up to eight times of J_c has been sent to induce magnetic reversal in the precession regime independent of thermally assisted fluctuation process [115]. The switching speed of the free layer in the MTJ is estimated from the macrospin model to be [116, 117]:

$$\tau \propto \frac{\ln(\pi/2\theta)}{J_c - J_{c0}}, \tag{39}$$

where θ is the initial misalignment between the free layer and pinned layer. For pulse width comparable or shorter than the thermal attempt time (1 ns), the STT induced switching process follows the adiabatic precessional model, almost independent of the thermal agitation [118].

In addition, the read and write current density should have a sufficiently wide margins to prevent accidental disturbances during operation. While it is inevitable to have dispersion due to variation in fabrication processes, the margin spreads should not overlap each other as shown in Fig. 9 in order to avoid soft errors. During the read operation, the bit resistance is compared to a reference value halfway between R_{AP} and R_P, which usually corresponds to a biasing voltage of no more than 300 mV.

4.1.3 Resistance Area Product

In what follows from the switching efficiency, one must also consider the breakdown voltage of the insulating barrier deployed in the magnetic tunnel junctions. Both TMR and TDDB are proportional to RA product, which comes at a cost to further downscaling as well as read access speed. In the case of MgO commonly chosen for

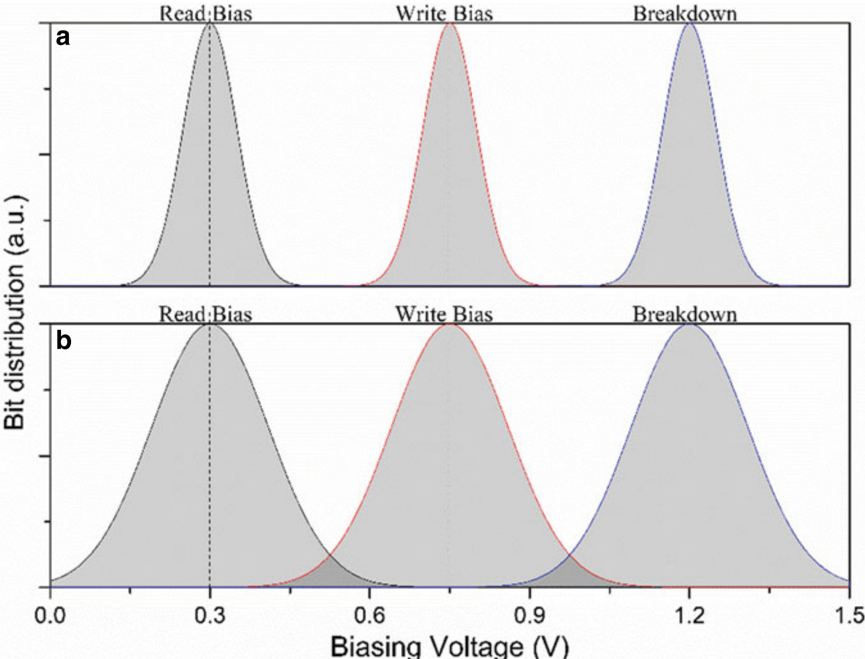


Fig. 9 **a** Ideal case of read and write margins of the MTJ bits. **b** Overlapping margins may lead to soft and hard errors

its enhanced spin polarization ability, the critical switching voltage has a thickness dependence, which should not exceed 1.2 V for 1 nm of MgO.

Tunability of the Resistance-Area (RA) product is highly desirable from the engineering viewpoint as it allows capacitance matching with the CMOS technology node [10, 119]. In a MTJ with double MgO tunnel barriers, the tunnel barrier between the reference layer and the free layer (herein referred to as tunnel barrier 1) is mainly responsible for the TMR ratio and interfacial PMA. The tunnel barrier at the other interface of the free layer is used to enhance interfacial PMA, and therefore its RA product is kept below 1 $\Omega\ \mu m^2$. To keep the RA product between 5 to 10 $\Omega\ \mu m^2$, the thickness of the MgO is typically reduced to 1 nm. This makes the MTJ vulnerable to hard errors, and is imperative to optimize and control the oxidation state of the tunnel barrier.

4.1.4 CMOS BEOL Integration Process

Finally, the realization of MRAM devices must include the integration with the CMOS BEOL processes. The deposition of dielectric passivation via PECVD technique occurs at 400 °C, while Cl-based RIE occurs at 350 °C, which may have a detrimental impact on the magnetic properties of the MTJ. The fabrication of the

MTJ nanopillars has also evolved from various methods, where etchback becomes the norm as it can allow for MTJ features to go below 40x. Due to the hydrophilic nature of MgO, in-situ encapsulation is done after etching to prevent the formation of Mg(OH)2. Since the free layer consisting of CoFeB alloys achieves PMA through interfacial effect and out-diffusion of the interstitial B atoms, a MTJ stack optimized for annealing 400 °C process does not necessarily yield the equivalent or better performance at lower processing temperature.

5 State-of-the-Art MTJ Stack Design

We have presented Fig. 2 as an ideal MTJ configuration earlier during the introduction, which has eliminated a lot of considerations for simplicity. However, the MTJ stack should consist of appropriate materials to address the concerns as mentioned in Sect. 4. As summarized in Fig. 10a, the key considerations in the design of pMTJ stacks revolve not only around material, magnetic and electrical properties, but must also contemplate on how each decision intertwines with the subsequent overlayers. In addition, certain magnetostatic effects are only desirable in a certain scenarios to achieve an objective, often at an expense to other parameters.

For instance, the hard layer has to be extremely stable in its magnetic configuration, but a high stray field may be induced. The stray field may offset the coercivity field of the soft layer via magnetostatic coupling, leading to an additional energy bias required in order to induce magnetization reversal. Biasing is undesirable as it leads to increase in energy consumption, decreased reliability (unintentional switching), decrease in TMR (voltage biasing), and additional stress to the MTJ stack (e.g. time-dependent dielectric breakdown). To resolve this problem, a synthetic anti-ferromagnetic (SAF) structure consisting of two hard layers coupled by ultrathin

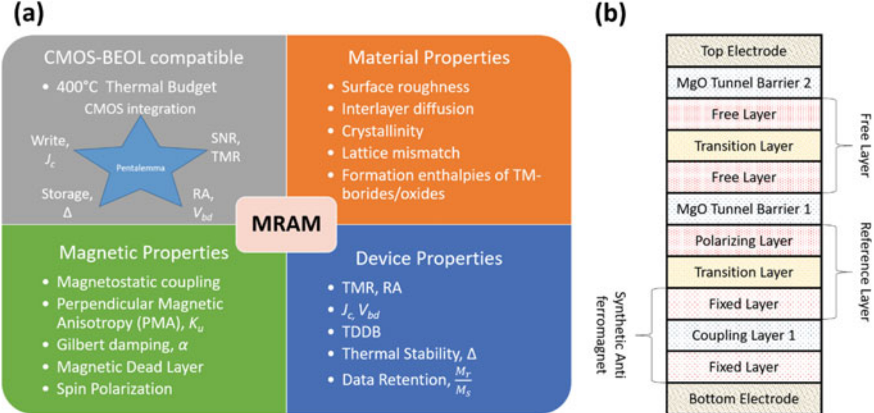


Fig. 10 **a** Summary of stack design considerations. **b** State-of-the-art MTJ stack design

intermediate layer is introduced to eliminate the stray field, but this increases the stack complexity.

Furthermore, the required magnetic properties and thermal robustness of hard layer limits the choice of materials to Co/*X* multilayers, where *X* is typically Pd, Pt or Ni that can provide high PMA through magnetocrystalline anisotropy. In addition to having low spin polarizing efficiencies required to create a spin current for STT switching, such hard layers with fcc(111) crystallinity would not be able induce bcc-MgO(100) growth required for high TMR. An additional ferromagnetic layer (referred to as polarizing layer) with high spin polarizing efficiency can circumvent the problem, but must be coupled to the SAF structure through amorphous texture-breaking layer, therefore adding complexity to the MTJ design.

Given the complexity involved, this chapter is devoted to the functions of each section within the full pMTJ stack design as shown in Fig. 10b. Each subsection will describe the rationale, compromises and limitations behind the current material choices, as well as future outlook and development plans. The discussion is focused on bottom-pinned pMTJ stack, wherein the stack begins with the reference section as the first set of ferromagnetic layers to be deposited.

5.1 Seed Layer

One of the major challenges with the fabrication of MRAM stacks is the tight control on the deposition process as the interfacial effects, crystallinity and surface roughness are critical to the performance of such complex multilayer structures. It is of paramount importance to have an initial ultra-smooth seed layer to provide the crystalline template for the reference layer. Since Co/Pt multilayers are one of the de facto choice as hard magnetic layers due to its high M_s, high H_c, high T_c and high tunability, the seed layer chosen should be in the fcc phase with minimal lattice mismatch.

The seed layer candidates for Co/Pt multilayers are usually Ru and Pt, as they have been demonstrated to promote the necessary crystallographic texture that retains PMA even after 400 °C for at least 30 min [119, 120, 121]. While Ta has also been reported before as a seed layer for Co/Pt multilayers [122, 123], it is also well known to lack the thermal robustness required for pMTJ to be compatible with CMOS BEOL processes [122, 123, 124, 125, 126]. A thick Pt seed layer is desirable to achieve high PMA in Co/Pt multilayers, but is typically reduced for practicality in pMTJ fabrication [127, 128]. Pt, being an unreactive noble metal, can pose a challenge when reactive ion etching is used in device patterning for fast throughput [120, 129, 130]. On the other hand, the redeposition as a result of Ru etching remains conductive even after post-etch surface treatment via oxidation. Although Co/Pt multilayers with high PMA has been achieved with 7 nm Ru seed layer, the risk of device failure due to sidewall redeposition shorting the MTJ pillar increases as devices scale down [120]. Other key considerations for seed layer include electrical conductivity, 400 °C thermal robustness and minimal thickness to allow larger process margin

for downstream processes (*e.g.* chemical mechanical polishing (CMP), etch). CMP can also be performed prior to seed layer deposition to reduce the initial surface roughness to help minimize defects and reduce Neel coupling [119, 131, 132].

5.2 Hard Layer

Although rare earth magnets such as NdFeB and TbCoFe are one of the strongest ferromagnetic materials, it has seen limited success in the earlier designs of MTJs due to its low Curie temperature and high corrosion factor [133, 134]. The typical minimum thickness required for ferrimagnetic rare earth-3d transition metal alloy (RE-TM) alloys to exhibit PMA is >10 nm, increasing the aspect ratio of the patterned MTJ devices [135].

Other candidates such as chemically ordered $L1_0$ CoPt and FePt alloys require at least 500 °C of substrate temperature, falling outside of the temperature window for typical CMOS BEOL processes [136, 137]. Meta-stable $L1_1$ phase CoPt with substrate temperatures between 250 to 400 °C have also been reported [138, 139, 140, 141]. However, this may limit throughput as it takes time for the sample to be heated up and cooled down within a vacuum chamber.

Other alternatives to the hard layer design include the inner SAF, which had a reduced HL2 thickness to allow smoother interface, as well as, enhanced control over the offset field induced to the free layer [142].

5.3 Coupling Layers

To reduce the stray field generated from hard layer 1, an additional thin layer of Ru (0.4–0.9 nm) is inserted between hard layer 1 and hard layer 2 to create a SAF section (see Sect. 3.3.3). To prevent accidental switching of the reference layer section, it is desirable for the induced exchange coupling field H_{ex} to be higher than the coercivity of the free layer. This can mean that the first peak of IEC with strong J_{ex} has to be deployed, thus increasing the process challenge.

The IEC effect is also used in subsequent sections within the pMTJ stack to couple magnetic layers together in a ferromagnetic manner. Examples include Co/Pt/*X*/CoFeB in the reference layer section and CoFeB/*X*/CoFeB free layer section, where *X* is the coupling layer. In these cases, the thicknesses of these coupling layers were tuned to enable ferromagnetic coupling instead. The main aim of such coupling layers would be to break the crystalline texture so as to minimize lattice strain that may affect the growth of bcc (001) CoFeB/MgO. Notable candidates include Ta, W and Mo [124, 125, 126, 143].

5.4 *Polarizing Layer*

In order to induce magnetization reversal within the free layer through electric current, spin polarization must first be enabled by the reference layer. The spin polarization is defined in Eq. (7), which can be quantified by using the superconductivity measurement technique described by Tedrow and Meservey [144, 145].

In addition, spin scattering effect must also be considered if there is non-ferromagnetic metal present as in the case of Co/*X* multilayers, where X = Pt or Pd are common materials as they are able to induce strong MCA due to their high spin–orbit coupling effect. While Co/Pt multilayer section is an excellent candidate for hard layers, it has a major drawback of having poor spin polarization efficiency (~0.46 due to large spin scattering effect at the non-magnetic layers, in contrast to ~ 90% for Co/Ni or 65% for amorphous $Co_{40}Fe_{40}B_{20}$. In addition, Co/Pt multilayers may translate its fcc crystal structure to the MgO tunnel barrier, which must be in the bcc (100) crystal structure in order to allow for enhanced coherent tunneling in the majority Δ_1 state. CoFeB is typically used to solve the above problems simultaneously as it achieves its PMA from the hybridization of iron 3d and oxygen 2p orbitals, amorphousness when deposited which, upon annealing, crystallizes into bcc (100) with small lattice mismatch MgO, as well as high spin polarization efficiency. The CoFeB can be coupled with the Hard Layer 2 through an ultra-thin Ta insertion layer, which is subsequently referred to as the reference layer (see Fig. 10).

5.5 *Tunnel Barrier Layer*

The tunnel barrier is arguably one of the most important layers alongside the free layer, and MgO is typically chosen for high TMR ratio. The stress voltage tolerable for MgO is typically 1.2 V, which is higher than AlOx barriers posing as an attractive advantage. To keep RA low, the thickness of MgO is kept to ~1 nm. This may result in pinhole formation, which can lead to current shunting or reduction of the dielectric endurance during device operation. As such, deposition techniques such as radio frequency sputtering of MgO dielectric or Mg deposition followed by controlled oxidation are often carefully optimized [98, 146].

Another key concern is the Néel coupling effect, which arises due to magnetostatic interactions between the hard and soft layer and may drastically reduce the TMR. Therefore, the underlayers leading up to the tunnel barrier should ideally have minimal surface roughness. To improve on the overall device performance, an optional *in-situ* annealing during deposition can help to obtain bcc-MgO(100) [21, 98]. This is then followed by an ex-situ annealing with magnetic field applied to enhance crystallization of both MgO and CoFeB [97, 147, 148].

In addition, having a second MgO over the CoFeB-based free layer can lead to an improvement in thermal stability as well as symmetrical current switching. This is due to the additional interfacial anisotropy induced on the other side of CoFeB,

although a thin coupling layer has to be included between the two CoFeB as a strain relief from crystallization from both sides of the MgO.

5.6 Free Layer

The free layer is also another key research focus as it is responsible for the storage of the bit information. There are various design concepts pertaining to the free layer to deal with specific issues with minor trade-offs. In general, CoFeB has been commonly used as the free layer in pMTJ due to its high degree of flexibility [95, 149], low Gilbert damping factor ranging from 0.0035 to 0.032 [149, 150, 151, 152, 153, 154], and amorphousness during deposition [50, 155]. CoFeB may crystallize in bcc(100) orientation upon annealing beyond 300 °C, resulting in a good lattice matching with MgO and subsequently high TMR [30, 156]. A significant amount of research has been focused on the annealing dependence of CoFeB and how its crystallization, boron and oxygen absorption by the adjacent layers (e.g. Ta, Mo, W and MgO), and magnetic dead layer effect due to rough interfaces and/or interlayer diffusion [29, 86, 126, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165]. The magnetic dead layer is another concern as it may reduce the effective thickness of the free layer, and therefore the thermal stability.

However, there is a limit in the maximum thermal stability achievable using a single CoFeB layer since its origin of PMA comes from interfacial effect. To increase the thermal stability, free layer with two CoFeB layers coupled by a thin non-magnetic layer (such as Ta, Mo or W) has been reported to achieve almost double the thermal stability while maintaining comparable TMR and J_c [86, 166, 167, 168]. Both sides of the CoFeB layers can be sandwiched by MgO tunnel barriers for maximum interfacial anisotropy effect as well as improving symmetrical switching during write operations from P $\rightarrow$ AP state. The standard seems to be leaning towards double MgO-based pMTJ stack as discussed above, but some novel concepts are also introduced below.

Liu et al. suggested to substitution one of the CoFeB layer within the free layer section with Co/Ni multilayers instead as they bear similarities in some of the key magnetic properties [169]. Co/Ni has a slightly higher gilbert damping value than CoFeB, but has a much higher M_s and PMA generated from magnetocrystalline anisotropy (MCA, see Sect. 3.3) [169, 170, 171, 172, 173, 174]. Since Co/Ni multilayers do not rely on interfacial PMA, the magnetic volume can now scale according to the demands. That having said, strong MCA is observed when Co/Ni multilayers are in the (111) crystalline direction, which means that the coupling layer between Co/Ni and CoFeB must be sufficiently thick or amorphous to prevent the transfer of fcc crystalline texture to the CoFeB free layer. Furthermore, the MCA of Co/Ni is heavily reliant on the contiguous layers to transfer the crystalline texture, which is restricted to only one side of the interface since the other side would be the texture breaking coupling layer. This means that double-MgO will not be deployed in MTJs with such hybrid free layer.

Another novel development is the abandonment of interfacial efforts entirely and instead focused on shape anisotropy effect to induce magnetization in the out-of-plane direction. High TMR and high thermal stability is achieved even with diameters less than 7 nm, which could provide further scaling down to the 1× node and beyond [175, 176]. This would however, lead to very high aspect ratio of the free layer section and subsequently the pMTJ stack, which may lead to additional integration challenges.

6 Experimental Techniques

6.1 Ferromagnetic Resonance Spectroscopy

The concept of ferromagnetic resonance (FMR) was first experimentally observed in the 1946 by Griffiths [177] and subsequently theorized by Kittel [178], similar to other nuclear resonance phenomena such as electron paramagnetic resonance and nuclear magnetic resonance [179]. Under an external applied magnetic field, the magnetic moment that arises from the orbital and spin angular momentum precess at a particular frequency proportional to its gyromagnetic ratio. For the case of ferromagnetic materials, additional intrinsic contributions, such as magneocrystalline anisotropy or demagnetization energy, are to be considered in addition to the Zeeman effect due to the external magnetic field, summing up to form an effective field H_{eff}. A resonance effect occurs when the precessing moment is subjected to an alternating field of the same frequency, typically in the microwave range for ferromagnetic materials as the gyromagnetic ratio is in the order of GHz per Tesla. An absorption spectrum of the Lorentzian form is observed as the microwave is absorbed during the process.

In spin wave resonance, FMR is a unique phenomenon where the spins precess uniformly at the same phase, i.e. the wave vector $k = 0$ [180]. The collection of spins is considered to be a single entity (i.e. macrospin approximation), and therefore can be derived from the Landau-Lifshitz-Gilbert (LLG) equation. The following example draws its solution laid out by Bilzer [181], but examines the case of a thin film with PMA. The conditions set for the material of interest has an OOP easy axis along the z-axis. Assuming that a strong external magnetic field H_{ext} is applied in the z-axis that aligns all magnetic moments to achieve saturation magnetization M_s, the addition of a small alternating field h_{ac} in the x-axis (through an AC microwave source) will perturb the magnetization M, which can be expressed as:

$$M = M_s m \approx m_x\hat{x} + m_y\hat{y} + M_s\hat{z}, \tag{40}$$

where $\hat{x}$, $\hat{y}$ and $\hat{z}$ are unit vectors in Cartesian form. H_{eff} is further defined as

$$H_{eff} = h_{ac}\hat{x} + (H_{ext} - M_{eff})\hat{z} \tag{41}$$

where $M_{eff} = M_s - \frac{2K_\perp}{\mu_0 M_s t}$ is the contraction of the demagnetization and PMA terms which are along the same axis. Therefore, the LLG equation can be expressed as:

$$\frac{dm_x}{dt} = -\gamma\mu_0\big(H_{ext} - M_{eff}\big)m_y - \alpha \times \frac{\delta m_y}{\delta t}. \tag{42}$$

$$\frac{dm_y}{dt} = -\gamma\mu_0\big[-\big(H_{ext} - M_{eff}\big)m_y + M_s h_{ac}\big] + \alpha \times \frac{\delta m_x}{\delta t} \tag{43}$$

$$\frac{dm_z}{dt} \approx 0 = -\gamma\mu_0 h_{ac} m_y - \frac{\alpha}{M_s}\left(m_x \frac{\delta m_y}{\delta t} - m_y \frac{\delta m_x}{\delta t}\right). \tag{44}$$

Since the magnetization components in the x- and y-axis are in precessional motion due to the presence of h_{ac}, the ansatz $m_{x,y} = m_{x,y}e^{i\omega t}$ is used to further linearize the set of equations into the following form:

$$i\omega m_x = -(\omega_H + i\omega\alpha)m_y \tag{45}$$

$$-(\omega_H + i\omega\alpha)m_x + \omega_M h_{ac} + i\omega m_y = 0, \tag{46}$$

where the following set of notations $\omega_H = \gamma\mu_0\big(H_{ext} - M_{eff}\big)$ and $\omega_M = \gamma\mu_0 M_s$ are used for convenience. Equations (45) and (46) can be linearized into the form $m = \chi h$, where χ is the susceptibility in the form of a rank-2 tensor due to presence of magnetic anisotropy. Equation (44) can be dropped as the only non-zero term is an energy term containing the product of m_y and h_{ac}. The set of linear equation from Eqs. (45) and (46) can therefore be expressed into the following matrix form:

$$\begin{pmatrix} \omega_H + i\omega\alpha & -i\omega \\ i\omega & \omega_H + i\omega\alpha \end{pmatrix}\begin{pmatrix} m_x \\ m_y \end{pmatrix} = \begin{pmatrix} \omega_M h_{ac} \\ 0 \end{pmatrix}, \tag{47}$$

which can be rewritten as the following expression;

$$\begin{pmatrix} m_x \\ m_y \end{pmatrix} = \begin{pmatrix} \chi_{xx} & \chi_{xy} \\ \chi_{yx} & \chi_{yy} \end{pmatrix}\begin{pmatrix} h_{ac} \\ 0 \end{pmatrix}. \tag{48}$$

The component of the susceptibility tensor of interest is χ_{xx} as the h_{ac} is applied in the x-direction, and can be expressed as:

$$\chi_{xx} = \frac{m_y}{h_{ac}} = \frac{\omega_M(\omega_H + i\omega\alpha)}{(\omega_H - i\alpha\omega)^2 - \omega^2} = \frac{\omega_M\omega_H\big(\omega_H^2 - \omega^2\big) - i\alpha\omega\omega_M\big(\omega_H^2 + \omega^2\big)}{(\omega_H^2 - \omega^2)^2 + (2\alpha\omega_H\omega)^2}, \tag{49}$$

where all the terms containing $\alpha^2\omega^2$ are excluded, since it leads to zero in the case for $\alpha << 1$ in common ferromagnetic materials (CoFeB ~ 0.014). The imaginary component follows the form of a general Lorentzian function $L(x) = \frac{K}{(\omega-\omega_0)^2+\omega^2(\Delta\omega)^2}$ with a frequency half width $\Delta\omega = 2\alpha\omega_H$. Furthermore, the maximum of the curve occurs when $\omega = \omega_H$, in agreement with resonance condition set by the Kittel formula:

$$f = \frac{\mu_0\gamma}{2\pi}\left(H_{ext} - M_{eff}\right). \quad (50)$$

In the case of a material with in-plane magnetization, applying the same procedure but with $m_x = M_s$ and h_{ac} in the y-axis results in the following equation [181]:

$$f = \frac{\gamma}{2\pi}\sqrt{(H_{ext} + H_k)\left(H_{ext} + 4\pi M_{eff}\right)}, \quad (51)$$

where H_k is the in-plane anisotropy field. For the occurrence of FMR, h_{ac} must be orthogonal to H_{ext}. In addition, h_{ac} would be typically small in order for resonance to occur in the linear regime.

A variety of physical properties can be deduced from FMR spectroscopy, making it one of the most powerful tools available for magnetization characterization. In addition to the quantification of M_{eff} as described in Eqs. (50) and (51), the full-width-half-maximum (FWHM) of the Lorentzian curve can also yield the phonological effective gilbert damping parameter α_{eff} via the relation:

$$\Delta H = \frac{4\pi\alpha_{eff}}{\gamma} f + \Delta H_0, \quad (52)$$

where ΔH_0 is the inhomogeneous linewidth broadening dependent on the film quality affecting the local resonance field. In turn, α_{eff} is related to the spin mixing conductance via the expression:

$$\alpha_{eff} = \alpha_0 + \frac{g\mu_B}{4\pi M_s} g_{\uparrow\downarrow}\frac{1}{t}, \quad (53)$$

where α_0 is the bulk damping constant of the magnetic material with gyromagnetic ratio g and thickness t, μ_B is the Bohr magneton and $g_{\uparrow\downarrow}$ is the spin mixing conductance as a consequence of spin pumping from the ferromagnetic to a non-magnetic layer [182, 183]. Another intrinsic material property that can be deduced from in-plane FMR spectroscopy would be the exchange stiffness A_{ex}, which can be expressed as:

$$f_n^2 = \left(\frac{\gamma\mu_0}{2\pi}\right)^2\left[H_{ext} + M_{eff} + \frac{2A_{ex}}{M_{eff}}\left(\frac{n\pi}{t}\right)^2\right] \times \left[H_{ext} + \frac{2A_{ex}}{M_{eff}}\left(\frac{n\pi}{t}\right)^2\right], \quad (54)$$

where n is the index of the perpendicular standing spin wave [152].

The rudimentary form of a FMR spectrometer would require a microwave source, a microwave detector, an external magnetic field generated by an external magnetic field, and a transmission line or a cavity to allow interaction between the sample to and the microwave. In earlier versions of the experimental setup, a microwave resonant cavity was used to produce a standing wave with a uniform ac field of the microwave source h_{ac} for the sample to be placed within [179, 184]. The cavity is placed under an external magnetic field which is modulated with a Helmholtz coil operating within a few tens of Hz. The modulated signal is picked up by the lock-in amplifier, which measures the in-phase and quadrature voltage. Due to the physical limitation of the length of the cavity to form uniform standing modes, relatively few data points can be extracted.

Recent advancements have led to the usage of vector network analyzer (VNA) along with a transmission line to perform broadband microwave spectroscopy. Due to the quasi-static TEM mode characteristic of transmission lines such as grounded coplanar waveguide (GCPW) and microstrip, h_{ac} encircles the signal trace as shown in Fig. 11b. In order to minimize reflection losses at frequencies above 30 GHz, GCPW design is preferred over microstrip configuration. The curvature of the GCPW is typically large enough to prevent shorting when a thin film MTJ sample is placed on top of the signal trace in a flip-chip configuration. The main difference between the two configurations is having two ground planes and the signal trace s on the same plane for the GCPW design. By including these grounds, the impedance matching can be achieved by tuning the gap distances between the s and signal trace width w. Additionally, through vias, with spacing $d = \frac{\lambda}{8}$, can be connected to a bottom ground plane to improve the transmission behavior. Through this configuration, the frequencies of the microwave source can be varied instead of the external magnetic field, leading to a significant increase in the measurement throughput without sacrificing resolution.

The strength of h_{ac} is typically assumed to be uniform across w and approximated by Ampere's law $h_{ac} = \frac{I}{2w}$ [185]. However, h_{ac} could be smaller due to the microwave attenuation down the transmission line, imperfect current distribution or

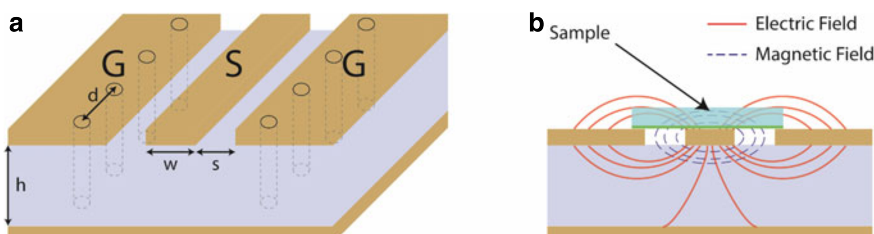


Fig. 11 **a** Schematic of the typical grounded coplanar waveguide (GCPW). **b** The cross-sectional view of the GCPW, with the sample placed on top of the GCPW in a flip-chip configuration (magnetic film shown in green). The magnetic field is assumed to be uniform in an ideal scenario

presence of air gap between the sample and GCPW [186]. Nonetheless, the magnitude of h_{ac} is appropriately small to excite the magnetization within the linear regime [187].

One drawback of such frequency sweeping mode is the challenges associated with achieving matching RF impedance over the entire bandwidth. The design of GCPW or microstrip follows the microwave transmission theory and a dielectric material with high permittivity ε_r is desirable so as to minimize dielectric loss. Impedance matching can be achieved by tuning the gap distances between s and w. In addition, tapering s may also help to tune the inductance and capacitance of the coplanar waveguide, and the introduction of via holes can help to reduce radiative losses. Most importantly, the GCPW should not be plated using electroless nickel/immersion gold (ENIG), a method that is commonly deployed in PCB finishing, due to the presence of the ferromagnetic nickel. Instead, immersion silver is deployed to ensure low insertion loss, even at higher frequencies.

The end launch connectors and coaxial cables should be free of any magnetic impurities to avoid the influence of external magnetic fields [179]. The VNA should also be placed at a sufficient distance from the electromagnet, as the microwave source is a YIG-based oscillator that might be influenced by external magnetic fields. If air-spaced coaxial cables are used, over bending of the cables may lead to irreversible damage. The dielectric within the cable is air for minimal dielectric loss, and the inner conductor is supported by spacers placed at regular intervals to prevent the inner conductor from touching the outer ground shield.

The importance of FMR spectroscopy in the development of MRAM arises from its ability to decouple the anisotropy contributions from the various ferromagnetic sections in the MTJ. Since the effective fields generated from intrinsic contributions are different for the soft layer and hard layers, the ferromagnetic resonance conditions will differ for each layer within the MTJ, as shown in Fig. 12. For example, the material choice for the soft layer (CoFeB) typically has a lower PMA in contrast to the hard layers, and therefore resonates at a lower microwave frequency of ~5 GHz. The free layer is also shown to have resonance points that are symmetric about zero external magnetic field. Due to the antiferromagnetic coupling, the FMR of HL1 and HL2 will switch at the coercivity points of the SAF structure. In contrast, the M-H loop measurements of the hard axis (e.g. VSM, MOKE, SQUID) will be the total contribution from the entire MTJ stack as the all magnetic moments will rotate coherently towards its hard axis. Since crystallinity, interlayer diffusion and surface roughness plays an important role in K_{eff}, truncating out the free layer section is not a good representative.

6.2 Current In-Plane Tunneling (CIPT)

Four-point probe measurements (also known as the Kelvin technique) have been commonly used to accurately monitor the sheet resistance of materials. It has been

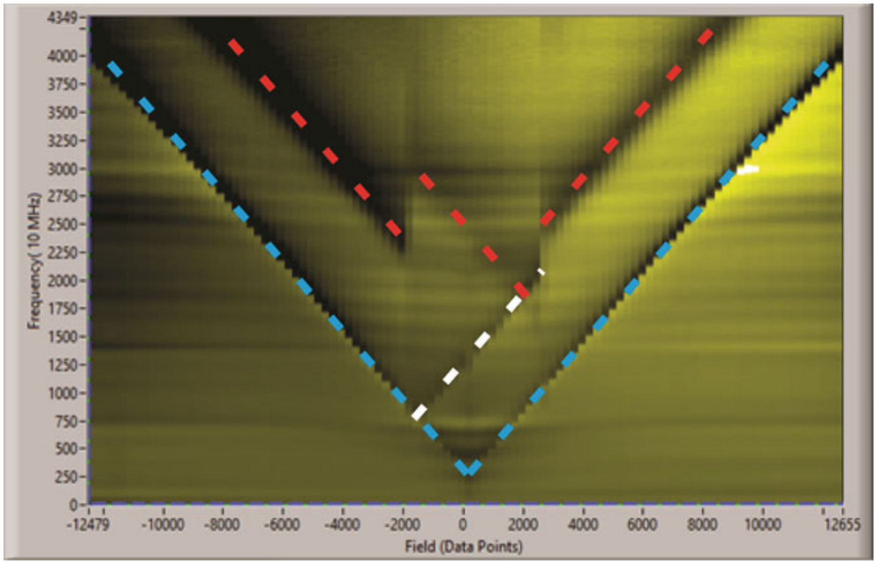


Fig. 12 Frequency-swept FMR spectra of a typical pMTJ stack presented in a contrast plot. The y-axis is the microwave frequencies swept at a step size of 10 MHz, while the x-axis is the external magnetic field at coarser step size. The color contrast indicates the intensity of the transmitted microwave power, where the black corresponds to the absorption of microwave due to FMR. Dashed lines in blue, red and white are guides to the eye which corresponds to the FMR conditions for FL, HL2 and HL1 respectively

commonly deployed as a non-destructive inline metrology tool, allowing for thickness calibration and uniformity check. Since the fabrication of patterned MTJ devices is extremely time and resource intensive (ranging from weeks to months depending on the level of complexity and tool availability), the CIPT technique can be utilized in thin film MTJ stacks to characterize local distribution of the electrical transport properties, or to perform full wafer mapping to check for film uniformity. Another advantage of this thin film characterization is the process variations inherent to device patterning can be excluded, allowing for systematic and efficient troubleshooting. CAPRES, in junction with IBM and Infineon Technologies, have developed a sophisticated CIPT metrology system to simultaneously measure the TMR, RA and sheet resistances of the top electrode R_T and bottom electrodes R_B. In the case MTJ stack analysis by CIPT, the probe spacings would be narrower to ensure tunneling through the MTJ.

Within the probe head, a multiplexer is used to select a combination of 4 out of the 12 collinear probes with ultra-narrow spacing (on the order of a few μms). A small current I is typically sent through the outer probes while the voltage V is measured by the inner probes connected to a high impedance electrometer. To further eliminate error contribution from the contact resistance, a small AC current in conjunction with a lock-in amplifier can be used. In the CAPRES model, a permanent magnet was used to perform the magnetic field sweep which is calibrated prior to measurements.

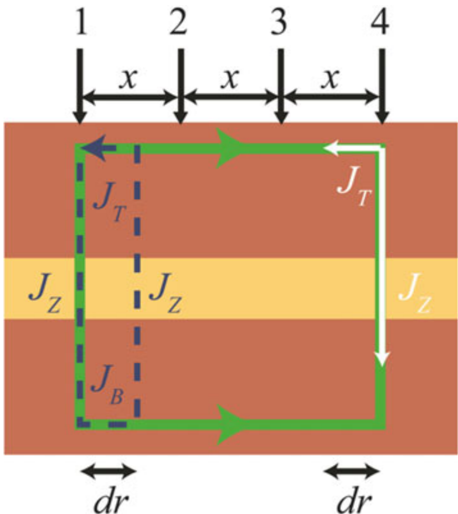


Fig. 13 Schematic of the CIPT model, illustrating in blue the theoretical breakdown of current flowing through probe 1. Modelled after reference [188]. In blue, used to model the Eq. (56) to conserve voltage. In white, used to model Eq. (58) to decompose the components of current flow. In green, used to model Eq. (60) to conserve current

The theory behind sheet resistance R_s measurements using four-point probe will be briefly discussed in this section. It is defined as $R_s = \frac{\rho}{t}$, where ρ is the resistivity of a material and t is the thickness. However, the geometric correction factors, thickness and spacing between the probes should also be represented. The current injected through a probe can be considered to flow radially outwards through the material in the form of a cylinder with radius r and thickness t. Therefore, the current density J is defined as $J = \frac{I}{2\pi rt}$.

Furthermore, the current also induces an electric field $E = \frac{dV}{dr} = -\rho J$ for an ohmic material, in which Ohm's law can be applied to obtain the voltage drop over a concentric radial distance r. From Fig. 13, the potential drop from point 2 to point 3 as a result of a current injected from point 1 can be calculated from the integration of the E over two probe distances, resulting in the following expression:

$$\int_{r_2}^{r_3} Edr = V_3 - V_2 = \frac{\rho}{t}\frac{I}{2\pi}\left(ln\frac{1}{r_3} - ln\frac{1}{r_2}\right) = R_s\frac{I}{2\pi}\left(ln\frac{r_3}{r_2}\right). \quad (55)$$

Figure 14 illustrates a four point probe measurement with equal probe spacing x. The drop in voltage measured between probes 2 and 3 is due to the current injected from probe 4, which can be seen as a negative current flowing in the opposite direction as compared to probe 1. Therefore, the vector sum of these two contributions leads to the overall potential drop $\Delta V = R_s\frac{I}{2\pi}\left(\ln\frac{r_3}{r_2}\right) - R_s\frac{I}{2\pi}\left(\ln\frac{r_2}{r_3}\right)$. In the case of equal probe spacing x or $R_s = \frac{\pi}{\ln 2}\frac{\Delta V}{I}$, the potential drop from the viewpoint of probe 1 can be expressed as $\Delta V = R_s\frac{I}{2\pi}(2\ln 2)$.

The concept of sheet resistance can also be applied to the MTJ model as described by Worledge et al. [188], in which the current passing through the tunnel barrier is assumed to be perpendicular to the infinitely extended film plane (CPP configuration).

In Fig. 13, by the conservation of voltage, the voltage drop across a closed loop with an infinitesimally small loop of radius dr should be zero. Starting from the current flowing from the top electrode through the tunnel barrier to the bottom electrode, then back to itself, the expression as per Kirchhoff's Second Law would be:

$$J_z(r)RA + t_B R_B J_B(r)dr - J_z(r+dr)RA - t_T R_T J_T(r)dr = 0, \tag{56}$$

where the subscripts z, T and B refers to the current flowing through the tunnel barrier in the z direction, the top electrode and the bottom electrode, respectively. Equation (56) can be further reduced by combining the first and third term $dJ_z(r) = J_z(r+dr) - J_z(r)$ and differentiating Eq. (56) by dr to give the following expression:

$$\frac{dJ_z(r)}{dr}RA - t_B R_B J_B(r) + t_T R_T J_T(r) = 0 \tag{57}$$

Furthermore, assuming the current injected into the ultrathin top electrode from a point contact source can only flow out into the tunnel barrier or radially along the top electrode (per cylindrical model), the total current flow can be decomposed into these two contributions to obtain the following expression:

$$2\pi r t_T J_T(r) - 2\pi r dr J_z(r) - 2\pi(r+dr)t_T J_T(r+dr) = 0, \tag{58}$$

which in the limit of $dr \to 0$, leads to:

$$J_z(r) + \frac{dJ_T}{dr}t_T - \frac{J_T}{r}t_T = 0. \tag{59}$$

Finally, a charge current flowing from probe 1 to probe 4 can either flow through both the top and the bottom electrodes of the MTJ, but would still obey the law of current conservation (Kirchhoff's first law) to give the following expression:

$$I - 2\pi r t_T J_T(r) - 2\pi r t_B J_B(r) = 0. \tag{60}$$

Equations (57), (59) and (60) can be used to solve as a set of simultaneous equations. Along with the expression of electric field flowing through the top electrode as $E_T(r) = R_T t_T J_T(r)$, J_T can then be determined by solving the above equations to form:

$$-\frac{RA}{R_T}\frac{d^2E_T(r)}{dr^2} - \frac{RA}{R_T r}\frac{dE_T(r)}{dr} + \left(1 + \frac{RA}{R_T r^2} + \frac{R_B}{R_T}\right)E_T(r) - \frac{IR_B}{2\pi r} = 0. \tag{61}$$

To fit the equation to the modified Bessel function of the first order [188], Eq. (61) is further simplified by the constant expression $\delta = R_T R_B I/2\pi\lambda(R_T + R_B)$, the characteristic length scale $\lambda = \sqrt{RA(R_T + R_B)}$ and its ratio with respect to the interprobe spacing $z = x/\lambda$. Using the same approach seen in the determination of

sheet resistance, the voltage measured can be extracted by integrating E_T over the probe distances.

In order to ensure that the current flows through both the top and bottom electrodes for TMR to be measured, a few considerations have to be made when using the CIPT tool. Firstly, the mean probe spacing distance x has to be chosen to ensure results follow a good fit to the model. A general rule of thumb described by CAPRES states that in order for the current flowing through the outer probes to be distributed amongst the top and bottom electrodes equally, the characteristic length scale $\lambda = \sqrt{\frac{RA}{R_T + R_B}}$ should be between $\lambda \leq x \leq 5\lambda$. Thus, ultra-small probe spacings are required for MTJ thin films which typically have very small RA values.

Secondly, the top and bottom electrodes have to be designed carefully as the R_T/R_B ratio is important. As current prefers to flow in the path of least resistance, the current effectively shunts through the top electrode if R_B is much higher than R_T, resulting in minimal tunneling through the tunnel barrier. This can be interpreted as a suppression of the measured TMR. In most electrode designs, the R_T/R_B ratio is set to be ~18.

Lastly, at least 7 nm of Ru is required as a capping layer in order for the MTJ stack to remain conductive after exposure to native oxidation. Otherwise, the software is programmed to repeatedly drive in the probes, resulting in probe breakage.

6.3 Sputtering

Magnetron sputtering deposition is a physical vapor deposition process used extensively in semiconductor industry for thin film growth, due to its ultra-smooth deposition, good film adhesion and high uniformity at an acceptable throughput. As shown in Fig. 14, a high voltage is applied to generate plasma by ionizing inert gas (usually Ar, Kr or Xe) introduced into the chamber, which is maintained at 20 mTorr. The Ar^+ ions created will accelerate towards the cathode, bombarding atoms from the high purity target (typically >99.9%) with high energy. The ejected atoms from the target will then impinge onto the surface of the substrate, provided that the mean free path of the material is longer than the target-to-substrate distance. At the same time, released electrons during ionization may also collide with additional Ar atoms, creating more secondary Ar^+ ions and free electrons to create a self-sustaining process. The chamber pressure is lowered to 2 mTorr before opening the shutter gate for the actual deposition. To avoid charge build-up on dielectric materials such as MgO, RF sputtering is performed at 13.56 MHz frequency and 100 W power.

The popularity of magnetron sputtering arises from its flexibility in tuning several parameters in order to optimize and ensure smooth and uniform coating of the target material. A low base chamber pressure is often required to minimize interfacial oxidation and H_2O contamination that may be detrimental to hydrophilic materials, such as MgO. Since the generation of plasma is usually inefficient (<0.1% ionization rate), permanent magnets can help to confine the secondary electrons into a helical

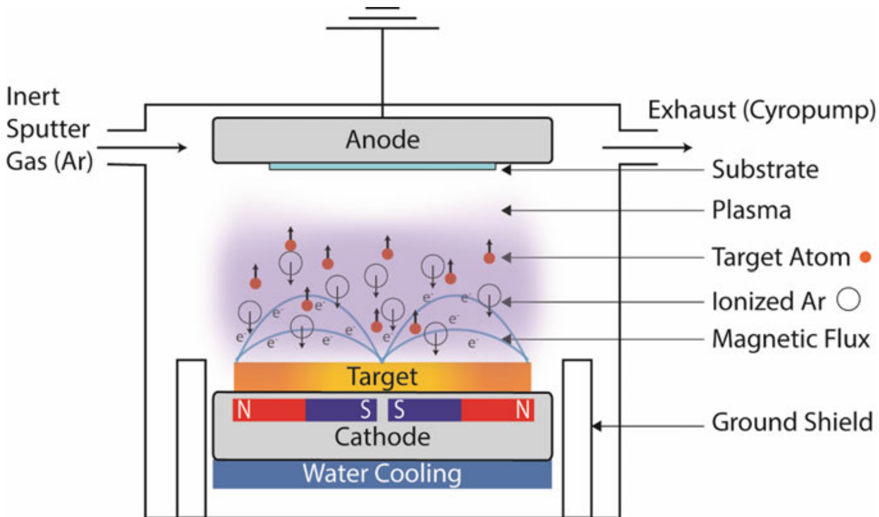


Fig. 14 Schematic diagram of the main chamber within the sputtering system

trajectory path along the magnetic field lines for further ionization events. This eliminates the need for high working gas pressure that may be unfavorable for film quality and efficient target utilization. The type of inert gas (Ar, Kr, Xe) used to generate plasma ions can also help to improve uniformity by ensuring efficient transfer of momentum. The sputtering pressure and applied power can control the deposition rate to ensure film adhesion and minimal stress. Finally, the spindle rotary speed in which the sample is mounted onto and the target to substrate distance allows for uniform distribution of sputtered material.

6.4 Vibrating Sample Magnetometer

Vibrating sample magnetometer (VSM) is a standard tool in magnetization characterization and crucial for MRAM development. With a detection range of down to 10^{-6} emu with a noise level of $<10^{-7}$ emu, VSM is suitable for measuring pre-patterned pMTJ thin film samples. The operating principle of VSM relies on the Faraday's law of induction, where an induced electromagnetic force (EMF) is generated from a magnetic sample when it is mechanically vibrating between a set of pick up coils. The voltage is detected by a lock in amplifier which takes reference to the frequency of vibration. A piece of Nickel standard is used as a calibration sample to establish a conversion ratio between V_{emf} picked up by a lock in amplifier and its known mass magnetization. The Nickel standard, like any samples to be tested, is to be placed on a clean quartz rod and secured with teflon tape, vacuum grease, double side tape or

wax. The quartz rod is then inserted into the collet of the vibrator shaft and held in place by tightening the nut.

As magnetic moment is proportional to magnetic volume, the magnetic signal of the samples are usually very weak (ranging between 10^{-4} and 10^{-6} emu) and can be easily drowned by the signal arising from the diamagnetic property of the quartz rod, Si substrate and teflon tape. Therefore, there is a need to perform additional background correction, by mathematically removing the slope and offset due to the diamagnetism. Unless otherwise specified, the magnetic moment m is saturated at very large applied field (>10 kOe) in order to ensure that all magnetic moment or spins are aligned in the same direction as the external applied magnetic field. M_s can then be determined by the following simple relation $M_s = \frac{m}{V}$, where V is the volume of the magnetic material.

By performing a magnetic field sweep, the M-H loop is able to provide information of the magnetic properties of the sample, such as remanence magnetization M_r, saturation magnetization M_s, coercivity field H_c and also its uniaxial anisotropy field H_k. An external applied magnetic field can also be swept under user-defined variable conditions (temperature, angle) in order to characterize the behavior of the magnetic sample at each field step size. The effective anisotropy energy *Keff* can then be determined from the enclosed area within the easy and the hard axes measurements $K_{eff}V = \left[\int_0^{M_s} H dm\right]_{Hard} - \left[\int_0^{M_s} H dm\right]_{Easy}$ [189, 190, 191]. This can be an alternative method to quantify K_{eff} when FMR signal is not detected due to large gilbert damping factor, when the magnetic material is considered as a single entity. Therefore, as noted in the previous section on FMR, this method is not applicable for full pMTJ stacks, as the effective magnetic anisotropy measured in the hard axis will be due to the contributions of all the magnetic layers.

The typical M-H loops of the pMTJ stack are as shown in Fig. 15a, b, commonly referred to as major and minor loops, respectively. The magnetization reversal of each section is indicated by colored arrows in the legend. Initially, under an extremely

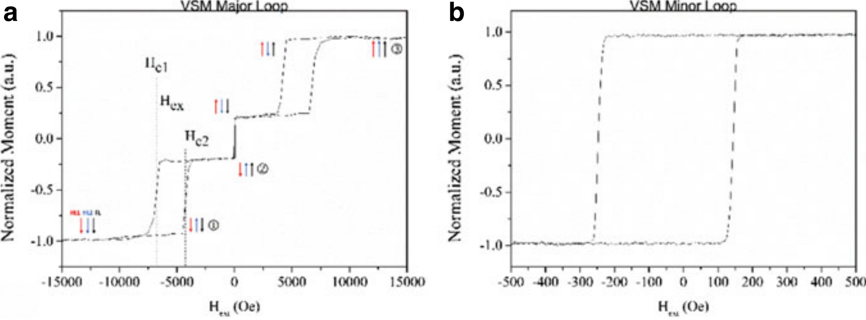


Fig. 15 **a** Major loop by sweeping a large external magnetic field. **b** Minor loop by sweeping a smaller external magnetic field to capture the areal moment of the free layer. Hysteresis is observed when the external magnetic field is reversed in the opposite direction

strong (negative) magnetic field, all magnetic moments within the pMTJ will align along the same direction as the external magnetic field. However, if the (negative) magnetic field is no longer able to favorably sustain the energy minimization along the same direction in the SAF configuration, the weaker of the ferromagnetic section within the SAF structure will switch first, as labelled as event ① occurring when $H_{ext} = H_{c2}$. Event ② occurs when H_{ext} is now applied in the positive regime such that H_{ext} is greater than the coercivity of the free layer. This event will trigger the magnetization reversal of the moments in the free layer, which were initially aligned in the negative direction. Upon further increasing the H_{ext} to saturation field in the positive direction, the energy configuration is now favorable for all three magnetic sections to be aligned in the positive field denoted as event ③. The reverse of the M-H loop (positive to negative H_{ext} sweep) can be similarly explained as above. The measurement does not retrace back on itself, due to hysteresis which is the basis of information storage at $H_{ext} = 0$ Oe. The centerline dotted across the midpoint between HL1 and HL2 as shown in Fig. 15 is the exchange field H_{ex} due to the interlayer exchange coupling mediated by the thin Ru spacer layer. The exchange coupling can be determined as $J_{ex} = H_{ex} M_s t$, where $M_s t$ is the areal moment of HL2.

In Fig. 15b, a magnetic field of smaller magnitude than the switching field of HL2 (~600 Oe) was applied to capture the magnetic moment of only the free layer. The $M_s t$ together with H_{eff} determined from FMR measurements, allows for the quantification of thermal stability of the MTJ.

6.5 *Magnetic Tunnel Junction Patterning*

6.5.1 Reactive Ion Etching (RIE)

The patterning process of MTJ consists of two steps; UV lithography patterning of photoresist and subsequent etching of the SiO_2 hard mask and MTJ material. The photoresist protects the area from being etched away. During etching, the material is physically and chemically attacked and eroded in the unprotected areas. The etch rate of material is a synergistic combination of both chemical and physical processes.

Vertical sidewalls and very accurate transfer of photoresist patterns to closely-packed high density of devices is feasible by anisotropic plasma etching, as wet etching has an issue of significant undercut relative to device size. Reactive ion etching (RIE) is done in a vacuum with a high density plasma source with a 13.6 MHz excitation frequency and 800 W RF power. The substrate holder has a bias power of 250–500 Wb, with an electrostatic chucking system and He backside cooling, as shown in Fig. 16. During etching, the temperature of the wafer was kept at 70 °C and the chamber pressure at 0.3 Pa. RIE is a combination of physical (bombardment) and chemical (reactive) process.

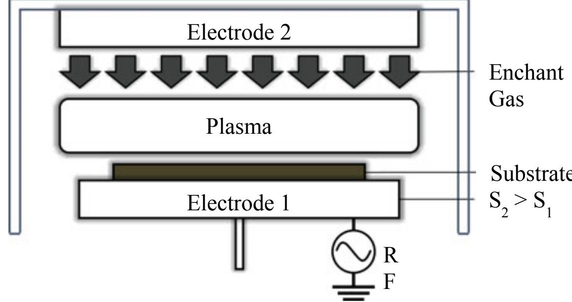


Fig. 16 Illustration of reactive ion etching system

6.5.2 Ion Beam Etching (IBE)

Ion beam etching (IBE) is the physical etching process whereby the ions of an inert gas (Ar, Kr, Xe) are accelerated from a wide ion beam source into the surface of a substrate to etch away all materials to a desired depth or under layer. Since the accelerated ions will etch all materials, it is challenging to obtain an appropriate non-eroding hard mask. Typical etch rates for ion milling by Ar^+ are ~20 nm/min for SiO_2, Si and photoresist, and 20–100 nm/min for metals. Therefore, IBE is insignificantly slower than RIE. The biggest advantage of IBE is the ability of the sample stage to tilt and rotation, and hence a sidewall cleaning at high IBE angle can be added after the main etch to remove any re-deposited materials. The primary Ar angular beam can be split into its vertical components, which serve as the main etching mechanism, while the horizontal components provide a cleaning effect counteracting the re-deposition rate [192].

The ion beam source is an Ar ion source with a radio frequency plasma generator connected to an antenna, as shown in Fig. 17. A 1.8 MHz oscillating current in the antenna produces an electromagnetic field, from which primary electrons obtain energy and are agitated. The main plasma is initiated by inelastic collisions between hot electrons and the inert (Ar, Kr, Xe) gas atoms, producing ions and electrons pairs.

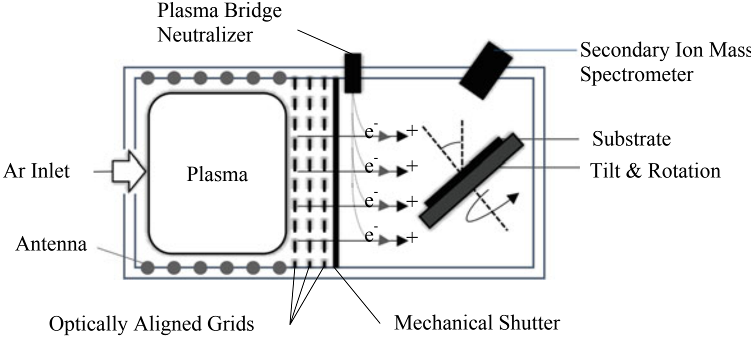


Fig. 17 Illustration of ion beam etching system

A set of optically aligned grids (Mo, W, graphite) acts as electrostatic aperture, which extracts and accelerates the ions to form a mono-energetic collimated beam. The inner grid (screen grid) is biased positive relative to the beam voltage, while the center grid (accelerator grid) is biased negative relative to the acceleration voltage, creating an electric field due to the potential difference. Ions in the plasma are extracted through the apertures and accelerated away from the source at high energies (300–1000 eV). The outer grid (decelerator grid) is placed at ground potential, which acts a collimator to reduce the beam divergence and also prevent the back-streaming of electrons from the plasma bridge neutralizer. The plasma bridge neutralizer emits electrons orthogonal to the grids to neutralize the incoming charged ion beam, and hence prevents space or surface charging of the sample. A secondary ions mass spectroscopy (SIMS) detector collects the ejected secondary ions from the sample to obtain the elemental composition of the etched materials and also provide a real-time end point detection system.

6.5.3 Prevailing Challenges

Non-volatile By-products.

Typically, plasma etching is based on the volatility of the etched by-products. Due to the thermodynamically-favored reaction, there is no ion bombardment necessary during etching, but only to induce directionality. The optimum photoresist sidewall is slightly sloped at 86°−88° for good device pitch and size control. However, the aggressive etching process of plasma etching requires the utilization of a hard mask that can tolerate long etch time or high bias power. The hard mask is etched with photoresist patterns, then the photoresist is stripped away and the etching of the MTJ is continued with the hard mask only. Si is easily etched by halogen etch chemistries, such as CF_4, Cl_2 or HBr, which forms volatile by-products at room temperature and low pressures [193], see Table 2. In most manufacturing fabrication process, SiO_2 hard mask is used and etched with CHF_3 and CF_4 gas chemistries as both etching chemistry are highly selectivity against the carbon mask. They provide F and C for etching which gives SiF_4 and CO_2 etch by-products, and polymer precursors of $CF_2{}^*$ radicals. Polymerization would occur on Si surfaces, while there are no polymerizations on SiO_2 surfaces due to the oxygen supply, resulting in CO_2 formation.

However, the ferromagnetic material in the MTJ stack, such as Co, Fe, Ni and alloys, does not readily form halogen compounds with high vapor pressures, and the

Table 2 Typical halogen-based etch gases [193]

Halogen	Etch gases	By-products
Fluorine	CF_4, CHF_3, SF_6, C_2F_6, C_4F_8	SiF_4
Chlorine	Cl_2, BCl_3, $SiCl_4$	$SiCl_4$
Bromine	HBr	$SiBr_4$

etched by-products are non-volatile, thus decreasing the etching rate and increases its difficulty of removal from the etching chamber [194, 195, 196]. Early RIE process chemistry involved Cl_2/Ar and BCl_3/Ar plasma, which showed re-deposition of residues such as Cl-based by-products, which significantly deteriorated the magnetic properties of the MTJ stack [196, 197, 198]. Figure 19 shows the SEM images obtained from pre-, post-Cl_2, post-Cl_2/H_2, and post-Ar plasma assisted etching on patterned TiN/CoFe/Ti on Si wafers [199]. The Ar-only plasma resulted in sidewall redeposition and/or TiN hard mask corrosion and causes a rough surface due to defect sites induced by the ion bombardment [200, 201], as shown in Fig. 18c, d. While, the Cl_2 plasma etched both CoFe and the TiN hard mask, with substantial re-deposition of non-volatile metal chlorides layers, as shown in Fig. 18e, f. The subsequent exposure to H_2 plasma visibly removed the re-deposited metal chlorides on the sidewalls, as shown in Fig. 18g, h. Any residual Cl_2 gas may form corrosive HCl on reaction with moisture/water vapor that will degrade the MTJ magnetic properties. Hence, a post-etch treatment with O_2 or H_2O plasma passivation step is mandatory before exposure to atmospheric conditions [202, 203].

This RIE process has been recently incorporated to etch CoFeB/MgO MTJ stack [204]. Although the removal of metal chlorides and restoration of magnetic properties of CoFeB was demonstrated by the H_2 plasma treatment, the risk of corrosion can be well avoided by using organic chemistries as an alternative, which has shown to synthesize volatile metallic organic precursors for many known magnetic materials.

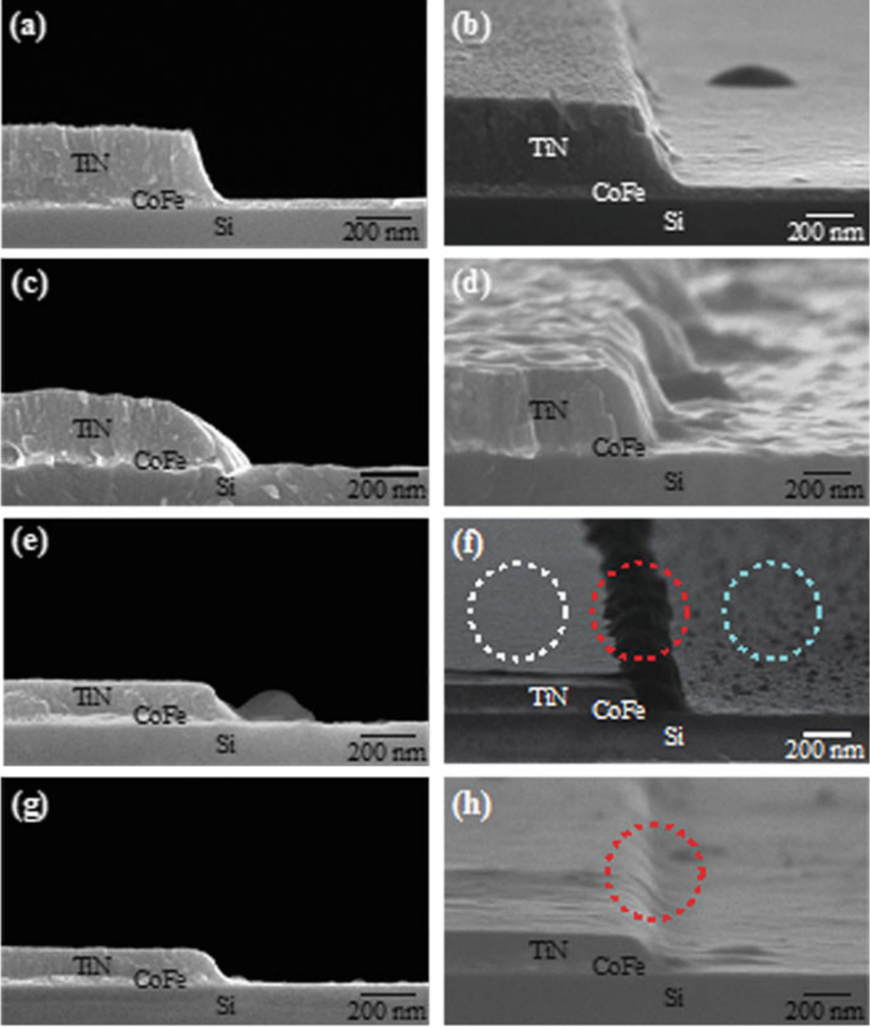


Fig. 18 SEM image of TiN/CoFe/Ti on Si wafers **a**, **b** before etching, **c**, **d** Ar plasma etching, **e**, **f** Cl_2 plasma etching, **g**, **h** subsequent H_2 plasma etching. [199] Adapted from and reprinted with permission from T. Kim, Y. Kim, J. K.-C. Chen, and J. P. Chang, "Viable chemical approach for patterning nanoscale magnetoresistive random access memory," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 33, p. 021,308 (2015)

Magnetic Film Degradation and Recovery.

In recent development, RIE plasma with non-corrosive and organic chemistries, such as CH_3COOH/Ar [205, 206], CO/NH_3 [207], C_2H_5OH [208] and Me-OH/Ar [209, 210] have been explored for MTJ etching. Although the aforementioned chemistries have displayed high selectivity between hard masks, such as Ti and Ta, and the MTJ

stack materials, resulting in a reduction in re-deposition and improvement in sidewall profiles, but magnetic degradation is still evident [209, 210].

Me-OH plasma was first proposed in 2004 by Osada et al., due to its high etch selectivity to Ta hard masks, attributed to the C–O-based etch chemistries [209, 211]. During etching, the hardening of the hard mask surface layer by nitradation, carbonization, or oxidation from the C–O-based plasma, known as hardening mask etching, provides a high anisotropy of ~80° without residues, sidewall re-deposition or corrosion. However, a major issue with C–O-based etch chemistries is the magnetic degradation of the MTJ magnetic properties due to oxidation from the pattern edge of the CoFeB free layer, leading to a significant reduction of TMR [210, 212]. Therefore, a recovery process using reductive He/H_2 plasma treatment has been proposed [209, 212]. The hydrogen radicals from the He/H_2 plasma will reduce the oxidized part of the pattern edge etched by the C–O-based etch chemistries. During the etching and recovery processes, both oxidation and reduction progressed from the sidewalls by diffusion without ion irradiation. The higher energies of the magnetic metal oxides leads to preferential reduction, with little impact on other parts of the MTJ, such as the MgO tunnel barrier.

An oxygen-based plasma is not favored for the MTJ etch process due to the risk of over-oxidation with reaction with the MgO tunnel barrier. However, the advantages of O_2 gas are isolation of damaged region from the patterned MTJ and the formation of a self-aligned passivation layer to enclose and protect the MTJ from magnetic degradation during BEOL integration processes. Therefore, a curing process involving a non-reactive oxygen treatment, known as oxygen showering post-treatment (OSP), which uses an ozone diffusion chamber to recover the etch damage by selective oxidation and also improve the electric and magnetic properties of the MTJs [213]. In addition, the OSP has been demonstrated to effectively recover electric short fails caused by the IBE process through selective oxidation and isolation of the damaged region from the MTJ [214] (Fig. 19).

Sidewalls Re-deposition.

The self-aligned integration scheme which encompasses a conducting hard mask between the MTJ and upper wiring layer offers a simple process with fewer processing steps. However, the tall conducting hard mask limits the beam angle and also leads to the shadowing of high density arrays of MTJ structures. A short conducting hard mask is more beneficial in reduction of sidewall re-depositions, but requires the inclusion of a metal via to connect the MTJ to the upper wiring layer. This short hard mask is only compatible for STT-MRAM with larger diameters MTJs due to the difficulty in via alignment and overlay tolerances. For STT-MRAM targeting sub-nanometer MTJs, the tall conducting hard mask with a tapered profile is more suitable and can help reduce sidewall re-deposition and MTJ shunting.

A post-etch treatment using a high-angle, low-energy, low-damage IBE cleanup step to remove the re-deposition materials and damaged layers from the MTJ has been proposed [201, 215]. M. Gajek et al. fabricated PMA MTJ with diameters of 20 nm by a combination of RIE and IBE, as shown in Fig. 20. The IBE was performed at

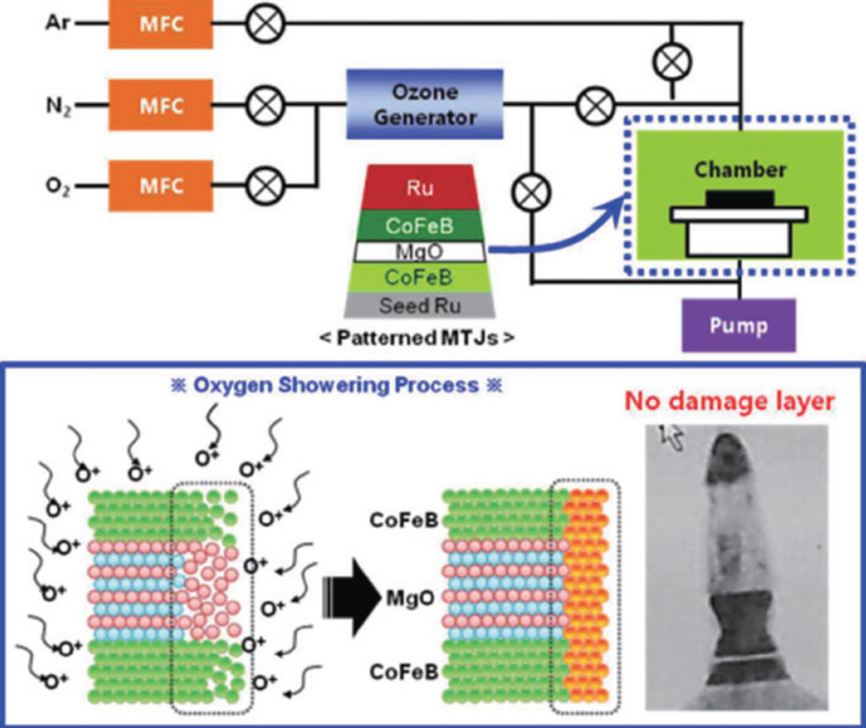


Fig. 19 The illustration of the oxygen showering setup and etch damage recovery mechanism by the oxygen showering post-treatment (OSP). Adapted from and reprinted with permission from J. Jeong and T. Endoh, "Novel oxygen showering process (OSP) for extreme damage suppression of sub-20 nm high density p-MTJ array without IBE treatment," in *VLSI Technology (VLSI Technology)*, pp. T158-T159: IEEE (2015)

glancing incidence of 80° to reduce the lateral size of the MTJ without compromising the aspect ratio.

The biggest advantage of IBE is the ability of the sample stage to tilt and rotation, and hence a sidewall cleaning at high IBE angle can be added after the main etch to remove any re-deposited materials. However, the typical etch rates for ion milling by Ar+ are ~20 nm/min for SiO_2, Si and photoresist, and 20–100 nm/min for metals, thus IBE is insignificantly slower than RIE. In addition, high beam angle etching may cause possible edge damage on the MgO barrier, degrading the MTJ performance, and the maximum beam angle is limited due to the shadowing of high density arrays of MTJ structures [192].

Even though the patterning of high density MTJs for STT-MRAM products is achievable with the current IBE manufacturing tools, the CMOS industry still heads towards the direction of RIE processing for device patterning due to the lack of experience and issue of beam divergence leading to wafer-to-wafer non-uniformity in 300 mm wafers processing. Nonetheless, the ferromagnetic material such as Co, Fe,

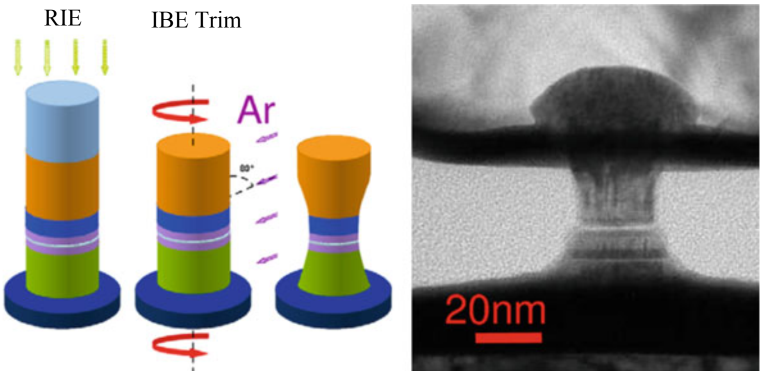


Fig. 20 TEM image of 20-nm-diameter MTJ fabricated by a combination of RIE and IBE processes. [201]. Adapted from and reprinted with permission from M. Gajek et al., "Spin torque switching of 20 nm magnetic tunnel junctions with perpendicular anisotropy," Applied Physics Letters, vol. 100, p. 132,408 (2012)

Ni and alloys used in the MTJ stack does not readily form halogen compounds with high vapor pressures, and the etched by-products are non-volatile, thus decreasing the etching rate and increases its difficulty of removal from the etching chamber [192]. Furthermore, the MTJ magnetic layers are easily damaged by heat, strain and residual chemical etchant [216]. Hence, a more effective fabrication process must be developed to avoid these problems for STT-MRAM to be production-ready.

6.6 Perspectives

MTJ have been intensively studied both fundamentally and experimentally for STT–MRAM applications [59]. STT–MRAM is well-suited as a storage technology due to its higher read/write speed, with good scalability for high density arrays and also higher endurance as compared to other RAMs. It is the only emerging memory technology that has demonstrated its capability to provide speed and endurance needed for the 1st tier in enterprise class storage systems. However, the challenges related to the BEOL technology for STT–MRAM products are difficult to overcome for mass production, especially the MTJ patterning process. Startup companies, such as Everspin Technologies, Avalanche and Spin-Transfer Technologies, are working together with large-scale manufacturers, e.g. GlobalFoundries, TSMC, NEC, IBM, Toshiba, Hynix, TDK, Micron and Samsung, and equipment vendors such as Canon Anelva, AMAT, LAM, TEL and Keysight Technologies on STT-MRAM for embedded applications, storage-class memories and cache memory. In recent years, the prominent MRAM alliances are GlobalFoundries–Everspin, IBM–TDK, Hynix–Toshiba and Qualcomm–TSMC.

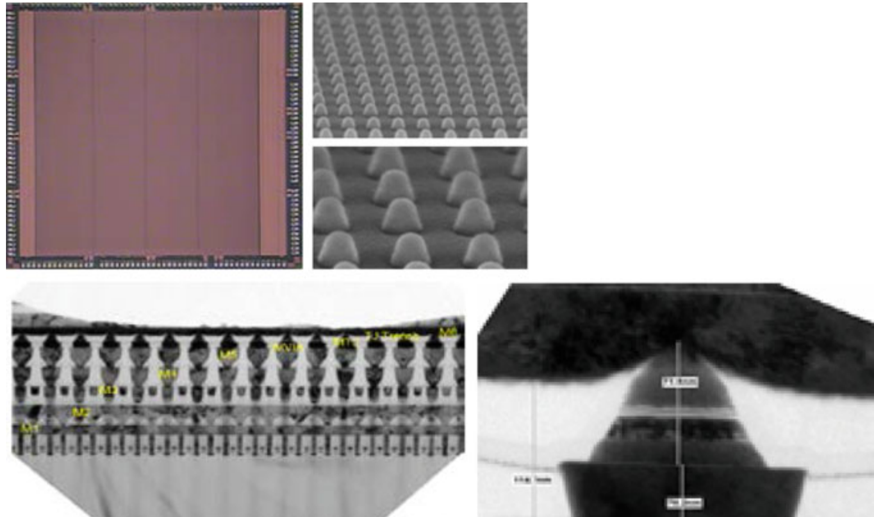


Fig. 21 SEM and TEM of 40 Mb STT-MRAM arrays for GP-MCU applications [217]. Adapted from and reprinted with permission from J. Wong et al., “CMOS-embedded STT-MRAM Arrays in 2× nm Nodes for GP-MCU applications,” (2017)

GlobalFoundries and Everspin Technologies have demonstrated 40 Mb MTJ arrays with reliable interconnects obtained from the smooth surface interfaces between MTJ films, bottom electrode and top electrode, as shown in Fig. 21. The MTJ sits directly on the bottom electrode to reduce the cell pitch and an in-situ post-etch encapsulation prevents oxidation after vacuum break. In 2018, Everspin Technologies announced production of its STT-MRAM products as an embedded 22 nm memory in GlobalFoundries 22FDX platform.

References

1. J. Åkerman, Toward a universal memory. Science **308**, 508–510 (2005)
2. R. Sbiaa, H. Meng, S.N. Piramanayagam, Materials with perpendicular magnetic anisotropy for magnetic random access memory. Physica Status Solidi (RRL) Rapid Res. Lett **5**, 413–419 (2011)
3. E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis et al., Advances and future prospects of spin-transfer torque random access memory. IEEE Trans. Magn. **46**, 1873–1878 (2010)
4. J. Slaughter, R. Dave, M. DeHerrera, M. Durlam, B. Engel, J. Janesky et al., Fundamentals of MRAM technology. J. Supercond. **15**, 19–25 (2002)
5. X. Dong, X. Wu, G. Sun, Y. Xie, H. Li, Y. Chen, Circuit and microarchitecture evaluation of 3D stacking magnetic RAM (MRAM) as a universal memory replacement, in *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE* (2008), pp. 554–559

6. J. DeBrosse, D. Gogl, A. Bette, H. Hoenigschmid, R. Robertazzi, C. Arndt et al., A high-speed 128-kb MRAM core for future universal memory applications. IEEE J. Solid-State Circuits **39**, 678–683 (2004)
7. L. Thomas, G. Jan, S. Le, S. Serrano-Guisan, Y.-J. Lee, H. Liu, et al., Probing magnetic properties of STT-MRAM devices down to sub-20 nm using spin-torque FMR, in *Electron Devices Meeting (IEDM), 2017 IEEE International* (2017), pp. 38.4.1–38.4.4
8. L. Thomas, J. Guenole, L. Son, S. Serrano-Guisan, L. Yuan-Jen, L. Huanlong, et al., Development of perpendicular STT-MRAM for last level cache applications
9. W. Xu, H. Sun, X. Wang, Y. Chen, T. Zhang, Design of last-level on-chip cache using spin-torque transfer RAM (STT RAM). IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **19**, 483–493 (2011)
10. R.C. Sousa, I.L. Prejbeanu, Non-volatile magnetic random access memories (MRAM). C R Phys. **6**, 1013–1021 (2005)
11. J.S. Meena, S.M. Sze, U. Chand, T.-Y. Tseng, Overview of emerging nonvolatile memory technologies. Nanoscale Res. Lett. **9**, 526 (2014)
12. S.A. Wolf, J. Lu, M.R. Stan, E. Chen, D.M. Treger, The promise of nanomagnetics and spintronics for future logic and universal memory. Proc. IEEE **98**, 2155–2168 (2010)
13. G. Binasch, P. Grünberg, F. Saurenbach, W. Zinn, Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. Phys. Rev. B **39**, 4828 (1989)
14. M.N. Baibich, J.M. Broto, A. Fert, F.N. Van Dau, F. Petroff, P. Etienne et al., Giant magnetoresistance of (001) Fe/(001) Cr magnetic superlattices. Phys. Rev. Lett. **61**, 2472 (1988)
15. K. Nagasaka, CPP-GMR technology for magnetic read heads of future high-density recording systems. J. Magn. Magn. Mater. **321**, 508–511 (2009)
16. C. Tsang, R.E. Fontana, T. Lin, D.E. Heim, V.S. Speriosu, B.A. Gurney et al., Design, fabrication and testing of spin-valve read heads for high density recording. IEEE Trans. Magn. **30**, 3801–3806 (1994)
17. K. Shimazawa, Y. Tsuchiya, T. Mizuno, S. Hara, T. Chou, D. Miyauchi et al., CPP-GMR film with ZnO-based novel spacer for future high-density magnetic recording. IEEE Trans. Magn. **46**, 1487–1490 (2010)
18. J. Daughton, GMR applications. J. Magn. Magn. Mater. **192**, 334–342 (1999)
19. J. Daughton, J. Brown, E. Chen, R. Beech, A. Pohm, W. Kude, Magnetic field sensors using GMR multilayer. IEEE Trans. Magn. **30**, 4608–4610 (1994)
20. S. Parkin, N. More, K. Roche, Oscillations in exchange coupling and magnetoresistance in metallic superlattice structures: Co/Ru, Co/Cr, and Fe/Cr. Phys. Rev. Lett. **64**, 2304 (1990)
21. H. Maehara, K. Nishimura, Y. Nagamine, K. Tsunekawa, T. Seki, H. Kubota et al., Tunnel magnetoresistance above 170% and resistance–area product of 1 Ω (μm)2 attained by in situ annealing of ultra-thin MgO tunnel barrier. Appl. Phys. Express **4**, 033002 (2011)
22. B. Engel, J. Akerman, B. Butcher, R. Dave, M. DeHerrera, M. Durlam et al., A 4-Mb toggle MRAM based on a novel bit and switching method. IEEE Trans. Magn. **41**, 132–136 (2005)
23. M. Julliere, Tunneling between ferromagnetic films. Phys. Lett. A **54**, 225–226 (1975)
24. S. Maekawa, U. Gafvert, Electron tunneling between ferromagnetic films. IEEE Trans. Magn. **18**, 707–708 (1982)
25. J. Nowak, J. Raułuszkiewicz, Spin dependent electron tunneling between ferromagnetic films. J. Magn. Magn. Mater. **109**, 79–90 (1992)
26. T. Miyazaki, N. Tezuka, Giant magnetic tunneling effect in Fe/Al2O3/Fe junction. J. Magn. Magn. Mater. **139**, L231–L234 (1995)
27. J.S. Moodera, L.R. Kinder, T.M. Wong, R. Meservey, Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. Phys. Rev. Lett. **74**, 3273 (1995)
28. D. Wang, C. Nordman, J.M. Daughton, Z. Qian, J. Fink, 70% TMR at room temperature for SDT sandwich junctions with CoFeB as free and reference layers. IEEE Trans. Magn. **40**, 2269–2271 (2004)
29. S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. Lee, K. Miura, H. Hasegawa et al., Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in Co Fe B/ Mg O/ Co Fe B pseudo-spin-valves annealed at high temperature. Appl. Phys. Lett. **93**, 082508 (2008)

30. W. Butler, X.-G. Zhang, T. Schulthess, J. MacLaren, Spin-dependent tunneling conductance of Fe| MgO| Fe sandwiches. Physical Review B **63**, 054416 (2001)
31. X.-G. Zhang, W. Butler, Large magnetoresistance in bcc Co/ Mg O/ Co and Fe Co/ Mg O/ Fe Co tunnel junctions. Physical Review B **70**, 172407 (2004)
32. J. Mathon, A. Umerski, Theory of tunneling magnetoresistance of an epitaxial Fe/MgO/Fe (001) junction. Physical Review B **63**, 220403 (2001)
33. J.J. Bean, M. Saito, S. Fukami, H. Sato, S. Ikeda, H. Ohno et al., Atomic structure and electronic properties of MgO grain boundaries in tunnelling magnetoresistive devices. Scientific Reports **7**, 45594 (2017)
34. Y. Ke, K. Xia, H. Guo, Oxygen-vacancy-induced diffusive scattering in Fe/MgO/Fe magnetic tunnel junctions. Phys. Rev. Lett. **105**, 236801 (2010)
35. M. Durlam, D. Addie, J. Akerman, B. Butcher, P. Brown, J. Chan*, et al.*, "A 0.18/spl mu/m 4Mb toggling MRAM," in *IEEE International Electron Devices Meeting 2003*, 2003, pp. 34.6.1–34.6.3.
36. Y. Zheng, J.-G. Zhu, Switching field variation in patterned submicron magnetic film elements. J. Appl. Phys. **81**, 5471–5473 (1997)
37. J. Shi, S. Tehrani, T. Zhu, Y. Zheng, J.-G. Zhu, Magnetization vortices and anomalous switching in patterned NiFeCo submicron arrays. Appl. Phys. Lett. **74**, 2525–2527 (1999)
38. L. Savtchenko, B. N. Engel, N. D. Rizzo, M. F. Deherrera, and J. A. Janesky, "Method of writing to scalable magnetoresistance random access memory element," ed: Google Patents, 2003.
39. J. H. Nickel and L. T. Tran, "Thermally-assisted switching of magnetic memory elements," ed: Google Patents, 2003.
40. Y. Huai and M. Pakala, "Method and system for providing heat assisted switching of a magnetic element utilizing spin transfer," ed: Google Patents, 2006.
41. J. Slaughter, R. Dave, M. Durlam, G. Kerszykowski, K. Smith, K. Nagel*, et al.*, "High speed toggle MRAM with MgO-based tunnel junctions," in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 2005, pp. 873-876
42. J. Akerman, P. Brown, M. DeHerrera, M. Durlam, E. Fuchs, D. Gajewski et al., Demonstrated reliability of 4-Mb MRAM. IEEE Trans. Device Mater. Reliab. **4**, 428–435 (2004)
43. W. Thomson, XIX. On the electro-dynamic qualities of metals: effects of magnetization on the electric conductivity of nickel and of iron. Proc. R. Soc. London **8**, 546–550 (1857)
44. T. McGuire, R. Potter, Anisotropic magnetoresistance in ferromagnetic 3d alloys. IEEE Trans. Magn. **11**, 1018–1038 (1975)
45. S. Tumanski, *Thin film magnetoresistive sensors* (CRC Press, 2001)
46. R. Sbiaa, Magnetoresistive read heads: fundamentals and functionality, in *Developments in Data Storage: Materials Perspective* (2011), p. 97
47. R. Hunt, A magnetoresistive readout transducer. IEEE Trans. Magn. **7**, 150–154 (1971)
48. D. Thompson, Magnetoresistive transducers in high-density magnetic recording, in *AIP Conference Proceedings* (1975), pp. 528–533
49. N.F. Mott, The resistance and thermoelectric properties of the transition metals. Proc. R. Soc. Lond. A **156**, 368–382 (1936)
50. S. Huang, T. Chen, C. Chien, Spin polarization of amorphous CoFeB determined by point-contact Andreev reflection. Appl. Phys. Lett. **92**, 242509 (2008)
51. W.H. Butler, Tunneling magnetoresistance from a symmetry filtering effect. Sci. Technol. Adv. Mater. **9**, 014106 (2008)
52. J.C. Slonczewski, Current-driven excitation of magnetic multilayers. J. Magn. Magn. Mater. **159**, L1–L7 (1996)
53. L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current. Phys. Rev. B **54**, 9353 (1996)
54. S.S. Parkin, M. Hayashi, L. Thomas, Magnetic domain-wall racetrack memory. Science **320**, 190–194 (2008)
55. M. Hayashi, L. Thomas, R. Moriya, C. Rettner, S.S. Parkin, Current-controlled magnetic domain-wall nanowire shift register. Science **320**, 209–211 (2008)

56. G. Tatara, H. Kohno, Theory of current-driven domain wall motion: spin transfer versus momentum transfer. Phys. Rev. Lett. **92**, 086601 (2004)
57. C. Kittel, On the theory of ferromagnetic resonance absorption. Phys. Rev. **73**, 155 (1948)
58. J. Osborn, Demagnetizing factors of the general ellipsoid. Phys. Rev. **67**, 351 (1945)
59. S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo et al., A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction. Nat. Mater. **9**, 721 (2010)
60. V. Naik, H. Meng, R. Sbiaa, Thick CoFeB with perpendicular magnetic anisotropy in CoFeB-MgO based magnetic tunnel junction. AIP Adv. **2**, 042182 (2012)
61. B.D. Cullity, C.D. Graham, *Introduction to magnetic materials* (Wiley, 2011)
62. P. Grünberg, R. Schreiber, Y. Pang, M. Brodsky, H. Sowers, Layered magnetic structures: evidence for antiferromagnetic coupling of Fe layers across Cr interlayers. Phys. Rev. Lett. **57**, 2442 (1986)
63. S.S. Parkin, Systematic variation of the strength and oscillation period of indirect magnetic exchange coupling through the 3d, 4d, and 5d transition metals. Phys. Rev. Lett. **67**, 3598 (1991)
64. D. Edwards, J. Mathon, R. Muniz, M. Phan, Oscillations of the exchange in magnetic multilayers as an analog of de Haas–van Alphen effect. Phys. Rev. Lett. **67**, 493 (1991)
65. P. Bruno, Theory of interlayer magnetic coupling. Phys. Rev. B **52**, 411 (1995)
66. P. Bruno, Theory of interlayer exchange interactions in magnetic multilayers. J. Phys. Condens. Matter **11**, 9403 (1999)
67. P. Bruno, C. Chappert, Ruderman-Kittel theory of oscillatory interlayer exchange coupling. Phys. Rev. B **46**, 261 (1992)
68. T.L. Gilbert, A phenomenological theory of damping in ferromagnetic materials. IEEE Trans. Magn. **40**, 3443–3449 (2004)
69. S.I. Kiselev, J. Sankey, I. Krivorotov, N. Emley, R. Schoelkopf, R. Buhrman, et al., Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380 (2003)
70. S. Kaka, M.R. Pufall, W.H. Rippard, T.J. Silva, S.E. Russek, J.A. Katine, Mutual phase-locking of microwave spin torque nano-oscillators. Nature **437**, 389 (2005)
71. D. Houssameddine, U. Ebels, B. Delaët, B. Rodmacq, I. Firastrau, F. Ponthenier et al., Spin-torque oscillator using a perpendicular polarizer and a planar free layer. Nat. Mater. **6**, 447 (2007)
72. S. Ishibashi, T. Seki, T. Nozaki, H. Kubota, S. Yakata, A. Fukushima et al., Large diode sensitivity of CoFeB/MgO/CoFeB magnetic tunnel junctions. Appl. Phys. Express **3**, 073001 (2010)
73. Y. Gui, Y. Xiao, L. Bai, S. Hemour, Y. Zhao, D. Houssameddine et al., High sensitivity microwave detection using a magnetic tunnel junction in the absence of an external applied magnetic field. Appl. Phys. Lett. **106**, 152403 (2015)
74. S. Miwa, S. Ishibashi, H. Tomita, T. Nozaki, E. Tamura, K. Ando et al., Highly sensitive nanoscale spin-torque diode. Nat. Mater. **13**, 50 (2014)
75. A. Tulapurkar, Y. Suzuki, A. Fukushima, H. Kubota, H. Maehara, K. Tsunekawa et al., Spin-torque diode effect in magnetic tunnel junctions. Nature **438**, 339 (2005)
76. B. Fang, M. Carpentieri, X. Hao, H. Jiang, J.A. Katine, I.N. Krivorotov et al., Giant spin-torque diode sensitivity in the absence of bias magnetic field. Nat. Commun. **7**, 11259 (2016)
77. S. Ishibashi, K. Ando, T. Seki, T. Nozaki, H. Kubota, S. Yakata et al., High spin-torque diode sensitivity in CoFeB/MgO/CoFeB magnetic tunnel junctions under DC bias currents. IEEE Trans. Magn. **47**, 3373–3376 (2011)
78. N. Rizzo, M. DeHerrera, J. Janesky, B. Engel, J. Slaughter, S. Tehrani, Thermally activated magnetization reversal in submicron magnetic tunnel junctions for magnetoresistive random access memory. Appl. Phys. Lett. **80**, 2335–2337 (2002)
79. S. Van Beek, K. Martens, P. Roussel, Y.C. Wu, W. Kim, S. Rao et al., Thermal stability analysis and modelling of advanced perpendicular magnetic tunnel junctions. AIP Adv. **8**, 055909 (2018)
80. M. Sharrock, Time dependence of switching fields in magnetic recording media. J. Appl. Phys. **76**, 6413–6418 (1994)

81. J.J. Kan, C. Park, C. Ching, J. Ahn, Y. Xie, M. Pakala et al., A study on practically unlimited endurance of STT-MRAM. IEEE Trans. Electron Devices **64**, 3639–3646 (2017)
82. W.C. Law, T. Tahmasebi, F. Tan, T. Jin, W. Gan, X. Zhu et al., High temperature ferromagnetic resonance study on pMTJ stacks with diffusion barrier layers. J. Phys. D Appl. Phys. **51**, 1–7 (2018)
83. S.Y. Jang, S. Lim, S. Lee, Magnetic dead layer in amorphous CoFeB layers with various top and bottom structures. J. Appl. Phys. **107**, 09C707 (2010)
84. S.Y. Jang, C.-Y. You, S. Lim, S. Lee, Annealing effects on the magnetic dead layer and saturation magnetization in unit structures relevant to a synthetic ferrimagnetic free structure. J. Appl. Phys. **109**, 013901 (2011)
85. K. Oguz, P. Jivrajka, M. Venkatesan, G. Feng, J. Coey, Magnetic dead layers in sputtered Co 40 Fe 40 B 20 films. J. Appl. Phys. **103**, 07B526 (2008)
86. H. Sato, M. Yamanouchi, S. Ikeda, S. Fukami, F. Matsukura, H. Ohno, Perpendicular-anisotropy CoFeB-MgO magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure. Appl. Phys. Lett. **101**, 022414 (2012)
87. Y. Huai, M. Pakala, Spin scattering and heat assisted switching of a magnetic element (Google Patents, 2006)
88. L.T. Tran, Thermal-assisted switching array configuration for MRAM (Google Patents, 2005)
89. I. Prejbeanu, M. Kerekes, R.C. Sousa, H. Sibuet, O. Redon, B. Dieny et al., Thermally assisted MRAM. J. Phys. Condens. Matter **19**, 165218 (2007)
90. J. Wang, P. Freitas, Low-current blocking temperature writing of double barrier magnetic random access memory cells. Appl. Phys. Lett. **84**, 945–947 (2004)
91. W. Kang, Y. Cheng, Y. Zhang, D. Ravelosona, W. Zhao, Readability challenges in deeply scaled STT-MRAM, in *Non-volatile Memory Technology Symposium (NVMTS), 2014 14th Annual* (2014), pp. 1–4
92. Y. Zhang, X. Wang, Y. Chen, STT-RAM cell design optimization for persistent and non-persistent error rate reduction: a statistical design view, in *Proceedings of the International Conference on Computer-Aided Design* (2011), pp. 471–477
93. W. Zhao, Y. Zhang, T. Devolder, J.-O. Klein, D. Ravelosona, C. Chappert et al., Failure and reliability analysis of STT-MRAM. Microelectron. Reliab. **52**, 1848–1852 (2012)
94. S. Ikeda, J. Hayakawa, Y.M. Lee, R. Sasaki, T. Meguro, F. Matsukura et al., Dependence of tunnel magnetoresistance in MgO based magnetic tunnel junctions on Ar pressure during MgO sputtering. Jpn. J. Appl. Phys. **44**, L1442 (2005)
95. Y. Lee, J. Hayakawa, S. Ikeda, F. Matsukura, H. Ohno, Effect of electrode composition on the tunnel magnetoresistance of pseudo-spin-valve magnetic tunnel junction with a MgO tunnel barrier. Appl. Phys. Lett. **90**, 212507 (2007)
96. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, K. Ando, Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. Nat. Mater. **3**, 868 (2004)
97. S. Yuasa, D. Djayaprawira, Giant tunnel magnetoresistance in magnetic tunnel junctions with a crystalline MgO (0 0 1) barrier. J. Phys. D Appl. Phys. **40**, R337 (2007)
98. X. Chen, P.P. Freitas, Magnetic tunnel junction based on MgO barrier prepared by natural oxidation and direct sputtering deposition. Nano-Micro Lett. **4**, 25–29 (2012)
99. S. Yuasa, A. Fukushima, H. Kubota, Y. Suzuki, K. Ando, Giant tunneling magnetoresistance up to 410% at room temperature in fully epitaxial Co/Mg O/Co magnetic tunnel junctions with bcc Co (001) electrodes. Appl. Phys. Lett. **89**, 042505 (2006)
100. L. Néel, Magnetisme-sur un nouveau mode de couplage entre les animantations de deux couches minces ferromagnetiques. Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences **255**, 1676–2000 (1962)
101. J. Moritz, F. Garcia, J. Toussaint, B. Dieny, J. Nozieres, Orange peel coupling in multilayers with perpendicular magnetic anisotropy: application to (Co/Pt)-based exchange-biased spin-valves. EPL (Europhys. Lett.) **65**, 123 (2004)
102. B. Schrag, A. Anguelouch, S. Ingvarsson, G. Xiao, Y. Lu, P. Trouilloud et al., Néel “orange-peel” coupling in magnetic tunneling junction devices. Appl. Phys. Lett. **77**, 2373–2375 (2000)

103. S.S. Parkin, C. Kaiser, A. Panchula, P.M. Rice, B. Hughes, M. Samant et al., Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. Nat. Mater. **3**, 862 (2004)
104. J. Zhang, R. White, Voltage dependence of magnetoresistance in spin dependent tunneling junctions. J. Appl. Phys. **83**, 6512–6514 (1998)
105. S. Zhang, P. Levy, A. Marley, S. Parkin, Quenching of magnetoresistance by hot electrons in magnetic tunnel junctions. Phys. Rev. Lett. **79**, 3744 (1997)
106. C.H. Shang, J. Nowak, R. Jansen, J.S. Moodera, Temperature dependence of magnetoresistance and surface magnetization in ferromagnetic tunnel junctions. Phys. Rev. B **58**, R2917 (1998)
107. V. Drewello, J. Schmalhorst, A. Thomas, G. Reiss, Evidence for strong magnon contribution to the TMR temperature dependence in MgO based tunnel junctions. Phys. Rev. B **77**, 014440 (2008)
108. S. Wang, R. Ward, G. Du, X. Han, C. Wang, A. Kohn, Temperature dependence of giant tunnel magnetoresistance in epitaxial Fe/MgO/Fe magnetic tunnel junctions. Phys. Rev. B **78**, 180411 (2008)
109. M. Jourdan, Revival of Heusler compounds for spintronics. Mater. Today **8**, 362–363 (2014)
110. T. Graf, C. Felser, S.S. Parkin, Simple rules for the understanding of Heusler compounds. Prog. Solid State Chem. **39**, 1–50 (2011)
111. R. Sbiaa, S. Lua, R. Law, H. Meng, R. Lye, H. Tan, Reduction of switching current by spin transfer torque effect in perpendicular anisotropy magnetoresistive devices. J. Appl. Phys. **109**, 07C707 (2011)
112. C. Lin, S. Kang, Y. Wang, K. Lee, X. Zhu, W. Chen, et al., 45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell, in *Electron Devices Meeting (IEDM), 2009 IEEE International* (2009), pp. 1–4
113. Z. Diao, D. Apalkov, M. Pakala, Y. Ding, A. Panchula, Y. Huai, Spin transfer switching and spin polarization in magnetic tunnel junctions with MgO and AlO x barriers. Appl. Phys. Lett. **87**, 232502 (2005)
114. X. Yao, H. Meng, Y. Zhang, J.-P. Wang, Improved current switching symmetry of magnetic tunneling junction and giant magnetoresistance devices with nano-current-channel structure. J. Appl. Phys. **103**, 07A717 (2008)
115. D. Worledge, G. Hu, D.W. Abraham, J. Sun, P. Trouilloud, J. Nowak et al., Spin torque switching of perpendicular Ta| CoFeB| MgO-based magnetic tunnel junctions. Appl. Phys. Lett. **98**, 022501 (2011)
116. J. Sun, Current-driven magnetic switching in manganite trilayer junctions. J. Magn. Magn. Mater. **202**, 157–162 (1999)
117. Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen et al., Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory. J. Phys. Condens. Matter **19**, 165209 (2007)
118. Y. Huai, Spin-transfer torque MRAM (STT-MRAM): challenges and prospects. AAPPS Bulletin **18**, 33–40 (2008)
119. M. Gottwald, J. Kan, K. Lee, X. Zhu, C. Park, S. Kang, Scalable and thermally robust perpendicular magnetic tunnel junctions for STT-MRAM. Appl. Phys. Lett. **106**, 032413 (2015)
120. J. Chatterjee, T. Tahmasebi, S. Mertens, G.S. Kar, T. Min, J. De Boeck, Seed layer effect on the magnetic properties of ultrathin Co/Pt multilayers with perpendicular magnetic anisotropy. IEEE Trans. Magn. **50**, 1–4 (2014)
121. K. Yakushiji, T. Saruya, H. Kubota, A. Fukushima, T. Nagahama, S. Yuasa et al., Ultrathin Co/Pt and Co/Pd superlattice films for MgO-based perpendicular magnetic tunnel junctions. Appl. Phys. Lett. **97**, 232508 (2010)
122. W. Peng, O. Keitel, R. Victoria, E. Koparal, J.H. Judy, Co/Pt superlattices with ultra-thin Ta seed layer on NiFe underlayer for double-layer perpendicular magnetic recording media. IEEE Trans. Magn. **36**, 2390–2392 (2000)

123. P. Chowdhury, P. Kulkarni, M. Krishnan, H.C. Barshilia, A. Sagdeo, S. Rai et al., Effect of coherent to incoherent structural transition on magnetic anisotropy in Co/Pt multilayers. J. Appl. Phys. **112**, 023912 (2012)
124. J. Chatterjee, R.C. Sousa, N. Perrissin, S. Auffret, C. Ducruet, B. Dieny, Enhanced annealing stability and perpendicular magnetic anisotropy in perpendicular magnetic tunnel junctions using W layer. Appl. Phys. Lett. **110**, 202401 (2017)
125. H. Almasi, D.R. Hickey, T. Newhouse-Illige, M. Xu, M.R. Rosales, S. Nahar et al., Enhanced tunneling magnetoresistance and perpendicular magnetic anisotropy in Mo/CoFeB/MgO magnetic tunnel junctions. Appl. Phys. Lett. **106**, 182406 (2015)
126. T. Liu, Y. Zhang, J.W. Cai, H.Y. Pan, Thermally robust Mo/CoFeB/MgO trilayers with strong perpendicular magnetic anisotropy. Sci. Rep. **4**, 5895 (2014)
127. S. Tang, P. Carcia, D. Coulman, A. McGhie, Scanning tunneling microscopy of Pt/Co multilayers on Pt buffer layers. Appl. Phys. Lett. **59**, 2898–2900 (1991)
128. C.-J. Lin, G. Gorman, C. Lee, R. Farrow, E. Marinero, H. Do et al., Magnetic and structural properties of Co/Pt multilayers. J. Magn. Magn. Mater. **93**, 194–206 (1991)
129. E.H. Kim, Y.B. Xiao, S.M. Kong, C.W. Chung, Investigation on etch characteristics of nanometer-sized magnetic tunnel junction stacks using a HBr/Ar plasma. J. Nanosci. Nanotechnol. **11**, 6616–6620 (2011)
130. E.H. Kim, T.Y. Lee, C.W. Chung, Evolution of etch profile of magnetic tunnel junction stacks etched in a CH3OH/Ar plasma. J. Electrochem. Soc. **159**, H230–H234 (2012)
131. S. Hassan, L. Xue, J. Anh, M. Pakala, G. Sin, M. Okazaki, STT-RAM device performance improvement using CMP process, in *Advanced Semiconductor Manufacturing Conference (ASMC), 2017 28th Annual SEMI* (2017), pp. 209–211
132. M. Wang, Y. Zhang, X. Zhao, W. Zhao, Tunnel junction with perpendicular magnetic anisotropy: status and challenges. Micromachines **6**, 1023–1045 (2015)
133. Z. Li, D.J. Smith, E. Marinero, J. Willett, Investigations of microstructure of thin TbFeCo films by high-resolution electron microscopy. J. Appl. Phys. **69**, 6590–6594 (1991)
134. N. Nishimura, T. Hirai, A. Koganei, T. Ikeda, K. Okano, Y. Sekiguchi et al., Magnetic tunnel junction device with perpendicular magnetization films for high-density magnetic random access memory. J. Appl. Phys. **91**, 5246–5249 (2002)
135. H.-M. Lee, Y.-C. Lee, H.-H. Chen, L. Horng, J.-C. Wu, C.-M. Lee, et al., The development of perpendicular magnetic tunneling junctions, in *Spin* (2012), p. 1230002
136. G. Kim, Y. Sakuraba, M. Oogane, Y. Ando, T. Miyazaki, Tunneling magnetoresistance of magnetic tunnel junctions using perpendicular magnetization L 1 0-Co Pt electrodes. Appl. Phys. Lett. **92**, 172502 (2008)
137. M. Yoshikawa, E. Kitagawa, T. Nagase, T. Daibou, M. Nagamine, K. Nishiyama et al., Tunnel magnetoresistance over 100% in MgO-based magnetic tunnel junction films with perpendicular magnetic L1 $_0 $-FePt electrodes. IEEE Trans. Magn. **44**, 2573–2576 (2008)
138. S. Iwata, S. Yamashita, S. Tsunashima, Perpendicular magnetic anisotropy and magneto-optical Kerr spectra of MBE-grown PtCo alloy films. IEEE Trans. Magn. **33**, 3670–3672 (1997)
139. M. Ohtake, D. Suzuki, M. Futamoto, F. Kirino, N. Inaba, Preparation of L 11-CoPt/MgO/L 11-CoPt tri-layer film on Ru (0001) underlayer. AIP Adv. **6**, 056103 (2016)
140. J.-C.A. Huang, A. Hsu, Y. Lee, T.-H. Wu, C. Lee, Influence of crystal structure on the perpendicular magnetic anisotropy of an epitaxial CoPt alloy. J. Appl. Phys. **85**, 5977–5979 (1999)
141. H. Sato, T. Shimatsu, Y. Okazaki, H. Muraoka, H. Aoi, S. Okamoto et al., Fabrication of L 1 1 type Co-Pt ordered alloy films by sputter deposition. J. Appl. Phys. **103**, 07E114 (2008)
142. G. Kar, W. Kim, T. Tahmasebi, J. Swerts, S. Mertens, N. Heylen, et al., Co/Ni based p-MTJ stack for sub-20 nm high density stand alone and high performance embedded memory application, in *Electron Devices Meeting (IEDM), 2014 IEEE International* (2014), pp. 19.1.1–19.1.4
143. J. Zhou, W. Zhao, Y. Wang, S. Peng, J. Qiao, L. Su et al., Large influence of capping layers on tunnel magnetoresistance in magnetic tunnel junctions. Appl. Phys. Lett. **109**, 242403 (2016)

144. R. Meservey, P. Tedrow, Spin-polarized electron tunneling. Phys. Rep. **238**, 173–243 (1994)
145. P. Tedrow, R. Meservey, Spin polarization of electrons tunneling from films of Fe Co, Ni, and Gd. Phys. Rev. B **7**, 318 (1973)
146. O. Se Young, C.-G. Lee, A.J. Shapiro, W.F. Egelhoff Jr., M.D. Vaudin, J.L. Ruglovsky, et al., "X-ray diffraction study of the optimization of MgO growth conditions for magnetic tunnel junctions. J. Appl. Phys. **103**, 07A920 (2008)
147. S. Yuasa, Y. Suzuki, T. Katayama, K. Ando, Characterization of growth and crystallization processes in CoFeB/ MgO/ CoFeB magnetic tunnel junction structure by reflective high-energy electron diffraction. Appl. Phys. Lett. **87**, 242503 (2005)
148. J. Hayakawa, S. Ikeda, F. Matsukura, H. Takahashi, H. Ohno, Dependence of giant tunnel magnetoresistance of sputtered CoFeB/MgO/CoFeB magnetic tunnel junctions on MgO barrier thickness and annealing temperature. Jpn. J. Appl. Phys. **44**, L587 (2005)
149. T. Devolder, P.-H. Ducrot, J.-P. Adam, I. Barisic, N. Vernier, J.-V. Kim et al., Damping of CoxFe80−xB20 ultrathin films with perpendicular magnetic anisotropy. Appl. Phys. Lett. **102**, 022407 (2013)
150. M.P.R. Sabino, S. Ter Lim, M. Tran, Influence of Ta insertions on the magnetic properties of MgO/CoFeB/MgO films probed by ferromagnetic resonance. Appl. Phys. Express **7**, 093002 (2014)
151. G. Jan, Y.-J. Wang, T. Moriyama, Y.-J. Lee, M. Lin, T. Zhong et al., High spin torque efficiency of magnetic tunnel junctions with MgO/CoFeB/MgO free layer. Appl. Phys. Express **5**, 093008 (2012)
152. C. Bilzer, T. Devolder, J.-V. Kim, G. Counil, C. Chappert, S. Cardoso et al., Study of the dynamic magnetic properties of soft CoFeB films. J. Appl. Phys. **100**, 053903 (2006)
153. X. Liu, W. Zhang, M.J. Carter, G. Xiao, Ferromagnetic resonance and damping properties of CoFeB thin films as free layers in MgO-based magnetic tunnel junctions. J. Appl. Phys. **110**, 033910 (2011)
154. E.C.I. Enobio, H. Sato, S. Fukami, F. Matsukura, H. Ohno, CoFeB thickness dependence of damping constants for single and double CoFeB-MgO interface structures. IEEE Magn. Lett. **6**, 1–3 (2015)
155. C. Park, J.-G. Zhu, M.T. Moneck, Y. Peng, D.E. Laughlin, Annealing effects on structural and transport properties of rf-sputtered CoFeB/MgO/CoFeB magnetic tunnel junctions. J. Appl. Phys. **99**, 08A901 (2006)
156. J. Swerts, S. Mertens, T. Lin, S. Couet, Y. Tomczak, K. Sankaran et al., BEOL compatible high tunnel magneto resistance perpendicular magnetic tunnel junctions using a sacrificial Mg layer as CoFeB free layer cap. Appl. Phys. Lett. **106**, 262407 (2015)
157. X. Kozina, S. Ouardi, B. Balke, G. Stryganyuk, G.H. Fecher, C. Felser et al., A nondestructive analysis of the B diffusion in Ta–CoFeB–MgO–CoFeB–Ta magnetic tunnel junctions by hard x-ray photoemission. Appl. Phys. Lett. **96**, 072105 (2010)
158. N. Miyakawa, D. Worledge, K. Kita, Impact of Ta diffusion on the perpendicular magnetic anisotropy of Ta/CoFeB/MgO. IEEE Magn. Lett. **4**, 1000104–1000104 (2013)
159. T. Miyajima, T. Ibusuki, S. Umehara, M. Sato, S. Eguchi, M. Tsukada et al., Transmission electron microscopy study on the crystallization and boron distribution of CoFeB/MgO/CoFeB magnetic tunnel junctions with various capping layers. Appl. Phys. Lett. **94**, 122501 (2009)
160. J.-H. Kim, J.-B. Lee, G.-G. An, S.-M. Yang, W.-S. Chung, H.-S. Park et al., Ultrathin W space layer-enabled thermal stability enhancement in a perpendicular MgO/CoFeB/W/CoFeB/MgO recording frame. Sci. Rep. **5**, 16903 (2015)
161. S. Couet, J. Swerts, S. Mertens, T. Lin, Y. Tomczak, E. Liu et al., Oxygen scavenging by ta spacers in double-MgO free layers for perpendicular spin-transfer torque magnetic random-access memory. IEEE Magn. Lett. **7**, 3103004 (2016)
162. A. Le Goff, R. Soucaille, T. Tahmasebi, J. Swerts, A. Furnemont, T. Devolder, Optimization of top-pinned perpendicular anisotropy tunnel junctions through Ta insertion. Jpn. J. Appl. Phys. **54**, 090302 (2015)
163. M.P.R. Sabino, T.L. Sze, M. Tran, Influence of Ta insertions on the magnetic properties of MgO/CoFeB/MgO films probed by ferromagnetic resonance. Appl. Phys. Express **7**, 093002 (2014)

164. S. Couet, T. Devolder, J. Swerts, S. Mertens, T. Lin, E. Liu et al., Impact of Ta and W-based spacers in double MgO STT-MRAM free layers on perpendicular anisotropy and damping. Appl. Phys. Lett. **111**, 152406 (2017)
165. T. Devolder, E. Liu, J. Swerts, S. Couet, T. Lin, S. Mertens et al., Ferromagnetic resonance study of composite Co/Ni - FeCoB free layers with perpendicular anisotropy. Appl. Phys. Lett. **109**, 142408 (2016)
166. J.-H. Park, Y. Kim, W. Lim, J. Kim, S. Park, W. Kim, et al., Enhancement of data retention and write current scaling for sub-20 nm STT-MRAM by utilizing dual interfaces for perpendicular magnetic anisotropy, in *2012 Symposium on VLSI Technology (VLSIT)* (2012), pp. 57–58
167. S.E. Lee, Y. Takemura, J.G. Park, Effect of double MgO tunneling barrier on thermal stability and TMR ratio for perpendicular MTJ spin-valve with tungsten layers. Appl. Phys. Lett. **109**, 182405 (2016)
168. S.E. Lee, T.H. Shim, J.G. Park, Perpendicular magnetic tunnel junction (p-MTJ) spin-valves designed with a top Co2Fe6B2 free layer and a nanoscale-thick tungsten bridging and capping layer. Npg Asia Mater. **8**, e324 (2016)
169. E. Liu, J. Swerts, S. Couet, S. Mertens, Y. Tomczak, T. Lin et al., [Co/Ni]-CoFeB hybrid free layer stack materials for high density magnetic random access memory applications. Appl. Phys. Lett. **108**, 132405 (2016)
170. E. Liu, J. Swerts, T. Devolder, S. Couet, S. Mertens, T. Lin et al., Seed layer impact on structural and magnetic properties of [Co/Ni] multilayers with perpendicular magnetic anisotropy. J. Appl. Phys. **121**, 043905 (2017)
171. S. Mizukami, X. Zhang, T. Kubota, H. Naganuma, M. Oogane, Y. Ando et al., Gilbert damping in Ni/Co multilayer films exhibiting large perpendicular anisotropy. Appl. Phys. Express **4**, 013005 (2011)
172. G. Daalderop, P. Kelly, F. Den Broeder, Prediction and confirmation of perpendicular magnetic anisotropy in Co/Ni multilayers. Phys. Rev. Lett. **68**, 682 (1992)
173. S. Girod, M. Gottwald, S. Andrieu, S. Mangin, J. McCord, E.E. Fullerton et al., Strong perpendicular magnetic anisotropy in Ni/Co (111) single crystal superlattices. Appl. Phys. Lett. **94**, 262504 (2009)
174. S. Fukami, T. Suzuki, H. Tanigawa, N. Ohshima, N. Ishiwata, Stack structure dependence of Co/Ni multilayer for current-induced domain wall motion. Appl. Phys. Express **3**, 113002 (2010)
175. N. Perrissin, S. Lequeux, N. Strelkov, A. Chavent, L. Vila, L.D. Buda-Prejbeanu et al., A highly thermally stable sub-20 nm magnetic random-access memory based on perpendicular shape anisotropy. Nanoscale **10**, 12187–12195 (2018)
176. K. Watanabe, B. Jinnai, S. Fukami, H. Sato, H. Ohno, Shape anisotropy revisited in single-digit nanometer magnetic tunnel junctions. Nat. Commun. **9**, 663 (2018)
177. J. Griffiths, Anomalous high-frequency resistance of ferromagnetic metals. Nature **158**, 670 (1946)
178. C. Kittel, Interpretation of anomalous Larmor frequencies in ferromagnetic resonance experiment. Phys. Rev. **71**, 270 (1947)
179. M. Farle, Ferromagnetic resonance of ultrathin metallic layers. Rep. Prog. Phys. **61**, 755 (1998)
180. C. Kittel, *Introduction to Solid State Physics*, vol. 8 (Wiley New York, 1976)
181. C. Bilzer, *Microwave Susceptibility of Thin Ferromagnetic Films: Metrology and Insight into Magnetization Dynamics* (Université Paris Sud-Paris XI, PhD, 2007)
182. Y. Tserkovnyak, A. Brataas, G.E. Bauer, Spin pumping and magnetization dynamics in metallic multilayers. Phys. Rev. B **66**, 224403 (2002)
183. J.M. Shaw, H.T. Nembach, T.J. Silva, Determination of spin pumping as a source of linewidth in sputtered Co 90 Fe 10/Pd multilayers by use of broadband ferromagnetic resonance spectroscopy. Phys. Rev. B **85**, 054412 (2012)
184. W. Yager, R. Bozorth, Ferromagnetic resonance at microwave frequencies. Phys. Rev. **72**, 80 (1947)

185. G. Counil, J. Kim, K. Shigeto, Y. Otani, T. Devolder, P. Crozat et al., Inductive measurement of the high frequency permeability of a Permalloy thin film. J. Magn. Magn. Mater. **272**, 290–292 (2004)
186. I. Neudecker, Magnetization dynamics of confined ferromagnetic systems (2006)
187. G. Counil, J.-V. Kim, T. Devolder, C. Chappert, K. Shigeto, Y. Otani, Spin wave contributions to the high-frequency magnetic response of thin films obtained with inductive methods. J. Appl. Phys. **95**, 5646–5652 (2004)
188. D. Worledge, P. Trouilloud, Magnetoresistance measurement of unpatterned magnetic tunnel junction wafers by current-in-plane tunneling. Appl. Phys. Lett. **83**, 84–86 (2003)
189. J.H. Jung, S.H. Lim, S.-R. Lee, Perpendicular magnetic anisotropy properties of CoFeB/Pd multilayers. J. Nanosci. Nanotechnol. **11**, 6233–6236 (2011)
190. T. Young Lee, D. Su Son, S. Ho Lim, S.-R. Lee, High post-annealing stability in [Pt/Co] multilayers, vol. 113. (2013), pp. 1–3
191. S.T. Lim, M. Tran, J.W. Chenchen, J.F. Ying, G. Han, Effect of different seed layers with varying Co and Pt thicknesses on the magnetic properties of Co/Pt multilayers. J. Appl. Phys. **117**, 17A731 (2015)
192. X. Peng, S. Wakeham, A. Morrone, S. Axdal, M. Feldbaum, J. Hwu et al., Towards the sub-50 nm magnetic device definition: Ion beam etching (IBE) vs plasma-based etching. Vacuum **83**, 1007–1013 (2009)
193. S. Franssila, *Introduction to microfabrication* (Wiley, 2010)
194. J.H. Hwang, G. Jin, X. Chen, Etching methods for a magnetic memory cell stack (Google Patents, 2004)
195. R.L. Comstock, Review modern magnetic materials in data storage. J. Mater. Sci. Mater. Electron. **13**, 509–523 (2002)
196. T. Kim, J.K.-C. Chen, J.P. Chang, Thermodynamic assessment and experimental verification of reactive ion etching of magnetic metal elements. J. Vacuum Sci. Technol. Vacuum Surf. Films **32**, 041305 (2014)
197. S.R. Min, H.N. Cho, K.W. Kim, Y.J. Cho, S.-H. Choa, C.W. Chung, Etch characteristics of magnetic tunnel junction stack with nanometer-sized patterns for magnetic random access memory. Thin Solid Films **516**, 3507–3511 (2008)
198. C. Fabrie, J. Kohlhepp, H. Swagten, B. Koopmans, M. Andriesse, E. van der Drift, Magnetization losses in submicrometer CoFeB dots etched in a high ion density Cl 2-based plasma. J. Vacuum Sci. Technol. B Microelectron. Nanometer Struct. Process. Measure. Phenomena **24**, 2627–2630 (2006)
199. T. Kim, Y. Kim, J.K.-C. Chen, J.P. Chang, Viable chemical approach for patterning nanoscale magnetoresistive random access memory. J. Vacuum Sci. Technol. Vacuum Surf. Films **33**, 021308 (2015)
200. K. Keizo, U. Hiroaki, S. Katsumi, H. Hiromitsu, S. Tadahiko, Etching magnetic tunnel junction with metal etchers. Japan. J. Appl. Phys. **49**, 08JB02 (2010)
201. M. Gajek, J. Nowak, J. Sun, P. Trouilloud, E. O'sullivan, D. Abraham, et al., Spin torque switching of 20 nm magnetic tunnel junctions with perpendicular anisotropy. Appl. Phys. Lett. **100**, 132408 (2012)
202. K. Jung, H. Cho, Y. Hahn, D. Hays, E. Lambers, Y. Park et al., Comparison of Cl2/He, Cl2/Ar, and Cl2/Xe plasma chemistries for dry etching of NiFe and NiFeCo. J. Electrochem. Soc. **146**, 1465–1468 (1999)
203. H. Cho, K.-P. Lee, K. Jung, S. Pearton, J. Marburger, F. Sharifi et al., Corrosion-free dry etch patterning of magnetic random access memory stacks: effects of ultraviolet illumination. J. Appl. Phys. **87**, 6397–6399 (2000)
204. N.D. Altieri, J.K.-C. Chen, L. Minardi, J.P. Chang, Plasma–surface interactions at the atomic scale for patterning metals. J. Vacuum Sci. Technol. A Vacuum Surf. Films **35**, 05C203 (2017)
205. A.A. Garay, J.H. Choi, S.M. Hwang, C.W. Chung, Inductively coupled plasma reactive ion etching of magnetic tunnel junction stacks in a CH3COOH/Ar gas. ECS Solid State Lett. **4**, P77–P79 (2015)

206. A.A. Garay, S.M. Hwang, J.H. Choi, B.C. Min, C.W. Chung, Inductively coupled plasma reactive ion etching of CoFeB magnetic thin films in a CH3COOH/Ar gas mixture. Vacuum **119**, 151–158 (2015)
207. A. Orland, R. Blumenthal, Metal etching with organic based plasmas. II. CO/NH 3 plasmas. J. Vacuum Sci. Technol. B Microelectron. Nanometer Struct. Process. Measure. Phenomena **23**, 1597–1602 (2005)
208. X. Gu, Y. Kikuchi, T. Nozawa, S. Samukawa, A new metallic complex reaction etching for transition metals by a low-temperature neutral beam process. J. Phys. D Appl. Phys. **47**, 322002 (2014)
209. K. Kinoshita, T. Yamamoto, H. Honjo, N. Kasai, S. Ikeda, H. Ohno, Damage recovery by reductive chemistry after methanol-based plasma etch to fabricate magnetic tunnel junctions. Japan. J. Appl. Phys. **51**, 08HA01 (2012)
210. K. Kinoshita, H. Utsumi, K. Suemitsu, H. Hada, T. Sugibayashi, Etching magnetic tunnel junction with metal etchers. Japan. J. Appl. Phys. **49**, 08JB02 (2010)
211. N. Matsui, K. Mashimo, A. Egami, A. Konishi, O. Okada, T. Tsukada, Etching characteristics of magnetic materials (Co, Fe, Ni) using CO/NH3 gas plasma for hardening mask etching. Vacuum **66**, 479–485 (2002)
212. K. Kinoshita, H. Honjo, S. Fukami, H. Sato, K. Mizunuma, K. Tokutome et al., Process-induced damage and its recovery for a CoFeB–MgO magnetic tunnel junction with perpendicular magnetic easy axis. Jpn. J. Appl. Phys. **53**, 103001 (2014)
213. J. Jeong, T. Endoh, Improvement of electric and magnetic properties of patterned magnetic tunnel junctions by recovery of damaged layer using oxygen showering post-treatment process. J. Appl. Phys. **117**, 17D906 (2015)
214. J. Jeong, T. Endoh, Ion beam etching process for high-density spintronic devices and its damage recovery by the oxygen showering post-treatment process. Jpn. J. Appl. Phys. **56**, 04CE09 (2017)
215. V. Ip, S. Huang, S.D. Carnevale, I.L. Berry, K. Rook, T.B. Lill et al., Ion beam patterning of high-density STT-RAM devices. IEEE Trans. Magn. **53**, 1–4 (2017)
216. C.-W. Chien, D.-Y. Wang, K.-H. Shen, S.-H. Huang, K.-M. Kuo, S.-Y. Yang et al., Evidences of reactive-ion-etching-induced damages to the ferromagnet of perpendicular magnetic tunnel junctions. IEEE Electron Device Lett. **34**, 241–243 (2013)
217. J. Wong, K. Wong, C. Wang, K. Lee, K. Yamane, V. Naik, et al., “CMOS-embedded STT-MRAM arrays in 2x nm nodes for GP-MCU applications (2017)

Current-Driven Domain Wall Dynamics in Magnetic Heterostructures for Memory Applications

Sachin Krishnia and Wen Siang Lew

Abstract Conventional semiconductor based data storage devices will ultimately fail to meet the increasing demand for the vast computation and storing capacities. In 2008, IBM scientists have developed a new concept of memory, which is based on driving of magnetic domain walls (DWs) along a nanowire using an electric current. In this chapter, we first discuss the efficient current-induced nucleation of DWs. We then review the mechanism of various driving forces for DWs such as spin-Hall effect, Rashba effect, Dzyaloshinskii-Moriya interaction etc., in out-of-plane ferromagnetic materials interfaced with heavy-metals. Changing the ferromagnetic material to synthetic antiferromagnets resulted in stable, faster and efficient DW dynamics. An additional driving force, Ruderman-Kittel-Kasuya-Yosida exchange coupling torque, drives the domain walls at very high speeds in synthetic antiferromagnetic wire. Finally, the thermal stability of DWs and spin–orbit torques in synthetic antiferromagnetic structures are discussed.

1 Introduction

Ferromagnetic materials contain groups of magnetic moments aligned in same direction even in the absence of external magnetic field. However, not all the magnetic moments are necessarily aligned in the same direction in a macroscopic piece of ferromagnet. The ferromagnetic materials may be demagnetized to minimize its net magnetization. In the demagnetized ferromagnetic material, there are regions within which all the magnetic moments are aligned in the same direction, called magnetic domains. These magnetic domains are separated by thin boundaries, called magnetic domain walls (DWs) [1–3]. Dynamics of the DWs plays an important role in the magnetization reversal of the magnetic materials [4]. The DWs dynamics often governs the operation of memory and logic devices, i.e. switching and shifting of “0” and “1” digital bits [5–10].

S. Krishnia · W. S. Lew (✉)
School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore
e-mail: wensiang@ntu.edu.sg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_3

Understanding the interactions between electric current and the DWs, a phenomenon that was first investigated more than 30 years ago in magnetic thin films [11, 12], has become of enormous interest over the past decade. Although, there are several possible ways in which the current can interact with the DWs, however, the interactions those can drive the DWs, are of the most interest to develop future spintronic devices. Amongst the key emerging memory technologies, spin-transfer torque magnetic random access memory (STT-MRAM) is the predominant non-volatile memory, expected to have infinite endurance along with the ability to scale down below 20 nm.[1]

Though there is no inherent magnetic degradation in STT-MRAM, there is however an electrical breakdown of the MgO tunnel barrier that causes rapid aging of the device. Also, the read and write paths are identical in the STT-MRAM devices that may cause the unintentional writing by read current. To avoid this issues, the write voltage must be kept low, below 400 mV [13]. A solution for this is the SOT-based MRAM—a three terminal device that separates the read and write path [14, 15]. However, SOT-MRAM poses scaling issue to low dimensions. Moreover, the fabrication and process steps in MRAM are complex.

A possible memory alternative which combines the positive attributes of MRAM along with having simplified fabrication steps and avoiding the usage of tunnel barrier is DW based memory devices. The first prototype of the DW based memory device was proposed by IBM scientists [16]. The information in DW devices are stored in the form of magnetic domains that are separated by DWs [7, 10, 17]. The DW devises also offer as a higher speed and reliable alternative of hard-disk drives (HDD) since it does not have moving parts. On one end the bits may be written by changing the magnetization directions in the nanowire. Operation of the DW devices is analogous to a non-volatile shift register. The bits can be pushed by using electric current pulses and read using a magnetic tunnel junction or Hall probe as the sensor. An operational mechanism of DW memory device is illustrated in Fig. 1. The device speed and reliability depend on how fast the DWs can be shifted in the nanowire i.e. the DW velocity, while the storage density is determined by how closely the DWs can be placed within a nanowire as well as how closely the nanowires can be placed. The DW devices offer a possibility to re-orient the structure in the vertical direction and store the bits in a three-dimensional geometry, thereby increasing storage density. Therefore, controlled nucleation, high speed motion and detection of the DWs are the key prerequisites for making the memory devices feasible.[2]

[1] *Handbook of Magnetism and Advanced Magnetic Materials*. Edited by Helmut Kronmüller and Stuart Parkin. Volume 5: *Spintronics and Magnetoelectronics.*2007, John Wiley & Sons, ISBN: 978–0-470–02,217-7.

[2] *Introduction to Magnetic Random-Access Memory*. Edited by Bernard Dieny, Ronald B. Goldfarb, and Kyung-Jin Lee. 2016, Wiley-IEEE Press, ISBN: 978–1-119–00,974-0.

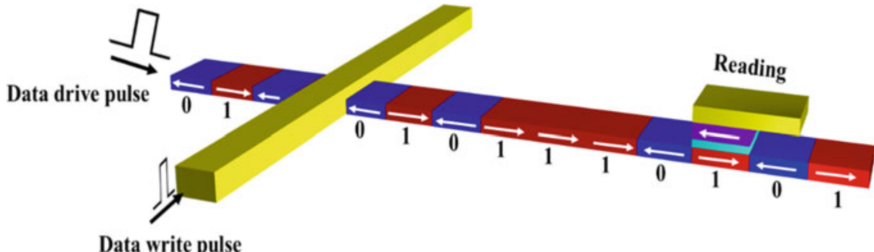


Fig. 1 Illustration of a magnetic DW memory device. The data '0' and '1' are stored in magnetic domains of opposite magnetization shown by blue and red colors, respectively. The DWs can be generated by applying a current pulse to the injection line: data write pulse. The DWs are then shifted towards the reading sensor by injecting spin-polarized currents into the nanowire: data drive pulse

2 Data Writing: Domain Wall Injection

Efficient and controlled DW injection into ferromagnetic wires is essential step for the realization of DW memory devices. In particular, the DWs separate two data bits in a memory device while motion of the DWs leads to reading and writing processes. In this section, we shall discuss the two main methods of DW injection into the DW memory devices: (1) Domain wall injection by local Oersted field and (2) In-line domain wall injection.

2.1 Domain Wall Injection by Local Oersted Field

Nucleation of a DW by generating local Oersted field in a current carrying strip-line is the most commonly used method in DW-related research and we name this method: Field Based Injection (FBI). This method is illustrated in Fig. 2. As can be seen from the figure, The FBI method requires the inclusion of an additional current-carrying line, called strip-line, to be patterned orthogonal to the data-carrying nanowire [18–21].

In this method, the data can be written into the nanowire by applying current pulses to the strip-line. The current in the strip-line then generates a strong local magnetic field around it which in turn the magnetization reversal underneath the strip-line. For instance, when the magnetization of the data nanowire is originally pointing in the $-z$ direction, it is possible to create a new data bit at the right side of the strip-line by applying current along the $-y$ direction, as shown in Fig. 2. This is because the $-y$ current will generate a counter-clockwise magnetic field around the strip-line. The magnetic field that is generated at the right side of the strip-line will point to $+z$ direction, which is opposite to the original direction of the local magnetization. The magnetic field will then flip the local magnetization at that area to create a new

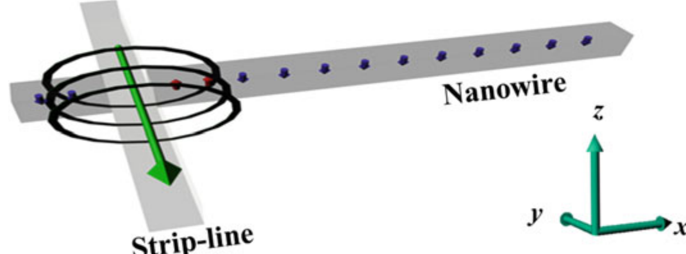


Fig. 2 Domain wall nucleation by passing current through a strip-line to generate Oersted field. The blue and red arrows represent the magnetization along +z direction and –z direction, respectively. The wire was initially saturated along –z direction. The current into the strip-line generates Oersted field and nucleate a domain of reverse magnetization which is along +z direction (red arrows)

data bit. The FBI method also work in the case when the data nanowire has in-plane magnetization.

Scanning electron microscope (SEM) image of such a device is shown in Fig. 3a. The film stack of the devices Ta(3)/Pt(3)/Co(0.25)/[Ni(0.5)/Co(0.25)]$_4$/Pt(3)/Ta(3), was sputter deposited on thermally grown SiO_2 substrate at a base pressure of 2

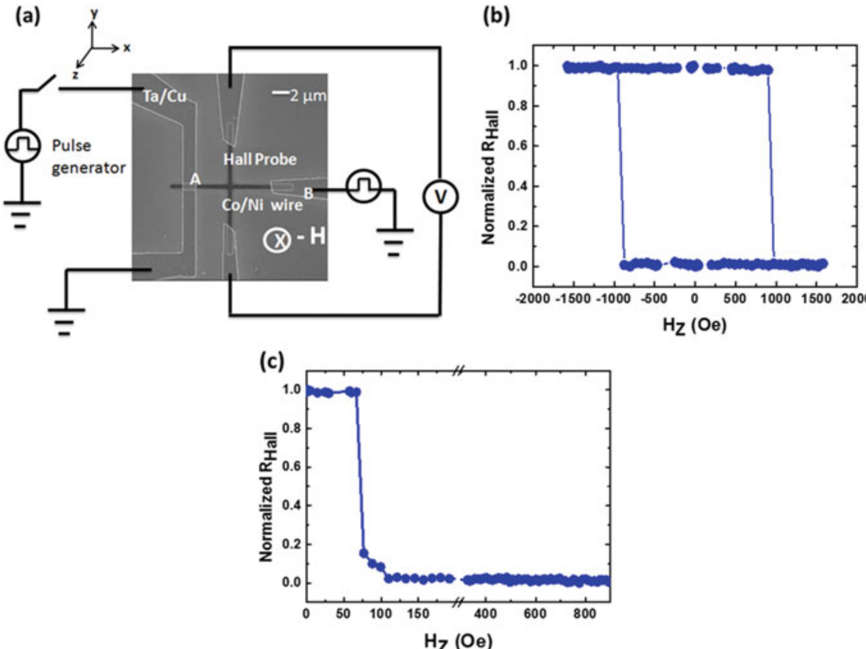


Fig. 3 **a** Scanning electron microscopy of the fabricated device with external circuit connections. **b** Normalized Hall resistance variation of the nanowire when out-of-plane external field is swept **c** Field induced domain wall driving. The Hall resistance drops at the de-pinning field [22]

$\times 10^{-8}$ Torr. The hard-axis anisotropy (H_K) of the thin film was 5 kOe. The width of the devices was kept 300 nm. The current pulses were injected through electrode (Ta(5 nm) /Cu (100 nm) /Au(10 nm)) to generate a local Oersted field. The successful nucleation of the DW was detected using anomalous Hall effect (AHE) by measuring Hall resistance (R_{Hall}) between the two electrodes. The Hall resistance is the proxy of perpendicular magnetization which is measured using a constant 50 μA bias current. The separation between electrodes A (B) and the Hall probe was kept 3.8 μm. Figure 3b shows the normalized R_{Hall} measured by sweeping an external magnetic field along the out-of-plane direction. As per our convention, the normalized $R_{Hall} = 1$ (0) corresponds to field saturation along the z-direction (−z-direction). A square hysteresis loop was observed, indicating a perpendicular easy axis of magnetization of the nanowire with a coercivity of 1 kOe [22].

In order to inject a DW, the nanowire is first saturated by applying a large global out-of-plane external field along z-direction, followed by application of a current pulse (85 mA, 50 ns) to electrode A. The local Oersted field generated by the pulsed current, nucleates a DW underneath the electrode A. The DW is then driven by an external magnetic field. Figure 3c shows the plot of normalized R_{Hall} obtained by sweeping an external magnetic field along +z direction. Two steps are observed in R_{Hall} as the field was gradually ramped. The first step at 60 Oe corresponds to DW propagation and pinning at the junction between the Hall probe and the nanowire. The second step at 100 Oe corresponds to DW depinning at the Hall probe and propagation through the nanowire and the Hall probe [22]. This method is commonly used to inject the DW locally.

2.2 In-line Domain Wall Injection

The second method is through in-line injection, which does not require an additional DW injector line to the device. Here, DW is injected to the nanowire using the same contact pad used to drive the DW. There are several variations to this method. In one variation, Sethi et al., have shown the DWs nucleation at a cross-bar structure within the nanowire [22]. The demagnetization energy at the Hall structure is higher which increases the susceptibility of the local magnetization and thus makes it possible for DWs to be nucleated. An SEM image of the devices is shown in Fig. 3a. The device parameters are discussed in the previous Sect. 2.1.

A pulsed current was applied between electrodes A & B to study the effect of in-plane current on the domain wall nucleation process. The nanowire was initially saturated in −z direction and electrical pulses of density varying from 6.8×10^{11} to 1.52×10^{12} A/m^2 were injected through electrodes A & B. The pulse width was kept constant at 50 ns. Subsequently, the external magnetic field was gradually increased in the +z-direction and change in the R_{Hall} was measured as shown in Fig. 4a. The change in R_{Hall} was not observed for low current densities, indicating no change in the magnetization of the nanowire. However, the R_{Hall} was dropped at external field strength of 150 Oe when a pulse of current density 8.7×10^{11} A/m^2, was injected

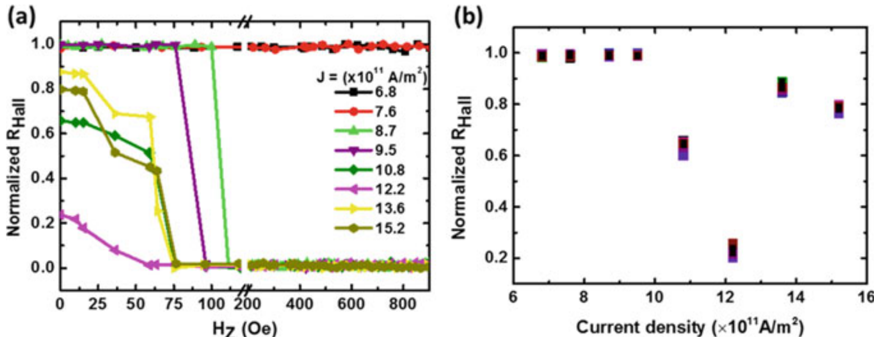


Fig. 4 Anomalous Hall effect (AHE) measurements on a device **a** Normalized R_{Hall} variation with external magnetic field as a function of current density. **b** Statistical distribution of R_{Hall} as a function of current density [22]

[22]. This indicates magnetization reversal at the Hall cross junction. At the Hall cross junction, the demagnetization energy is higher and the spins at the edge experience lesser exchange interaction thereby increasing the probability of DW nucleation in this region. The higher demagnetization energy generates a gradient in anisotropy much like the artificial generation of anisotropy gradient by Phung et al. [23]. As the current density increases, the drop in R_{Hall} occurs at lower field strength.

The spins at the edges of the Hall cross are canted due to the fringing field and are uncompensated, which are STT-driven and responsible for DW nucleation and propagation. The spin configuration at the Hall probe reverts to the original magnetization state and the R_{Hall} recovers to the original value after the DWs are de-pinned from it. The plot of normalized R_{Hall} with respect to the applied current density is shown in Fig. 4b. The spread of R_{Hall} indicates non-uniformity in reversed magnetization volume, which indicates the presence of stochasticity in the nucleation process. There is a possibility of multiple DW generation contributing to the stochasticity.

The Kerr images depicting direct observation of DWs at the Hall cross and nucleation of multiple DWs are shown in Fig. 5. Figure 5a shows the Kerr image of the device magnetized along the −z-direction, Fig. 5b shows the image after an in-plane current is injected. The dark contrast at the Hall junction represents the switched magnetic domains. Further set of trials lead to the nucleation of multiple domains on passing the current as shown in Fig. 5c. The images emphasize the event of DW depinning and propagation away from the Hall junction after the nucleation. The Kerr images are taken on 2 μm width wires to get better resolution. Figure 5d–e shows the images on 1.5 μm width wires. It is worth noting that although the electrical measurements are performed on 300 nm wide nanowires, the width of the nanowire and shape anisotropy would not play much role in perpendicular magnetic anisotropy (PMA) wires hence the nucleation process is similar in larger width nanowires.

Now we discuss how the electrical pulse width modulates the DW nucleation probability. Figure 6a shows the variation of R_{Hall} with respect to the magnetic field which is swept after application of current pulses of a density 1.22×10^{12} A/m^2

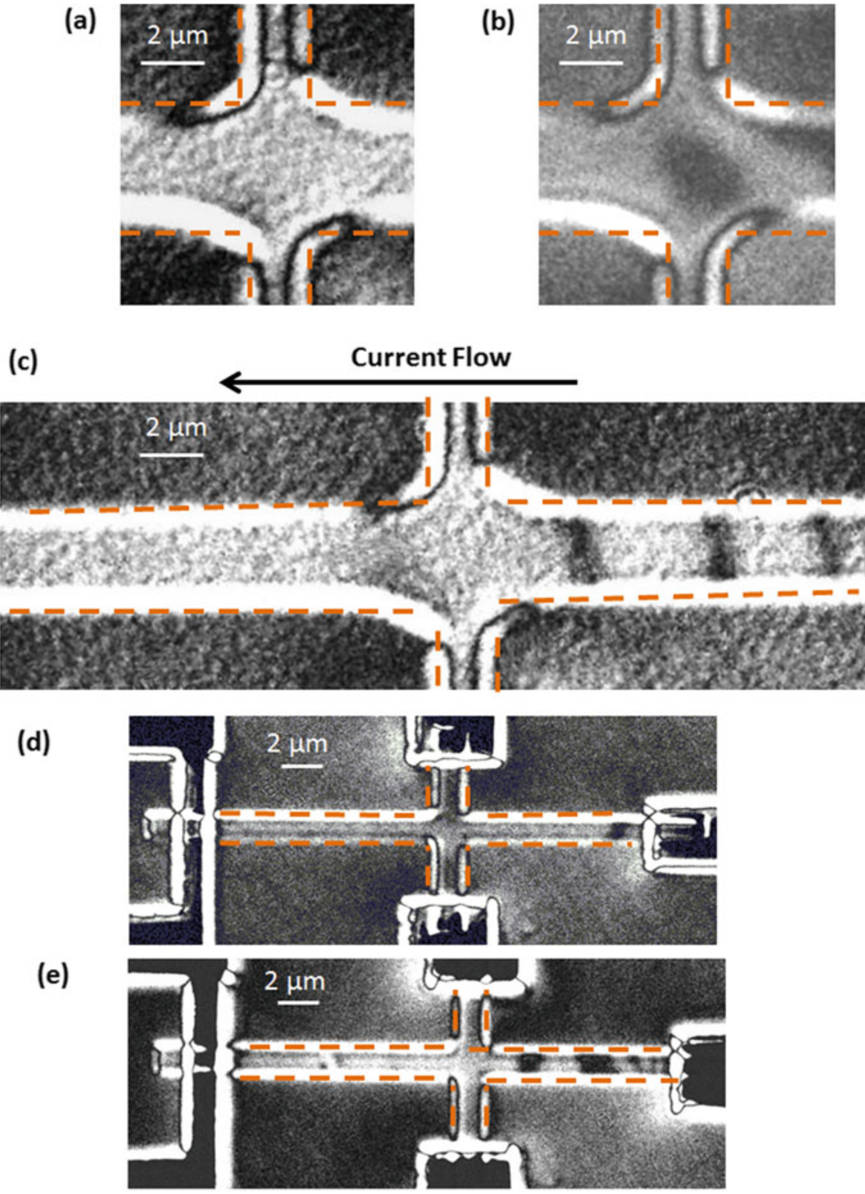


Fig. 5 **a** Kerr image of device saturated in −z-direction **b** Injection of in-plane current reverses magnetization in the Hall cross junction and contrast changes to dark **c** Nucleation and de-pinning of multiple domain walls (DWs) in 2 µm nanowire **d** In-plane current pulse application to 1.5 µm nanowire nucleates and drives a single DW **f** Similar magnitude of current pulse nucleates and de-pins multiple DWs in 1.5 µm nanowire

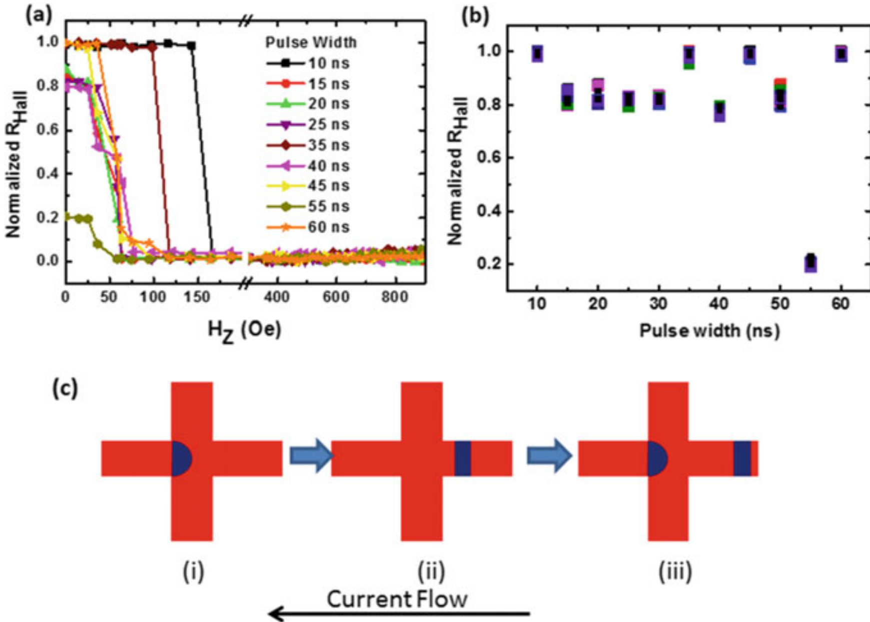


Fig. 6 AHE measurements on a device **a** Normalized R_{Hall} variation with external magnetic field as a function of pulse width for a current density 1.22×10^{12} A/m^2 **b** Statistical distribution of R_{Hall} as a function of pulse width. Each measurement was repeated 20 times. **c** Schematic depicting a possible scenario showing multiple DW nucleation and propagation across the Hall junction (i) DW nucleation via expansion of reversed magnetic domains at the Hall junction (ii) Depinning and propagation of DW away from the Hall junction (iii) Nucleation and expansion of second DW at the Hall junction [22]

with different pulse widths. The R_{Hall} drops at relatively larger magnetic fields for 10 ns and 15 ns pulse widths. This indicates that at lower operating power, current is insufficient to cause magnetization reversal and requires assistance from the field. Once DWs are nucleated, the de-pinning of DWs from the Hall junction occurs at around 75 Oe as shown in the Fig. 6a. A normalized R_{Hall} after the application of current pulse without applying any magnetic field is shown in Fig. 6b. When the pulse width lies in the range of 15–50 ns, the R_{Hall} varies between 0.8 and 1 indicating no or small reversal at the corners of the Hall junction. Maximum reversal at the Hall junction is observed for pulse width of 55 ns at which the R_{Hall} falls to its minimum value. The R_{Hall} recovers to its maximum value upon increasing the pulse width to 60 ns indicating depinning of the DWs from the Hall junction as shown in Fig. 5c–e. All these results indicate a possibility of multiple DW generation as illustrated by a schematic in the Fig. 5c.

Interestingly, the DW nucleation was not observed in nanowires without Hall cross geometry. This implies that Hall cross plays a crucial role in the DW nucleation process. Other factors, for instance Joule heating may assist the nucleation process but are not enough to cause the nucleation process on their own. The thermal activation

would be more dominant if $K_uV \leq 60k_BT$ [24]s, where K_UV is the product of perpendicular anisotropy constant and the magnetic volume. For our device, $K_U = 1.88 \times 10^5$ J/m^3, $V = 1.5 \times 10^{-6}$ (width) $\times$ 2.8×10^{-9} (thickness) $\times$ 8×10^{-9} (DW width), where DW width is estimated using $\Delta = \sqrt{A/K_U}$ and $A = 1 \times 10^{-11}$ J/m [25]. This gives the product $K_UV = 40$ eV, which is $1400k_BT$. Thus temperature alone cannot account for compensating the anisotropy.

3 Current-Induced Domain Wall Driving

In the previous section, we have discussed various methods of the DW injection in nanowires. These DWs are then shifted to the read sensor for the operation of a memory device. There are two main methods to drive the DWs: by magnetic fields and by electric currents. The magnetic field induces bi-directional motion of DWs i.e. the magnetic domain that are oriented parallel to the external magnetic field expand while the magnetic domains oriented antiparallel to the external magnetic field shrinks.[3] In the case of electric current, all the DWs move in the same direction that makes current-driven DW dynamics important to be studied for device application. Here, we describe the mechanisms of DW dynamics induced by current. Two torque transfer mechanisms have been proposed in literature namely, spin-transfer torque (STT) in conventional ferromagnets [26–28]. and spin–orbit torques (SOTs) in systems with heavy-metal and ferromagnetic interface[29–31]. The magnetization dynamics is governed by the Landau-Lifshitz-Gilbert (LLG) equation which is discussed in the following sections.

3.1 Spin-Transfer Torque Driven Domain Wall Dynamics

In this section, we will discuss how a DW is driven in order to transfer the data from the strip-line to the read sensor. The first method is through direct application of electrical current to the ferromagnetic nanowire [18, 32–34]. By doing so, the injected current will be spin polarized as it travels through the ferromagnetic nanowire due to the spin polarization effect. When the conduction electrons or unpolarized spins are injected into a ferromagnetic wire of a specific magnetization direction, the unpolarized spins are scattered by local magnetic moments through *s-d* [35] or *s-f* [36] exchange interactions. The spins which are parallel to the local magnetic moments are less scattered while the others are more scattered. The spin dependent scattering generates spin current in which majority of the spins of conduction electrons become aligned along the direction of the local magnetic moments. A schematic illustrating

[3] *Spin Dynamics in Confined Magnetic Structures III*. Edited by Burkard Hillebrands and Andre Thiaville. Series: *Topics in Applied Physics.* 2006, Springer, ISBN: 9,783,540,398,424.

spin polarization of the conduction electrons is shown in Fig. 7. When the spin-polarized current within the nanowire arrives at areas where the local magnetization is not uniform, i.e. where the DW is present, there will be an angular momentum transfer from the spin-polarized current to the local magnetization and vice versa. The transfer of angular momentum from the spin-polarized current to the local magnetization is known as adiabatic Spin Transfer Torque (STT). The DW magnetization is rotated along the spin-polarized currents to conserve the total angular momentum that leads to the DW motion. This concept was first proposed by Berger in 1984 [12]. He demonstrated the DW motion using microsecond long pulses in Permalloy ferromagnetic thin films [11, 12]. Later in 2000s, the concept was applied to move the DWs in ferromagnetic nano-strips, motivated by the possibility of realizing spintronics memory devices. Furthermore, the experimental demonstration of the DW motion in nano-strips by the application of electric currents of relatively low magnitude, stimulated further research in this area [37–39]. The STT-induced DW motion is mostly studied in in-plane Permalloy nanowires. The critical current density required to move the DWs in Permalloy nanowires is found to be of the order of 10^{12} A/m^2. Also, high DW velocity of > 200 m/s has been achieved by IBM in 2008 [8, 40]. The dynamics of the current induced DW motion can be understood by incorporating adiabatic STT term into the LLG equation.

$$\frac{\partial \mathrm{M}}{\partial t} = -\gamma \mathrm{M} \times \mathrm{H}_{eff} + \alpha \mathrm{M} \times \frac{\partial \mathrm{M}}{\partial t} - (\mathrm{u}.\nabla)M. \quad (1)$$

The third term on the right hand side in the above equation represents the adiabatic STT and u represents the spin drift velocity and is given by-

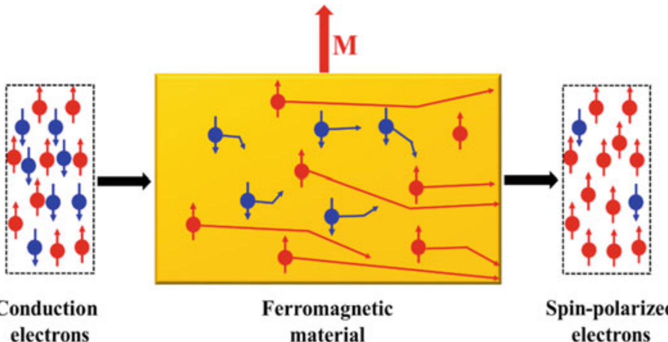


Fig. 7 A schematic showing spin current generation from charge current. The conduction electrons are injected into a FM of magnetization, M pointing along a specific direction (shown by a red arrow). The charge current has equal number of electrons with spin 'up' and spin 'down'. While travelling through the FM, spin 'up' and spin 'down' electrons experience different resistance due to magnetization dependent scattering that results into spin-polarized current

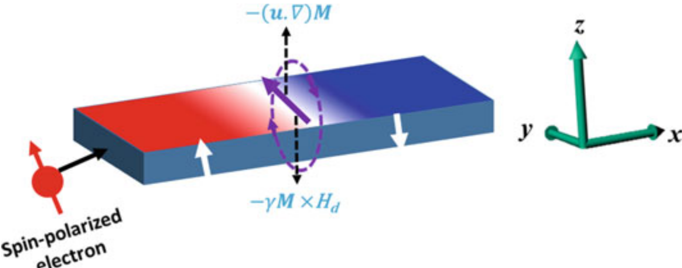


Fig. 8 Current driven DW dynamics for out-of-plane magnetized nanostrip through adiabatic STT. The DW moves by precession about the x-axis when the adiabatic torque can overcome the torque due to demagnetization field

$$u = \frac{g\mu_B P J_e}{2|e|M_s}, \tag{2}$$

where g is the Lande factor, P is the spin polarization, M_s is the saturation magnetization, J_e is the current density and μ_B is the Bohr magnetron.

The adiabatic STT acting on a DW would be proportional to the gradient of DW magnetization [41, 42]. Figure 8 shows the dynamics of a Bloch DW under the adiabatic STT in a perpendicularly magnetized wire. When the current is applied along $-x$ direction, the spins that are polarized along $+z$ direction, exert a damping-like torque on the DW, which tries to orient its magnetization along $+z$ direction. This motion induces the damping torque in the clock-wise direction that rotates the DW magnetization in-plane, thus deviating the DW from Bloch configuration. This initiates a demagnetizing torque $-|\gamma|\mathbf{m} \times \mathbf{H_d}$ along $-z$ direction countering the adiabatic STT torque. Here, $\mathbf{H_d}$ is the demagnetizing field. This torque in-turn generates a damping torque rotating magnetization along the counter-clockwise direction. Thus all the torques balance out and there is no motion of the DW. This is referred to as intrinsic pinning and there is a threshold current required to drive the DW [43].

The DW has an intrinsic pinning and a minimum current density is required to move the DW which is generally referred as a threshold current density. However, the experimentally observed threshold current densities to move the DWs in Permalloy nano-strips were found to be one order less than that predicated by theoretical calculations. The calculations of adiabatic STT suggest the DW depinning current density as high as 10^{13} A/m^2 [44, 45]. Moreover, the threshold current density was found to depend on the DW pinning due to the defects [43]. The deviation in experimentally observed and theoretically predicated threshold current densities can be explained by considering an additional STT term: non-adiabatic STT. The modified LLG equation can be expressed as [46, 47]:

$$\frac{\partial M}{\partial t} = -\gamma M \times H_{eff} + \alpha M \times \frac{\partial M}{\partial t} - (u.\nabla)M - \beta M \times [-(u.\nabla)M]. \tag{3}$$

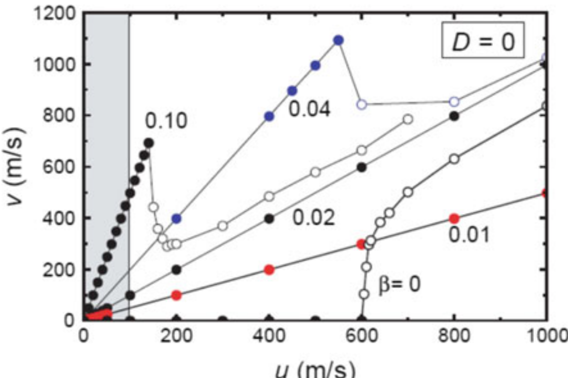


Fig. 9 Simulated DWs velocities for various values of non-adiabatic (β) spin transfer torques [27]

The fourth term on the right hand side in the above equation represents the non-adiabatic STT and β is the non-adiabatic parameter that defines the strength of the non-adiabatic torque.

The non-adiabatic STT term is orthogonal to the adiabatic STT term and is equivalent to a torque produced due to a magnetic field. The non-adiabatic torque makes the DW pinning extrinsic instead of the intrinsic and can lead to larger DW speeds without precessions. Since the symmetry of the non-adiabatic torque is that of a field, it is often called field-like STT. In the analogy to field driven DW dynamics, the non-adiabatic torque should drive the DW through rigid translation motion and precession motion for small and higher current densities, respectively and Walker breakdown should be observed in velocity-current curves for sufficiently large β values. All these experimental observed phenomena are well explained using Eq. (3) as shown in Fig. 9 [27, 42].

The presence of non-adiabatic component along with adiabatic STT makes the DW dynamics dependent on the ratio of β to α. When $\beta < \alpha$ and the current density is small, the field-like torque would be able to compensate the damping torque due to adiabatic STT and the demagnetizing field. The damping torque due to the non-adiabatic STT would compensate the torque due to non-adiabatic STT. Thus the contribution for DW motion would come from the torque due to demagnetizing field which would drive the DW forward without precession. On further increasing the current, the damping torque due to adiabatic STT is able to compensate the damping torque due to the demagnetizing field and the field-like torque i.e. the torque due to non-adiabatic STT. Thus the DW precesses, however, the velocity still increases since the precession generates a torque in the same direction as that due to demagnetizing field. When $\beta > \alpha$, the field-like torque becomes larger than damping torque due to the adiabatic STT and the demagnetizing field. This causes the DW to precess continuously and generate a torque to partially compensate the torque due to demagnetizing field. Thus, there is a decrease in DW velocity due to the precession. This corresponds to Walker breakdown[48–50] and the DW velocity decreases with increase in the current. The velocity of DW below the Walker breakdown is $\mathbf{v} = \beta\mathbf{u}/\alpha$ and approaches $\mathbf{u}$, when current increases beyond the Walker breakdown limit.

For the precessional regime of the DW motion, the DW velocity approaches $v \rightarrow u$ [27, 48].

The physical origin of non-adiabatic STT may be due to spin mistracking, i.e. when conduction electrons may get scattered from narrow DWs instead of tracking the moments and directly transfer linear momentum [44]. The ratio of non-adiabaticity parameter, β, to the Gilbert damping parameter, α, is crucial in ascertaining the DW dynamics. Some studies propose $\beta/\alpha \sim 1$ [51, 52], while more recent studies propose $\beta/\alpha \sim 10$ [53, 54].

It should be notated that the DWs are driven along the electron flow direction by STT when majority carriers are electrons with positive spin-polarization. This is in a sharp contrast with spin–orbit torque (SOT) induced DW dynamics. The SOT moves the DWs along either current flow or electron flow direction [31]. The direction of SOT driven DW motion depends on the signs of spin–orbit interactions, inversion symmetries etc., which are discussed in the following section.

3.2 Spin–Orbit Torque Driven Domain Wall Dynamics

Aside from the STT, the DWs can also be drive via spin–orbit torques that originate from spin–orbit interactions. These torques are often called spin–orbit torques (SOTs). An electron possesses orbital as well as spin angular momentum. These may be coupled to give rise to total angular momentum j, such that $m = \gamma j$ where γ is the gyromagnetic ratio. The basic principle is that in the rest frame of an electron, the nucleus revolves around it with speed v, generating a current loop $I = Zev/2\pi r$, where Z is the atomic number. This produces a magnetic field $\mu_0 I/2r$ at the center which is responsible for the spin–orbit coupling, $B_{so} = \mu_0 Zev \big/ 4\pi r^2$. The energy associated with the spin–orbit coupling can be expressed in terms of Bohr magneton and Bohr radius:

$$E_{so} \approx -\frac{\mu_0 \mu_B^2 Z^4}{4\pi a^3}. \tag{4}$$

The spin–orbit interaction becomes strong for heavy elements due to Z dependence. This is one primary reason for interfacial effects observed when heavy metals like Pt, Ta, W are used as underlayer or capping layer in the magnetic heterostructures.[4] In the following section, effects of various spin–orbit torques on the domain wall dynamics are discussed.

---

[4] *Magnetism and Magnetic Materials.* Edited by John Michael David Coey. 2010, Cambridge University Press, ISBN: 9,780,511,845,000.

3.2.1 Rashba Effect

Unlike the STT-induced magnetization switching, the SOT does not require polarization of spins, and thus it opens new prospects in magnetization manipulation to design energy efficient spintronics devices. In systems with broken inversion symmetry having a ferromagnet and heavy metal interface e.g. Pt/Co/AlO_x or Ta/CoFeB/MgO, there exists a source of SOT, namely, Rashba effect [55, 56]. These structures exhibit an asymmetric crystal field along the z-direction (out-of-plane) [57–59]. From the rest frame of electrons moving in the ferromagnet, the electric field is an equivalent magnetic field, which polarizes electron spin as shown in Fig. 10. The Rashba field would act as an in-plane field, favouring a Bloch configuration but not driving the DW forward.

In 2008, Moore et al. showed the DW speed of ~ 100 m/s against the electron flow in Pt/Co/Al_2O_3 nanowires [60]. The non-adiabatic component of the SOT in the DW (β) was estimated to be ~ 1 in a follow-up study [61]. These remarkable observations were first explained by Miron et al., by introducing current-induced field-like torque [55]. These torques are originated from the current flowing in adjacent heavy-metal layers. They observed asymmetric current-induced DW nucleation in the presence of external in-plane magnetic field in the structure with inversion asymmetry (Pt/Co/Al_2O_3). The DW nucleation behaviour in such structures was found to be dependent on relative directions of the in-plane field and the applied current as shown in Fig. 11.

The amplitude of the field-like effective field (H_{FL}) or Rashba field was estimated from the quantitative analysis of the DW nucleation and it was found to be 1000 Oe/10^{11} A/m^2. Later, numerous studies on the DW motion against the current flow, have been reported in the asymmetric structures: Pt/Co/Ni/Co/TaN [30, 62] and Pt/CoFe/MgO[31]. The torque from the Rashba field is given by:

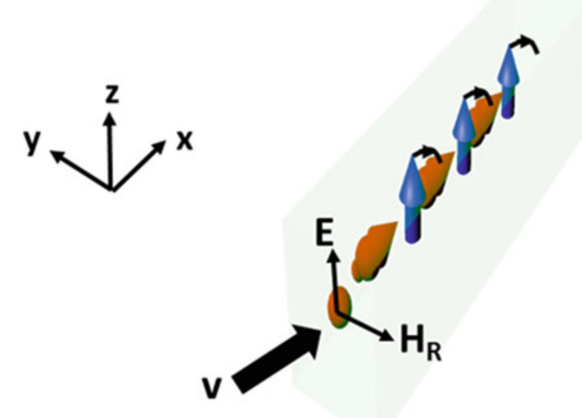


Fig. 10 Schematic illustrating Rashba effect in an out-of-plane magnetized nanostrip. The electrons with velocity 'v' see the electric field due to crystal potential which acts as Rashba magnetic field (H_R) in their rest frame. This tilts the moments slightly out-of-plane due to s-d interactions

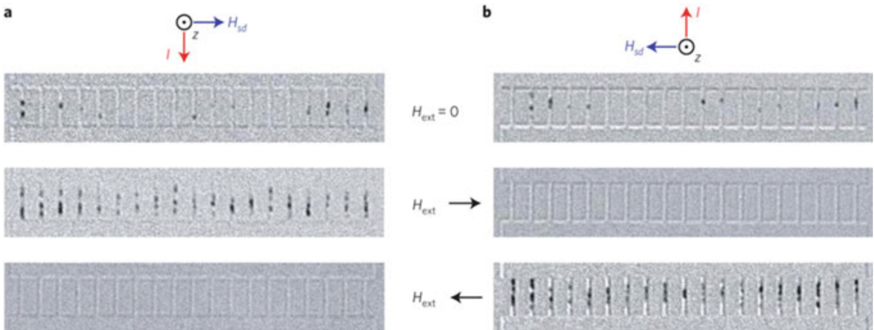


Fig. 11 Nucleation of asymmetric DWs *via* current injection into the wires in the presence of external magnetic field. For a specific current direction the DWs injection probability can be enhanced or suppressed by changing the direction of external magnetic field. This is interpreted as evidence of the presence of the field-like torque (H_{FL}) perpendicular to the current directions that either adds or subtracts to H_{ext}. For an opposite current, the action of H_{ext} on the domain nucleation is opposite [55]

$$\tau_{FL} = -\mathrm{m} \times \mathrm{H}_{FL}, \tag{5}$$

where m is the magnetization and H_{FL} is the Rashba field. The Rashba field is given by-

$$\mathrm{H}_{FL} = -2\frac{\alpha_R m_e}{\hbar|e|M_s} P|J_e|\left(\hat{z} \times \hat{\jmath}_e\right), \tag{6}$$

where α_R is the Rashba parameter, $\boldsymbol{J_e}$ is the direction of electron flow, and m_e is the mass of electron [55, 63]. The effective H_{FL} is an in-plane magnetic field that acts along the transverse direction of the wire and therefore, H_{FL} cannot drive the DWs in a PMA wire. However, the origin of higher DW speeds in Pt/Co/Al_2O_3 wires has been attributed to the combination of Rashba field and larger negative non-adiabatic STT. In addition, the Rashba field enhances the DW fidelity and consequently, relatively high DW speeds of ~ 400 m/s in the direction of current flow have been achieved [56]. The magnetic force microscopy images and DW schematics together with the H_{FL} direction are shown in Fig. 12.

Apart from large DW velocities, Miron et al., reported that the direction of DW motion was against the electron flow direction i.e. opposite to that expected from the STT model. Two possible scenarios were proposed to explain the experimental observations: (1) Both the adiabatic and non-adiabatic torques are negative and (2) the non-adiabatic torque is negative while the adiabatic torque is positive. The former explanation could be possible in case of negative spin polarization. The negative non-adiabatic torque could cause the DW motion along the current flow only below the Walker breakdown.

In 2012, Thiaville et al., proposed a new mechanism combining the damping-like torque from the SOT and an asymmetric exchange interaction originated from

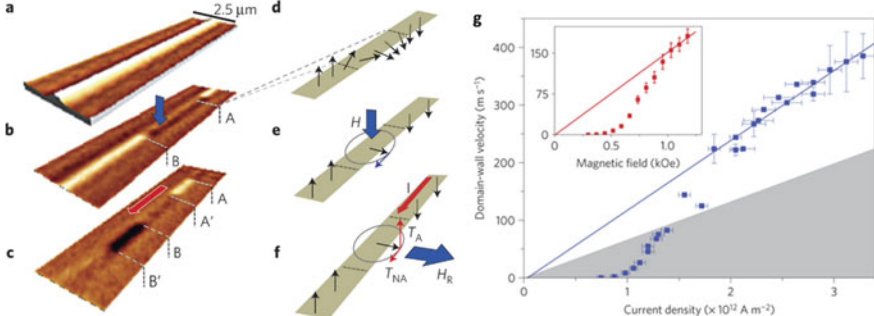


Fig. 12 **a**–**c** Magnetic contrast of a PMA wire measured using differential Kerr microscopy. **d**–**f** Schematic showing the directions of the field like torque (T_{NA}) due to Rashba field. **g** High speed DW velocity in PMA wire with current density. The current-induced DW motion exhibits creep and flow regimes similar to the field-driven DW motion. The solid line is the liner fit to the DW flow regime. Inset shows the field-driven DW velocity to compare the field-like nature of the current-induced Rashba field [56]

the spin–orbit coupling due to broken inversion asymmetry [64]. The asymmetric exchange interaction is often called Dzyaloshinskii-Moriya interaction (DMI). The proposed mechanism explains the direction and large velocities of the DWs without any signature of Walker breakdown. The mechanism of DMI is explained in next section.

3.2.2 Dzyaloshinskii-Moriya Interaction

The Dzyaloshinskii-Moriya interaction (DMI) is a three-site antisymmetric exchange interaction and known to originate when magnetic thin films are interfaced with a strong spin–orbit coupling materials [65–67]. It is an indirect interaction between two atomic spins S_i and S_j through an atom of high spin–orbit coupling, located in the adjacent heavy metal layer. The energy associated with the DM interaction, the DMI energy (E_{DM}) is often given by-

$$E_{DM} = \mathrm{D}_{ij} \cdot \left(\mathrm{S}_i \times \mathrm{S}_j\right), \tag{7}$$

here, D_{ij} is the DM interaction vector and its direction depends on the studied system. For a ferromagnetic thin film that is grown on heavy metal of high spin–orbit coupling, the DMI constant is given by-

$$\mathrm{D}_{ij} = d\ \mathrm{u}_{ij} \times \mathrm{z}, \tag{8}$$

where, u_{ij} is the unit distance vector between the two spins S_i and S_j, z is the unit vector perpendicular to the thin film plane from heavy metal to ferromagnetic thin film and d is a coefficient proportional to the spin–orbit coupling [68]. The DM

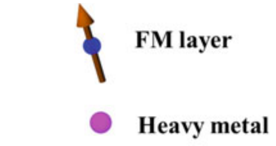
interaction produces a field known as the DMI field ($\boldsymbol{H}_{DMI}$), which is given by-

$$\mathrm{H}_{DMI} = -\frac{\partial E_{DM}}{\partial \mathrm{M}}. \tag{9}$$

The direction of the DMI field is collinear to the u_{ij} vector that introduces chiral magnetic textures into the ferromagnetic structures [69, 70]. In ferromagnetic structures with perpendicular magnetic anisotropy, the DMI field favours the stabilization of Néel DWs over Bloch with a set chirality [64]. The DMI field, $\mathbf{H_{DMI}}$ can be expressed in terms of $\mathbf{D}$ and the domain wall width, Δ [71, 72],

$$\boldsymbol{H}_{\boldsymbol{DMI}} = \frac{\boldsymbol{D}}{\mu_0 M_S \Delta} \tag{10}$$

The above equation can be used to estimate the DMI coefficient. There exists a minimum critical value of $\mathrm{H_{DMI}}$ such that it is able to overcome the magnetostatic energy and favour Néel DWs. Consider $\mathrm{H_D}$ as the DW anisotropy field such that, $\mathrm{H_D} = 4\mathrm{K_D}/\pi\mu_0\mathrm{M_S}$, with $\mathrm{K_D} = \mathrm{N_x}\mu_0\mathrm{M_S}^2/2$ being the magnetostatic DW anisotropy and $\mathrm{N_x} = \mathrm{t_f}\ \ln(2)/\pi\Delta$ is the demagnetization coefficient for magnetic film of thickness $\mathrm{t_f}$ [73]. Then Néel DWs are preferred if $\mathrm{H_{DMI}} > \mathrm{H_D}$ else Bloch DWs are stabilized in the thin films [71, 74]. A sketch representing the DMI vector D_{ij} due to the antisymmetric exchange interaction among spins S_i, S_j and an atom of high spin–orbit coupling is shown in Fig. 13 [75].



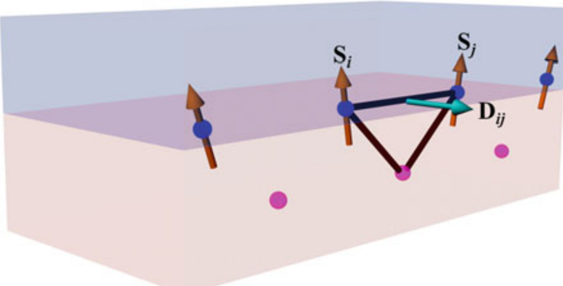


Fig. 13 A schematic showing the Dzyaloshinskii-Moriya interaction at the interface between a ferromagnetic layer and a heavy metal layer of high spin–orbit coupling

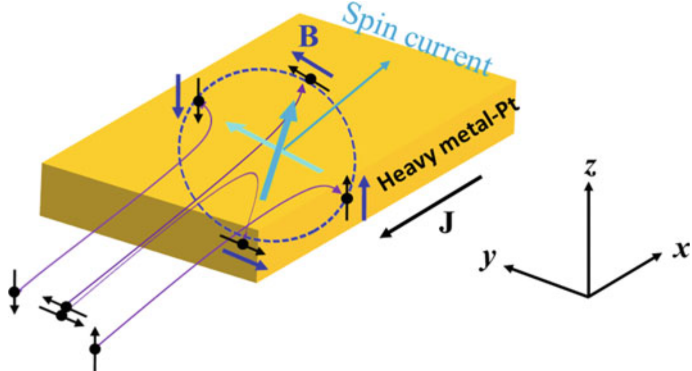


Fig. 14 Schematic showing the generation of spin current in heavy metals [79]

3.2.3 Spin-Hall Effect

In 2012, another source to generate the SOT via spin Hall effect (SHE) phenomenon was proposed by Liu et al. [76, 77]. They demonstrated that the ferromagnetic structures can be switched by applying the current to an adjacent heavy-metal layer of high spin–orbit coupling. When charge current is injected into the heavy-metal along $-x$ direction, the electrons move along $+x$ direction and the positive ions would move in $-x$ direction from moving frame of electron. The relativistic effect will generates an Oersted field. Consequently, the Oersted field sets up scattering preference for the conduction spins whose directions are parallel to the field direction as shown in Fig. 14. The conduction electrons of a particular spin orientation are scattered in a direction that is orthogonal to the current flow direction whereas the other spins are scattered in the opposite direction[78, 79]. The mechanism of the spin dependent scattering is shown by a schematic in Fig. 14.

The scattered electrons of the two different spin orientations accumulate at opposite interfaces of the heavy metal: this is called spin current. The subsequent spin current then diffuses into the adjacent ferromagnetic layer and excite the magnetic precession by transferring it's the angular momentum to the ferromagnetic layer. The torque from the spin Hall effect (τ_{SL}) is Slonczewski-like in nature and is given by [31]

$$\tau_{SL} = -\gamma_0 \mathrm{m} \times \mathrm{H}_{SL}, \tag{11}$$

where ***m*** is the magnetization vector and H_{SL} is the Slonczewski-like (SL) effective field which is given by-

$$\mathrm{H}_{SL} = -\frac{\hbar \theta_{SH} j_a}{2\mu_0 |e^-| M_s t} \mathrm{m} \times \hat{\mathrm{u}}_y, \tag{12}$$

where J_a is the applied current density in $-x$ direction, M_s is the saturation magnetization, t is the thickness of magnetic layer, $\boldsymbol{m}$ is the magnetization vector, $\hat{u}_y$ is a unit vector in y-direction, $|e^{-}|$ is the absolute value of the electron charge and θ_{SH} is the effective spin Hall angle. The spin Hall angle is the ratio of spin current (I_s) to charge current (I_e), i.e. $\theta_{SH} = {}^{I_s}/_{I_e}$. However, according to Eqs. (11) and (12), the torque from SHE can not drive the Bloch DWs. The DW configuration must be stabilized in Néel to drive it by the SHE torque. Haazen et al. have demonstrated the SHE induced DW motion by forcing the Bloch DW into Néel configuration using external in-plane magnetic fields [80]. Whether the magnetization switching is due to Rashba effect or spin Hall effect is still under debate. Liu et al. have shown that magnetization switching in Pt/Co/AlO_x stacks [76, 77] similar to that studied by Miron et al. [56], is due to the strong spin Hall effect in Pt heavy metal layer. Following these works, various experimental methods have been proposed to measure the Rashba field and SL-field in such structures and the SL field is found to be dominated over the Rashba field [81–84].

The combined LLG equation incorporating the effect of field, STT, SOT and DMI can be written as follows:

$$\frac{\partial \boldsymbol{m}}{\partial t} = -|\gamma|\boldsymbol{m} \times \boldsymbol{H}_{eff} + \alpha \boldsymbol{m} \times \frac{\partial \boldsymbol{m}}{\partial t} - (\boldsymbol{u} \cdot \nabla)\boldsymbol{m} - \beta \boldsymbol{m} \times [-(\boldsymbol{u} \cdot \nabla)\boldsymbol{m}] - |\gamma|\boldsymbol{m} \times [\boldsymbol{m} \times |H_{SL}|(\hat{z} \times \widehat{\boldsymbol{J}_e})] \quad (13)$$

where, H_{eff} includes Zeeman, magnetostatic, exchange, anisotropy and DMI fields, the third and fourth terms are torques due to STT and SOT, respectively. Here, the Rashba effect and field like term is excluded since its contribution is small [68].

3.3 *Chiral Domain Walls Dynamics Under Spin–Orbit Torques*

In a PMA ferromagnetic wire, the DWs and their chirality can be fixed by making use of DMI field as described in Sect. 3.2.2. The direction of DMI field is along the wire long axis that helps in stabilizing Néel type of DWs over Bloch DWs irrespective of the wire width[64]. In addition, the DMI also fixes the chirality of the Néel DWs, therefore the DWs magnetization for up-down domains is opposite to that for down-up domains as shown by a schematic in Fig. 15.

When current is applied to the wire, the magnetization of Néel DWs rotates into the transverse direction of the wire due to the torque from the spin Hall effect in heavy metal layer. Since the DMI field is always along the wire long axis, this rotation of the DW from the Néel configuration induces an angle between the DMI field and the DW magnetization, consequently inducing an out-of-plane torque on the DW. Sign of the DMI torque is opposite for up-down domains to that of the down-up domains

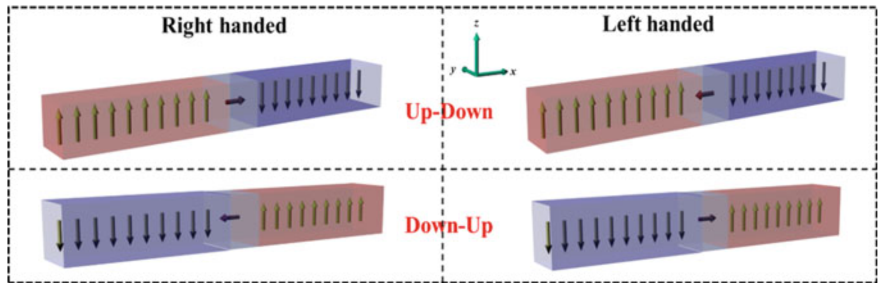


Fig. 15 DMI field-induced chirality of the DWs. The direction of DMI field in FM layers that are grown on Pt underlay, is along the x-axis that stabilize the DW magnetization into Néel configuration. In addition, the DMI field also introduce chirality to the Néel DW. The direction of DW magnetization for right-handed and left-handed chiral DWs are shown

and therefore, DWs in both the configurations: up-down and down up, move in the same direction irrespective of the DW magnetization direction. The motion of a left-handed chiral Néel DW under the SOT is shown in Fig. 16. The SOT-driven chiral DWs motion is observed in PMA FM wires that are grown on heavy metals such as Pt, Ta, Ir and significant variations in the DW velocities are observed due to the differences in the DMI field and SOT strengths [30].

The first indirect evidence of the role of DMI on DW dynamics in multilayer structures was measured by Ryu. et al. [30] and Emori et al. [31]. They measured the current-induced DW dynamics in the presence of in-plane magnetic fields on Pt/Co/Ni, Pt/CoFeB/MgO and Ta/CoFeB/MgO structures. The in-plane magnetic field modifies the DW velocity as shown in Fig. 17. Depending on the direction of H_x, whether it is parallel (or antiparallel) to the domain wall magnetization, the

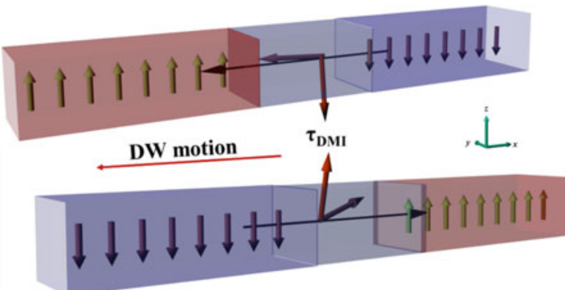


Fig. 16 DMI field driven left-handed chiral Néel DW dynamics in up-down (upper) and down-up (lower) configurations. The direction of DMI field is shown by a black arrow from down domain to up domain. The SOT drives the Néel DWs magnetization transverse to the wire (y-axis) that induces an angle between the DW magnetization and DMI field and simultaneously exerts DMI torque (red arrow). The direction of the DMI torque is in such a way that the DWs in both up-down and down-up configurations move in the same direction

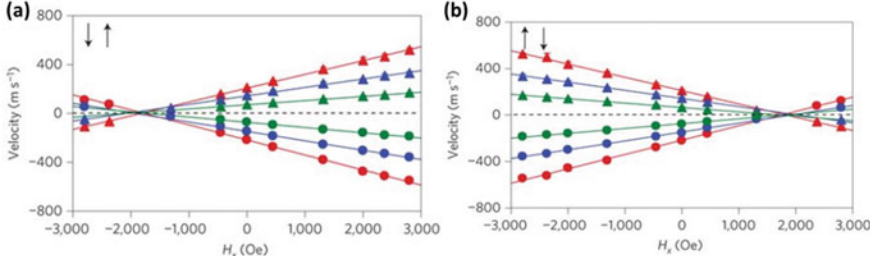


Fig. 17 Velocity of **a** down-up and **b** up-down DWs with external magnetic field, H_x for various currents in Pt/Co/Ni multilayer structures. The down-up DW velocity becomes zero at around $H_x = -2$kOe and the up-down DW at around +2kOe. At this field the Néel DW transforms into Bloch DW [30]

velocity of up-down (down-up) DW increase (decrease). This indicate that the magnetization of two consecutive DWs (up-down and down-up) is opposite. Moreover, the DWs velocities becomes zero at a certain longitudinal magnetic fields. This field is considered as the field required to overcome the DMI field and it reverses the DW configuration from Néel to Bloch. Observation of DW motion in opposite directions in Pt/CoFeB/MgO and Ta/CoFeB/MgO structures suggests that the sign of DMI in the two multipliers have same direction and sign of SHE is opposite in Pt and Ta [31].

To summarize, depending on the signs of DMI and SOT, the DWs in multilayer structures can be driven along either current flow or electron flow directions. The DW velocities in such structures are observed to be higher than the STT driven DWs and can be larger than 300–400 m/s, but the fringing field from the DWs limits the density of the storage devices.

4 Domain Wall Dynamics in Synthetic Antiferromagnetic Wires

Despite the higher efficiency and versatility of DW dynamics driven by a combined torque: induced from spin Hall effect and DMI in PMA wires, the magnetic dipole fringing field that each DW produces, limits the data storage density in the DW memory devices. Purnama et al., have shown that the DW stray field can be minimized in bilayer systems by making use of coupled Néel DWs of opposite magnetization [85]. Yang et al., have explored current-induced DW dynamics in nanowires formed from artificial antiferromagnetic systems or synthetic antiferromagnetic structures (SAF) to minimize the dipolar coupling between the DWs [86]. These structures are formed of a thin ferromagnetic layer coupled with another ferromagnetic layer through an ultrathin nonmagnetic spacer layer *e.g.* Ru, Ta, Cu etc. The two ferromagnetic magnetic layers are mirror image of each other as shown in Fig. 18.

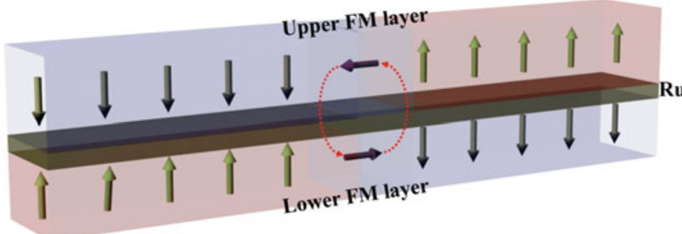
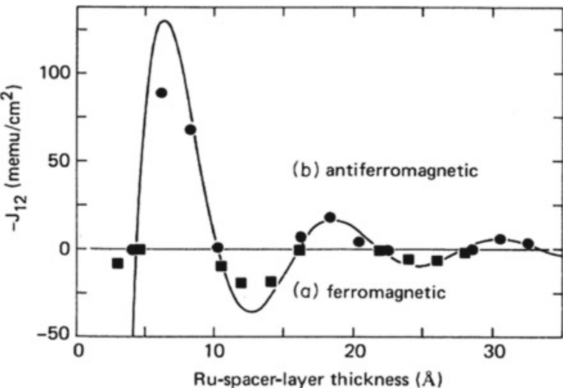


Fig. 18 Schematic representing domains and DWs in antiferromagnetically coupled bilayer structures. Minimization of dipolar coupling between the DWs via flux closure in SAF structures. (Red arrows)

Fig. 19 Dependency of the exchange coupling on Ru thickness in Co/Ru/Co trilayers structures [87]



The coupling between the two ferromagnetic layers originates from long-range oscillatory interaction that is known as Ruderman-Kittel-Kasuya-Yosida (RKKY). The RKKY interaction is an indirect exchange interaction that decreases in magnitude with thickness of the spacer layer. The sign of exchange coupling constant (J_{ex}) depends on the nonmagnetic spacer layer thickness thus exhibiting an oscillatory coupling between ferromagnetic and antiferromagnetic as a function of spacer thickness [87, 88]. The sign and strength of coupling constant as a function of Ru thickness in Co/Ru/Co trilayers is shown in Fig. 19. The net magnetization of the SAF structures can be tuned to zero by carefully calibrating the magnetization of the two ferromagnetic layers.

4.1 Functionality of Exchange Torque in Synthetic Antiferromagnetic Wires

First we explain how the interlayer exchange torque improves the DW dynamics in SAF wires. Let us consider two DWs having magnetic moments M_L and M_U that are

coupled to each other via exchange interaction J_{ex} as shown in Fig. 18. We assume that M_L and M_U are single spins. The energy of the system (E_{ex}) can be written as:

$$E_{ex} = -2J_{ex}\vec{M}_L.\vec{M}_U = -2J_{ex}M_L M_U \cos\theta, \tag{14}$$

where 'θ' is the angle between the two DWs and 'J_{ex}' is the interlayer exchange coupling constant. If $J_{ex} > 0$, the two DWs are aligned parallel to each other, *i.e.* $\theta = 0$, whereas if $J_{ex} < 0$, the two DWs are aligned antiparallel to each other, *i.e.* $\theta = \pi$. In both the cases, energy of the system is minimum according to Eq. (14), thus does not experience any torque. However, when this equilibrium position is perturbed through an external source such as current and its effect, the two collinear DWs are deviated away from $\theta = 0$ ($J_{ex} > 0$) and $\theta = \pi$ ($J_{ex} < 0$) equilibrium states, generating a torque. From the exchange theory of two spins, the magnetic moment of upper DW spin can be written as $\vec{M}_U = -g\mu_U\vec{S}_U$(For lower layer $\vec{M}_L = -g\mu_L\vec{S}_L$). The magnetic field experienced (H_{ex}^U)) by the top DW can then be expressed as[5]:

$$\vec{H}_{ex}^U = -\frac{\partial E_{ex}}{\partial M_U} = -\frac{2J}{g\mu_U}\vec{S}_L, \tag{15}$$

Hence, the torque on the top DW due to the exchange field is given by:

$$\tau_{ex}^U = -\vec{M}_U \times \vec{H}_{ex}^U = -2J\left(\vec{S}_U \times \vec{S}_L\right). \tag{16}$$

In the similar way, the exchange torque for a DW in lower ferromagnetic layer can be given by

$$\tau_{ex}^L = -\vec{M}_L \times \vec{H}_{ex}^L = -2J\left(\vec{S}_L \times \vec{S}_U\right). \tag{17}$$

For antiferromagnetic coupling, 'J_{ex}' is negative and constant, therefore, the direction of the torque depends on the projection of $\left(\vec{S}_U \times \vec{S}_L\right)$, which is always out-of-plane in the case of antiferromagnetically coupled Néel DWs in the PMA SAF wires [89]. From Eqs. (16) and (17), one can conclude that (1) the exchange torque is in out-of-plane direction, (2) the magnitude of τ_{ex}^U is same to that of τ_{ex}^L (3) τ_{ex}^U and τ_{ex}^L are opposite in direction and (4) they become maximum when the DWs in upper and lower ferromagnetic layers are perpendicular to each other.

[5] *Introduction to Solid State Physics*. Edited by Charles Kittel. 2004, Wiley, ISBN: 978–0-471–41, 526-8.

4.2 *Current-Induced Domain Wall Motion in Synthetic Antiferromagnetic Wires*

As discussed in the previous section, the exchange coupling torque can improve the DWs velocities significantly in SAF wires which is also observed in experiments. In this section, we will discuss how the SHE induced perturbation in antiferromagnetic alignment of the DWs enhances the DW velocity in SAF wires.

Krishnia et al., have demonstrated high speed current-induced DW motion in Pt/SAF/Ta wires[89]. An out-of-plane hysteresis loop of a SiO_2/Ta(3)/Pt(3)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ru(0.8)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ta(3) SAF thin film is shown in Fig. 20a. In the thin film stack, lower (M_L) and upper (M_U) ferromagnetic layers are formed from Co/Ni/Co trilayers and are coupled through ultrathin Ru spacer layer. Here, numbers in parenthesis represent the layer thickness in 'nm'. The magnetization loop of the thin film exhibits double switching at 320 Oe and 6500 Oe. The hysteresis loop is categorized into five regions. At zero magnetic field, the two ferromagnetic layers are coupled in antiferromagnetic manner as marked by region I. Regions II and IV show the flipping of a layer along the external magnetic field. In regions III and V, the exchange coupling is broken and the two ferromagnetic layers are saturated along the external magnetic fields. The magnetic field at which the exchange coupling is broken and the two ferromagnetic layers are aligned in same direction, is termed as exchange field (H_{ex}). The spin configurations of the two layers in all five regions are represented by the blue and purple arrows. The double switching in the hysteresis confirm the presence of RKKY antiferromagnetic coupling between the

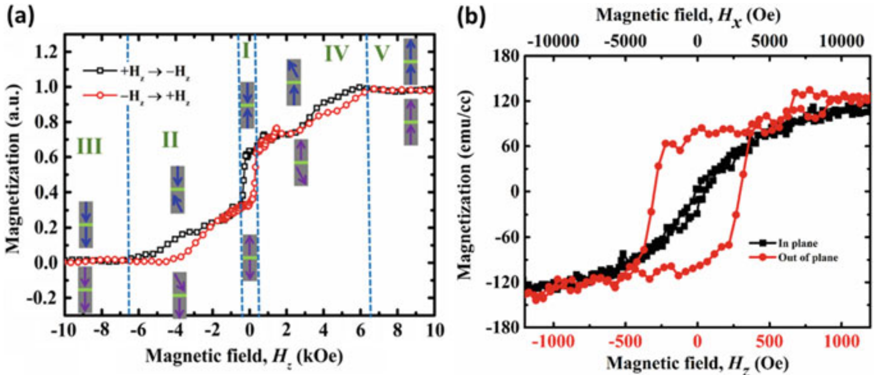


Fig. 20 **a** Out-of-plane VSM hysteresis loop measurement of the Ta(3)/Pt(3)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ru(0.8)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ta(3) SAF stack. The double switching behavior in the measurement indicates the antiferromagnetic coupling between two magnetic trilayer Co(0.4)/Ni(0.7)/Co(0.4) structures. Each region drawn in the hysteresis loop is illustrated with the relevant magnetic configurations of the SAF structure **b** In-plane (black) and out-of-plane (red) hysteresis loops of the SAF stack[89]

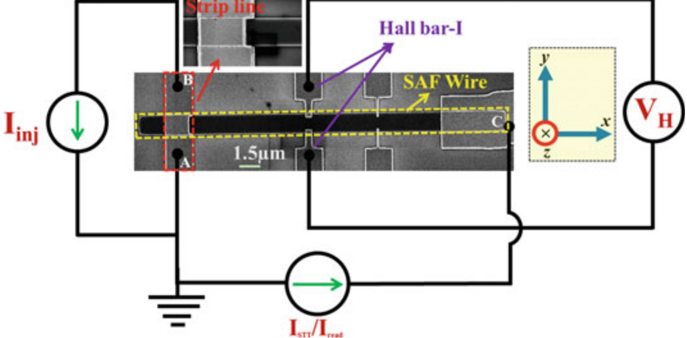


Fig. 21 Scanning electron microscope image with DW injection, driving and detection circuit schematic. Inset shows the close-up of strip-line

two magnetic layers. The energy associated with the antiferromagnetic exchange coupling is calculated as $E_{ex} = M_s t H_{ex} = 0.54$ erg/cm^2. Here, M_s is the saturation magnetization, H_{ex} is the interlayer exchange field, and t is the thickness of thin film [90].

In-plane and out-of-plane hysteresis loops of the SAF thin film are shown in Fig. 20b. The hard axis anisotropy (H_K) of the SAF thin film was found to be H_K ~ 5.5 kOe. The magnetic properties of the SAF thin films highly depends on exchange field as well as anisotropy field. Here, we note the exchange field is larger than the anisotropy field *i.e.* $H_{ex} > H_K$, therefore, the spin-flop states are permitted and step-wise hysteresis is not observed in regions II–III and IV–V.

Shown in Fig. 21, is an experimental set up to inject, drive and read the DWs in SAF wires. The SAF thin film stack was patterned using electron-beam lithography and Ar ion milling technique. Ta/Cu/Au electrodes were deposited using magnetron sputtering after a reverse sputtering process to ensure good Ohmic contacts. The length and width of fabricated nanowires were 30 and 1.5 μm, respectively. In the SAF nanowires, two ferromagnetic (FM) layers are coupled antiferromagnetically via RKKY coupling that makes the DW injection process problematic. Therefore, a Ta/Cu/Au π-shaped strip-line is used to inject the DWs into SAF wires (contact A → B). The current generates Oersted field locally and nucleate a domain of reverse magnetization underneath the strip line as described in the Sect. 2.1. The π-shaped injection line concentrates the current distribution and therefore, generates strong magnetic field locally at low current densities if compared to the conventional strip-line.

The current distribution in conventional and π-shaped strip-lines are compared in Fig. 22a. As can be seen in the Fig. 22a, a large local magnetic field initiates the domain nucleation which then spread out to form stable domain of reversed magnetization. The threshold currents required to inject the DW at different temperatures using conventional and shaped strip-lines are shown in Fig. 22b. The threshold injection current decreases with the temperature due to the Joule heating. For all the

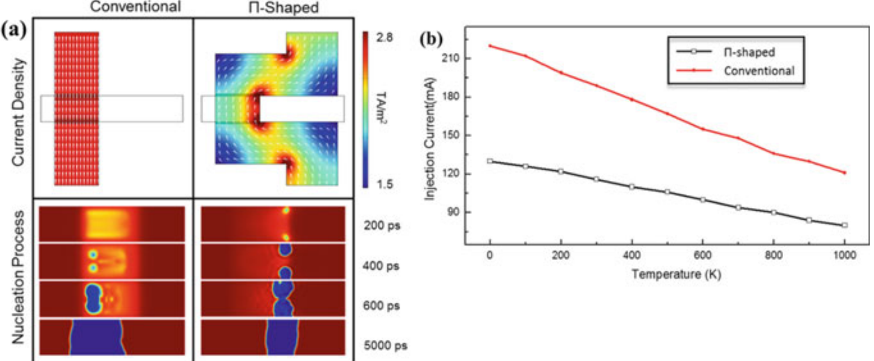


Fig. 22 **a** Comparison of simulated current density distribution of the strip-line and time-resolved magnetization between the conventional (left) and the Π-shaped strip line (right). **b** Threshold injection current density to inject a DW for both strip lines as a function of temperatures [19]

temperature, the threshold current required to inject the DWs are lower for π-shaped strip-line, thus, allowing for energy efficient DWs injection [19].

The resistance of strip-line was 60 Ω and the width and thickness were width of 2.5 μm and a 150 nm, respectively. Initially, the wire should be saturated along a fixed direction by applying a large global magnetic field. Here, the wire is saturated along $+z$-direction. Current pulses of several amplitudes and duration are applied to the strip-line by using a picosecond pulse generator (Picosecond 10300B). The applied current then generates local Oersted field which nucleates a domain of reversed magnetization under the strip-line. After the DW has been injected, an out-of-plane magnetic field in the opposite direction, i.e. $-z$ direction, is applied and the anomalous Hall effect (AHE) signal was detected at Hall bar-1, simultaneously. The drop in the R_H from 1 to 0 at a magnetic field strength of −320 Oe indicates the successful injection of the DW and termed as the DW depinning field (H_{dep}), as shown in Fig. 23a. The effect of current pulse amplitude and duration on DW nucleation process is shown in Fig. 23b. The threshold current to inject the DW in the SAF wire was found to be 0.23 Amps. The probability of the DW injection for all the currents increase with the pulse duration. Also, the probability of DW injection shifts towards lower pulse duration with increase in the current.

Prior to the DW driving measurements, the DWs were injected into the wire with 100% probability using a current pulse of 0.3 A and 60 ns and then driven by applied electrical pulses of different amplitudes and duration. The injected DWs are then driven by applying electrical pulses of different amplitudes and duration between contacts A and C of the device as shown in Fig. 21. Two Hall bars are also patterned on the wire to detect the DWs by using anomalous Hall effect (AHE). Keithley 2400 DC current source is used to supply a low amplitude current density ($I_{READ} = 6 \times 10^9$ A/m^2) between contacts 'A' and 'C' to measure the Hall voltage across the Hall bar-1. The spacing between the strip-line and the Hall bar-1, which served as the primary DW detector, is kept as 8 μm. A picosecond pulse generator: Picosecond

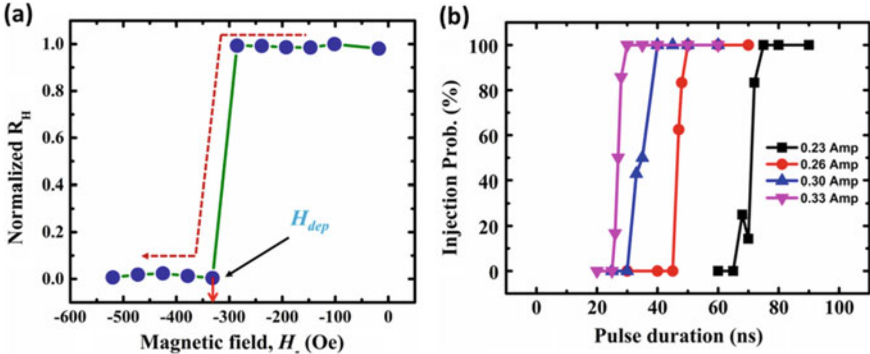


Fig. 23 **a** Normalized Hall resistance of the wire with the negative perpendicular magnetic field sweep, measured after a DW has been injected into the wire, saturated along positive z-direction. The DW is shown to be depinned from its original position at ~ 320 Oe **b** the probability of the DW injection as a function of current pulse duration for various current amplitudes [89]

10300B, is used to drive the DWs. A bias tee is used to separate the pulse generator and the Keithley 2400 DC source. After successful injection of the DWs in SAF wires, the DWs are driven by current pulses of several amplitudes. The position of DWs can be probed using Kerr microscope and also by detecting Hall voltage. The Kerr microscope images of a DW driving are shown in Fig. 24a.

The DW velocities for different current densities in the SAF wires are then measured and plotted in Fig. 24b. The threshold current density for our stack is found to be 6.58 × 10^{11} A/m^2. The DW velocity at current density of 1.04 × 10^{12} A/m^2 is found to be ~ 320 m/sec which is 1.5 times higher than the reported values for the DW dynamics in nanowires with perpendicular anisotropy at such low current

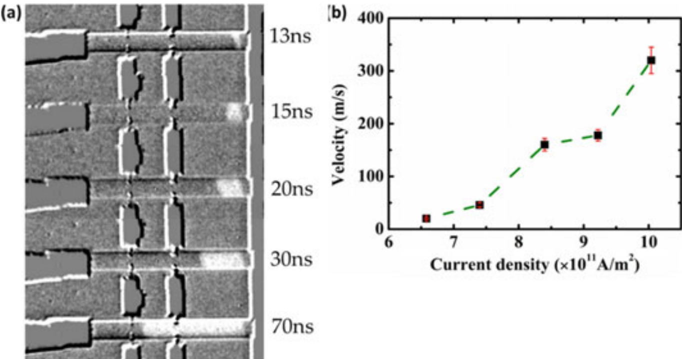


Fig. 24 **a** Kerr images of SAF device captured after successive current pulses of amplitude 9.2 × 10^{11} A/m^2. The pulse width is labeled for each DW position. Here, bright and dark contrast correspond to down and up magnetizations, respectively. **b** The measured DW velocity at various current densities in the SAF wires [89]

densities [30]. The higher DW speed in the SAF wires is due to presence of interlayer antiferromagnetic exchange torque due to the SHE-induced perturbation in DWs antiferromagnetic alignment [86, 89]. In the next section, lets us investigate the role of exchange torque on current-driven DW dynamics in SAF wires using micro-magnetic simulations.

In the SAF wire, the two ferromagnetic layers are interfaced with heavy metals which are the sources of spin currents. Since the SAF structures are grown on Pt heavy metal, the DMI interaction is naturally involved. In addition to the DMI, the heavy metals also act as a spin current generators. In the absence of current, the DWs in lower ferromagnetic layers is Néel type. A Néel DW of similar chirality is stabilized in the upper ferromagnetic layer due to antiferromagnetic exchange coupling. The exchange torque is zero in the absence of current as both the DWs were perfectly antiparallel to each other. A schematic illustrating domains and DWs in SAF wires in the absence of current is presented in Fig. 25a. When the current is injected into the SAF wires, spin currents are generated in bottom Pt and top Ta layers due to the SHE. The spin currents are then diffuse into the two FM layers and exert torques on local magnetization. The spin Hall torque rotates the DWs into transverse direction of the wire. Direction and magnitude of the SOT on a DW is given according to Eqs. (11) and (12). given by:

Now we describe effect of SOT on the two DWs. It is assumed that lower Néel DW magnetization is pointed along $+x$ direction and spin current from Pt layer attenuates across the lower ferromagnetic layer. The spin Hall angle of Pt is positive and an up-down DW is stabilized in the lower ferromagnetic layer. Therefore, the effective field $(H_{SL})_L$ that acts on the lower DW due to SHE can be written as:

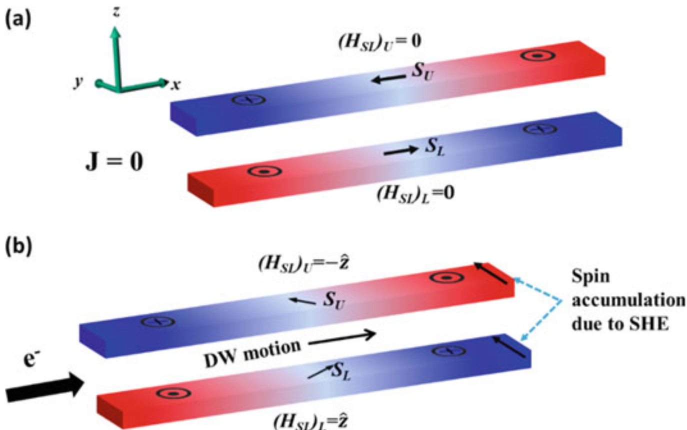


Fig. 25 Schematics of the SAF wire; showing the magnetic domains and DWs magnetization directions in the **a** absence and **b** presence of the current. The direction of SHE induced magnetic fields that act on the DWs, are also shown. The SOT and exchange torques are zero in the absence of current

$$\left(\vec{H}_{SL}\right)_L = \frac{\hbar\theta_{SH} j_a}{2\mu_0|e^-|M_s t}(\vec{m}_x) \times \vec{u}_y = \frac{\hbar\theta_{SH} j_a}{2\mu_0|e^-|M_s t}(\hat{z}). \quad (18)$$

According to Eq. (18), the SOT field is along the positive out-of-plane direction (+z), which helps to grow the "Up" domain and results in the DW propagating along the positive x-direction. Because of the AFM coupling, magnetization of the DW in the top FM layer is pointed along $-x$ axis. The upper ferromagnetic layer is interfaced with Ta heavy metal layer. Though Ta generates spin current of opposite sign to that of Pt, but it is interfaced at the opposite side therefore, the spin Hall angle of Ta can be considered as of same sign to that of Pt. The SOT field experienced by upper DW $(H_{SL})_U$ is given by:

$$\left(\vec{H}_{SL}\right)_U = \frac{\hbar\theta_{SH} j_a}{2\mu_0|e^-|M_s t}(-\vec{m}_x) \times \vec{u}_y = \frac{\hbar\theta_{SH} j_a}{2\mu_0|e^-|M_s t}(-\hat{z}). \quad (19)$$

The SOT field is in the negative out-of-plane direction, which favours the growth of "Down" domain and results in the DW propagating along the positive x-direction. In both ferromagnetic layers, the DWs propagate along positive x-direction due to the SOTs. The direction of SOT fields and DWs rotations are illustrated by a schematic in Fig. 25.

Now we explain how the interlayer exchange torque effects the DW dynamics in SAF wires. As discussed earlier, the SOT rotates the DWs into transverse direction of the wire. The SOT driven rotation of the DWs perturbs the antiferromagnetic alignment and consequently the exchange torque. The exchange torque rotates the DWs out-of-plane direction thereby driving DWs in perpendicularly magnetized SAF wires. It is difficult to vary interlayer exchange constant without compromising with other material parameters unchanged. Micromagnetic approach provides insightful analysis to understand how the interlayer exchange coupling affects DW dynamics in SAF wires [91]. A mesh size of 5 × 5 × 0.8 nm^3 is used. The thicknesses of the bottom and the top ferromagnetic layers are fixed at 1.6 nm and 2.4 nm, respectively. The two ferromagnetic layers are coupled in antiferromagnetic manner through a 0.8 nm thick Ru spacer layer. The antiferromagnetic coupling strength is varied correspond to three different exchange fields (Hex): 8440 Oe, 7000 Oe and 5550 Oe. The simulations are carried out for two DMI values (D) = -1.2×10^{-3} J/m^2 & D = -0.5×10^{-3} J/m^2. The exchange field term is included in the effective field (H_{eff}) term of the LLG Eq. (13). Instead of the usual 6 nearest-neighbor small-angle approximation for exchange interaction, the influence from the next nearest top and bottom magnetic moment is also considered. The modified algorithm allows for the calculation of the exchange coupling between two ferromagnetic materials even when they sandwich a non-magnetic Ruthenium (Ru) spacer. The exchange field that a moment 'm' experiences due to its neighbor 'm_i' is given by [92]:

$$\mathrm{H}_{ex} = 2\frac{A_{ex}}{M_{sat}}\sum_i C_i \frac{(\mathrm{m}_i - \mathrm{m})}{\Delta_i^2}, \quad (20)$$

where A_{ex} is the exchange stiffness, M_{sat} is the saturation magnetization, Δ_i is the separation distance between the two moments, and C_i is an arbitrary scaling factor which determines the strength of the exchange coupling interaction and is equals to 1 for nearest neighbors.

Micromagnetic simulation scheme of the DW dynamics in SAF wires, having a coupled Néel DW at the center is shown in Fig. 26a. Spin polarized currents of various amplitudes are then injected into the wire along the x-axis direction and the DW velocities for various interlayer exchange couplings are shown in Fig. 27. The

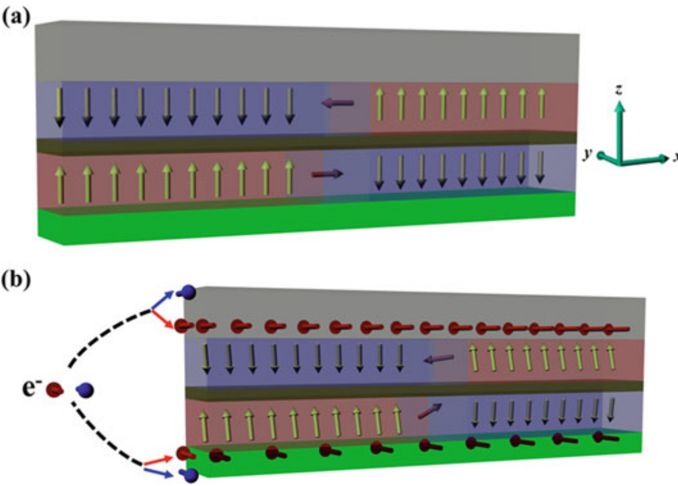


Fig. 26 **a** Schematic diagram of the SAF wire at t = 0 s and $J_a = 0$, employed in our micromagnetic model. Both the layers are coupled in antiferromagnetic manner. A DW is nucleated at the middle of the wire. **b** The schematic diagram of the spin configurations in SAF wire at t = x sec and $J_a \neq 0$. Scattering of spin currents (red and blue arrows) in top Ta and bottom Pt layers is also shown [89]

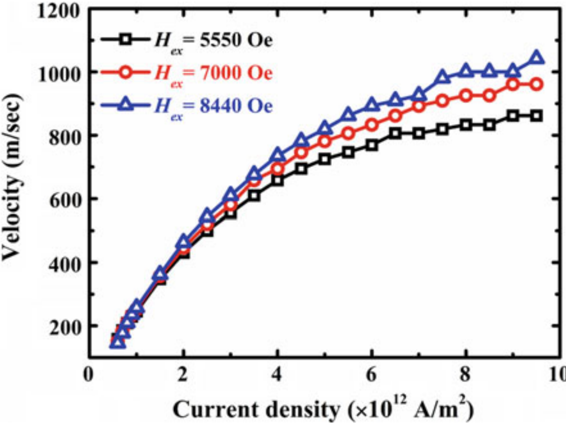


Fig. 27 Plot of the DW velocities as a function of current density for various RKKY exchange coupling strengths [89]

DW velocities increase with the current for all the interlayer exchange couplings. The effect of exchange torque on the DW dynamics can then be seen from the increases of the DW velocity with higher exchange coupling strength. As shown in the Fig. 27, for a fixed current density ~ 8×10^{12} A/m^2, the DW velocity is increased by ~ 190 m/s when H_{ex} is increased from 5550 to 8440 Oe.

As discussed in Sect. 4.1, the τ_{ex} greatly depends on the angle between the two DWs. Hence, the perturbation in the antiferromagnetic coupling due to SHE plays an important role to stimulate the exchange coupling torques. The τ_{ex} is zero in the absence of current as both the DWs were perfectly antiparallel to each other. When current is applied, both the DWs are rotated in the same direction (+y direction) due to SHE and that induces a perturbation in antiferromagnetic coupling. The DW rotation into wire transverse direction increase with current. However, the antiferromagnetic coupling opposes the rotation of the two DWs in same direction, leading to exert a torque according to Eqs. (16) and (17). The simulation results showed that the perturbation increases with higher current densities, and the two DWs are perpendicular to each other at higher current densities. The magnitude of τ_{ex} is maximum for a given magnitude of M_L and M_U and it functions in such a way that the two DWs are driven in same direction in the case of SAF wire. The DWs configurations and the directions of τ_{ex} in the (a) absence and (b) presence of current are shown by schematics in Fig. 28.

To explore the more details on the contribution of exchange torque on DW dynamics in SAF wires, the y-component of the DWs magnetization with current density is plotted in Fig. 29. It shows that the rotation of DWs magnetization got

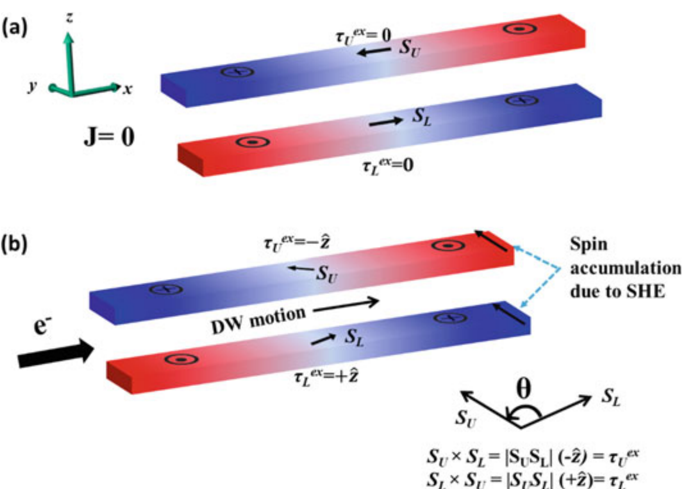


Fig. 28 Schematics showing the exchange coupling torque directions on the DWs in the SAF wire in the **a** absence and **b** presence of current. **c** The y-component of the DW magnetization as a function of current densities. The side-view simulation snapshots at low current density (point 'A') and high current density (point 'B') are also shown

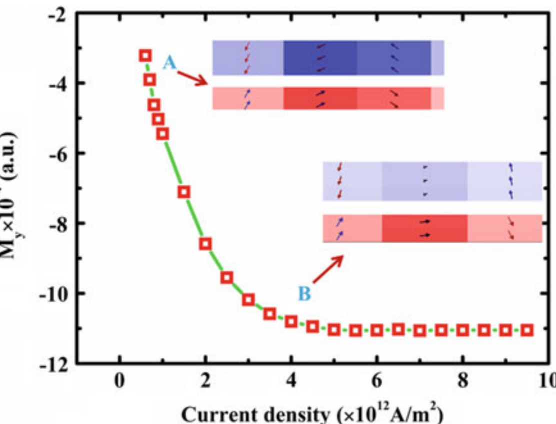


Fig. 29 The y-component of the DW magnetization as a function of current densities. The side-view simulation snapshots at low current density (point 'A') and high current density (point 'B') are also shown

saturated at higher current density and the angle between the two DWs is fixed, which results in the saturation of the exchange torque. The saturation of exchange torque results in the saturation of DWs velocities at higher current densities. Therefore, the SAF structures should be tuned in such a way that the upper and lower ferromagnetic layers experience SHE in same direction to achieve the highest DW velocities. If the magnetization rotation of the two DWs is opposite in directions, the perturbation in antiferromagnetic coupling will be smaller compared to the previous case and that will result in smaller DWs velocities. The higher perturbation in antiferromagnetic alignment has been achieved by Krishnia et al., by placing Ta and Pt at opposite interfaces of the SAF wires [89]. Based on the calculations, effects of capping and seed layers on the DW dynamics in SAF wires are summarized in a Table 1.

As discussed in Sect. 3.2.2, the DMI stabilize the DWs into Néel configuration. Figure 30a shows the DW velocity as a function of current density for two different DMI values -0.5 mJ/m^2 (black) and -1.2 mJ/m^2 (red). A significant increment in the DW velocity ~ 310 m/s is observed at a current density of 9×10^{11} A/m^2, when the DMI is increased from -0.5 to -1.2 mJ/m^2. Higher values of the DMI is shown to increase the M_x components of the DW which gives higher DW speeds due to the spin Hall effect. Figure 30b shows the normalized x and y components of the DW magnetization for two DMI values: -0.5 mJ/m^2 (solid lines) and -1.2 mJ/m^2 (dash lines). The increase in x-component with the higher DMI value indicates the stabilization of the Néel DW. The results show that wires with high DMI values will be helpful for the realization of high speed magnetic memory devices.

We have discussed the effects of various torques on DW dynamics in SAF wires. The exchange torque plays a vital role in driving the DWs in SAF wires. Now we explore how the DW dynamics is affected near magnetization compensation in such devices. The net magnetization of the SAF layers can be tuned by varying upper layer thickness. The hysteresis loops and the DWs velocities are shown in Fig. 31 for a series of SAF thin films, each with identical lower ferromagnetic layer but upper ferromagnetic layer is varied. The DW velocities are found to increase with

1 Role of spin Hall effect from capping and seed layers to stimulate the exchange coupling torques

	DW configuration	Direction of SHE	RKKY torque
J=0	S_t S_b	SHE=0	Cross product of two spins is zero. $\tau_{RKKY}=0$
$J_a>0$	Pt (t) S_t S_b Pt(t)	Inversion symmetry. SHE at both the DWs is same in magnitude but opposite in direction.	Cross product of two spins is zero. $\tau_{RKKY}=0$
	Pt (t_2) S_t S_b Pt(t_1)	SHE at both the DWs is different in magnitude but opposite in direction. Here $t_1 \neq t_2$.	Cross product of two spins is non-zero. $(\tau_{RKKY})_{(Pt/Pt)}$ is nonzero
	Ta S_t S_b Pt	Inversion asymmetry SHE at both the DWs is different in magnitude but same in direction.	Cross product of two spins is non-zero. $(\tau_{RKKY})_{(Pt/Ta)} > (\tau_{RKKY})_{(Pt/Pt)}$

M_U/M_L ratio. It is remarkable to note that the DW velocities in more compensated wires are observed up to 750 m/s. The DWs move much faster as the ratio of M_U/M_L approaches to 1. Moreover, the 1D model for the SAF structures predicts the DW velocity more than 1000 m/s in fully compensated wires [86].

4.3 Determination of Data Retention in Synthetic Antiferromagnetic Devices

In the DW based memory devices, the digital data '0 and '1' are stored in the form of magnetic domains within ferromagnetic wires. However, the magnetic domains have a finite probability to overcome the energy barrier (E_G) between the '0 and '1' states due to thermal excitations. The switching probability of the data bits between the two states due to the thermal fluctuation can be calculated using Boltzmann factor as below[93–95]:

$$P \propto \exp\left(\frac{-E_G}{k_B T}\right), \tag{21}$$

Fig. 30 **a** DWs velocity as a function of current density for two different DMI values -0.5 mJ/m^2 (black) and -1.2 mJ/m^2(red). **b** Normalized x and y components of the DW magnetization at 0.3 ns for DMI (D) = -0.5 mJ/m^2 (solid lines) and -1.2 mJ/m^2 (dash lines)

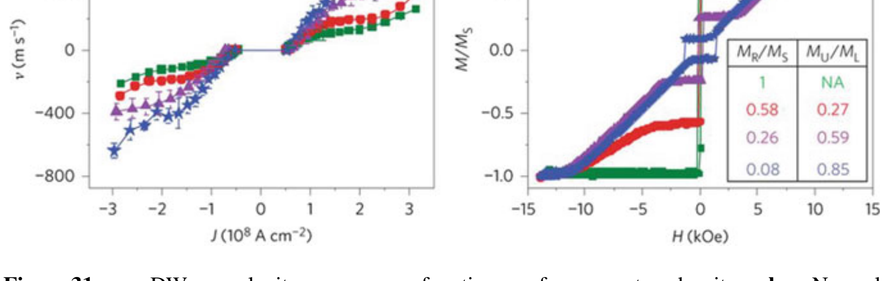


Fig. 31 a DWs velocity as a function of current density **b** Normalized hysteresis loops of a series of SAF thin films formed from TaN(2)/Pt(1.5)/Co(0.3)/Ni(0.7)/Co(0.15)/Ru(t_{Ru})/Co(t_a)/Ni(0.7)/Co(t_b)/TaN(5), where t_{Ru} = 0.2, t_a = 0.15, t_b = 0.15 (green squares), t_{Ru} = 0.8, t_a = 0.15, t_b = 0.15 (red circles), t_{Ru} = 0.8, t_a = 0.3, t_b = 0.15 (violet triangles), t_{Ru} = 0.8, t_a = 0.3, t_b = 0.3 (blur stars) [86]

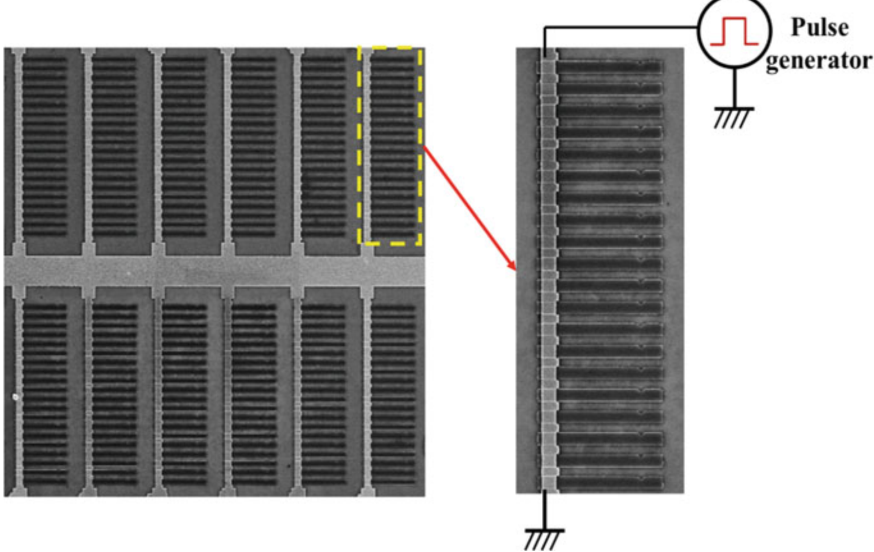


Fig. 32 Scanning electron microscope (SEM) image of the devices for the data retention experiments. 240 wires were fabricated on each device to calculate the statistical distribution of the data error. The zoom-in image of a row of wires is shown. The π-shaped strip-line is used to inject the DWs in the wire using a pulse generator. Using this method, the DWs in all the wire of a row are injected in a single pulse

where K_B is the Boltzmann factor and T is the absolute temperature. The magnetic devices are required to retain the data for at-least 10 years at operating temperature ranging from 80 to 120 °C. In this section, we discuss the data retention and thermal stability experiments in SAF based DW memory devices.

The SAF thin film stacks patterned into several wires using EBL and ion-milling techniques are shown in Fig. 32. In each device, 240 wires are patterned to calculate the statistical distribution of errors in the data. Cu/Au lines are also fabricated at the edge of each wire to inject the DWs by applying electrical pulses. After successful injection of DWs in the wires, the DWs are driven towards the center of wires using pulsed magnetic field. The positions of the DWs can imaged using Kerr microscopy as shown in Fig. 33. The devices are then baked at various temperatures over a range of time spans. During the baking, the devices are mounted on glass slides which was kept inside an oven. Positions of the DWs in the wires in the post-baked devices are measured using the Kerr microscopy.

The temperature excites the magnetic moments and causes the DWs to displace from their original positions. The displacements in the DWs position were measured by using Kerr microscopy. Kerr images of a device before and after baking at 210 °C for 23 h, are shown in Fig. 33a and b, respectively. A visible displacement of the DW in a wire is marked in the yellow circle. The wires were divided into sub-cells of 500 nm along the length for accurate measurements of the DWs displacement.

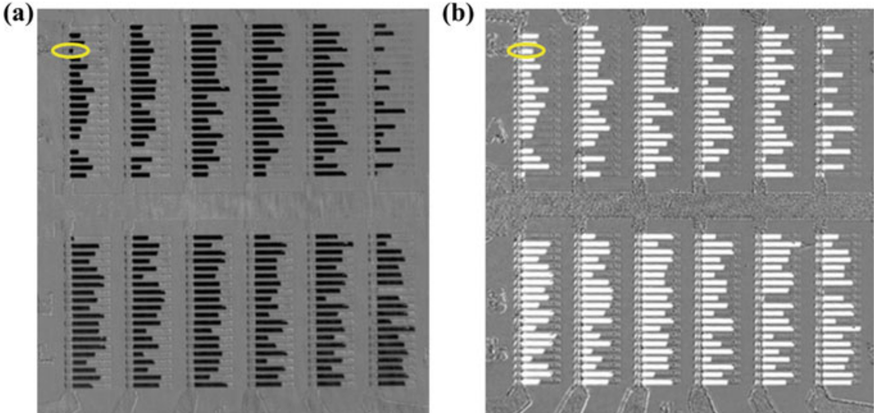


Fig. 33 **a** Kerr microscope image of a devices to measure the DWs positions in each wire at room temperature. **b** Kerr microscope images of the devices, baked at 210 °C for 23 h. Displacement of the DW position due to thermal excitations in a wire is marked in the circle area

Each sub-cell was considered as a bit. If the DW displacement is more than 500 nm, we consider that the bit has been flipped from 0 (1) to 1(0) and can be counted as an error. In these experiments, temperature-induced DW movements were observed in the devices those were baked at temperatures 190 °C and above.

The DW memory devices are annealed at different temperatures: 190 °C, 210 °C, 220 °C, 230 °C and 250 °C over several time spans ranging from 30 min to few days. The difference in the DWs position before and after baking is calculated using an image processing method. The length of each wire was kept 20 μm, therefore, each wire can be divided into 40 bits. Also, the DW injection process is stochastic in nature, therefore the wire with unsuccessful DW injection were not considered into the calculations. The error rate at a temperature can be counted using following relation-

$$\text{Error rate } (\%) = \frac{\text{bit error}}{\text{Total No. of bits}} \times 100, \tag{22}$$

where total no of bits are: number of wire with successful DW injections × 40. The error rate in the DWs positions with the baking time at an elevated temperature 190 °C is shown in Fig. 34a. Similarly, the error rates can be calculated using Eq. (22) for all the baking temperatures.

The random DW displacement in several wires due to the thermal excitation is regarded as data dissipation. The thermal stability factor of the devices can be calculated using following relation:

$$P(t) \approx ft \exp(-\Delta), \tag{23}$$

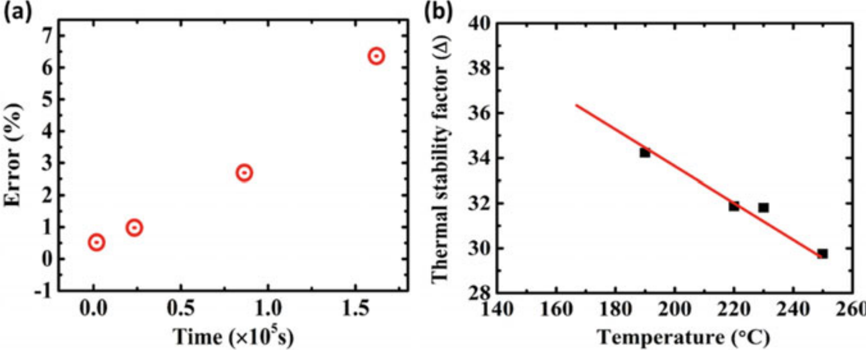


Fig. 34 **a** The bit error rate versus baking time measured at an elevated temperature 190 °C. The error rate is found to be proportional to the device baking time for all elevated temperature. **b** Thermal stability factor of the DW memory at high temperature

where P is the probability of switching of the magnetic bits between 0 and 1 states (or error rate), f ~ 1 GHz is a characteristic frequency, t is the time, $\Delta = E_B / k_B T$ is the thermal stability factor, k_B is the Boltzmann constant and T is the temperature. The thermal stability factor for various temperatures can be calculates using Eq. (23).

The thermal stability factors for various temperatures are derived from the experiments and shown in Fig. 34b. Values of the thermal stability factors exhibit linear dependency on the baking temperature. The SAF DW memory devices are found to have Δ ~ 34 at an elevated temperature 190 °C.

5 Evaluation of Spin–Orbit Torques in Synthetic Antiferromagnetic Structures

In the previous sections, we have discussed that the exchange torque provides a novel driving mechanism to DWs in SAF wires. The exchange torque accounts for high current-induced DW velocities compared to single ferromagnetic wire. The amplitude of exchange torque greatly depends on perturbation in antiferromagnetic coupling due to spin–orbit torques. In this section, we discuss the strength of spin–orbit torques in SAF devices.

Experimental measurements of SOT probe the effect of the electric current on the magnetization, e.g. by inducing oscillations, switching, or DW motion etc. There are several techniques to characterize the SOT such as harmonic Hall voltage measurement [83, 96], spin-torque ferromagnetic resonance [97], magneto-optical Kerr effect (MOKE) [98] etc. Here, we will describe the characterization of SOT in SAF structures using harmonic Hall voltage measurement technique.

In this technique, an alternating current is applied across the nanowire and harmonic response of the magnetization is detected at a low frequency, typically

up to few kHz. The harmonic Hall voltage response is recorded while sweeping an external magnetic field either in the longitudinal or transverse direction to current flow. Using analytical methods, the effective SL field and field-like (FL) fields are estimated from the voltage response curves. The harmonic measurement setup with coordinate axes, polar angle (θ) and azimuthal angle (φ) of magnetization is shown in Fig. 35.

Application of the current into the wire induces magnetic fields and thus modulates the magnetization angle from equilibrium by an amount $\Delta\theta$ and $\Delta\varphi$. The Hall resistance (R_{XY}) of the wire can be expressed as [81, 83]:

$$R_{XY} = \frac{1}{2}\Delta R_A \cos\theta + \frac{1}{2}\Delta R_P \sin^2\theta \sin 2\varphi, \tag{24}$$

where, ΔR_A and ΔR_P represent changes into the Hall resistance due to anomalous Hall effect (AHE) and planar Hall effect (PHE), respectively. The Hall voltage which is the product of Hall resistance and current can be expressed as:

$$V_{XY} = R_{XY} I. \tag{25}$$

When a sinusoidal current ($I = I_0 \sin\omega t$) is injected into the wire, the current-induced effective fields also oscillate in sync with the sinusoidal current. The magnetization angles are also modulated as $\Delta\theta\sin\omega t$ and $\Delta\varphi\sin\omega t$. Substituting the modified Eq. (24) into Eq. (25) yields the modulated Hall voltage which can also be expressed in terms of the applied signal frequency as [83].

$V_{XY} = V_0 + V_\omega \sin\omega t + V_{2\omega}\cos 2\omega t,$
$V_0 = \frac{1}{2}\left(B_\theta + B_\varphi\right)\Delta I,$

$$V_\omega = A\Delta I \tag{26}$$

$$V_{2\omega} = -\frac{1}{2}\left(B_\theta + B_\varphi\right)\Delta I \tag{27}$$

where, $A = \frac{1}{2}\Delta R_A \cos\theta_0 + \frac{1}{2}\Delta R_P sin^2\theta_0 \sin 2\varphi_0,$

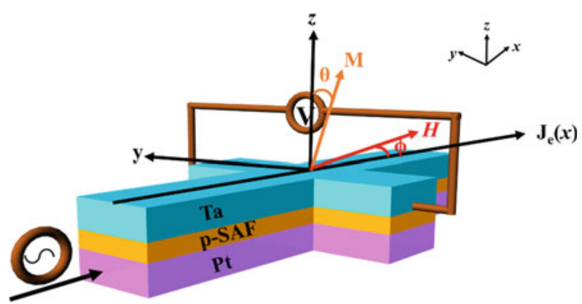


Fig. 35 Schematic of the Pt/SAF/Ta Hall structure to measure the SOT using harmonic Hall voltage measurements. Definition of spherical coordinate system is illustrated together with direction of the magnetization, M and field, H

$$B_{\theta} = \frac{1}{2}(-\Delta R_A \sin\theta_0 + \Delta R_P \sin 2\theta_0 \sin 2\varphi_0)\Delta\theta$$

$$B_{\varphi} = \Delta R_P sin^2\theta_0 \cos 2\varphi_0 \Delta\varphi$$

The angular term variations $\Delta\theta$ and $\Delta\varphi$ depend on the current induced effective fields, H_{SL} along the longitudinal direction and H_{FL} along the transverse direction. The second harmonic Hall voltage, $V_{2\omega}$ shows the dependency on $\Delta\theta$ and $\Delta\varphi$ and thus contains information about the effective fields through $\Delta\theta$ and $\Delta\varphi$ terms. Following the derivation in Ref. [83], the respective curvature and slope of V_{ω} and $V_{2\omega}$ versus the external field are calculated to obtain the ratios H_{SL} and H_{FL}, defined as $H_{SL} = (\frac{\partial V_{2\omega}}{\partial H_X}/\frac{\partial^2 V_{\omega}}{\partial H_X^2})$ and $H_{FL} = (\frac{\partial V_{2\omega}}{\partial H_Y}/\frac{\partial^2 V_{\omega}}{\partial H_Y^2})$. As discussed earlier, the measured Hall voltage also contains contributions from planar Hall effect (PHE) that leads to mixing of the H_{SL} and H_{FL}. We define the field generated in longitudinal and transverse direction as Corr_H_{SL} and Corr_H_{FL}, respectively. By defining $\xi = \frac{\Delta R_P}{\Delta R_A}$, finally it can be shown that.

$$Corr_H_{SL} = -2\frac{H_{SL} \pm 2\xi H_{FL}}{1 - 4\xi^2}, \tag{28}$$

$$Corr_H_{FL} = -2\frac{H_{FL} \pm 2\xi H_{SL}}{1 - 4\xi^2}. \tag{29}$$

The $\pm$ sign corresponds to M pointing along the $\pm$ z.

The harmonic measurements are performed on the same SAF structures discussed in Sect. 4.2. The in-plane magnetic field was swept in the longitudinal (H_L) and transverse (H_T) directions to the current of frequency 133 Hz, to quantify Slonczewski-like (SL), and Field-like (FL) effective fields, respectively. Variation in V_{ω} and $V_{2\omega}$ with H_L for at a current density 1.04×10^{11} A/m^2 is shown in Fig. 36. The red and black curves are corresponding to net $M_Z > 0$ (net 'up') and $M_Z < 0$ (net 'down') magnetizations, respectively. The first harmonic of the Hall voltage can be fitted to an equation-

$$V_{\omega} = AH_x + BH_x^2, \tag{30}$$

where, A and B are the polynomial coefficients of the equation. Similarly, linear fitting can be performed on the second harmonic Hall voltage to get an equation of the form

$$V_{2\omega} = CH_x + D, \tag{31}$$

where, C and D are the slope and intercept of the linear curve, respectively. Now, the longitudinal effective fields can be obtained using following equation -

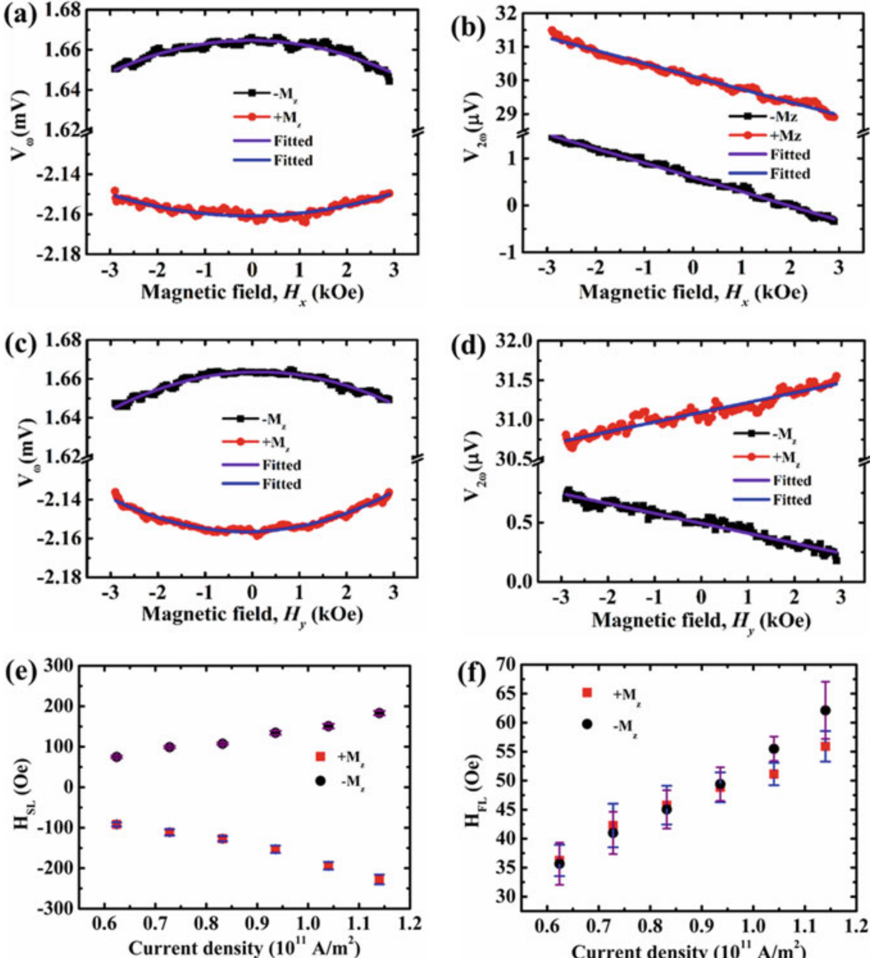


Fig. 36 **a** First and **b** second harmonics of the Hall voltage for net 'up' (red) and 'down' (black) magnetizations versus longitudinal field, measured for a current density 1.04×10^{11} A/m^2. **c** First and **d** second harmonics of the Hall voltage for net 'up' (red) and 'down' (black) magnetizations versus transverse field, measured for a current density 1.04×10^{11} A/m^2. **e** SL fields for 'up' (red) and 'down' (black) net magnetizations with current densities. **f** FL fields for 'up' (red) and 'down' (black) net magnetizations with different current densities. All the measurements were performed on the device fabricated on thin film stack Ta(3)/Pt(3)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ru(0.8)/[Co(0.4)/Ni(0.7)/Co(0.4)]/Ta(3)

$$H_{SL(FL)} = \frac{C}{B} = -2\left(\frac{\partial V_{2\omega}}{\partial H_{L(T)}}\right) \Bigg/ \left(\frac{\partial^2 V_{\omega}}{\partial H_{L(T)}^2}\right). \tag{32}$$

The longitudinal effective field increases linearly with applied current density for both 'up' (red) and down (black) net magnetizations as shown in the Fig. 36. Also, it

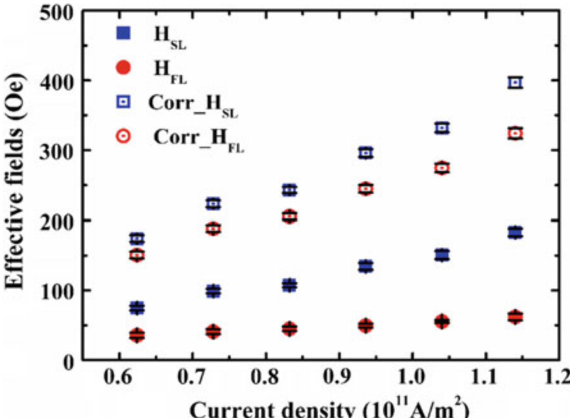


Fig. 37 The SL (blue squares) and FL (red circles) effective fields without (solid) and with (open) planar Hall effect (PHE) corrections for sample S1 at different current densities

is noted that the magnitude of H_{SL} is similar for 'up' and 'down' net magnetizations. However, sign of the H_{SL} is found to depend on the direction of the net magnetization.

In the similar way, the V_{ω} and $V_{2\omega}$ can be measured while sweeping the magnetic fields transverse to the applied current direction. The second harmonic voltage varies monotonically with H_T similar to that for H_L. However, sign of the slopes are opposite for 'up' and 'down' magnetization states when the magnetic field is swept along the transverse direction. The H_{FL} increases linearly with the applied current density for both 'up' (red) and down (black) net magnetization states and irrespective of the H_{SL}, the sign of the H_{FL} is independent of the magnetization state.

The contribution of PHE can be evaluated by measuring AHE at different field angles and a ratio of AHE to PHE resistance: $\xi = \Delta R_A/\Delta R_P \approx 0.33$ is calculated for these SAF structures. The PHE corrected longitudinal ($Corr_H_{SL}$) and transverse ($Corr_H_{FL}$) fields (using Eqs. 28 and 29) together with H_{SL} and H_{FL} are shown as a function of current density in Fig. 37. The corrected effective fields plotted versus the applied alternating current density and found to vary linearly. H_{SL} and H_{FL} are evaluated as 320 Oe per 10^{11} A/m^2 and 260 Oe per 10^{11} A/m^2, respectively, from the slopes of the plots.

6 Summary

In this chapter, we have presented an overview on new mechanisms of current-driven domain wall motion in ferromagnetic and synthetic antiferromagnetic nanowires. The fundamental characteristics and operating mechanism of domain wall based spintronic devices are reviewed. The current-induced torques such as adiabatic and

non-adiabatic spin-transfer torques, spin-Hall effect, Rashba effect, Dzyaloshinskii-Moriya interaction drive the domain walls in ferromagnetic and synthetic antiferromagnetic wire. Distinct domain wall dynamics in synthetic antiferromagnetic materials interfaced with heavy metals have shown to be much more efficient. A novel chiral torque: combination of DMI, SHE and exchange coupling torque, drives the domain walls at very high speeds in synthetic antiferromagnetic wire. These remarkable current-induced domains wall dynamics in ferromagnetic and synthetic antiferromagnetic wires are not only of great interest on fundamental physics point of view but also have potential use to design domain wall based memory and logic devices.

References

1. R.C O' Handley, *Modern Magnetic Materials: Principles and Applications* (John Wiley Sons, New York; Chichester Weinheim Brisbane, Singapore, Toronto, 2000)
2. C. Kittel, *Introduction to Solid State Physics* (2004)
3. B.D. Cullity, C.D. Graham, *Introduction to Magnetic Materials,* 2nd edn (2009)
4. A. Thiaville, Y. Nakatani, *Domain-Wall Dynamics in Nanowiresand Nanostrips* (Spin Dyn. Confin. Magn. Struct, III, 2006)
5. D.A. Allwood, G. Xiong, C.C. Faulkner, D. Atkinson, D. Petit, R.P. Cowburn, Magnetic domain-wall logic. Science(80-). **309**(5741), 1688–1692 (2005)
6. D.A. Allwood et al., Submicrometer ferromagnetic NOT gate and shift register. Science (80-). **296**(5575), 2003–2006 (2002)
7. J.H. Franken, H.J.M. Swagten, B. Koopmans, Shift registers based on magnetic domain wall ratchets with perpendicular anisotropy. Nat. Nanotechnol. **7**(8), 499–503 (2012)
8. M. Hayashi, L. Thomas, R. Moriya, C. Rettner, S.S.P. Parkin, Current-controlled magnetic domain-wall nanowire shift register. Science (80-) (2008)
9. L. O'Brien, D. E. Read, H. T. Zeng, E. R. Lewis, D. Petit, and R. P. Cowburn, "Bidirectional magnetic nanowire shift register," *Appl. Phys. Lett.*, vol. 95, no. 23, 2009.
10. S.S.P. Parkin, M. Hayashi, L. Thomas, Magnetic domain-wall racetrack memory. Science (2008)
11. C.Y. Hung, L. Berger, Exchange forces between domain wall and electric current in permalloy films of variable thickness. J. Appl. Phys. (1988)
12. L. Berger, Exchange interaction between ferromagnetic domain wall and electric current in very thin metallic films. J. Appl. Phys (1984)
13. T. Min et al., A study of write margin of spin torque transfer magnetic random access memory technology. in *IEEE Transactions on Magnetics* (2010)
14. M. Cubukcu et al., Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction. Appl. Phys. Lett. (2014)
15. G. Prenat et al., Ultra-fast and high-reliability SOT-MRAM: from cache replacement to normally-off computing. IEEE Trans. Multi-Scale Comput. Syst (2016)
16. A.J. Annunziata et al., Racetrack memory cell array with integrated magnetic tunnel junction readout. in *Technical Digest—International Electron Devices Meeting.* IEDM (2011)
17. Y. Zhang, W. Zhao, J.O. Klein, D. Ravelsona, C. Chappert, Ultra-high density content addressable memory based on current induced domain wall motion in magnetic track. IEEE Trans. Magn. (2012)
18. L. Bocklage, F.U. Stein, M. Martens, T. Matsuyama, G. Meier, Time structure of fast domain wall creation by localized fields in a magnetic nanowire. Appl. Phys. Lett. (2013)
19. S.F. Zhang et al., Highly efficient domain walls injection in perpendicular magnetic anisotropy nanowire. Sci. Rep. (2016)

20. C. Guite, I.S. Kerk, M.C. Sekhar, M. Ramu, S. Goolaup, W.S. Lew, All-electrical deterministic single domain wall generation for on-chip applications. Sci. Rep. (2014)
21. M. Muñoz, J.L. Prieto, Suppression of the intrinsic stochastic pinning of domain walls in magnetic nanostripes. Nat. Commun. **2**(1), 562 (2011)
22. P. Sethi, C. Murapaka, G.J. Lim, W.S. Lew, In-plane current induced domain wall nucleation and its stochasticity in perpendicular magnetic anisotropy Hall cross structures. Appl. Phys. Lett. (2015)
23. T. Phung et al., Highly efficient in-line magnetic domain wall injector. Nano Lett. (2015)
24. T.W. McDaniel, Ultimate limits to thermally assisted magnetic recording. J. Phys. Conden. Matt. (2005)
25. S. Fukami, T. Suzuki, N. Ohshima, K. Nagahara, N. Ishiwata, Micromagnetic analysis of current driven domain wall motion in nanostrips with perpendicular magnetic anisotropy. J. Appl. Phys. (2008)
26. D.C. Ralph, M.D. Stiles, Spin transfer torques. J. Magn. Magn. Mater. (2008)
27. A. Thiaville, Y. Nakatani, J. Miltat, and Y. Suzuki, "Micromagnetic understanding of current-driven domain wall motion in patterned nanowires," *Europhys. Lett.*, 2005.
28. L. Berger, "Exchange interaction between electric current and magnetic domain wall containing Bloch lines," *J. Appl. Phys.*, 1988.
29. I. M. Miron *et al.*, "Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection," *Nature*. 2011.
30. K.S. Ryu, L. Thomas, S.-H. Yang, S.S.P. Parkin, Chiral spin torque at magnetic domain walls. Nat. Nanotechnol. (2013)
31. S. Emori, U. Bauer, S.M. Ahn, E. Martinez, G.S.D. Beach, Current-driven dynamics of chiral ferromagnetic domain walls. Nat. Mater. (2013)
32. M. Hayashi, L. Thomas, C. Rettner, R. Moriya, Y.B. Bazaliy, S.S.P. Parkin, Current driven domain wall velocities exceeding the spin angular momentum transfer rate in permalloy nanowires. Phys. Rev. Lett. (2007)
33. L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current. Phys. Rev. B - Condens. Matter Mater. Phys. (1996)
34. J.C. Slonczewski, Current-driven excitation of magnetic multilayers J. Magn. Magn. Mater. (1996)
35. D. Langreth, J. Wilkins, Theory of spin resonance in dilute magnetic alloys. Phys. Rev. B **6**(9), 3189–3227 (1972)
36. B. Coqblin, J.R. Schrieffer, Exchange interaction in alloys with cerium impurities. Phys. Rev. (1969)
37. M. Tsoi, R.E. Fontana, S.S.P. Parkin, Magnetic domain wall motion triggered by an electric current. Appl. Phys. Lett. (2003)
38. J. Grollier et al., Switching a spin valve back and forth by current-induced domain wall motion. Appl. Phys. Lett. (2003)
39. A. Yamaguchi, T. Ono, S. Nasu, K. Miyake, K. Mibu, T. Shinjo, Real-space observation of current-driven domain wall motion in submicron magnetic wires. Phys. Rev. Lett. (2004)
40. S.S.P. Parkin, M. Hayashi, L. Thomas, Magnetic racetrack memory. Science (80-) (2008)
41. A. Mougin, M. Cormier, J.P. Adam, P. J. Metaxas, J. Ferré, Domain wall mobility, stability and Walker breakdown in magnetic nanowires. EPL (2007)
42. A. Thiaville, Y. Nakatani, J. Miltat, N. Vernier, Domain wall motion by spin-polarized current: a micromagnetic study. J. Appl. Phys. (2004)
43. T. Koyama et al., Observation of the intrinsic pinning of a magnetic domain wall in a ferromagnetic nanowire. Nat. Mater. **10**(3), 194–197 (2011)
44. G. Tatara and H. Kohno, "Theory of Current-Driven Domain Wall Motion: Spin Transfer versus Momentum Transfer," *Phys. Rev. Lett.*, 2004.
45. M. Kläui et al., Domain wall motion induced by spin polarized currents in ferromagnetic ring structures. Appl. Phys. Lett. (2003)
46. G. Tatara, T. Takayama, H. Kohno, J. Shibata, Y. Nakatani, H. Fukuyama, Threshold current of domain wall motion under extrinsic pinning, β-term and non-adiabaticity J. Phys. Soc Japan (2006)

47. S. Zhang, Z. Li, Roles of nonequilibrium conduction electrons on the magnetization dynamics of ferromagnets. Phys. Rev. Lett. (2004)
48. G.S.D. Beach, C. Nistor, C. Knutson, M. Tsoi, J.L. Erskine, Dynamics of field-driven domain-wall propagation in ferromagnetic nanowires. Nat. Mater. (2005)
49. A. Vanhaverbeke, A. Bischof, R. Allenspach, Control of domain wall polarity by current pulses. Phys. Rev. Lett. (2008)
50. J.C. Slonczewski, Theory of Bloch-line and Bloch-wall motion. J. Appl. Phys. (1974)
51. G. Meier, M. Bolte, R. Eiselt, B. Krüger, D.H. Kim, P. Fischer, Direct imaging of stochastic domain-wall motion driven by nanosecond current pulses. Phys. Rev. Lett. (2007)
52. R. Moriya, L. Thomas, M. Hayashi, Y.B. Bazaliy, C. Rettner, S.S.P. Parkin, Probing vortex-core dynamics using current-induced resonant excitation of a trapped domain wall. Nat. Phys. (2008)
53. M. Eltschka et al., Nonadiabatic spin torque investigated using thermally activated magnetic domain wall dynamics. Phys. Rev. Lett. (2010)
54. L. Heyne et al., Direct determination of large spin-torque nonadiabaticity in vortex core dynamics. Phys. Rev. Lett. (2010)
55. I.M. Miron et al., Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. Nat. Mater. (2010)
56. I.M. Miron et al., Fast current-induced domain-wall motion controlled by the Rashba effect. Nat. Mater. (2011)
57. C. Ast et al. Giant spin splitting through surface alloying. Phys. Rev. Lett. (2007)
58. A. Manchon, H.C. Koo, J. Nitta, S.M. Frolov, R.A. Duine, New perspectives for Rashba spin-orbit coupling. Nat. Mater. (2015)
59. V.M. Edelstein, Spin polarization of conduction electrons induced by electric current in two-dimensional asymmetric electron systems. Solid State Commun. (1990)
60. T.A. Moore et al., High domain wall velocities induced by current in ultrathin Pt/Co/AlOx wires with perpendicular magnetic anisotropy. Appl. Phys. Lett. (2008)
61. I. M. Miron, P.J. Zermatten, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, Domain wall spin torquemeter. Phys. Rev. Lett. (2009)
62. K.S. Ryu, L. Thomas, S.H. Yang, S.S.P. Parkin Current induced tilting of domain walls in high velocity motion along perpendicularly magnetized micron-sized Co/Ni/Co racetracks. Express Appl. Phys. (2012)
63. A. Manchon, S. Zhang, Theory of nonequilibrium intrinsic spin torque in a single nanomagnet. Phys. Rev. B Condens. Matter Mater. Phys. (2008)
64. A. Thiaville, S. Rohart, É. Jué, V. Cros, A. Fert, Dynamics of Dzyaloshinskii domain walls in ultrathin magnetic films. EPL (2012)
65. I. Dzyaloshinskii, Thermodynamic theory of 'weak' ferromagnetism in antiferromagnetic substances Sov. Phys JETP (1957)
66. T. Moria, Anisotropic superexchange iteraction and weak ferromagnetism. Phys. Rev. (1960)
67. M. Bode et al., Chiral magnetic order at surfaces driven by inversion asymmetry. Nature (2007)
68. A.V. Khvalkovskiy et al. Matching domain-wall configuration and spin-orbit torques for efficient domain-wall motion. Phys. Rev. B Condens. Matter Mater. Phys. **87**, 020402 (2013)
69. E. Martinez, S. Emori, N. Perez, L. Torres, G.S.D. Beach, Current-driven dynamics of Dzyaloshinskii domain walls in the presence of in-plane fields: Full micromagnetic and one-dimensional analysis. J. Appl. Phys. (2014)
70. S. Rohart, A. Thiaville, Skyrmion confinement in ultrathin film nanostructures in the presence of Dzyaloshinskii-Moriya interaction. Phys. Rev. B Condens. Matter Mater. Phys. (2013)
71. R. Lavrijsen et al., Asymmetric magnetic bubble expansion under in-plane field in Pt/Co/Pt: Effect of interface engineering. Phys. Rev. B Condens. Matter Mater. Phys. (2015)
72. R. Soucaille et al., Probing the Dzyaloshinskii-Moriya interaction in CoFeB ultrathin films using domain wall creep and Brillouin light spectroscopy. Phys. Rev. B (2016)
73. S.V. Tarasenko, A. Stankiewicz, V.V. Tarasenko, J. Ferré, Bloch wall dynamics in ultrathin ferromagnetic films. J. Magn. Magn. Mater. (1998)

74. S.G. Je, D. H. Kim, S.C. Yoo, B.C. Min, K.J. Lee, S.B. Choe, Asymmetric magnetic domain-wall motion by the Dzyaloshinskii-Moriya interaction. Phys. Rev. B Condens. Matter Mater. Phys. (2013)
75. A. Fert, V. Cros, J. Sampaio, Skyrmions on the track. Nature Nanotechnology (2013)
76. L. Liu, O.J. Lee, T.J. Gudmundsen, D.C. Ralph, R.A. Buhrman, Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from the spin hall effect. Phys. Rev. Lett. (2012)
77. L. Liu, C.F. Pai, Y. Li, H.W. Tseng, D.C. Ralph, R.A. Buhrman, Spin-torque switching with the giant spin hall effect of tantalum. Science (80-) (2012)
78. J. Sinova, S.O. Valenzuela, J. Wunderlich, C.H. Back, T. Jungwirth, Spin Hall effects. Rev. Mod. Phys. (2015)
79. S.H. Yang, S. Parkin, Novel domain wall dynamics in synthetic antiferromagnets. J. Phys. Conden. Mat. (2017)
80. P.P.J. Haazen, E. Murè, J.H. Franken, R. Lavrijsen, H.J.M. Swagten, B. Koopmans, Domain wall depinning governed by the spin Hall effect. Nat. Mater. (2013)
81. J. Kim et al., Layer thickness dependence of the current-induced effective field vector in Ta|CoFeB|MgO. Nat. Mater. (2013)
82. P. Sethi, S. Krishnia, S.H. Li, W.S. Lew, Modulation of spin-orbit torque efficiency by thickness control of heavy metal layers in Co/Pt multilayers. J. Magn. Magn. Mater. (2017)
83. M. Hayashi, J. Kim, M. Yamanouchi, H. Ohno, Quantitative characterization of the spin-orbit torque using harmonic Hall voltage measurements. Phys. Rev. B Condens. Matter Mater. Phys. (2014)
84. F. Luo et al. Simultaneous determination of effective spin-orbit torque fields in magnetic structures with in-plane anisotropy. Phys. Rev. B (2017)
85. I. Purnama, I.S. Kerk, G.J. Lim, W.S. Lew, Coupled Néel domain wall motion in sandwiched perpendicular magnetic anisotropy nanowires. Sci. Rep. (2015)
86. S. H. Yang, K.S. Ryu, S. Parkin, Domain-wall velocities of up to 750 m s-1driven by exchange-coupling torque in synthetic antiferromagnets. Nat. Nanotechnol. (2015)
87. S.S.P. Parkin, D. Mauri, Spin engineering: Direct determination of the Ruderman-Kittel-Kasuya-Yosida far-field range function in ruthenium. Phys. Rev. B (1991)
88. S.S.P. Parkin, R. Bhadra, K.P. Roche, Oscillatory magnetic exchange coupling through thin copper layers. Phys. Rev. Lett. (1991)
89. S. Krishnia et al., Role of RKKY torque on domain wall motion in synthetic antiferromagnetic nanowires with opposite spin Hall angles. Sci. Rep. (2017)
90. Y.J. Chang, A. Canizo-Cabrera, V. Garcia-Vazquez, Y.H. Chang, T.H. Wu, Perpendicular magnetic tunnel junctions with synthetic antiferromagnetic pinned layers based on [Co/Pd] multilayers. J. Appl. Phys. (2013)
91. A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, B. Van Waeyenberge, The design and verification of MuMax3. AIP Adv. (2014)
92. M.J. Donahue, D.G. Porter, Exchange energy formulations for 3D micromagnetics. Phys. B Condens. Matt. (2004)
93. W.F. Brown, Thermal fluctuations of a single-domain particle. Phys. Rev. (1963)
94. H. Guchang, H. Jiancheng, S.C. Hin, M. Tran, L.S. Ter, Switching methods in magnetic random access memory for low power applications. J. Phys. D. Appl. Phys. (2015)
95. L. Thomas, G. Jan, S. Le, P.K. Wang, Quantifying data retention of perpendicular spin-transfer-torque magnetic random access memory chips using an effective thermal stability factor method. Appl. Phys. Lett. (2015)
96. K. Garello et al., Symmetry and magnitude of spin-orbit torques in ferromagnetic heterostructures. Nat. Nanotechnol. (2013)
97. W. Zhang, W. Han, X. Jiang, S.H. Yang, S.S.P. Parkin, Role of transparency of platinum-ferromagnet interfaces in determining the intrinsic magnitude of the spin Hall effect. Nat. Phys. (2015)
98. C. Stamm et al.,Magneto-Optical Detection of the Spin Hall Effect in Pt and W Thin Films. Phys. Rev. Lett. (2017)

Electric-Field-Controlled MRAM: Physics and Applications

James Lourembam and Jiancheng Huang

Abstract This chapter provides a comprehensive overview of the electric-field (or voltage) control of magnetic anisotropy, an emerging concept for the next-generation memory device. This approach has many technological appeals as it can enable ultra-low-latency data transfer and ultra-low power electronics. The underlying mechanisms governing magnetization switching from an applied electric-field are interfacial spin-charge coupling and Larmor precession. This allows for electric-field-driven MRAM as opposed to current-driven in conventional spin-torque MRAM. An exhaustive discussion in particular relevance to industry-friendly materials is provided in this chapter. The challenges in implementation and possible solutions including field-free approach are discussed. The chapter summarises the experimental and theoretical progress in electric-field-controlled MRAM, discusses our current understanding, and finally presents the prospects of utilising this approach.

1 Introduction

Magnetoresistive random-access memory (MRAM) utilises magnetic tunnel junctions (MTJ), where data bits are written and stored by manipulating the relative magnetization orientation of the ferromagnetic layers, which are separated by an oxide barrier. Manipulating the storage layer in MTJ was initially achieved through Oersted field generated from the current passing through the bit line [1]. However, such a scheme though still in production now does not favor device scaling. The next generation of MRAM was realized through spin-transfer-torque (STT) where writing is achieved by sending a current through the MTJ along with transfer of

J. Lourembam (✉)
Institute of Materials Research and Engineering, Agency for Science Technology and Research (A*STAR), #08-03, 2 Fusionopolis Way, Innovis 138634, Singapore
e-mail: james_lourembam@imre.a-star.edu.sg

J. Huang
Taiwan Semiconductor Manufacturing Company, 168 Park Ave. 2, Hsinchu Science Park, Hsinchu 300-75, Taiwan ROC
e-mail: huangjia@tsmc.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_4

spin momentum which enables magnetization switching [2–6]. STT effect was first proposed in theory in 1996 independently by Slonczewski and Berger [2, 6]. Eventually, practical STT switching was demonstrated, first in MTJ with AlO_x devices in 2004 then on MgO-based MTJ in 2005 [7–9]. These were MTJs with in-plane magnetization that supports scaling but required a large writing current. In 2010, the successful demonstration of perpendicular MTJ based on CoFeB-MgO heterostructures allowed reduction in current switching accelerating industry demand for STT-MRAM [10–14].

STT-MRAM belongs to a unique group of emerging non-volatile memories (NVMs), which can eliminate standby power, a major bottleneck against scaling in existing semiconductor memory solutions. Compared to other non-magnetic NVMs, STT-MRAM is superior in terms of lower power consumption and faster writing speed. However, compared to existing semiconductor solutions writing energy is large—still in the sub-pJ/bit level at best [15–18]. This dynamic power consumption predominantly comes from ohmic dissipation and is dependent on the large current required for switching [19]. For applications in cache-memory or memory-intensive computing, frequent rewriting is required for which STT is still not the best solution. An alternative solution for ultra-low power writing within the MRAM family is through electric-field (or voltage) controlled magnetic anisotropy [19–35]. In this scheme, the ohmic loss is significantly reduced since the current flowing through the MTJ is significantly smaller. This approach has gained much momentum in recent years because of the demonstration of write energies at two orders of magnitude lower than the traditional STT approach [19, 36].

The potential benefits for E-field MRAM cannot be understated. Firstly, compared to STT-MRAM, which requires an electrical current, an ideal MRAM based on the electric-field would only consume the amount of energy need to charge and discharge the MgO insulator. MTJ scalability and thicker MgO can both be implemented in this scheme. The resulting small capacitance implies sub fJ/bit level charging energies [23], putting it on a level with conventional CMOS SRAM or DRAM (1–10 fJ/bit) [37]. Secondly, existing current consumption for STT-MRAM is directly proportional to the area of the MTJ. This implies that variation in the programming current is a squared function of the linear variation in the critical-dimension (CD). E-field MRAM would eliminate this issue and reduce process-induced variations back to a single dimension (the thickness of the MgO). Thirdly, conventional STT-MRAM scaling is limited by the selected transistor strength. Typically the transistor pitches are much larger than the MTJ in order to supply the amount of current needed for STT-MRAM to operate. E-field MRAM could allow for more aggressive scaling, opening up more avenues for MRAM to be used in applications that require extremely low power consumption. As shown in Fig. 1a Electric Field-MRAM (EF-MRAM) can provide superior performance both in writing energy and writing time. It is also particularly suitable for cache memory (L1, L2) applications where writing speeds faster than 10 ns are required (Fig. 1b).

In the following sections, we will discuss in detail, electric-field writing mechanism describing the physics, material dependence and their close relationship with device operation.

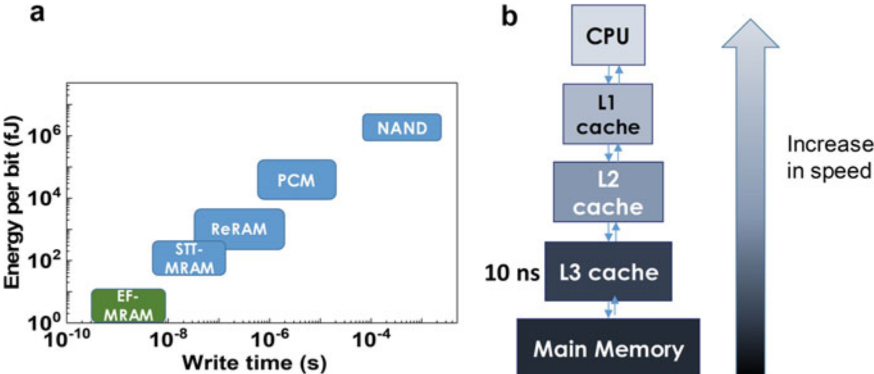


Fig. 1 **a** Comparison of various memory technologies in the plot of writing energy per bit versus writing speed. **b** Memory-compute hierarchical architecture showing various levels of writing speed requirements

2 Materials and Techniques

2.1 First-Principles Calculations

Well-motivated by the practicality of electric-field based writing mechanism, there has been intensive studies in applying this approach on popular MRAM 3*d* ferromagnet/oxide heterostructures such as CoFeB/MgO [33, 38–40]. In these systems, the perpendicular magnetic anisotropy (PMA) arises from strong hybridization between the interfacial (Co)Fe–3*d* and the O–2*p* orbitals and also exhibits strong spin–orbit coupling (SOC) [10, 41, 42]. The introduction of the electric-field in these systems is argued to be purely an interfacial effect [43], which is why significant modulation of anisotropy by an electric field is observed in thin films even though it is screened at the ferromagnet/oxide interface. This phenomenon is also widely known as voltage-controlled magnetic anisotropy (VCMA), and its origin is generally discussed as the manifestation of relative change in the occupancy of the 3*d* orbitals associated with capacitive charging and discharging effect at the ferromagnet/oxide interface.

Since orbital physics plays a vital role in the physical origin of this effect, first-principles calculations can help to understand the electric- field effect as well as accelerate materials discovery to achieve high electric-field efficiency. In particular, by looking at the *K*-resolved magnetic anisotropy, determined from first-principles calculations, one can see a very significant difference in the two-dimensional Brillouin zone with and without electric-field for Ta/CoFe/MgO [44]. There have been several works on this topic using density functional theory, particularly in heavy metal/ferromagnet/oxide heterostructure because it is more straightforward to implement this stack structure in MTJs. Studies include materials combined with various ferromagnets, heavy metals, and even oxides other than MgO. Using different ferromagnets it was shown that the electric-field could tune Co/MgO, Fe/MgO and

CoFe/MgO structures with a Co-rich interface having higher efficiency [45, 46]. First-principles calculations predicted that the electric-field efficiency in heterostructures Fe/Co/MgO and FeCo/MgO values were larger by a factor of 10.9 and 6.75, respectively, as compared to that of Fe/MgO, which was reported to be $+130$ fJV^{-1} m^{-1}. In this case, the direction of the electric-field points towards the MgO at the interface. In this convention, the perpendicular magnetic anisotropy energy (MAE) increases as the electron density decreases (by applying a positive voltage) at the free layer/MgO interface. For the sake of simplicity, we will adopt this convention for the rest of the chapter unless otherwise stated.

Interestingly, while the electric-field tuning of magnetic anisotropy is thought to be limited to the ferromagnet/oxide interface, the underlayer or the cap layer was also reported to influence the electric-field efficiency [47, 48]. Depending on the heavy metal used as an underlayer, the sign of the efficiency may change from positive to negative. Interestingly, Pt as the underlayer was found to show substantially more efficiency compared to Ta or Au [47, 48]. An exhaustive first-principles investigation of different heavy materials to achieve high VCMA is still lacking. The choices of heavy metal could be expanded to Ir, Nb, Hf, Zr etc. Hf underlayer on CoFe/MgO structures could see electric field efficiency as high as -387 fJV^{-1} m^{-1} [44].

First-principles calculations can also guide us in the invention of new stack structures such as the insertion of an oxide or a metal layer between the ferromagnet and the barrier oxide [49]. Insertion of an oxide monolayer can significantly affect the redistribution of *d*-electrons near the Fermi level. By studying monolayer insertions with oxides such as FeO, CoO, NiO, PdO and ZnO, Minggang et al. found that ZnO insertion shows the highest efficiency amongst these systems reaching E-field efficiency of $+166$ fJV^{-1} m^{-1} [50]. VCMA effect under select monolayer insertions are shown in Fig. 2, and Magnetocrystalline anisotropy (MCA) shows quasi-linear dependence with electric-field. Here, the direction of the positive electric-field points away from the MgO. Such oxide insertion schemes can be easily realised practically by introducing an oxidation step in the MTJ stack deposition process. In the next section, experimental results and methods on electric-field efficiency will be described in detail.

2.2 Experiments—Single Ferromagnet Structures

One of the popular methods of determining E-field efficiency is by isolating the free-layer from the full MTJ stack and studying it independently, e.g. gated-Hall bar structure. Even though VCMA can be determined directly from MTJs, the fabrication of these devices can be costly and time-consuming especially for the discovery of new materials. At the same time, the electric-field can also induce secondary effects which can complicate the understanding of materials dependent E-field efficiency. Let us first discuss the gated-Hall bar approach, the schematic of which is given Fig. 3a and the VCMA calculation method is described in the following paragraphs.

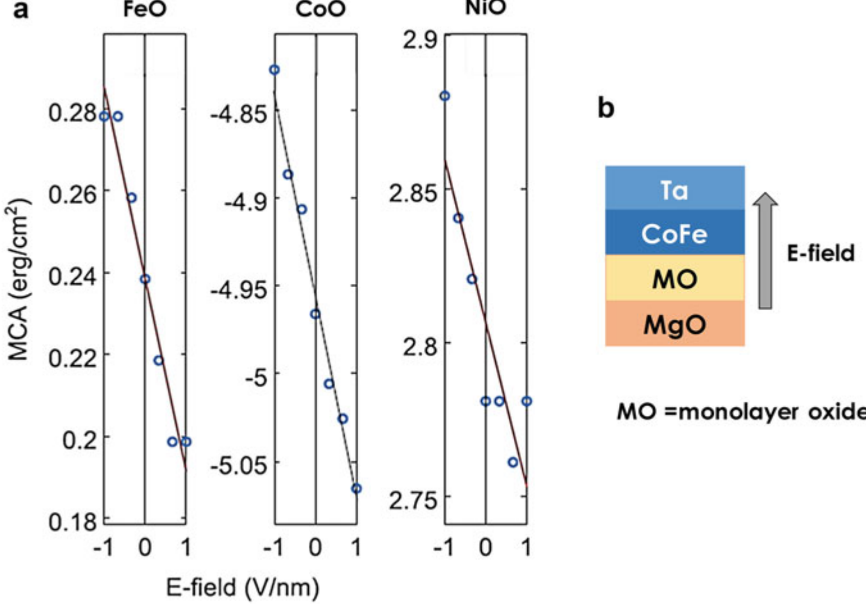


Fig. 2 Electric-field tuning of magnetic anisotropy of Ta/CoFe/MO/MgO heterostructures where MO is a monolayer oxide. **b** The direction of the electric-field used in the simulations is given on the right. Reproduced from [50] © 2018, with the permission of AIP Publishing

Under an external applied magnetic field (H), Hall resistance (R_{H}) measurements on these ferromagnetic heterostructures will have an anomalous Hall effect (AHE) contribution besides the ordinary Hall effect (OHE) which is given by:

$$R_H = R_{OHE} + R_{AHE} = A\mu_0 H + BM_{\perp} \tag{1}$$

where A and B are ordinary Hall and anomalous Hall coefficients respectively. $M_{\perp}$ represents the magnetization component along the perpendicular direction. The MAE along the perpendicular direction can then be determined using the "area under the curve" as follows:

$$MAE = -\frac{M_S}{R_{AHE}(\max)} \int_0^{R_{AHE}(\max)} H dR \tag{2}$$

where M_S is the saturation magnetization, and $R_{AHE}(\max)$ is the maximum anomalous Hall effect resistance contribution at saturation. This integral is similar to the integral of the R-H curve in Fig. 7b, albeit with normalized units. Using the gated-Hall measurements, one can determine different MAE at different gate voltages. The electric-field efficiency ξ for a ferromagnetic thickness, t_{FM} and applied electric-field E, is given by:

$$\xi = \frac{\Delta MAE.t_{FM}}{E}. \quad (3)$$

By probing the electric-field effect on the anomalous Hall signal, it was reported that Ta/CoFeB/MgO showed an efficiency of + 33 fJV^{-1} m^{-1} for $t_{\mathrm{FM}} = 1.33$ nm [51]. Further investigation by Lau et. al. showed that in this system, ξ has a strong thickness dependence and the magnitude decreases by ~ 50% with decreasing ferromagnetic thickness from ~0.5 nm to ~1nm [52]. On the other hand, for W/CoFeB/MgO heterostructures, ξ varies by ~ 50% with thickness [52]. Utilising a similar technique, it was also found that annealing decreases electric-field efficiency in Mo/CoFeB/MgO, which showed ξ value of + 40 fJV^{-1} m^{-1} at an annealing temperature of 430 °C [39].

Besides, the anomalous Hall approach, electric-field efficiency can also be experimentally determined from magneto-optic Kerr (MOKE) measurements (Fig. 3b), which also relies on the "area under the curve" approach to determined MAE at different gate voltages. Using MOKE, a large value of $\xi = +\ 602$ fJV^{-1} m^{-1} was found in FePd/MgO [53].

The last method in this section is the ferromagnetic resonance spectroscopy (FMR), where the ferromagnetic stack is processed into circular mesas, and the device is then mounted into a TE_{011} microwave cavity [54]. Here, the effective anisotropy can be extracted by fitting the FMR condition [54],

$$f = \frac{g\mu_0\mu_B}{2\pi\hbar}\sqrt{H_1 H_2} \quad (4)$$

and

$$\begin{aligned} H_1 &= H_R \cos(\theta - \theta_M) + H_{K1}^{eff} \cos^2\theta_M - H_{K2}\cos^4\theta_M, \\ H_2 &= H_R \cos(\theta - \theta_M) + H_{K1}^{eff} \cos 2\theta_M - \frac{H_{K2}}{2}(\cos 2\theta_M + \cos 4\theta_M) \end{aligned}. \quad (5)$$

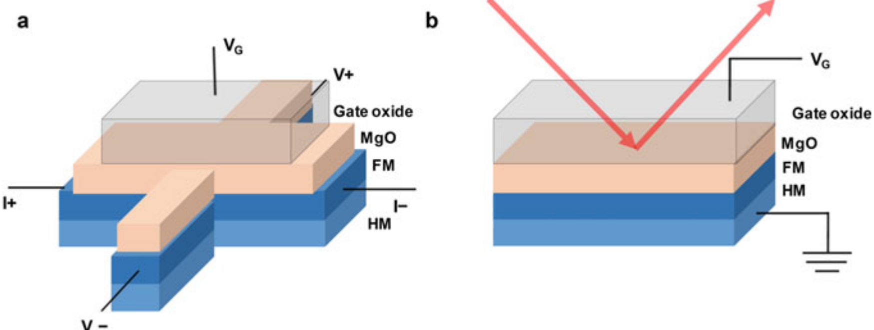


Fig. 3 **a** Schematic structures for determining electric-field efficiency from anomalous hall measurements in gated-hall bar structures and, **b** similar structure for gated polar-MOKE measurements

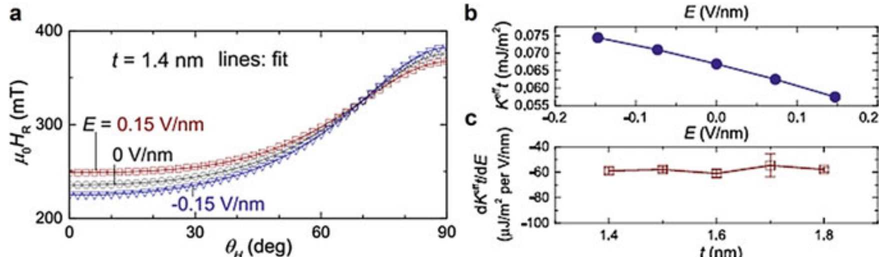


Fig. 4 **a** Resonant field with and its dependence with the applied field angle for Ta/CoFeB/MgO. The response and the resulting fits for $H_{K1}{}^{eff}$ change as the applied field to the sample is changed. **b** Dependence of areal effective magnetic anisotropy on electric-field for $t_{FM} = 1.4$ nm as deduced from the FMR approach. **c** Plot of thickness dependent VCMA where the y-scale is quantitatively equal to ξ in the units of 10^{-9} μJ/V-m as determined from the FMR condition. Reproduced from [54] © 2014, with the permission of AIP Publishing

Here g is the Landé factor, μ_B, the Bohr magneton, $\hbar$ the Dirac constant, $H_{kI}{}^{eff}$ and H_{k2} are the first- and second-order anisotropy field constants respectively and are extracted from fitting Eq. 5 into the experimental data. θ is the direction of the applied field and θ_M is the direction of the magnetization determined by minimizing the energy density [54]. An example is shown in Fig. 4a for Ta/CoFeB/MgO, which shows how the resonant field H_R (and hence the effective anisotropy) changes as a voltage is applied across the sample. The calculated change in anisotropy per unit area is given in Fig. 4b. Here, the positive electric-field direction is away from the MgO, and hence a negative slope is obtained (similar to Fig. 2). Correspondingly, ξ is also negative (Fig. 4c) because of this convention. It may be noted here that the strength of ξ is slightly larger than that determined from anomalous Hall measurements. The origin of this difference in strength may be due to sample preparation methods and will be discussed in Sect. 2.4.

2.3 Experiments—Magnetic Tunnel Junctions

While the previous sections described devices with only a ferromagnetic layer and an insulator, the electric-field efficiency, ξ, may also be measured in MTJs directly. This approach is reviewed in this section. The section will then be followed by a review of ξ in different materials when integrated with various materials.

An initial check on the strength of electric-field tuning of anisotropy in MTJs can come from analyzing coercivity, H_C versus applied voltage bias (V). As shown in Fig. 5b, taking the example of a Ta/CoFeB/MgO free layer MTJ, H_C shows quasi-linear dependence with V and the slope can indicate trends for ξ. In this case, the direction of the positive electric-field is away from MgO. While one can measure the change in the H_C of the MTJ by sweeping the applied bias, this measurement does not give us a direct change in the anisotropy energy (ξ) which needs to be determined

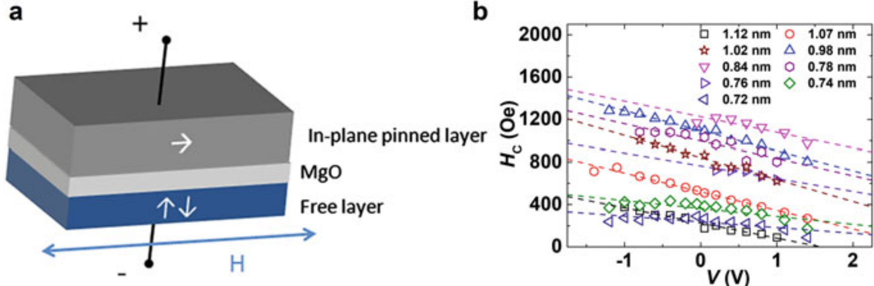


Fig. 5 **a** A typical orthogonal stack adopted on a top-pinned MTJ used for the measurement of the VCMA coefficient or efficiency, ξ. The field applied in the horizontal direction (hard axis) direction of the free layer causes the free layer to cant. This canting can be measured by TMR to deduce the strength of the anisotropy. **b** Coercivity tuning of an MTJ based on Ta/CoFeB/MgO free layer by voltage application. Quasi-linear tuning is shown for various thicknesses. Reproduced from [24] © 2018, with the permission of AIP Publishing

in terms of the change in the anisotropy field, H_K. ξ follows from Eq. (3) as:

$$\xi = \frac{\Delta H_K M_S t_{FM} t_{ox}}{2\Delta V}. \tag{6}$$

Here V is the change in voltage, t_{ox} the thickness of the barrier oxide layer. Note that Eq. 6 has been rewritten using the change in the electric-field, $\Delta E = \Delta V / t_{ox}$. The most common method for measuring ξ is through the "area method", which involves using an orthogonal anisotropy stack with the easy axes of the free and the reference layer orthogonal to each other. This allows the measurement of magnetoresistance in the hard axis direction of the free layer. The easiest way to do this for a PMA free layer is to use a thick in-plane reference layer [29, 55]. A typical orthogonal stack is shown in Fig. 5a. By controlling its thickness, the free layer can either be made in-plane (Fig. 5a) or perpendicular. In most studies, PMA stacks are often used since the vast majority of target applications require perpendicular MTJs. For a PMA free layer, the external field is applied along the device surface to measure its hard axis magnetization. This is shown in Fig. 5a, where the field, H, is applied in the horizontal direction, orthogonal to the free layer's easy axis.

The resistance measurement appears similar to that shown in Fig. 6a for a typical hard axis measurement. To convert this into a magnetization curve, the Slonczewski's model [56] for TMR is adopted, whose simplified version [29] is:

$$\frac{M_{in-plane}}{M_S} = \frac{R_P}{R} \frac{(R_{\downarrow\rightarrow} - R)}{(R_{\downarrow\rightarrow} - R_P)} \tag{7}$$

where $M_{in\text{-}plane}$ is the in-plane magnetization component, R_P is the MTJ resistance when both the free and reference layer are in-plane, R is the measured resistance, $R_{\downarrow\rightarrow}$ is the resistance when no field is applied, and the reference and free layer are

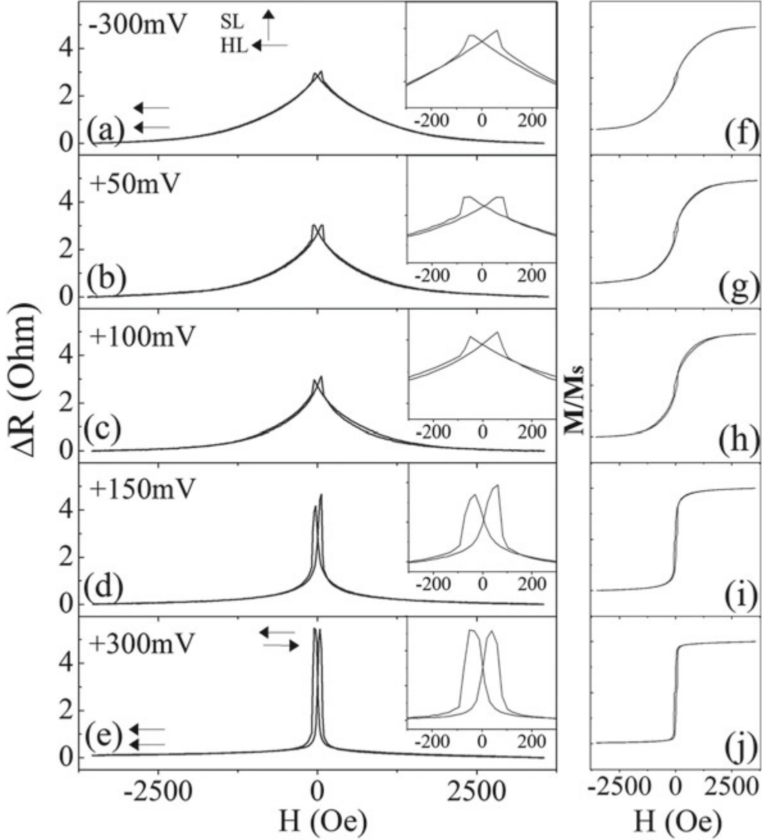


Fig. 6 **a–e** When a negative voltage is applied, the resistance sweep exhibits the typical hard axis response shown by taking the example of an MTJ stack of V/Fe/MgO/Fe. **f–j** This can be converted to the regular *M-H* curve using Eq. 7. When a positive voltage is applied, the hard axis now becomes the easy axis as shown by the M-H curve. Reproduced from [29] © 2013, with the permission of AIP Publishing

orthogonal. Taking V/Fe/MgO/Fe as an example, one can observe from Fig. 6a–e that as the voltage is changed from negative to positive, the magnetization becomes increasingly in-plane. Here, a positive voltage is applied to the bottom V/Fe soft layer. Figure 6f–j shows the magnetization calculated using Eq. 7. To obtain the change in the anisotropy, the normalized area along the magnetization is computed (Fig. 7b) which is similar to the previous discussions on anomalous Hall tuning using Eq. 2. An analogous expression can be used to calculate the anisotropy field, H_k. The expression in Eq. 2 only uses positive magnetization but implicitly assumes that the magnetization is symmetric. The area computed by integrating $H\mathrm{d}M$ is shown by the shaded region in Fig. 7b. Alternatively, if the magnetization follows an ideal hard axis straight line, one can also use the saturation field, H_{sat}, to replace H_K and directly [43] determine ξ using Eq. 6.

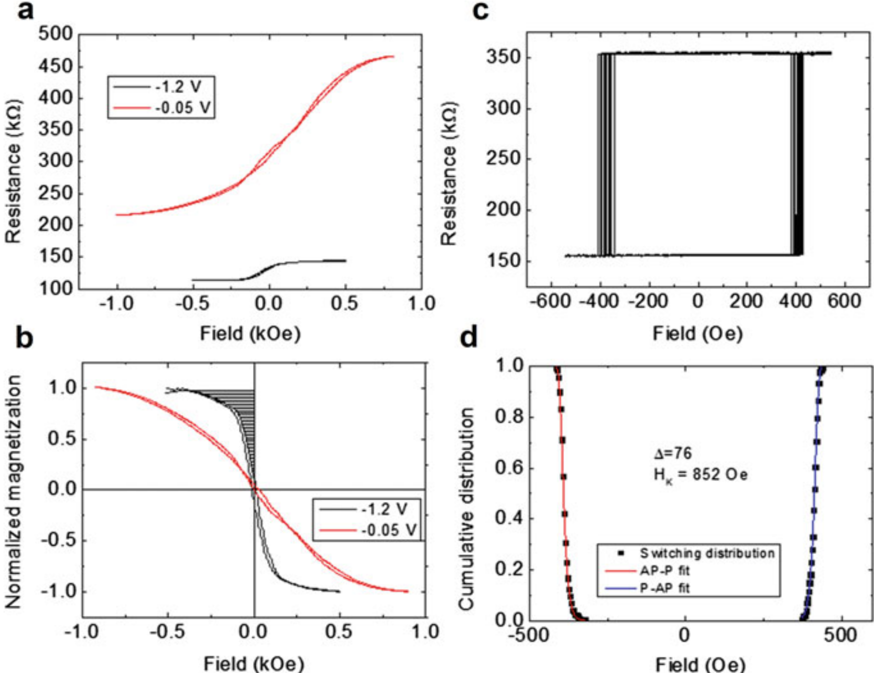


Fig. 7 **a** Plot showing resistance with an applied field in the direction parallel to the device surface. The stack is as follows: PtMn pinned layer/MgO/CoFeB 1.6 nm/Ta. **b** The resistance data in Fig. 7a is converted to conductance and normalized to give the in-plane magnetization. To obtain the anisotropy perpendicular anisotropy energy the triangular area is calculated using Eq. 3. **c** Another way to estimate the change in energy is by using the switching field distribution. The figure shows multiple sweeps of the hysteresis loop using a fixed sweep rate. **d** Using Eq. 8, one can then extract the change in anisotropy from the data in Fig. c. Here the MTJ shows $\Delta \approx 76$ and $H_k \approx 852$ Oe. This is the average value from both AP → P and P → AP transitions

As an alternative, one can also fix the pinned layer in-plane by using an antiferromagnet such as PtMn. The *R-H* from such a stack is shown in Fig. 7a. The difference in the resistance values and the TMR scale for −1.2 V and −0.05 V originates from the voltage-dependent resistance of the MTJ. Similar to the case with the in-plane pinned layer, the resistance shown in Fig. 7a has to be converted to conductance (Fig. 7b) since conductance is proportional to the in-plane component of the free layer [56]. The disadvantage of this method is the exchange bias is typically weaker than the demagnetization field of a thick in-plane pinned layer. This causes the pinned layer to tilt at a lower field, limiting the measurement range for both the voltage and the magnetic field sweeps. Other than this difference in converting resistance to magnetization, both MTJ and gated-Hall bars yield curves similar to that shown in Figs. 6f–j and 7b.

The "area method" is suitable in cases when the pinned layer is assumed not to affect the magnetic properties of the free layer. However, in cases where mechanical effects and the related strain can cause changes to perpendicular anisotropy [57], it

becomes necessary to estimate the VCMA effect within the actual perpendicular MTJ stack where both the reference and the free layer show PMA. This estimation can be done by measuring the distribution of the switching fields in the MTJ using the following equation,

$$P = 1 - \exp\left[\frac{-H_k f_0 \sqrt{\pi}}{2R\sqrt{\Delta}} \mathrm{erfc}\left(\sqrt{\Delta}\left(1 - \left|\frac{H - H_0}{H_k}\right|\right)\right)\right] \tag{8}$$

where f_0 is the attempt frequency (assumed to be 1 GHz), R, the sweep rate. H_o is the median offset field and, Δ is the thermal stability. The change in anisotropy can then be extracted when H_k is determined at various applied voltages. The determination of H_k is illustrated in Fig. 7c,d taken for a single applied voltage. Figure 7c shows the hysteresis that is obtained by sweeping an external field multiple times. This gives us a distribution of the switching fields (left and right P → AP and AP → P). As shown by Fig. 7d, this distribution is then fitted into the cumulative distribution, P, given by Eq. 8, which gives us values for H_k and Δ [58].

We have discussed in Sect. 2.2, the determination of ξ from the absorption spectrum using the field-sweep FMR in a resonant cavity. However, other forms of FMR can also be used, such as the homodyne detection method where the device employed is not a single ferromagnetic stack but rather an MTJ [27, 59]. For this method, the DC resistance of the MTJ is measured while the AC voltage is applied to the MTJ. The FMR signal is then read out via the rectified DC voltage across the device.

2.4 Summary of Materials Dependence

Table 1 shows a list of heterostructure stacks (not exhaustive) and the corresponding VCMA efficiencies.

The large efficiencies of certain materials, e.g. the V/Fe/MgO stack, have been attributed to the electromigration of defects in the MgO barrier. For VCMA originating from electromigration, the mechanical motion of chemical migration limits the switching speed and may be susceptible to reliability issues similar to resistive random access memory (RRAM) [60]. Note that for the typical Ta/CoFeB/MgO heterostructure, a range of values, 30–100 fJV^{-1} m^{-1}, have been reported. It is generally believed that the typical value is around 30 fJV^{-1} m^{-1} and values at the high or low end correspond to a difference in the level of oxidation of the Co/Fe oxide at the CoFeB/MgO interface [61].

As indicated by Table 1 and eluded in Sect. 2.1, the underlayer plays an integral part in the magnitude of VCMA. It has been reported by Barnes et al. [62] that there is an efficiency inversion between 4d and 5d metals. However, apart from a single study with Ru [63], most studies have not been able to replicate this successfully [59], indicating perhaps stringent preparation conditions. It is also noteworthy that VCMA generally decreases with increasing annealing temperature [59]. On the other hand, W and Mo underlayer structures do not show a strong dependence

Table 1 A table of E-field efficiencies and the associated free layer/underlayer heterostructure. The origin of the coefficient is also stated

Heterostructure	E-field efficiency (fJV^{-1} m^{-1})	Origin	Reference
Ta/CoFeB/MgO	33,85,100	Electronic	[64, 23, 61]
Ir/CoFeB/MgO	100	Electronic	[65]
Ru/CoFeB/MgO	−18	Electronic	[63]
W/CoFeB/MgO	20–50	Electronic	[52, 59]
Mo/CoFeB/MgO	40	Electronic	[39]
MgO/$Fe_{80}B_{20}$/MgO	108 (+ve bias), 24 (−ve bias)	Electronic	[66]
Ir-doped Fe/Cr	320	Electronic	[67]
V/Fe/MgO	1150	Charge trapping/electromigration	[29]
Ta/CoFeB/MgO/PZT	20	Electronic	[68]
FePd/MgO	602	Electronic/charge trapping	[53]

with annealing temperature and may be suitable for high-temperature applications of EF-MRAM [39, 59].

From a practical standpoint, for any material system to gain traction, there is an absolute minimum ξ required. This lower bound can be estimated based on a few limits of MRAM and are as follows— (1) the thermal stability required for 10 years of operation, usually 60 k_bT (k_b is the Boltzmann constant and T is the temperature, which ensures less than 1 ppm of bits flipping in that time frame), (2) the MgO breakdown voltage, and (3) the maximum thermal stability possible with the nucleation diameter constraints [69]. Thus, combining Eq. 6 with the formula for thermal stability, $\Delta = \frac{M_S H_K V}{2K_B T}$(where V is the volume, K_B the Boltzmann constant and T the temperature), and using a nucleation diameter of about 40 nm [69] and the approximate breakdown voltage of 1 V/nm [70] give us an approximate ξ of at least 200 fJV^{-1} m^{-1}. This is the lower limit and does not include any buffers necessary in the design of actual memory to ensure a large enough sigma separation between write and breakdown voltage in Mbit sized arrays. This implies that most materials fall short of this requirement. One silver lining is that the time applied for programming is typical of the order of ns and the MTJ is not expected to breakdown at high voltage if the pulse is kept short. This implies that any material can still be expected to last its full (10 years) lifetime usage as long as the total accumulated time-to-failure is within limits. As an example, if the breakdown voltage at sub-ns pulses doubles, then this will likely reduce the lower limit of ξ by half.

The last thing to note is that for certain materials, PMA modulation is not linear with voltage [67] and the values quoted in Table 1 usually refer to the region where the highest modulation was observed (i.e. the steepest slope, see [67]). As such, an ideal figure of merit should be the effective slope over the breakdown voltage of MgO. Since MgO typically breakdowns at ~ 0.8–1 V/nm for thick MgO [71, 70],

this implies that the effective slope should be calculated over this entire voltage range instead. One extreme example of non-linear modulation of PMA is from the double MgO free layer [66]. Here, the change in anisotropy follows a V-shape centered with the minimal around 0 V. For both negative and positive voltages, the anisotropy increases, making the MTJ harder to write regardless of the voltage polarity.

In the next section, we will discuss how VCMA and its related effects have been utilized in MRAM.

3 Device Operation

3.1 MTJ Switching

Since the change in perpendicular anisotropy can be translated into a change in magnetization direction, Shiota and Kanai et al. have utilized this to control the spin dynamics of the MTJ [22, 32] via voltage resulting in precessional switching. The basic principles are illustrated in Fig. 8.

As seen in Fig. 8a, the VCMA effect is used to reduce the magnetization component in the z-direction. An external magnetic field is then required in the x (or y) direction to provide an axis about which the magnetization can precess. The dynamics of this precessional effect is given by the Larmor frequency (or period τ) [23],

$$\tau \approx 2\pi/\gamma\mu_0 H_x \tag{9}$$

where γ is the gyromagnetic constant, μ_0 the permittivity of free space and H_x the field in the x-direction. Here, we have ignored the effect of the electric-field on the reference layer but depending on the strength of the voltage this may be ignored [23]. We have also ignored the damping constant term α in Eq. 7 of Ref. [72] since this term is small. Figure 9a and b show typical colormap plots of the switching probability against both the field and the width of the electric-field pulse for magnetization reversal from AP to P and from P to AP respectively. The ridges or regions where the switching probability approaches 1, basically correspond to half a period τ defined in Eq. 9. As shown in Fig. 8c, which is a typical cross-section of the plots in Fig. 9 at a certain H_x, pulses equivalent to even multiples of half a period result in the magnetization returning to its starting point while pulses which are odd multiples of half a period corresponding to 180° rotation resulting in the switching of the initial magnetization. This can also be understood from Fig. 8b. Two features can generally be noted in Fig. 9a and b. First, along the contour given by Eq. 9 (i.e. assuming a fixed number of precessional cycles), it can be seen that the probability is low when the pulse width is high or the amplitude of H_x is low. This can be understood as follows. A high H_x will cause the magnetization to settle in the H_x direction. This implies that the probability for both transitions tends to 0.5. This is shown by the narrowing and gradually decreasing contour when H_x is high. As for long pulses

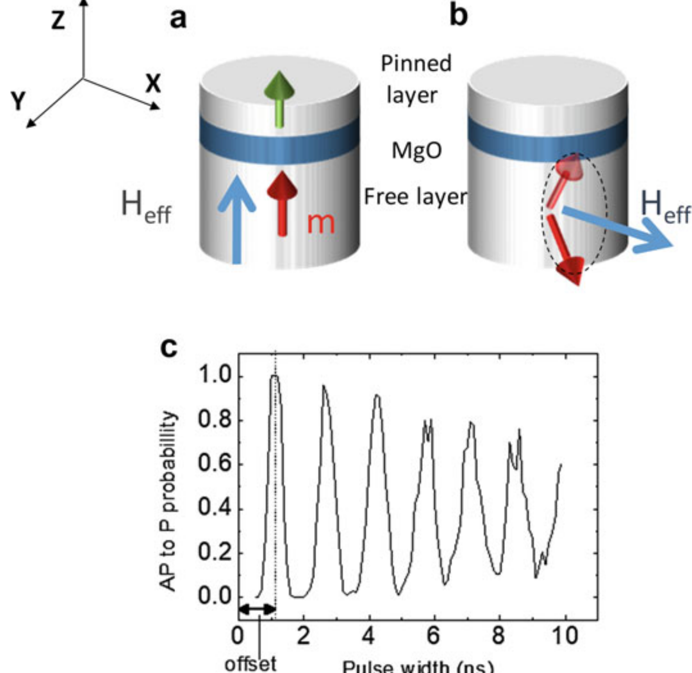


Fig. 8 **a, b** In precessional switching, a voltage is applied to bring the magnetization in-plane. This then causes the magnetization to precess about the applied magnetic field in the x-direction. Half a Larmor period causes the m_Z to flip while a full period causes m_Z to return to its original direction. **c** Switching probability of MRAM in precessional switching. Shown here is anti-parallel to parallel switching. The switching probability decreases with the higher-order periods. Different MTJs also show random changes to the starting position or offset.

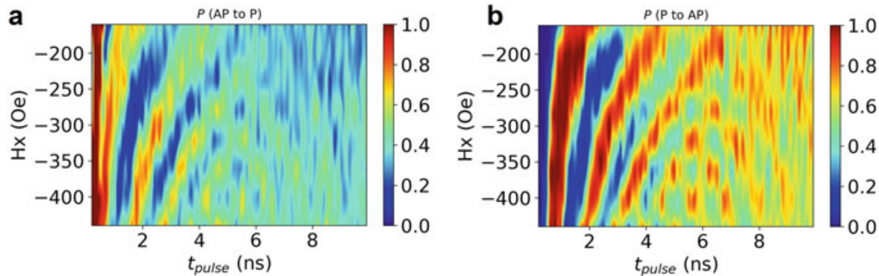


Fig. 9 **a** The switching probability contour plot showing precessional switching from AP to P, **b** The switching probability from P to AP. For a certain number of precessional cycles, the Larmor period (τ) shortens on increasing the amplitude of H_x. This is shown by the ridges in the colormap for both Fig. a and b. Reproduced from [24] © 2018, with the permission of AIP Publishing

(long τ), Kanai et al. have shown through simulations that longer pulse widths are more susceptible to thermal effects. Thermally-induced random stray fields from the pinned layer can cause phase shifts to the switching probability, again causing transition probabilities to tend towards 0.5. This is shown by the contour that becomes lower in intensities and more diffused with large pulse widths. This decoherence due to thermal effects also affects the higher-order peaks (more precessional cycles) in these contours more pronouncedly—the switching probability decreases with the second, third peaks and so on in the contour. This behavior is also evident in Fig. 8b. While precessional switching is fast (<1 ns), there are several drawbacks. Firstly, as seen from Fig. 8b, the switching probability has a sinusoidal dependence with pulse width which implies an accurate pulse width is required to switch with high accuracy. Secondly, an external in-plane field is required. Thirdly, the probability distributions may not have the same starting offset (Fig. 8b). Experimentally, it can be seen that different devices may have different offsets. This contributes to a further increase in errors for the fully precessional scheme. We will address some of these concerns and EF-MRAM's application in actual circuits in the next section.

3.2 Interplay of Electric-Field and Spin-Transfer Torque

Proposals that use only E-field is limited by the unidirectionality of the VCMA effect, i.e. typically only when a positive voltage is applied to the free layer does the surface anisotropy decrease. There are 2 documented ways to get this to work with while using a unidirectional current. The first scheme, proposed by Alzate et al. [73] relies on the fact that the spin-torque effect can still be relevant when the current is large.

Alzate et al. reported that the introduction of a small STT leakage current can act as a small effective magnetic field. The STT current, in this case, is not able to induce switching on its own. Since this effect is opposite of what is expected of the antidamping torque, this was attributed to some non-trivial effect of the field-like torque in STT acting in conjunction with the EF. This scheme should be highly dependent on the exact spin dynamics or the shape of the voltage pulse— there is no reason why the P state does not revert to the AP state as the voltage is reduced after writing to the P state. This may explain the high write error rates (WER). Lastly, the slight shift in the bias field required also implies an imbalance in the stability of the P and AP states. This might be a bottleneck of this scheme at elevated temperatures where the coercivity of the MTJ and hence its thermal stability is expected to shrink [74].

The second scheme was proposed by Kanai et al. [21] and is depicted in Fig. 10. As shown in Fig. 10a, two pulses are applied to the MTJ, the first one, with higher amplitude, reduces the anisotropy and causes precession while the second one with lower amplitude stabilizes the final magnetization with STT. Since STT is directional, the 2nd pulse is opposite in direction to first for AP $\rightarrow$ P switching as shown by Fig. 10b. Lower WER was reported in this scheme compared to pure precessional switching alone since the STT pulse flattens the oscillatory switching probability of

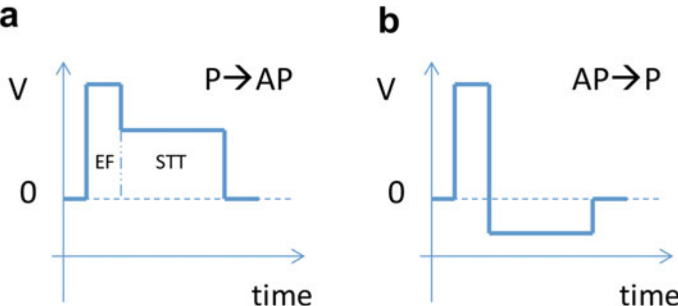


Fig. 10 Pulse waveforms for a switching scheme using EF + STT. Following Ref. [21], Fig. **a** shows scheme for P to AP switching while Fig. **b** shows scheme for AP to P switching with the aid of an additional STT pulse

the EF pulse. Similar to the first scheme, this scheme should technically be dependent on the exact spin dynamics and pulse timings. This is especially so for AP → P switching where the polarity of the STT pulse is opposite to the EF pulse and should cause sudden dampening of the precession due to the increase in PMA.

3.3 Secondary Electric-Field Effects

Performance of MRAM based on electric-field tuning of magnetic anisotropy can be compromised from possible secondary effects of the application of an external electric-field. Particularly in an MTJ structure, electric-field may also modulate the anisotropy of the reference layer [33]. In this scenario, one has to design a sufficiently strong reference layer whose magnetization will not switch during the device operation.

Another effect that needs to be considered is the electric-field tuning of Gilbert damping. It was reported that for Ta/CoFeB/MgO structure at CoFeB thickness of 1.4 nm, the damping could be modulated by $\sim -21\%$ by an applied electric-field of 1 V/nm [54]. Damping has significant effects on the precessional dynamics.

Additionally, it has been found that in CoFeB/GdO_x/CoFeB MTJs electric-field can tune coupling between the two magnetic layers from FM to AFM paving the way for an alternative approach to VCMA for magnetization switching [75].

3.4 Write-Error-Rate

As mentioned earlier, proposals that employ precessional switching or are dependent on pulse timings/amplitude show a high write-error-rate (WER), with extensive studies performed by Wang et al. [76]. To its benefit, the voltage pulses required in EF-MRAM typically have lower sigmas compared to that for current pulses — both the amplitude and rise and fall times of the voltage pulse are either similar

or tighter than the current pulses used in STT-MRAM [76]. However, even with moderate increases in temperature (room to 50 °C), the WER increases significantly from 10^{-8} to 10^{-4} [76]. The authors attributed this to the reduction in magnetization, which reduces the in-plane demagnetization field. This causes an increase in the Larmor frequency when the same pulse amplitude is used. While this can be solved with various circuit workarounds like pre-reads (writing only if the bit is in the wrong state) and the multiple writes (correcting for bits that fail to write correctly on the first pulse), this poses a challenge since consumer specs require a range of at least 0 to 70 °C [74].

Unlike conventional CMOS-based memory, MRAM is inherently probabilistic due to the thermal energy barrier. For STT-MRAM, one can increase write pulse widths such that one is always writing in the thermal activation regime. This brings the BER down to levels that can be covered by ECC [77]. However, for E-field MRAM schemes that use precessional switching primarily, the narrow pulse width switching windows make designs difficult without additional circuit workarounds. As shown by the switching probably in Fig. 8c, It can be seen that the switching probability drops drastically, the further the pulse width is away from the peak. The switching error rate of a single bit has been further experimentally studied by Shiota et al. [72, 78]. As shown by the switching probability in Fig. 11a, due to the sinusoidal write probabilities, the WER can increase from its minimum (~3×10^{-3}) to 0.5 in an interval of less than 0.5 ns [72]. It was shown that with a sufficiently high Δ, it was possible to obtain a WER of 10^{-5} or lower. The upper bound was solely limited by the breakdown voltage being lower than the critical voltage (voltage whereby H_k or Δ goes to zero) as described in a previous section. Simulations also showed that the WER could be further decreased with a lower damping constant, α, as shown in Fig. 11b and a (iii). In Fokker–Planck simulations, the thermal agitation is almost proportional to α [79] which explains the lower WER with lower α. It is to be noted that these studies typically use a single MTJ and the WER for an array is expected to be worse since each MTJ is expected to have its own optimal programming pulse width.

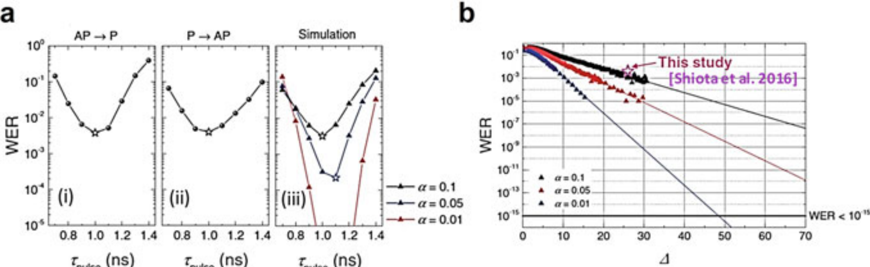


Fig. 11 **a** Figures show the error rate as a function of pulse width for both (i) AP → P and (ii) P → AP transitions. Simulation results are shown in (iii). **b** Also, from Ref. [72]. Write-error-rate (WER) simulation of a bit written at critical voltage and a pulse width set at half a period. It can be seen that a high Δ and a low damping constant, α contribute to a low WER. Here it is assumed that ξ is sufficient to bring the device's Δ to zero at a voltage not exceeding breakdown. Reprinted with permission from [72] © 2016 The Japan Society of Applied Physics

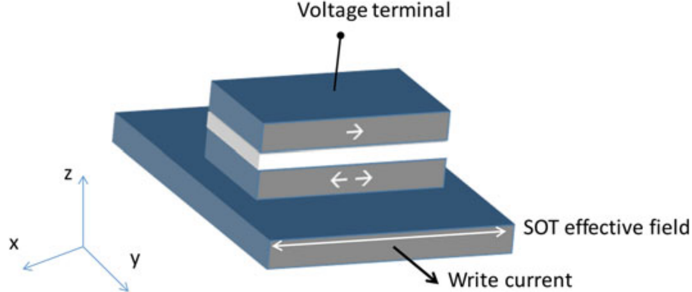


Fig. 12 A VoCSM (voltage-controlled spintronic memory) memory stack. The top voltage terminal is used to lower the anisotropy of the free layer. An effective field is then produced by the SOT writing current, which causes deterministic switching

One of the proposed schemes for lowering WER is to combine both the advantages of VCMA and spin–orbit-torque writing schemes. This is known as the VoCSM (voltage-controlled spintronic memory) and Fig. 12 illustrates this approach. Basically, voltage is applied across the MgO, functioning somewhat like a gate [similar to Fig. 3a], lowering the PMA of the selected MTJ bit. A current is then applied in the SOT channel, which follows the transverse path to switch the bit either via spin Hall effect or Rashba [80]. This scheme retains the structure (and also the added fabrication complexity) of the 3 terminal structure needed for SOT-based MRAM. There are several features to this approach. Firstly, conventional materials (e.g. Ta, W) compatible with CFB typically have small spin Hall angles (a measure of charge-to-spin current conversion efficiency) and cannot easily be used, on their own to switch the high thermal stability structures used in MRAM. Thus, EF and SOT effects are complementary in this scheme. Secondly, this scheme is not inherently dependent on the shape of the pulse-like those used for precessional or STT + VCMA MRAM discussed earlier [73, 21] and as such will be expected to have larger operating margins and lower WER. Thirdly, VCMA effects can also be used to strengthen the PMA during read (by the application of an appropriate voltage polarity), safeguarding the MRAM from read disturb. One disadvantage is that due to the complexity, 3 terminal structures are not expected to scale easily.

3.5 *Schemes for Field-Free Switching*

One of the major challenges of MTJ switching via the electric field is the requirement of an in-plane magnetic field component to realize switching. Besides the previously discussed EF + STT scheme for field-free switching, simulations have shown that this can be done by either inserting an in-plane layer [81] (Fig. 13a) or by using a conical free layer [82] (Fig. 13b). As its name suggest, with an in-plane magnetic layer, a thick ferromagnetic layer is placed adjacent to the MTJ separated by a metallic spacer. This supplies the required H_{bias} needed for precessional switching.

Fig. 13 **a** Following Ref. [81], an in-plane layer is used to provide the bias field for precessional switching. **b** Following Ref. [82], the free layer is already tilted due to the competition between the first and second-order anisotropies. Both schemes shown in Fig. a and b require an elliptical MTJ to fix the precessional axis

As for conical magnetization free layers, they have experimentally been found in Co/Pt multilayers [83] as well as on Co with Pt and Pd substrates [84]. Conical free layers are obtained when the 1st order anisotropy constant, H_{k1}, is negative and the 2nd order anisotropy constant, H_{k2}, is positive. Stacks with conical free layer have been shown in simulations to be beneficial to STT MRAM— lower switching current and faster switching time. Both the in-plane and conical free layer use an elliptical MTJ to fix the precession axis. One setback of both these proposals is the reduced thermal stability due to canted free layer. For instance, with an external field H_x, the effective thermal stability is modified as follows $\Delta = \Delta_0\left(1 - \frac{H_x}{H_k}\right)^2$, where Δ_0 is the zero-field thermal stability. It can be seen that in the presence of an applied field, the effective Δ is smaller. Yet another setback is the slightly reduced TMR since the free and pinned layers are not perfectly antiparallel/parallel to one another. In addition, for the in-plane layer approach, since the field is local to the MTJ, these conditions (e.g. shape variations) may vary from MTJ-to-MTJ within an array, introducing a process-induced variation and degrading WER. Lastly, both these proposals still switch via precession which means the pulse width constraints discussed in Sect. 3.1 remain.

4 Applications and Integration with CMOS

One upside to the EF writing approach is the ability to use a crossbar architecture [73]. This is the tightest architecture known which allows a $4F^2$ cell size [85] instead of the $6F^2$ seen for 1T-1MTJ scheme [86]. This is made possible because the voltage can be made unipolar. The simplicity of the crossbar compared to conventional architecture is shown in Fig. 14. On the other hand, limitations to the crossbar scheme include the leakage to the unselected cells. This limits the array size. Another downside is that diodes are inherently slower than MOS selectors. On the positive side, diodes

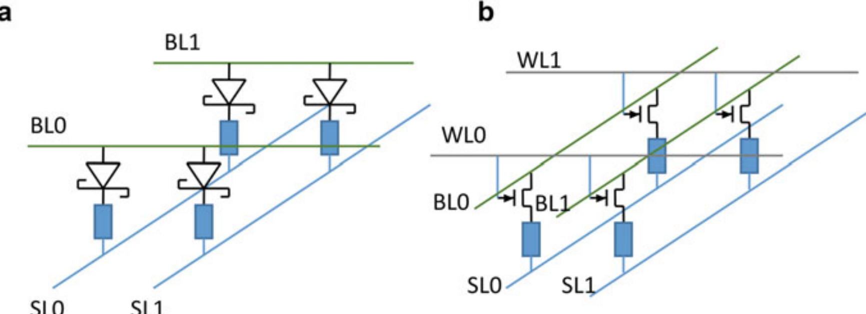


Fig. 14 **a** A crossbar architecture in an MTJ array. **b** A conventional 1T-1MTJ scheme. BL, SL, WL refer to bitline, sourceline and wordline respectively. Here, the blue rectangular structures represent MTJs

Table 2 A table of the various EF-MRAM schemes

Scheme	Pros	Cons	Reference
Conventional precessional	Switching within 2 ns Unipolar voltage	Requires external field Sensitive to write voltage/pulse width	[32, 22]
STT + EF (unipolar)	Unipolar voltage	Higher WER Requires external field	[73]
STT + EF (bipolar)	Lower WER compared to pure precessional switching	Bipolar voltage required Requires external field	[21]
VoCSM	No external field required Switching independent of pulse timings	Complicated structure	[80]
Conical free layer (proposal)	No external field required Faster switching times	Reduced thermal stability and TMR Sensitive to write voltage/pulse width	[82]
In-plane layer (proposal)	No external field required Faster switching times	Reduced thermal stability and TMR Sensitive to write voltage/pulse width	[81]

are capable of supplying larger current [87] with large on/off ratios, allowing more aggressive cell size shrinkage compared to CMOS-based selectors.

We will end this section with a table of the EF-MRAM schemes highlighting their pros and cons (Table 2).

5 Summary

E-field MRAM is one of the best solutions for next-generation ultra-fast and low-power memory applications. However, there are two bottlenecks for this technology

that need to be addressed immediately, viz (i) the high BER and (ii) the need for an in-plane magnetic field. The various schemes introduced in this chapter show how this two issues can be overcome in some scenarios but more work is needed to show it can truly replace existing STT-MRAM. Materials development is going to be an essential pursuit in this direction. Particularly, first principle calculations have shown a wealth of materials dependence and can potentially discover even higher EF efficiencies in unexplored material systems.

References

1. D. Apalkov, B. Dieny, J. Slaughter, Magnetoresistive random access memory. Proc. IEEE **104**, 1796–1830 (2016). https://doi.org/10.1109/JPROC.2016.2590142
2. L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current. Phys. Rev. B **54**, 9353–9358 (1996). https://doi.org/10.1103/PhysRevB.54.9353
3. A. Brataas, A.D. Kent, H. Ohno, Current-induced torques in magnetic materials. Nat. Mater. **11**, 372–381 (2012). https://doi.org/10.1038/nmat3311
4. C. Chappert, A. Fert, D.F.N. Van, The emergence of spin electronics in data storage. Nat. Mater. **6**, 813–823 (2007). https://doi.org/10.1038/nmat2024
5. Z. Diao, M. Pakala, A. Panchula et al., Spin-transfer switching in MgO-based magnetic tunnel junctions (invited). J. Appl. Phys. **99**, 08G510 (2006). https://doi.org/10.1063/1.2165169
6. J.C. Slonczewski, Current-driven excitation of magnetic multilayers. J. Magn. Magn. Mater. **159**, L1–L7 (1996). https://doi.org/10.1016/0304-8853(96)00062-5
7. Z. Diao, D. Apalkov, M. Pakala et al., Spin transfer switching and spin polarization in magnetic tunnel junctions with MgO and AlO_x barriers. Appl. Phys. Lett. **87**, 232502 (2005). https://doi.org/10.1063/1.2139849
8. J. Hayakawa, S. Ikeda, Y.M. Lee et al., Current-driven magnetization switching in CoFeB/MgO/CoFeB magnetic tunnel junctions. Jpn J. Appl. Phys. **44**, L1267 (2005). https://doi.org/10.1143/JJAP.44.L1267
9. H. Kubota, A. Fukushima, Y. Ootani et al., Evaluation of spin-transfer switching in CoFeB/MgO/CoFeB magnetic tunnel junctions. Jpn J. Appl. Phys. **44**, L1237–L1240 (2005). https://doi.org/10.1143/JJAP.44.L1237
10. S. Ikeda, K. Miura, H. Yamamoto et al., A perpendicular-anisotropy CoFeB—MgO magnetic tunnel junction. Nat. Mater. **9**, 721–724 (2010). https://doi.org/10.1038/nmat2804
11. S. Mangin, D. Ravelosona, J. Katine, a, et al., Current-induced magnetization reversal in nanopillars with perpendicular anisotropy. Nat. Mater. **5**, 210–215 (2006). https://doi.org/10.1038/nmat1595
12. H. Meng, J. Wang, Spin transfer in nanomagnetic devices with perpendicular anisotropy. Appl. Phys. Lett. **88**, 172506 (2006). https://doi.org/10.1063/1.2198797
13. J.Z. Sun, S.L. Brown, W. Chen et al., Spin-torque switching efficiency in CoFeB-MgO based tunnel junctions. Phys. Rev. B **88**, 104426 (2013). https://doi.org/10.1103/PhysRevB.88.104426
14. D.C. Worledge, G. Hu, D.W. Abraham et al., Spin torque switching of perpendicular Ta | CoFeB | MgO -based magnetic tunnel junctions. Appl. Phys. Lett. **98**, 022501 (2011). https://doi.org/10.1063/1.3536482
15. H. Liu, D. Bedau, D. Backes et al., Ultrafast switching in magnetic tunnel junction based orthogonal spin transfer devices. Appl. Phys. Lett. **97**, 242510 (2010). https://doi.org/10.1063/1.3527962
16. J. Lourembam, B. Chen, A. Huang et al., (2018a) A non-collinear double MgO based perpendicular magnetic tunnel junction. Appl. Phys. Lett. **113**, 022403 (2018). https://doi.org/10.1063/1.5038060

17. S. Matsunaga, J. Hayakawa, S. Ikeda et al., Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions. Appl. Phys. Express **1**, 0913011–0913013 (2008). https://doi.org/10.1143/APEX.1.091301
18. F. Ren, D. Markovic, True energy-performance analysis of the MTJ-based logic-in-memory architecture (1-bit full adder). IEEE Trans. Electron. Devices **57**, 1023–1028 (2010). https://doi.org/10.1109/TED.2010.2043389
19. S. Kanai, F. Matsukura, H. Ohno, Electric-field-induced magnetization switching in CoFeB/MgO magnetic tunnel junctions with high junction resistance. Appl. Phys. Lett. **108**, 192406 (2016). https://doi.org/10.1063/1.4948763
20. P.K. Amiri, J.G. Alzate, X.Q. Cai et al., Electric-field-controlled magnetoelectric RAM: progress, challenges, and scaling. IEEE Trans. Magn. **51**, 1–7 (2015). https://doi.org/10.1109/TMAG.2015.2443124
21. S. Kanai, Y. Nakatani, M. Yamanouchi et al., Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field-effect. Appl. Phys. Lett. **104**, 212406 (2014). https://doi.org/10.1063/1.4880720
22. S. Kanai, Y. Nakatani, M. Yamanouchi et al., In-plane magnetic field dependence of electric field-induced magnetization switching. Appl. Phys. Lett. **103**, 072408 (2013). https://doi.org/10.1063/1.4818676
23. S. Kanai, M. Yamanouchi, S. Ikeda et al., Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. Appl. Phys. Lett. **101**, 122403 (2012). https://doi.org/10.1063/1.4753816
24. J. Lourembam, J. Huang, S.T. Lim, E.F. Gerard, Role of CoFeB thickness in electric field controlled sub-100 nm sized magnetic tunnel junctions. AIP Adv. **8**, 055915 (2018). https://doi.org/10.1063/1.5006368
25. F. Matsukura, Y. Tokura, H. Ohno, Control of magnetism by electric fields. Nat. Nanotechnol. **10**, 209–220 (2015). https://doi.org/10.1038/nnano.2015.22
26. T. Nozaki, A. Koziol-Rachwal, W. Skowronski, et al., Large voltage-induced changes in the perpendicular magnetic anisotropy of an MgO-based tunnel junction with an ultrathin fe layer. Phys. Rev. Appl. **5**, 044006 (2016). https://doi.org/10.1103/PhysRevApplied.5.044006
27. T. Nozaki, Y. Shiota, S. Miwa et al., Electric-field-induced ferromagnetic resonance excitation in an ultrathin ferromagnetic metal layer. Nat. Phys. **8**, 491–496 (2012). https://doi.org/10.1038/nphys2298
28. T. Nozaki, Y. Shiota, M. Shiraishi et al., Voltage-induced perpendicular magnetic anisotropy change in magnetic tunnel junctions. Appl. Phys. Lett. **96**, 022506 (2010). https://doi.org/10.1063/1.3279157
29. A. Rajanikanth, T. Hauet, F. Montaigne et al., Magnetic anisotropy modified by electric field in V/Fe/MgO(001)/Fe epitaxial magnetic tunnel junction. Appl. Phys. Lett. **103**, 062402 (2013). https://doi.org/10.1063/1.4817268
30. T. Seki, M. Kohda, J. Nitta, K. Takanashi, Coercivity change in an FePt thin layer in a Hall device by voltage application. Appl. Phys. Lett. **98**, 212505 (2011). https://doi.org/10.1063/1.3595318
31. Y. Shiota, S. Miwa, T. Nozaki et al., Pulse voltage-induced dynamic magnetization switching in magnetic tunneling junctions with high resistance-area product. Appl. Phys. Lett. **101**, 102406 (2012). https://doi.org/10.1063/1.4751035
32. Y. Shiota, T. Nozaki, F. Bonell et al., Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses. Nat. Mater. **11**, 39–43 (2012). https://doi.org/10.1038/nmat3172
33. W.-G. Wang, M. Li, S. Hageman, C.L. Chien, Electric-field-assisted switching in magnetic tunnel junctions. Nat. Mater. **11**, 64–68 (2012). https://doi.org/10.1038/nmat3171
34. W.G. Wang, C.L. Chien, Voltage-induced switching in magnetic tunnel junctions with perpendicular magnetic anisotropy. J. Phys. D Appl. Phys. **46**, 074004 (2013). https://doi.org/10.1088/0022-3727/46/7/074004
35. M. Weisheit, S. Fähler, A. Marty et al., Electric field-induced modification of magnetism in thin-film ferromagnets. Science **315**, 349–351 (2007). https://doi.org/10.1126/science.1136629

36. C. Grezes, F. Ebrahimi, J.G. Alzate et al., Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product. Appl. Phys. Lett. **108**, 012403 (2016). https://doi.org/10.1063/1.4939446
37. H. Cai, W. Kang, Y. Wang et al., High performance MRAM with spin-transfer-torque and voltage-controlled magnetic anisotropy effects. Appl. Sci. **7**, 929 (2017). https://doi.org/10.3390/app7090929
38. K. Kita, D.W. Abraham, M.J. Gajek et al., Electric-field-control of magnetic anisotropy of Co0.6Fe0.2B0.2/oxide stacks using reduced voltage. J. Appl. Phys. **112**, 033919 (2012). https://doi.org/10.1063/1.4745901
39. X. Li, G. Yu, H. Wu et al., Thermally stable voltage-controlled perpendicular magnetic anisotropy in Mo|CoFeB|MgO structures. Appl. Phys. Lett. **107**, 142403 (2015). https://doi.org/10.1063/1.4932553
40. P.V. Ong, N. Kioussis, D. Odkhuu et al., Giant voltage modulation of magnetic anisotropy in strained heavy metal/magnet/insulator heterostructures. Phys. Rev. B **92**, 020407(R) (2015). https://doi.org/10.1103/PhysRevB.92.020407
41. S. Kanai, M. Tsujikawa, Y. Miura et al., Magnetic anisotropy in Ta/CoFeB/MgO investigated by x-ray magnetic circular dichroism and first-principles calculation Magnetic anisotropy in Ta/CoFeB/MgO investigated by x-ray magnetic circular dichroism and first-principles calculation. Appl. Phys. Lett. **105**, 222409 (2014). https://doi.org/10.1063/1.4903296
42. T. Maruyama, Y. Shiota, T. Nozaki et al., Large voltage-induced magnetic anisotropy change in a few atomic layers of iron. Nat. Nanotechnol. **4**, 158–161 (2009). https://doi.org/10.1038/nnano.2008.406
43. S. Zhang, Spin-dependent surface screening in ferromagnets and magnetic tunnel junctions. Phys. Rev. Lett. **83**, 640–643 (1999). https://doi.org/10.1103/PhysRevLett.83.640
44. M. Zeng, J. Lourembam, S.T. Lim, Electric-field Effect on Magnetic Anisotropy of MgO/CoFe/capping structures. J. Appl. Phys. **126**, 153902 (2019). https://doi.org/10.1063/1.5082610
45. K.H. He, J.S. Chen, Y.P. Feng, First principles study of the electric field effect on magnetization and magnetic anisotropy of FeCo/MgO(001) thin film. Appl. Phys. Lett. **99**, 072503 (2011). https://doi.org/10.1063/1.3626598
46. M.K. Niranjan, C.-G. Duan, S.S. Jaswal, E.Y. Tsymbal, Electric field effect on magnetization at the Fe/MgO(001) interface. Appl Phys Lett **96**, 222504 (2010). https://doi.org/10.1063/1.3443658
47. X.W. Guan, X.M. Cheng, T. Huang et al., Effect of metal-to-metal interface states on the electric-field modified magnetic anisotropy in MgO/Fe/non-magnetic metal. J. Appl. Phys . **119**, 133905 (2016). https://doi.org/10.1063/1.4945025
48. M. Tsujikawa, S. Haraguchi, T. Oda et al., A comparative ab initio study on electric-field dependence of magnetic anisotropy in MgO/Fe/Pt and MgO/Fe/Au films. J. Appl. Phys. **109**, 07C107 (2011). https://doi.org/10.1063/1.3540677
49. M. Tsujikawa, S. Haraguchi, T. Oda, Effect of atomic monolayer insertions on electric-field-induced rotation of magnetic easy axis. J.. Appl. Phys. **111**, 083910 (2012). https://doi.org/10.1063/1.3703682
50. M. Zeng, J. Lourembam, L.S. Ter, Large electric field modulation of magnetic anisotropy in MgO/CoFe/Ta structures with monolayer oxide insertion. Appl. Phys. Lett. **113**, 192404 (2018). https://doi.org/10.1063/1.5043443
51. M. Endo, S. Kanai, S. Ikeda et al., Electric-field effects on thickness dependent magnetic anisotropy of sputtered MgO/$Co_{40}Fe_{40}B_{20}$/Ta structures. Appl. Phys. Lett. **96**, 212503 (2010). https://doi.org/10.1063/1.3429592
52. Y. Lau, P. Sheng, S. Mitani et al., Electric field modulation of the non-linear areal magnetic anisotropy energy in CoFeB/MgO. Appl. Phys. Lett. **110**, 022405 (2017). https://doi.org/10.1109/INTMAG.2017.8008024
53. F. Bonell, S. Murakami, Y. Shiota et al., Large change in perpendicular magnetic anisotropy induced by an electric field in FePd ultrathin films. Appl Phys. Lett. **98**, 232510 (2011). https://doi.org/10.1063/1.3599492

54. A. Okada, S. Kanai, M. Yamanouchi et al., Electric-field effects on magnetic anisotropy and damping constant in Ta/CoFeB / MgO investigated by ferromagnetic resonance. Appl. Phys. Lett. **105**, 052415 (2014). https://doi.org/10.1063/1.4892824
55. Y. Shiota, S. Murakami, F. Bonell et al., Quantitative evaluation of voltage-induced magnetic anisotropy change by magnetoresistance measurement. Appl. Phys. Express **4**, 043005 (2011). https://doi.org/10.1143/APEX.4.043005
56. J.C. Slonczewski, Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier. Phys. Rev. B **39**, 6995–7002 (1989). https://doi.org/10.1103/PhysRevB.39.6995
57. V.B. Naik, H. Meng, J.X. Xiao, et al., Effect of electric-field on the perpendicular magnetic anisotropy and strain properties in CoFeB/MgO magnetic tunnel junctions. Appl. Phys. Lett. **105**, 052403 (2014). https://doi.org/10.1063/1.4892410
58. J. Huang, M. Tran, L.S. Ter et al., Determination of the electric field induced anisotropy change in sub-100 nm perpendicularly magnetized devices. AIP Adv. **6**, 055805 (2016). https://doi.org/10.1063/1.4942822
59. W. Skowroński, T. Nozaki, D.D. Lam et al., Underlayer material influence on electric-field controlled perpendicular magnetic anisotropy in CoFeB/MgO magnetic tunnel junctions. Phys. Rev. B **91**, 184410 (2015). https://doi.org/10.1103/PhysRevB.91.184410
60. A.J. Lohn, P.R. Mickel, M.J. Marinella, Mechanism of electrical shorting failure mode in resistive switching. J. Appl. Phys. **116**, 034506 (2014). https://doi.org/10.1063/1.4890635
61. X. Li, K. Fitzell, D. Wu et al., Enhancement of voltage-controlled magnetic anisotropy through precise control of Mg insertion thickness at CoFeB|MgO interface. Appl. Phys. Lett. **10**, 052401 (2017). https://doi.org/10.1063/1.4975160
62. S.E. Barnes, J. Ieda, S. Maekawa, Rashba spin-orbit anisotropy and the electric field control of magnetism. Sci. Rep. **4**, 1–5 (2014). https://doi.org/10.1038/srep04105
63. Y. Shiota, F. Bonell, S. Miwa et al., Opposite signs of voltage-induced perpendicular magnetic anisotropy change in CoFeB|MgO junctions with different underlayers. Appl. Phys. Lett. **103**, 082410 (2013). https://doi.org/10.1063/1.4819199
64. T. Inokuchi, H. Yoda, Y. Kato et al., Improved read disturb and write error rates in voltage-control spintronics memory (VoCSM) by controlling energy barrier height. Appl. Phys. Lett. **110**, 252404 (2017). https://doi.org/10.1063/1.4986923
65. W. Skowroński, T. Nozaki, Y. Shiota et al., Perpendicular magnetic anisotropy of Ir/CoFeB/MgO trilayer system tuned by electric fields. Appl. Phys. Expr. **8**, 053003 (2015). https://doi.org/10.7567/APEX.8.053003
66. T. Nozaki, K. Yakushiji, S. Tamaru et al., Voltage-induced magnetic anisotropy changes in an ultrathin FeB layer sandwiched between two MgO layers. Appl. Phys. Express. **6**, 073005 (2013). https://doi.org/10.7567/APEX.6.073005
67. T. Nozaki, A. Kozioł-Rachwał, M. Tsujikawa et al., Highly efficient voltage control of spin and enhanced interfacial perpendicular magnetic anisotropy in iridium-doped Fe/MgO magnetic tunnel junctions. NPG Asia Mater. **9**, 1–10 (2017). https://doi.org/10.1038/am.2017.204
68. D. Chien, X. Li, K. Wong et al., Enhanced voltage-controlled magnetic anisotropy in magnetic tunnel junctions with an MgO/PZT/MgO tunnel barrier. Appl. Phys. Lett. **108**, 112402 (2016). https://doi.org/10.1063/1.4943023
69. H. Sato, M. Yamanouchi, K. Miura et al., Junction size effect on switching current and thermal stability in CoFeB/MgO perpendicular magnetic tunnel junctions. Appl. Phys. Lett. **99**, 042501 (2011). https://doi.org/10.1063/1.3617429
70. C. Yoshida, M. Kurasawa, M.L. Young, et al., A study of dielectric breakdown mechanism in CoFeB/MgO/CoFeB magnetic tunnel junction. in *IEEE International Reliability Physics Symposium Proceedings,* Montreal, QC, p 139–142 (2009). https://doi.org/10.1109/IRPS.2009.5173239
71. M. Schäfers, V. Drewello, G. Reiss et al., Electric breakdown in ultrathin MgO tunnel barrier junctions for spin-transfer torque switching. Appl. Phys. Lett. **95**, 232119 (2009). https://doi.org/10.1063/1.3272268

72. Y. Shiota, T. Nozaki, S. Tamaru et al., Evaluation of write error rate for voltage-driven dynamic magnetization switching in magnetic tunnel junctions with perpendicular magnetization. Appl. Phys. Expr. **9**, 013001 (2016). https://doi.org/10.7567/APEX.9.013001
73. J.G. Alzate, P.K. Amiri, P. Upadhyaya et al., Voltage-induced switching of nanoscale magnetic tunnel junctions. in *Technical Digest—International Electron Devices Meeting*, IEDM. San Francisco, CA p 29.5.1–29.5.4 (2012). https://doi.org/10.1109/IEDM.2012.6479130
74. J.M. Iwata-Harms, G. Jan, H. Liu et al., High-temperature thermal stability driven by magnetization dilution in CoFeB free layers for spin-transfer-torque magnetic random access memory. Sci. Rep. **8**, 14409 (2018). https://doi.org/10.1038/s41598-018-32641-6
75. T. Newhouse-Illige, Y. Liu, M. Xu, et al., Voltage-controlled interlayer coupling in perpendicularly magnetized magnetic tunnel junctions. Nat. Commun. **8**, 15232 (2017). https://doi.org/10.1038/ncomms15232
76. S. Wang, H. Lee, F. Ebrahimi et al., Comparative evaluation of spin-transfer-torque and magnetoelectric random access memory. IEEE J. Emerg. Sel. Top Circuits Syst. **6**, 134–145 (2016). https://doi.org/10.1109/JETCAS.2016.2547681
77. A.V. Khvalkovskiy, D. Apalkov , S. Watts, et al., Basic principles of STT-MRAM cell operation in memory arrays. J. Phys. D Appl. Phys. **46**, 074001 (2013). https://doi.org/10.1088/0022-3727/46/7/074001
78. Y. Shiota, T. Nozaki, S. Tamaru et al., Reduction in write error rate of voltage-driven dynamic magnetization switching by improving thermal stability factor. Appl. Phys. Lett. **10**(1063/1), 4990680 (2017). https://doi.org/10.1063/1.4990680
79. H. Cheng, N. Deng, Influence of thermal agitation on the electric field induced precessional magnetization reversal with perpendicular easy axis. AIP Adv. **3**, 122124 (2013). https://doi.org/10.1063/1.4858423
80. S. Shirotori, H. Yoda, Y. Ohsawa, et al., Voltage-control spintronics memory with a self-aligned heavy-metal electrode. IEEE Trans. Magn. **53**, 3401104 (2017). https://doi.org/10.1109/TMAG.2017.2691764
81. J. Deng, G. Liang, G.Gupta, Ultrafast and low-energy switching in voltage-controlled elliptical pMTJ. Sci. Rep. **7**, 16562 (2017). https://doi.org/10.1038/s41598-017-16292-7
82. R. Matsumoto, T. Nozaki, S. Yuasa, H. Imamura, Voltage-Induced precessional switching at zero-bias magnetic field in a conically magnetized free layer. Phys Rev Appl **9**, 14026 (2018). https://doi.org/10.1103/PhysRevApplied.9.014026
83. H. Stillrich, C. Menk, R. Frömter, H.P. Oepen, Magnetic anisotropy and the cone state in Co/Pt multilayer films. J. Appl. Phys. **105**, 07C308 (2009). https://doi.org/10.1063/1.3070644
84. J.W. Lee, J.R. Jeong, S.C. Shin et al., Spin-reorientation transitions in ultrathin Co films on Pt(111) and Pd(111) single-crystal substrates. Phys. Rev. B **66**, 172409 (2002). https://doi.org/10.1103/PhysRevB.66.172409
85. R. Dorrance, J.G. Alzate, S.S. Cherepov et al., Diode-MTJ crossbar memory cell using voltage-induced unipolar switching for high-density MRAM. IEEE Electron. Device Lett. **34**, 753–755 (2013). https://doi.org/10.1109/LED.2013.2255096
86. A. Driskill-Smith, D. Apalkov, V. Nikitin et al., Latest advances and roadmap for in-plane and perpendicular STT-RAM. in *2011 3rd IEEE International Memory Workshop IMW 2011*, pp. 1–3 (2011). https://doi.org/10.1109/IMW.2011.5873205
87. J.H. Oh, J.H. Park, Y.S. Lim, et al., Full integration of highly manufacturable 512Mb PRAM based on 90nm technology. in *Technical Digest—International Electron Devices Meeting*, IEDM, 1-4 (2006). https://doi.org/10.1109/IEDM.2006.346905

Chiral Magnetic Domain Wall and Skyrmion Memory Devices

Kyujoon Lee, Dong-Soo Han, and Mathias Kläui

Abstract In the chapter, we have reviewed the fundamental physics for designing magnetic domain wall memories, especially domain wall racetrack memories. An overview of how the racetrack has been functionally improved and the fundamental physics behind the operating mechanism has developed is shown. Material wise, the design of the racetrack has changed from using in-plane magnetic materials to out-of-plane magnetic materials. The process of changing the material design resulted in new physics such as the spin-orbit torques (SOTs) and the Dzyaloshinskii-Moriya interaction (DMI) which resulted in domain wall motion with higher efficiency and stability. The SOT is the main mechanism in moving the domain walls efficiently by utilizing the spin Hall effect (SHE) and the inverse spin galvanic effect (ISGE) which have shown to be more efficient than the spin-transfer torque (STT) in current induced domain wall motion. The exact physics behind the SOTs is still not well known, but it was well demonstrated that the SOTs show higher efficiency for domain wall (DW) motion. However, this SOT requires additionally a chiral symmetry breaking such as due to DMI in order to act on the DWs. The DMI generates a certain chirality for the domain walls, especially forcing a chiral Néel type DW. The Néel DW is required for the SOT to act as a driving force of the DWs. The different sections of the chapter have reviewed the different physics and evidence of the SOT and DMI with the different experimental methods to quantify the SOT and DMI. Furthermore, as an outlook for the racetrack memory, we have reviewed the new exciting skyrmion racetrack memory which can be a future implementation of the racetrack memory.

1 Overview

The development of information technology is pushing the research efforts to find a new technology that can go beyond the complementary metal-oxide semiconductor (CMOS) for storage and computing. Spintronics is one of the most promising candidates, due to its fundamental aspect of using the additional degree of freedom of the

K. Lee · D.-S. Han · M. Kläui (✉)
Johannes Gutenberg University of Mainz, Mainz, Germany
e-mail: Klaeui@Uni-Mainz.de

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_5

spin angular momentum. Particularly, memory storage technologies using magnetic domain walls (DWs) have been under considerable interest for decades due to their promising technological aspects, advantageous over the conventional memory technologies with high energy consumption, the volatility of static and dynamic random access memory (RAM), and slow reading and writing operations. One example of such an approach is a magnetic domain-wall racetrack memory which was first proposed by IBM (see Fig. 1) [1]. In the magnetic domain-wall racetrack memory device, the magnetization state of the magnetic domains is used as "0" and "1" bits. At the writing element, the bits are written as illustrated in Fig. 1d. The written bits

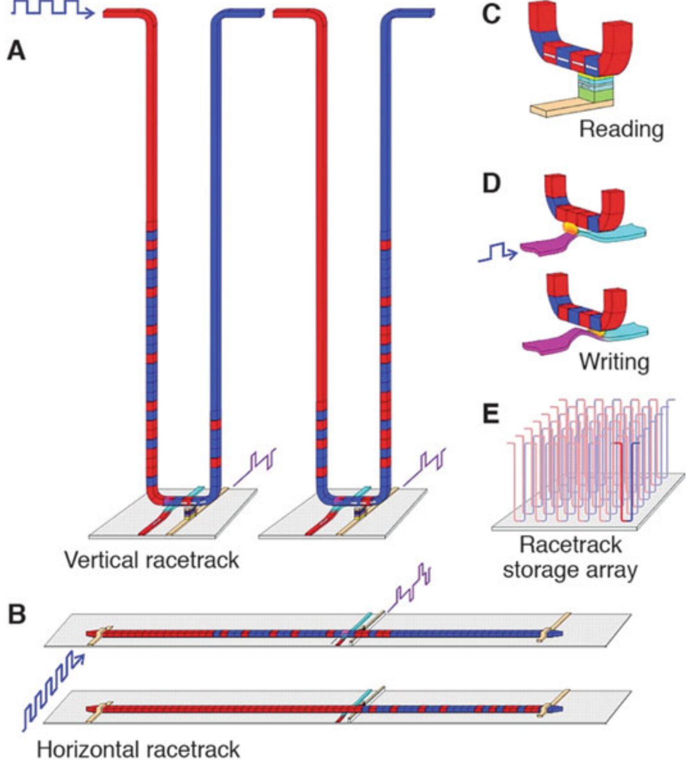


Fig. 1 The magnetic domain-wall racetrack memory as proposed by IBM [1]. **a** A vertical-configuration racetrack offers the highest storage density by storing the pattern in a U-shaped nanowire normal to the plane of the substrate. The two cartoons show the magnetic patterns in the racetrack before and after the DWs have moved down one branch of the U, past the read and write elements, and then up the other branch. **b** A horizontal configuration uses a nanowire parallel to the plane of the substrate. **c** Reading data from the stored pattern is done by measuring the tunnel magnetoresistance of a magnetic tunnel junction element connected to the racetrack. **d** Writing data is accomplished, for example, by the fringing fields of a DW moved in a second ferromagnetic nanowire oriented at right angles to the storage nanowire. **e** Arrays of racetracks are built on a chip to enable high-density storage. Adapted with permission from [1]

can be pushed forward or backward by shifting all DWs in the same direction by the use of an external magnetic field [2] or electric currents [1]. Thus moving these domain walls result in moving the bits as desired. Then the encoded bits can be read by the stationary reading element which is a magnetoresistive tunnel junction (MTJ) on the track (see Fig. 1c). The racetrack memories have many advantages compared to the conventional memory devices. Compared to the conventional hard disk drive there are no mechanically moving parts, thus possibly the racetrack memory can be faster beyond the mechanical limitations, and the problems due to mechanical failure do not exist. Another possible advantage of this racetrack memory is that it may allow one to construct three-dimensional devices orienting the racetrack upwards and store bits in a larger 3-dimensional space (see Fig. 1e), which may result in a high-storage density while achieving high speed, robustness, and low power consumption similar to the charge-based solid-state memory devices. To apply the racetrack memories, first a full understanding of DW motion must be obtained

Given the promising technical aspects, great progress in the racetrack memory has been made since its first proposal. A proof of concept for the racetrack memory using fields was demonstrated in soft ferromagnets with an in-plane magnetization at their remanent state [2, 3]. However, the DW motion driven by magnetic fields has severe obstacles, especially in the aspect of applications. First, there is a limitation to localize magnetic fields in small confined areas. This becomes more challenging to manipulate a DW in a device while keeping the neighboring DWs intact, as the devices are scaled down. Furthermore, when the DWs are driven by an in-plane magnetic field, the two adjacent DWs move in opposite directions. Therefore, the

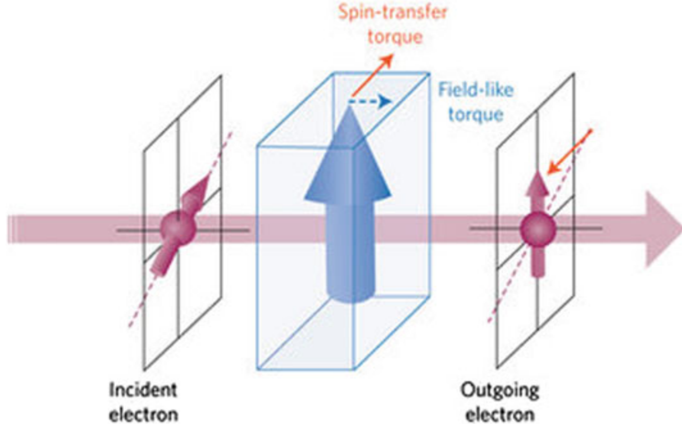


Fig. 2 Illustration of current-induced spin transfer torques taken from [6]. A spin-polarized current enters a ferromagnet. The interaction between the spin-polarized current and the magnetization causes a change in the spin direction of the outgoing electron compared with the incident electron, transferring its angular momentum to the localized magnetic moment in ferromagnets. Adapted with permission from [6]

racetrack memory, essentially requiring unidirectional motion of all DWs, is very difficult to implement with a field-driven scheme.

To overcome these obstacles, spin-transfer torque (STT), originally proposed by Berger and Slonczewski [4, 5], was suggested as an alternative means to drive DW motion. A simple concept of STT is illustrated in Fig. 2. When conduction electrons pass through a magnetic metal with a uniform magnetization, they become spin-polarized in the direction of the magnetization of the ferromagnet. As these spin-polarized electrons cross a DW, their spins are flipped as they track the locally varying magnetization in the DW, resulting in a change in the spin angular momentum of the electrons. Since the total angular momentum should be conserved in the system, the change in the spin angular momentum of the conduction electrons results in a change in the angular momentum of the magnetic moments.

In addition to the adiabatic STT, there exist an additional STT arising from a deviation from the adiabatic process. When the electron spins are passing through a DW, for example, with a relatively narrow DW width, the electrons' spins cannot adiabatically follow the magnetization profile within the DW anymore. Consequently, it may lead to the spatial mistracking of spins between the conduction electrons and the local magnetization. This mistracking of electrons' spins then introduces another torque namely the non-adiabatic STT, which has an effect similar to that of a magnetic field.

The DWs driven by the STTs can be understood by using the Landau-Lifshitz-Gilbert (LLG) equation incorporated with the STTs acting on the local magnetization ($\mathbf{m}$) within the DWs [7, 8]:

$$\dot{\mathbf{m}} = -\gamma_0 \mathbf{m} \times \mathbf{H} + \alpha \mathbf{m} \times \dot{\mathbf{m}} - (\mathbf{u} \cdot \nabla)\mathbf{m} + \beta \mathbf{m} \times [(\mathbf{u} \cdot \nabla)\mathbf{m}] \quad (1)$$

Here, γ is the gyromagnetic ratio, α the Gilbert damping parameter, $\mathbf{H}$ the effective magnetic field, and $\mathbf{u}$ the spin drift velocity which is $\mathbf{u} = g\mu_B P \mathbf{J_e}/2|e|M_S$, where g is the g-factor, μ_B Bohr magnetron, P spin polarization, e the electron charge, M_s the saturation magnetization, and $\mathbf{J}_e$ the current density vector with its direction in the electron flow. The first and second terms on the right side of the equation are the general torques exerted by the effective magnetic field and damping, respectively. The third and fourth terms represent the adiabatic and non-adiabatic STTs. For the case of the DW motion driven by the adiabatic STT, it is known that this always requires a threshold current density due to the intrinsic energy barrier originating from its mechanism to drive the DW. Figure 3a. illustrates how the adiabatic STT can drive the DW motion. For an in-plane magnetized thin film, when the electric current is applied into $+x$ direction, for example, the adiabatic STT is exerted on the local magnetic moment oriented in the transverse direction of the wire axis pushing toward the $-x$ direction. The precession of $\mathbf{m}$ driven by the STT initiates a damping torque perpendicular to the STT ($+z$), and this leads to the canting of the magnetization into the out of plane direction. The canting creates magnetic surface charges at the film surface, generating a demagnetizing field $\mathbf{H_d}$ against to the creation of the magnetic surface charges. The demagnetization field provides an additional precession torque ($+x$) and a damping torque ($-z$). In this case, these four torques all cancel out.

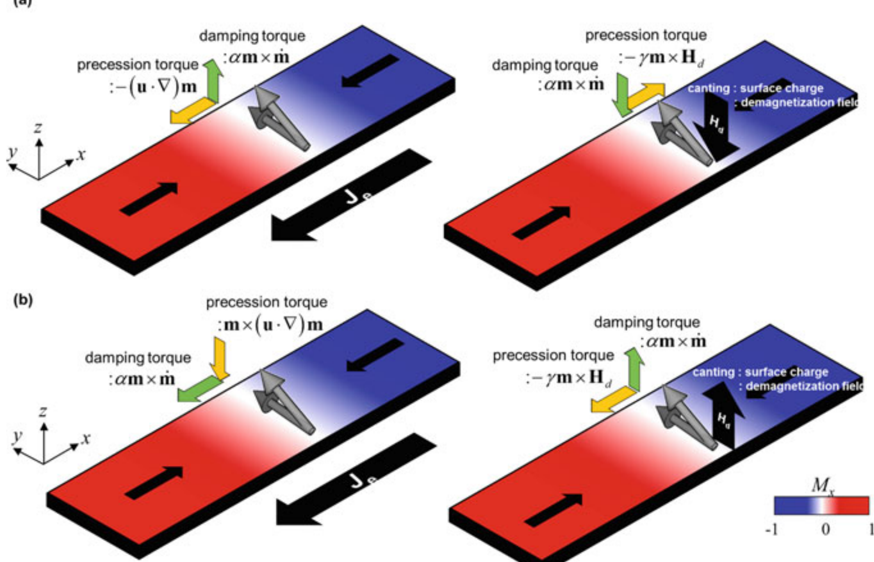


Fig. 3 DW dynamics driven by adiabatic **a** non-adiabatic spin-transfer torque **b** in an in-plane magnetized thin film. The colors on the strip represent the *x* component of the magnetization. The orange—and green-colored arrows indicate the precession and damping torques, respectively, induced by the STTs (left panel) and the demagnetization field (right panel). The cone-arrows on the plane represent the local magnetization (m) within DWs

Therefore, adiabatic STT cannot drive a DW directly over long distances. Instead, the adiabatic STT above a critical threshold current density—*Walker threshold*—high enough to overcome the energy barrier from the magnetostatic energy can propel the DW through precession around the *x*-axis.

On the other hand, non-adiabatic STT can drive a DW via rigid translation even at a very small current density**.** As displayed in Fig. 3b, the torques introduced by the non-adiabatic STT and demagnetization field are not canceled out but has an effective torque pointing to a certain direction, i.e., here $-x$ axis when the current is applied to the $+x$ axis. Consequently, the DW can move in the direction of the electron flow even at a very low current density [7].

Despite its great fundamental and practical interest, the STT-induced DW motion in the in-plane magnetized wires also faces practical issues. The STT-induced DW motion requires a high threshold current in moving DWs resulting in stability issues due to thermal effects, and also reliability issues of the domain wall motion existed. The typical current density to initiate current-induced DW motion and obtain high velocities of ~ 10–100 m/s is ~10^{12} A/m^2. This high current causes large energy dissipation, mostly accompanied by excessive Joule heating, which in turn hinders the power-efficiency of devices as well as reliable retention of stored information. Furthermore, the widths of DWs in in-plane magnetized thin film wires are relatively large, which is on the order of the width of the nanowires, i.e., typically hundreds

of nm [9]. In these DWs, a large width of the DWs limits the maximum information storage density.

To alleviate the listed issues, a different class of magnetic materials that is with out-of-plane magnetization has been investigated. In contrast to the relatively large DW width in the in-plane magnetized thin films, the DW width in an out-of-plane magnetized film is much narrower (~10 nm) [10]. Therefore, the out-of-plane magnetized films are considered to be technically more promising than the in-plane magnetized counterparts. In this respect, DWs motion in the out-of-plane magnetized thin films have been extensively studied, for example, in ferromagnetic multilayers, i.e., Co/Ni, Co/Pt, and Co/Pd, where a strong perpendicular magnetic anisotropy is present at their interfaces through spin–orbit coupling [7, 11].

However, experimental results of the current-driven DW motion in the out-of-plane magnetization have been controversial. In Co/Ni multilayers and TbFeCo alloys with rather thick magnetic films (~> 1 nm), it has been commonly observed that adiabatic STT plays an important role in DW motion [12], although the effect of non-adiabatic STT remains controversial. For example, Koyama et al. [12] have experimentally demonstrated that the current-induced DW velocity in Co/Ni multilayer is rather independent of DW pinning strength, indicating that the adiabatic STT plays a more dominant role in driving DW motion rather than the non-adiabatic STT. On the contrary, in ultrathin ferromagnetic multilayers (<1 nm) like Co/Pt multilayers, largely scattered experimental observations were reported. Ravelosona et al. [13] show a significant reduction in the required field to depin a DW in a Co/Pt multilayer with increasing current density which suggests that the DW is dominated by non-adiabatic STT. Furthermore, Boulle et al. [11] have reported a similar behavior with a high efficiency, presumed to be a consequence of strong non-adiabatic STT. However, note that other torques can lead to similar behavior as discussed in the next section. In contrast to these reports, a few experimental studies reported that the efficiency of STTs are very small in Co/Pt [14], and, even in some cases, the DW motion was measured with the DW direction opposite the direction of the electron flow, which is controversial to the expected results from the conventional STT [16, 20, 25]. These broad experimental discrepancies in similar magnetic multilayers imply that a radically new mechanism to explain the current-induced DW motion in these multilayers is required.

In this chapter, we will discuss the recent progress of chiral domain wall devices [19]. In the first Sect. 2, we will give an overview of the new mechanisms that have been revealed for the new phenomena leading to current-induced DW motion. A basic theoretical introduction about the mechanism and some experimental verification will be shown. In Sect. 3, we will discuss the chiral domain walls and the mechanism behind the induced chirality in the DWs. A basic theoretical explanation and some experimental results about the chiral DWs will be introduced. Furthermore, in the last section, we will give a short outlook of memory devices beyond domain wall memory devices based on skyrmions.

2 Spin–Orbit Torques Driven Magnetic Domain Wall Motion

In addition to the anomalous behaviors found in Co/Pt multilayers, remarkably efficient DW motion of which velocity is about 100 m/s and moving direction is against to the electrons flow was experimentally observed in ultrathin Co sandwiched between a nonmagnetic heavy metal and insulator [20]. This anomalously efficient current-induced DW motion suggested possible contribution from inverse spin galvanic effect (ISGE) [21, 22] or so-called Rashba-Edelstein effect. The ISGE is an interfacial spin–orbit coupling present at a surface or interface of materials which lacks inversion symmetry. This inversion asymmetry then result in an electric potential—arising from the discontinuity in electronic structure at the interface—along the surface or interface axis. When electrons flow on the interface or surface of a ferromagnetic layer, the electric field from the asymmetric potential transforms into an effective magnetic field in the electron's rest frame. The conduction electrons spins experience an effective magnetic field transverse to the electron flow direction, accordingly leading to partial spin polarization along the direction of the effective magnetic field. The conduction electrons spins interact with the magnetization in the ferromagnetic layer through *s*-*d* exchange coupling, exerting torques on the magnetization acting as an effective magnetic field (see Fig. 4b) [23]. This effect is generally believed to play a similar role to a non-adiabatic torque in the conventional STTs (exert a so-called field-like torque).

However, the mechanism based on the ISGE cannot fully account for the experimentally observed current induced DW motion [24]. Furthermore, some semi-classical transport theories predicted that the STTs in the ultrathin Co (<1 nm) is

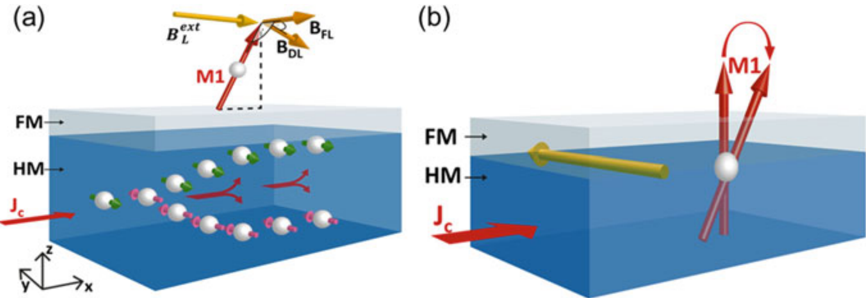


Fig. 4 **a** shows schematically the mechanism of the SHE that converts a longitudinal electron current, J_c, in the heavy metal (HM) into a transverse spin current by separating spin-left and spin-right electrons. The generated spin current results then in a spin accumulation at the HM/FM interface that diffuses across the interface into the FM and exerts torques (indicated by the effective fields of the damping-like torque, B_{DL}, and the field-like torque, B_{FL} that rotate a spin M1 that can furthermore be affected by an external field B_L^{ext} for instance longitudinal to the current). **b** shows the effective field generated by the inverse spin galvanic effect at the interface, which can lead to a switching of the magnetization

vanishingly small [26], because electrons flow mostly in the Pt than Co layer and the conduction electrons spins largely depolarize from interfacial scattering in ultrathin ferromagnets. Moreover, DWs in Pt/Co/oxide moving against the electron flow is contrary to the action of the STT.

Instead, a series of experiments by Liu et al. [27], have proposed an alternative mechanism based on the spin Hall effect (SHE) [28]. The SHE is a phenomenon which appears in systems with spin–orbit coupling. When an electrical current is applied, a spin current is generated in the transverse direction of the charge current. This spin current results in an accumulation of spin on the edges of the material. The SHE has been known to be realized in materials with strong spin–orbit coupling. It was first proposed by Dyakonov and Perel in 1979 [29]. The actual term SHE was first introduced by Hirsch [30]. By the semi-classical theory, the SHE could be explained mainly by three different mechanisms: intrinsic [31, 32], skew scattering [33, 34] and side-jump mechanisms [35]. The intrinsic SHE is dependent only on the band structure of the material. The skew scattering mechanism is proportional to the transport lifetime τ. Since the skew scattering term is proportional to the transport life time τ, the skew scattering spin Hall term is proportional to the longitudinal electrical conductivity. The side jump mechanism appears as the contribution to SHE other than the intrinsic and the skew scattering. The theoretical concept is when a Gaussian wave packet scatters from spherical impurities, the incident wave vector will undergo a transverse displacement. This is the cause of a SHE due to the side jump. With all these mechanisms combined in a system with spin–orbit coupling, the SHE can result in the non-equilibrium spin accumulation on edges of the nonmagnetic metal, consequently leading to the generation of a transverse spin current. As shown in Fig. 4a, the SHE effect occurs in the heavy metal (HM) underlayer accumulating spins at the interface of the heavy metal and ferromagnet (FM). The accumulated spins flow into the ferromagnetic layer exerting a torque on the magnetization of the ferromagnet. The SHE is known to be a bulk effect and is expected to be responsible for the Slozenski-like torque (or so-called damping-like torque).

Figure 5 displays how the SOTs and STTs differently act on the DW motion. The DW motion by the STTs in a perpendicularly magnetized thin film can be understood as the same mechanism described in the previous section. The only difference here is that the demagnetization field in the perpendicularly magnetized thin films is mainly attributed to the magnetic volume charge instead of the surface charge. By considering all the torques acting on the local magnetization, one can notice that the DW can move in the electron flow direction mostly attributed to the non-adiabatic torque—for the adiabatic torque, the torques from the adiabatic torque and the demagnetization field from the canting of the local magnetization (**m**) are all canceled out as like the in-plane magnetized counterpart, but, for the case of the non-adiabatic torque, the net torque is pointing downward, leading to the DW motion in a direction expanding the magnetic domains saturated downward (see Fig. 5a). In contrast, for the case of SOTs, the Bloch DW cannot move by the SOTs, as the local magnetization is always parallel to the spin polarization direction—the spin polarization direction is transverse to the electron flow direction. Therefore, a Néel DW is needed in order to explain the DW motion by SOT (the physical origin to stabilize the Néel DW will

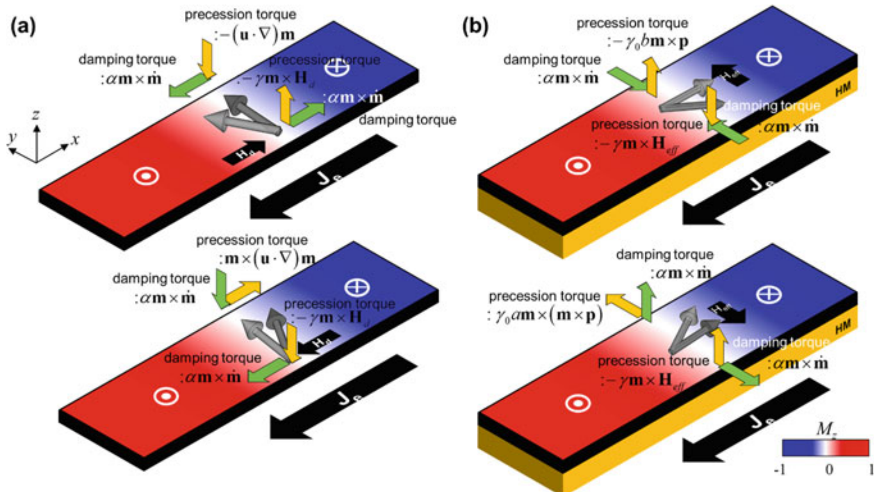


Fig. 5 DW dynamics driven by STTs (**a**), and SOTs (**b**) in an out-of-plane magnetized thin film. The colors on the strip represent the *z* component of the magnetization. The top and bottom panels of **a** represent the torques arising from the adiabatic and non-adiabatic STTs, respectively. The top and bottom panels of **b** represent the torques arising from the field—and damping-like SOTs, respectively. The $\mathbf{H}_d$ in **a** is attributed to the demagnetization field against to the increase of the magnetic volume charge. The $\mathbf{H}_{eff}$ in **b** is from the interfacial Dzyaloshinskii-Moriya interaction that prefer the Néel wall than the Bloch wall. The orange—and green-colored arrows indicate the precession and damping torques, respectively, induced by the STTs (or SOTs) and $\mathbf{H}_d$ ($\mathbf{H}_{eff}$). The cone-arrows on the plane represent the local magnetization (**m**) within DWs

be discussed later). As seen in Fig. 5b, when the field-like torque is exerted on the local magnetization, all torques are canceled out, therefore, this torque cannot solely move the DWs as like the adiabatic STT. On the contrary, the damping-like torque can drive the DW motion even at a very small current density, and the DW can move against to the electron flow—this depends on the sign of a spin Hall angle (or a ISGE coefficient) of the heavy metal adjacent to the ferromagnet and the chirality of the DW.

Although these two mechanisms cannot fully explain the experimental data, it has been experimentally shown that these two spin–orbit torques (SOTs), field-like and damping-like torques, are indeed present in ferromagnetic multilayers, particularly consisting of heavy metals and ferromagnetic metals. These SOTs play an important role in driving DWs or switching magnetization in ferromagnetic thin films. Therefore quantifying the SOT in different systems is critical to fully understand the DW motion and magnetization switching. In the following sections, we will introduce the experimental techniques to quantify these SOTs.

2.1 Experiments for Measuring SOTs

2.1.1 Spin-Torque Ferromagnetic Resonance Measurement

One of the most widely used methods to quantify the SOTs in various systems is to use the spin torque-ferromagnetic resonance (ST-FMR) technique. Conventionally, this method was used to measure the spin Hall angle in various systems. The magnetization dynamics is probed in a microstrip [27, 36, 37] as shown in Fig. 6a. An rf current generates a resonant precession of the magnetization in the bilayer, and the anisotropic magnetoresistance voltage shows an FMR spectrum. A modulated rf signal is used as an input and a lock-in amplifier is used to detect the voltage for the FMR spectrum. The in-plane fields are swept at a fixed rf frequency with an angle of 45° with respect to the applied current axis. One can measure the FMR spectrum as in Fig. 6b. The measured voltage of the FMR spectrum can be fitted to a Lorentzian curve of the form.

$$V_{mix} = S\frac{W^2}{(\mu_0 H - \mu_0 H_{FMR})^2 + W^2} + \mathrm{A}\frac{W(\mu_0 H - \mu_0 H_{FMR})}{(\mu_0 H - \mu_0 H_{FMR})^2 + W^2}, \quad (2)$$

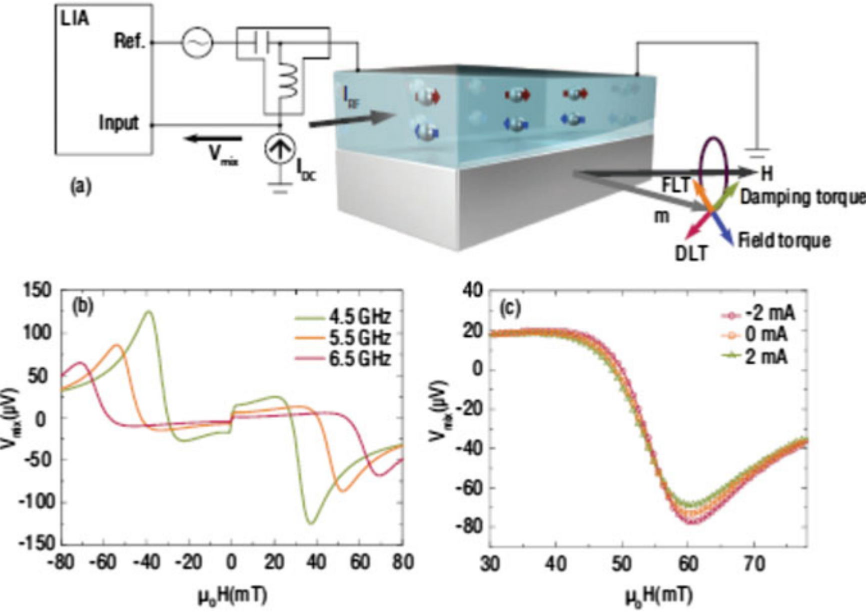


Fig. 6 **a** An illustration of the measurement scheme of the spin torque ferromagnetic resonance. **b** FMR spectrum of the ST-FMR measurement on Ta(3 nm)/$Ni_{80}Fe_{20}$(2.5 nm)/Pt(4 nm)and **c** the FMR spetra when an additional dc current is applied [27]. Adapted with permission from [27]

where W is the half width half maximum line width, H_{FMR} is the resonance field, S is the symmetric Lorentzian coefficient, and A is the antisymmetric Lorentzian coefficient. The ratio of S to A can be used to evaluate the ratio of the damping-like torque to the effective field from the Oersted field and the field-like torque. In order to decouple these two the magnitude of the rf current in the bilayer has to be known. For a more reliable quantification of the SOTs an additional dc-current modulation and thickness dependence of magnetic thin films should be measured [36, 38, 39]. By measuring the ST-FMR with an additional dc current, the damping-like torque can be quantified with the dependence of the linewidth on the dc current. Additionally, the change in the resonance field yields a direct measurement of the field-like torque.

2.1.2 Higher-Order Harmonics Measurements

Another widely used method in measuring SOTs is using the second harmonic measurements of electrical transport. By probing the Hall measurements using an ac current one can extract the effective SOT fields. As shown in Fig. 7a, in a Hall cross an ac current is applied and the Hall voltage is measured using a lock-in amplifier. When the Hall voltage is measured in a PMA system, because of the anomalous Hall effect, one can measure the out of plane component of the magnetization (m_z). The change in the anomalous Hall voltage will be proportional to the change in the m_z. If one assumes that the change in magnetization due to the SOT will be linear to the current applied, the change in the m_z due to SOT ($=\Delta m_z$) will be proportion ac current applied. Therefore, if an ac current ($I = I_0 \sin(\omega t)$) is applied, then the change in magnetization by SOT will be $\Delta m_z \propto I_0 \sin(\omega t)$. The Hall voltage can be described as following,

$$V_H = R_{AHE} I_0 \sin(\omega t)(m_z + \Delta m_z) \tag{3}$$

where R_{AHE} is the anomalous Hall resistance. Here, by substituting $\Delta m_z \propto I_0 \sin(\omega t)$ one can derive,

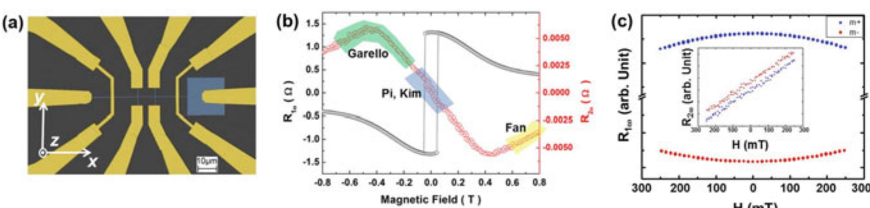


Fig. 7 **a** shows the SEM image of the Hall cross structure. **b** the first and second harmonic resistance measured while applying the magnetic field in-plane [45]. **c** the first and second harmonic resistance while applying low in-plane magnetic fields. Adapted with permission from [45]

$$V_H \propto R_{AHE} I_0 \sin(\omega t)(m_z + I_0 \sin(\omega t)) = R_{AHE}\left(I_0 \sin(\omega t) m_z + (I_0 \sin(\omega t))^2\right)$$
$$= R_{AHE}(I_0 \sin(\omega t) m_z + (I_0 \cos(2\omega t))) \quad (4)$$

Therefore, by measuring the first and second harmonics of the ac Hall voltage one can measure the effective fields generated by the SOT. In order to measure the damping-like and field-like effective fields one must measure the second harmonics while sweeping the magnetic field in the x and y direction. There are different analysis methods depending on which angle of magnetization tilt regime is needed to be analyzed [36, 40–43]. Here, we will concentrate only on the regime where the out of plane tilt of magnetization is very small. As shown in Fig. 7c, by applying a magnetic field in either x or y direction of the Hall cross the First and Second harmonic Hall voltage can be measured. The effective SOT fields can be described as follows,

$$B_{x,y} = \left(\frac{\partial V_{2\omega}}{\partial H_{x,y}} / \frac{\partial^2 V_{\omega}}{\partial H_{x,y}^2}\right) \quad (5)$$

where B_x and B_y will be the effective fields in the x and y direction, respectively and H_x and H_y is the applied magnetic field also in the x and y direction. When considering the planar Hall effect, a correction in the effective fields need to be calculated. The corrected effective fields would be,

$$H_{DL} = -\frac{2\frac{(B_X \pm 2\varepsilon B_Y)}{1-4\varepsilon^2}}{j},\ H_{FL} = -\frac{2\frac{(B_Y \pm 2\varepsilon B_X)}{1-4\varepsilon^2}}{j}, \quad (6)$$

where H_{DL} and H_{FL} is the damping-like effective field and the field-like effective field, respectively and $\xi = R_{PHE}/R_{AHE}$. In order to apply the effective fields to the domain walls there is a conversion factor that needs to be considered [44, 45].

2.1.3 Current-Field Equivalence Measurement and Domain Wall Motion

Measuring the spin–orbit torques by the current-field equivalence method is also a well-known method [46–48]. The current-field equivalence method is measuring the effect of an applied current on the depinning measurements [47, 49]. As shown in Fig. 8, the depinning measurements are done in the Hall cross. First, the Hall cross is saturated in a certain magnetization and by passing a current through the Oersted line (yellow strip in Fig. 8a) a domain wall is formed. By applying a magnetic field (H_p in Fig. 8b), the DW is driven into the Hall cross so it gets pinned in the Hall cross. Here, one has to measure two different fields, one is the depinning field without a current, i.e. H_c^*, the second is the depinning field with a current applied (H_{dep}). The difference between the two values will give the field equivalent to the effective field by the current. In systems where SOTs are present, the effective field due to the current

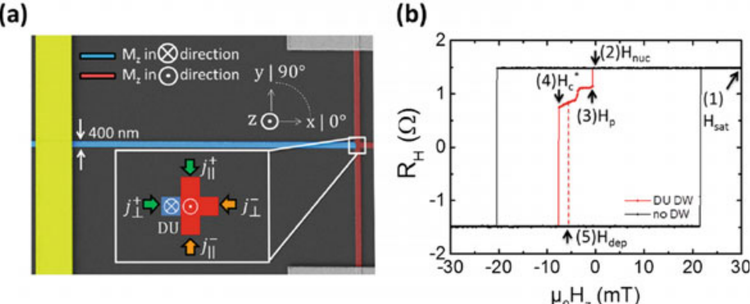


Fig. 8 **a** A schematic illustration of the Hall cross when the domain wall is pinned in the Hall cross. **b** graph of the anomalous Hall effect illustrating the different field values for the current-field equivalence method. (1) H_{sat}: field in which the magnetization of the Hall cross was saturated. (2) H_{nuc}: field value to nucleate a domain to create a domain wall. (3) H_p: field value to pin the domain wall into the Hall cross. (4) $H_c{}^*$: field to depin the domain wall with no current applied. (5) H_{dep}: depinning field value with a current applied [49]. Adapted with permission from [49]

will be the effective field generated by the SOTs. The current-field equivalence can be also measured for the change of the switching in the magnetization [46]. By measuring the difference in the switching with an applied current the effective fields could be extracted. In addition, the SOTs could also be extracted through the current induced domain wall motion measurements [20, 50, 51]. When measuring the current induced domain wall velocity while applying an in-plane field, one could extract the Dzyaloshinskii-Moriya interaction which will be explained in detail in the later sections. Additionally, by applying a one dimensional model, the spin–orbit torque could be extracted.

3 Chiral Magnetic Domain Walls Stabilized by Interfacial Dzyaloshinskii-Moriya Interaction

While the SHE or the ISGE can provide fundamentals to understand the DW motion moving against the electron flow, it is not able to explain how the damping-like torque (or field-like torque) can drive the DW with a Bloch wall spin structure, which is expected to be present in most out-of-plane magnetized thin films. In contrast to the STTs, the SOTs are effective only for the Néel wall. Accordingly, in order to account for the DW motion driven by the SOTs, the Néel DW should be taken into account instead of the Bloch DW. Indeed, it has been experimentally demonstrated that DW motion driven by the SHE, in out-of-plane magnetized Pt/Co/Pt, with the Néel DW initially configured by applying an in-plane bias field [50]. Such Néel DWs in nanostripes were attributed to Dzyaloshinskii-Moriya interaction (DMI) which prefers chiral Néel walls rather than Bloch walls. Recent observation of the built-in DW chirality due to the DMI enables one to accomplish the current-driven DW

motion without a magnetic field. Furthermore, two DWs with the opposite polarities (up-down and down-up) have the same chirality and this leads to the behaviour that two DWs move in the same direction when the current is applied.

3.1 Physical Origin of the Interfacial DMI

The DMI has been introduced to explain weak ferromagnetism in an antiferromagnet arising from a misalignment of the sublattices with respect to the totally antiparallel configuration. This anomalous behavior was at first introduced by Dzyaloshinskii [52] who explained the phenomenon with the non-centrosymmetric property of the system. After his phenomenological explanation, Moriya [53] connected his name to this term as he analytically found that the mechanism behind this interaction partially arises from spin–orbit coupling. This interaction was studied in a compound that lacks inversion symmetry like B20 group [54, 55] in the beginning. However, it has been recently recognized that the DMI may appear in ultrathin film systems where a structural inversion symmetry from interfaces and a large spin–orbit coupling exist.

The DMI between two spins in a magnetic material is mediated via the spin–orbit coupling of a heavy-metal atom with strong spin–orbit coupling, as shown schematically in Fig. 9c. Taking into account a spin–orbit coupling$\left(\lambda \vec{L} \cdot \vec{S}\right)$) and expanding in powers of λ, Moriya [53] found that the effective Hamiltonian for the interaction between two spins M_1 and M_2 contains a term $\vec{\boldsymbol{D}}_{12} \cdot (\vec{\mathbf{M}}_1 \times \vec{\mathbf{M}}_2)$. As shown schematically in Fig. 9c, in multilayer stacks, where the inversion symmetry is broken by sandwiching a ferromagnetic layer (grey) in between two different non-magnetic layers (bottom light blue layer is shown), this coupling is mediated by a heavy atom (blue) in one of the non-magnetic layers. The resulting DMI vector, $\mathbf{D}_{12}$,

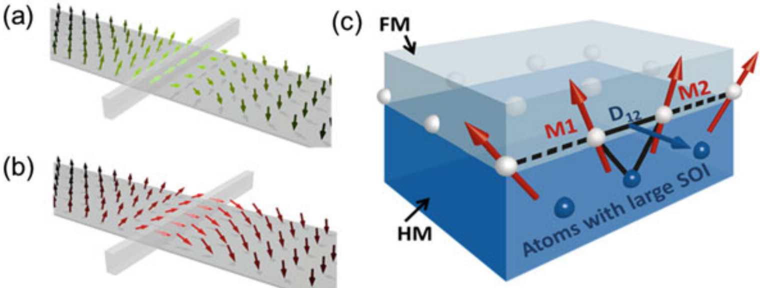


Fig. 9 **a** Schematic of a Bloch domain wall and **b** a Néel wall (from Ref. [57]). The DMI changes the angle of the magnetization in the domain wall and thus generates chiral Néel walls even for structures where Bloch walls would prevail without DMI. **c** In multilayer stacks with broken inversion symmetry, chiral coupling between two spins M1 and M2 is mediated by a heavy atom (blue) in one of the non-magnetic layers. The sign and strength of the resulting DMI ($\mathbf{D}_{12}$) are interface/materials properties that lead to one favored chirality of the spin structure. adapted from Ref. [58]

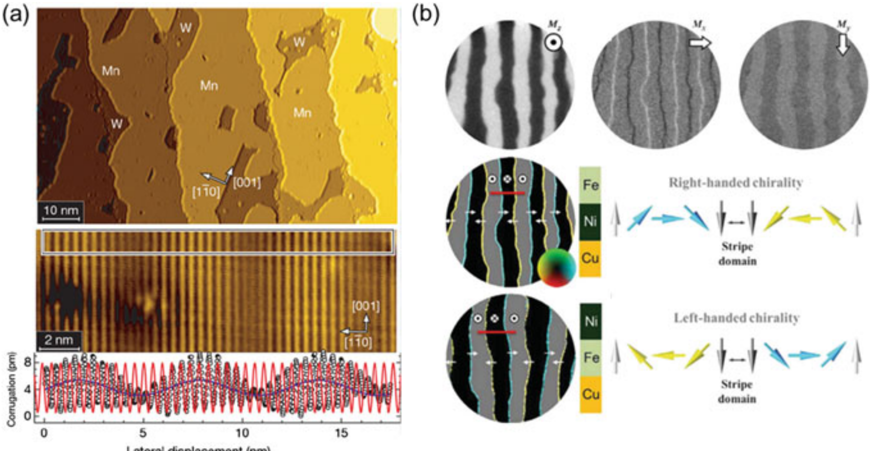


Fig. 10 **a** High-resolution constant-current image of Mn monolayers as probed by spin-polarized tunnelling. Comparison with a sine wave (red), expected for perfect AFM order, reveals a phase shift of pi between adjacent antinodes. In addition, there is an offset modulation (blue line), which we attribute to a varying electronic structure owing to spin–orbit coupling [59] **b** SPLEEM images of 2.5 ML Fe/2 MLNi/Cu(001) mapping orthogonal magnetization components: Compound image constructed from the SPLEEM images highlighting the DW. White arrows indicate the in-plane spin orientations inside the DWs [60]. Adapted with permission from [59, 60]

points in the plane of the layers and the magnitude and sign are interface/materials properties that lead to a favored chirality of a spin spiral state. For a sufficiently strong DMI, a spin spiral state is favored with the spiral axis lying in the plane and a spiral period of $4\pi A/D$ [56], where A is the strength of the Heisenberg exchange interaction and D is the strength of the DMI (Fig. 10).

3.2 *Experimental Measurement of the Interfacial DMI*

3.2.1 Imaging of Chiral Magnetic Domain Walls

First experimental evidence of the existence of the interfacial DMI at ultrathin layers was at first reported by Bode et al. [59], who observed magnetic order of a specific chirality in a single atomic layer of manganese on a tungsten (110) substrate by using spin-polarized scanning tunneling microscopy (STM). The results revealed that adjacent spins antiferromagnetically coupled are not perfectly canceled each other but slightly canted into a certain direction, leading to a spin spiral structure with a period of ~ 10 nm. Chen et al. [60], also reported direct imaging of a chiral Néel wall in ferromagnetic Fe/Ni bilayers epitaxially grown on Cu(100) by using spin-polarized low energy electron microscopy (SPLEEM). They found that the chirality

of the Néel wall is determined by the growth order of the Fe/Ni thin films, supporting the inversion-symmetry breaking mechanism of the DMI at the Fe/Ni interface.

3.2.2 Asymmetric Domain Wall Dynamics

In addition to the imaging techniques, there have been a lot of experimental effort to discover the interfacial DMI and quantify its strength through other measurement techniques. The most common way to measure the effect up to now is using asymmetric DW dynamics driven by a magnetic field or current. For example, Je et al. [61], reported an asymmetric DW motion of two opposite polarities under an in-plane bias field. (see Fig. 11a) Such asymmetrical DW velocities are attributed to the interfacial DMI which prefers a certain chiral configuration and differs the DW energy between DWs of two opposite polarities under the in-plane field [18, 61–63].

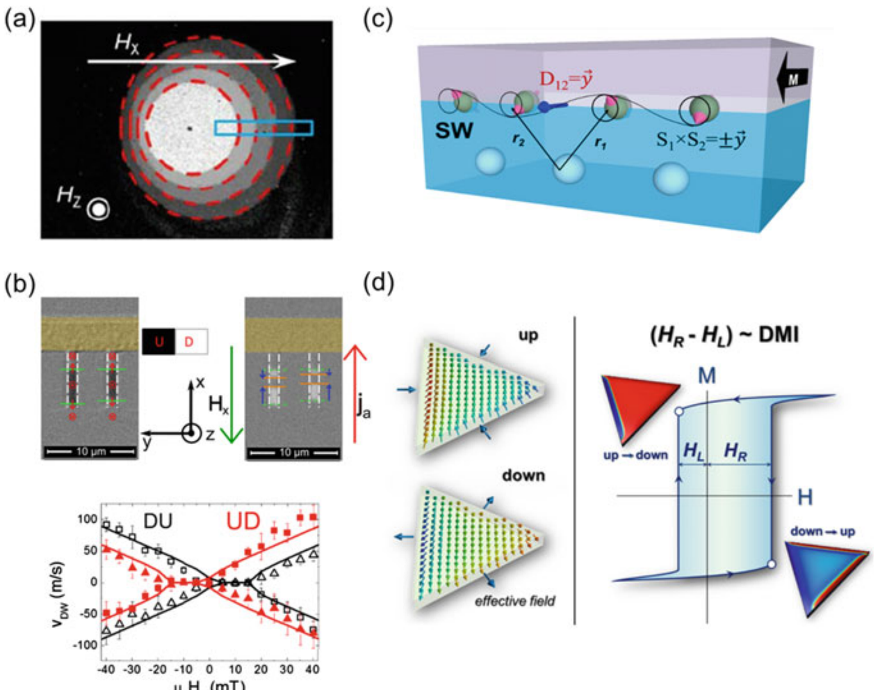


Fig. 11 **a** Circular DW expansion driven by an out-of-plane magnetic field Hz under an in-plane magnetic field Hx. Each image represents four sequential images with a fixed time step, which are captured using a magneto-optical Kerr effect microscope. The images were adapted from [61]. The white arrow and the symbols indicate the directions of each magnetic field. **b** Asymmetric DWs motion in nanowires under H_x [51]. **c** Asymmetric behavior of spin waves in the presence of the interfacial DMI. **d** Schematic diagram for an asymmetric magnetic hysteresis loop under H_x in a laterally asymmetric structure [78]. Adapted with permission from [51, 61, 78]

Lo Conte et al. [51], have reported asymmetric current-induced DW's velocity under the in-plane bias field in Ta/CoFeB/MgO nanowires. In the measurement, the DMI can be estimated by the field where DWs stop moving when a complete Bloch configuration is introduced by the in-plane field [64, 65]. However, the methods using DW dynamics have been pointed out that it may entail some drawbacks quantifying the interfacial DMI, as analysis can be hampered by possible additional effects [66, 67] from spin structure deformations and/or chiral damping [68], in addition to DMI. (see Fig. 11b).

3.2.3 Asymmetric Spin Wave Dispersion Relation

More recently, the measurements of the strength of the interfacial DMI via non-reciprocal spin waves' behavior have also been performed by using Brillouin Light Scattering (BLS) [69–72], time-resolved magneto-optical Kerr effect (TR-MOKE) [73], spin-polarized electron energy loss spectroscopy (SPEELS) [74], etc [75]. For the case of spin waves, the chirality of non-collinear spin configurations induced by spin waves is determined by their propagation direction (or wave vectors). For a Damon-Eshbach spin wave mode [76], where the wave vector is normal to the surface of magnetic thin films as well as the magnetization direction (see Fig. 11c), the total energy of two spin waves propagating into the opposite direction are differed by the interfacial DMI—the cross products of two neighboring spins vector, $\mathbf{S}_1 \times \mathbf{S}_2$, for both spin waves are all on the axis of the interfacial DMI vector, **D**, but antiparallel to each other. The main advantage of this technique for measuring the interfacial DMI is that it does not require an exchange stiffness for quantifying the strength of the interfacial DMI that is difficult to experimentally address in ultrathin ferromagnetic films. However, the accuracy of this approach is also contested, as asymmetric anisotropies at the two interfaces of the ferromagnetic materials (FM) can lead to similar asymmetries of the dispersion relation, making the interpretation for the DMI determination rather complex.

3.2.4 Other Measurement Techniques

Together with the techniques using spin dynamics, the measurements based on the statics of magnetic entities have been also developed. For example, Woo et al. [77], estimated the magnitude of the interfacial DMI by measuring the domain width of labyrinth stripe domain structures that is determined by the competition among micromagnetic energies, i.e., the interfacial DMI, magnetic anisotropic energy, exchange energy, dipolar energy, and Zeeman energy. Very recently, asymmetric magnetic hysteresis loop in laterally asymmetric structures under the application of an in-plane field is demonstrated by Han et al. (see Fig. 11d) [78].

4 Beyond Domain Wall Memories: Skyrmion Race-Track Memory Concepts

4.1 Skyrmions

Magnetic skyrmions are a particle-like chiral spin structure which is stabilized in systems with DMI. Skyrmions have gathered significant interest most recently in the spintronics community due to its exciting physics and possibility for applications in memory and logic devices [77, 79–83]. The topologically protected skyrmions are well known for the topological stability and predicted to have efficient current induced motion. These properties predict that skyrmions are promising for future applications. A skyrmion shows a unique spin structure where the center of the skyrmion is up (down) and the edges are down (up) and depending on the continuous spin structure the skyrmion is classified as a Néel type and a Bloch type skyrmion (see Fig. 12a, b). Skyrmions have been observed in out-of-plane magnetized materials, whether the out of plane magnetization was achieved via magnetic field or perpendicular magnetic anisotropy or a combination. Skyrmions are stabilized as a result of a trade-off between the exchange and the magnetostatic energy, and the bulk or interfacial DMI. The Bloch skyrmions are predominantly found in bulk materials while the Néel skyrmions are a characteristic of the interfacial DMI systems[20, 79, 84, 85]. The aforementioned bulk DMI is due to crystal inversion asymmetry and high spin–orbit coupling in ferromagnetic alloys (see Fig. 12c), such as B20 materials [54, 55], while the interfacial DMI has its origin in multilayers of a ferromagnet and a large spin–orbit material, such as Pt or Ir. In the multilayer systems, a structural inversion asymmetry exists, which is required for the DMI. In these skyrmion systems, there have been many different proposals for device applications. Especially, skyrmion racetrack memories which is analogous to the domain wall racetrack memory, the idea is to use a single skyrmion as a bit (see Fig. 12d, e). Depending on the topological number, when the topological winding number $Q = 1$ the bit could be defined as "1" and when $Q = 0$ the bit is "0". For the skyrmion racetrack to function properly, controlling the distance between the skyrmions is crucial, thus moving the skyrmions synchronously is crucial. As an advanced version of the skyrnion racetrack memory, multilane racetrack memory has been proposed [80] (See Fig. 12f). By micromagnetic simulations, it was shown that the by electrical currents the skyrmions could change lanes. Furthermore, a skyrmion racetrack memory was proposed in synthetic antiferromagnets where more efficient motion of skyrmions is predicted (Fig. 12g). In order to use these skyrmions as a racetrack memory device, similar to the domain wall racetrack memory the means to write, read, and shift the skyrmions is essential. In the following sections we will briefly give an overview of the studies in skyrmions. We will specially emphasize on the skyrmions in multilayer systems for these systems are directly applicable to the skyrmion racetrack memories.

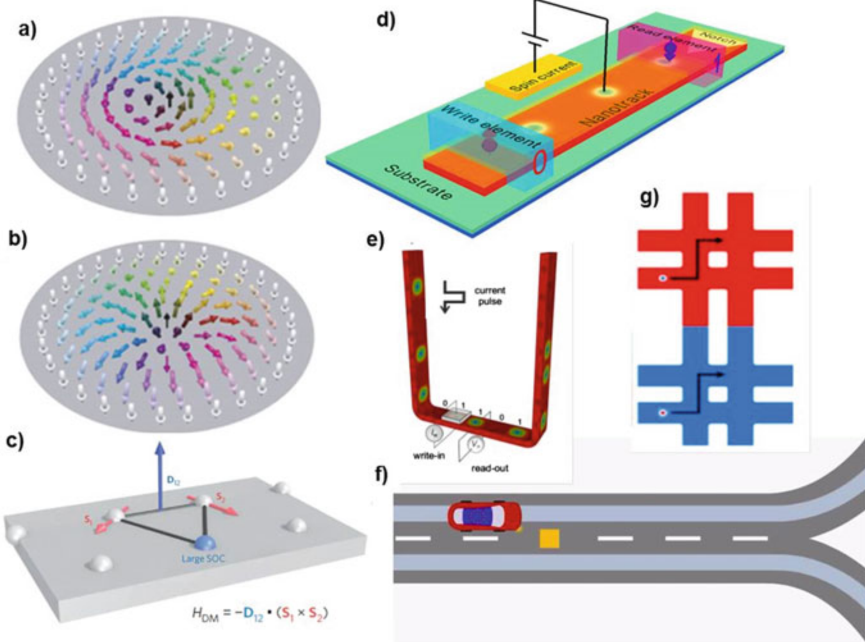


Fig. 12 The images are taken from [81] for **a** and **b**, [79] for **c**, [86] for **d**, [87] for **e**, [88] for **g**. Spin configuration in a skyrmion. a,b, Skyrmions in a 2D ferromagnet with uniaxial magnetic anisotropy along the vertical axis. The magnetization is pointing up on the edges and pointing down in the center. Moving along a diameter, the magnetization rotates by 2π around an axis perpendicular to the diameter **a** and by 2π around the diameter **b**, which corresponds to different orientations of the Dzyaloshinskii–Moriya vector. **c** schematic drawing of the Dzyaloshinskii-Moriya interaction in a bulk DMI system. **d**, **e** schematic illustration of the skyrmion based racetrack memory. The skyrmions are used as a bit, and the reading and writing are done by electrical means. **f** Multilane racetrack. Schematic of a skyrmion that can "drive" in matrix devices where skyrmions can be guided on complex trajectories with different lanes thus implementing non-volatile reconfigurable and synaptic logic. **g** Micromagnetic simulation of a skyrmion in a synthetic antiferromagnet (both layers are shown independently) moving on a complex trajectory due to spin torques Adapted with permission from [79, 81, 86–88]

4.2 Skyrmion Writing

Several studies have been done on the nucleation of skyrmions facilitating different means. Jiang et al. [89] demonstrated that a Néel skyrmion could be obtained through the conversion of a DW driven by spin Hall effect through a geometrical constriction (see Fig. 13a) in Ta/CoFeB/TaO_x multilayers. This mechanism resembles the formation of bubbles blown through a straw. Later, in Ir/Co/Pt [90] and Pt/Co/MgO [77] multilayers nucleation of skyrmions were demonstrated by changing the out-of-plane magnetic fields. Boulle et al. [91] showed that a single skyrmion could be nucleated at zero magnetic fields. Finally, the nucleation of skyrmions was achieved

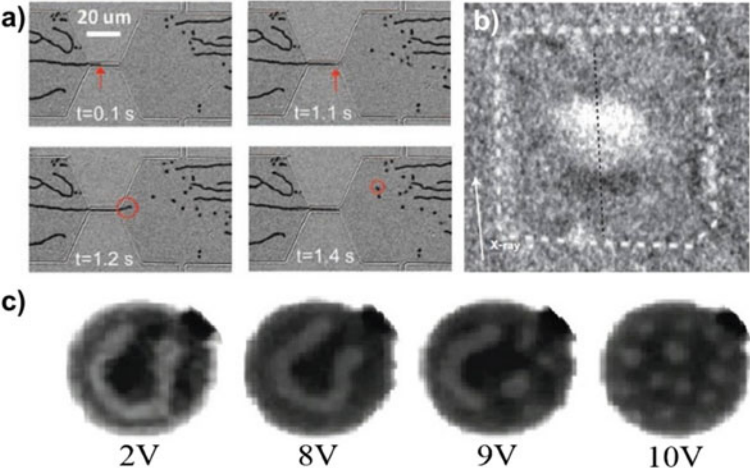


Fig. 13 **a** magnetic bubble skyrmions nucleated by spatially inhomogeneous currents [89]. **b** nucleation of single skyrmion in a confined geometry of Pt/Co/MgO [91]. **c** STXM images of skyrmion nucleation with bipolar field pulse in Pt/Co/Ta multilayers [77]. Adapted with permission from [77, 89, 91]

in out-of-plane field pulses from labyrinth stripe domains. Over a critical field value the stripe domains break down into a skyrmion lattice (see Fig. 13c). However, the nucleation of single skyrmions at room temperature is still a challenge. There are many predictions and calculations showing promising single skyrmion nucleation by utilizing STT, spin polarized currents, and microwave field. However, none have been demonstrated experimentally yet. For annihilating skyrmions, high out-of-plane magnetic fields or dc out-of-plane spin polarized currents where the magnetic field and spin polarization has to be opposite the core of the skyrmion.

4.3 Skyrmion Reading

The reading of skyrmions can be achieved by electrical detection via magnetic tunnel junctions (MTJ). By attaching tunnel junction to the skyrmion racetrack, it is possible to distinguish the existence of the skyrmion below the MTJ. In a MTJ, the resistance of the MTJ relies on the relative magnetization direction of the top and bottom ferromagnet of the MTJ. When the magnetization configuration of the two ferromagnets of the MTJ is parallel (or antiparallel) then the resistance will read low (or high). So if we use the racetrack as the bottom ferromagnet at the point of reading, by reading the resistance at the MTJ, the presence of a skyrmion could be read. Assuming the racetrack's magnetization is in the up state, if the top magnetization is in the up direction of the MTJ, the resistance of the reading element will be low. However, if you have a skyrmion which has a core in the down direction, when a skyrmion

is shifted below the MTJ, the resistance reading at the MTJ will be high. Another option is to measure the topological Hall effect at the reading element.

4.4 Skyrmion Shifting

For racetrack memory applications of skyrmions, the manipulation of skyrmions by in-plane currents is crucial. Current induced motion of skyrmions has been intensely studied both theoretically and experimentally. In 2012, the current-induced motion of a Bloch skyrmion lattice was demonstrated using Lorentz microscopy [92]. Schulz et al. [93] and Everschor et al. [94] analyzed the skyrmion motion using Berry phase and Thiele's equation. The first numerical calculation on the dynamics of the Bloch-type skyrmion lattice driven by the STT has been demonstrated [95]. The results reveals that the skyrmion can depin at a very low current density (~10^6A/m^2) and its motion for closely packed lattices is insensitive to defects. Later, the dynamics of Néel-type skyrmions was calculated by Sampaio et al. [84], particularly, in a perpendicular ultrathin ferromagnet strip. There were numerous studies of current induced skyrmion motion theoretically [85, 96]. In skyrmions stabilized by the interfacial DMI, the current-driven skyrmion motion can arise from two different effects. One is the spin polarized current passing through the ferromagnet, providing the STT effect and the second is the spin polarized current generated by the heavy metal layer – that is the SOT – as explained in the prior sections. Consequently, the skyrmion motion can be described by using Thiele's equation incorporated with these two effects. For example, by assuming a topological number $Q = 1$, the equation of motion of a skyrmion is given by: [8, 85, 94, 96, 97]

$$-M\ddot{R} + \mathbf{G} \times \left(\dot{\mathbf{R}} - \mathbf{u}\right) - \tilde{D}\left(\alpha\dot{\mathbf{R}} - \beta\mathbf{u}\right) + 4\pi B\tilde{R}(\Psi)\mathbf{j}_{\mathrm{HM}} + \mathbf{F} = 0 \quad (7)$$

where $\mathbf{R}$ is the position of the skyrmion at a given reference frame, the M is the effective mass of the skyrmion, $\tilde{D} = \mathrm{T}M_s\gamma \int dxdy(\partial_x\mathbf{m})^2$ is the dissipation constat, α is the Gilbert damping, β is the non-adiabaticity parameter, $\mathbf{G}$ is the gyro-coupling vector, $B = \frac{\gamma^2\hbar\theta_{SH}}{2e}I$ (with the reduced Planck constant $\hbar$, the spin Hall angle θ_{SH}, and $\mathrm{I} = \frac{1}{4}\int_0^\infty dr(\sin\theta\cos\theta + r\frac{d\theta}{dr})$ is a coefficient that depends on the spin configuration), $\tilde{R}(\Psi) = \begin{pmatrix} \cos\Psi & \sin\Psi \\ \sin\Psi & \cos\Psi \end{pmatrix}$ is the rotation matrix corresponding to domain wall angle ψ, $\mathbf{j}_{\mathrm{HM}}$ is the current density in the heavy metal, and $\mathbf{F}$ is a driving force acting on the skyrmion. Mostly, the skyrmion dynamics satisfy this Thiele equation however, there are details of skyrmion motion that are only apparent in micromagnetic simulations. Due to the gyro-term of the Thiele equation, the x and y components of $\mathbf{R}$ are canonically conjugate variables [96]. Therefore, when there is no confining potential ($\mathbf{F} = 0$), the steady state motion of a skyrmion should exhibit a linear trajectory with a certain skyrmion Hall angle, an angle between the x and y components of $\mathbf{R}$. Considering a spin current applied into the x-axis [$\mathbf{u}$ = (u,0,0)], the velocities of the

skyrmion along these two axes are given as:

$$\dot{R}_x = \left(\frac{\beta}{\alpha} + \frac{\alpha - \beta}{\alpha^3 \left(\tilde{D}/G \right)^2 + \alpha} \right) u \tag{8}$$

$$\dot{R}_y = \frac{(\alpha - \beta)(\tilde{D}/G)}{\alpha^2 \left(\tilde{D}/G \right)^2 + 1} u \tag{9}$$

For the case of $\alpha \ll 1$ and $\tilde{D}/G \leq 1$, the Eqs. 8 and 9 will be $\dot{R}_x \approx u$ and $\dot{R}_y \approx (\alpha - \beta)(\tilde{D}/G)$ [96]. In this case, the velocity of the skyrmion along the current flow does not depend on the material parameters, α and β, but, only the skyrmion Hall angle ($\tan(\xi) = \dot{R}_x/\dot{R}_y$) depends on the material parameters. Furthermore, the skyrmion Hall angle scales inversely with Q, as G is proportional to Q,

When the skyrmion motion is driven by the spin current $\mathbf{j}_{\mathrm{HM}} = (j_{HM},0,0)$ arising from the SOTs instead of the STTs, the skyrmion motion shows distinct behavior depending on the chirality of the interfacial DMI. The skyrmion is dragged to a direction where the spins of the domain walls are parallel to the injected spins [85]. For a skyrmion with $Q = 1$ the velocity will be,

$$\dot{R}_x = \left[\frac{\alpha \check{D} B}{G^2 + \alpha^2 \tilde{D}^2} \cos(\Psi) + \frac{GB}{G^2 + \alpha^2 \tilde{D}^2} \sin(\Psi) \right] j_{HM}, \tag{10}$$

$$\dot{R}_y = \left[-\frac{\alpha \check{D} B}{G^2 + \alpha^2 \tilde{D}^2} \sin(\Psi) + \frac{GB}{G^2 + \alpha^2 \tilde{D}^2} \cos(\Psi) \right] j_{HM}, \tag{11}$$

This implies that the Néel skyrmion with a low damping will move transverse to the spin current and the Bloch skyrmion moves in the direction of the spin current [96]. In the case of a confined potential, the motion of the skyrmion changes drastically. The skyrmion moves into a diagonal direction with respect to the current direction until it reaches the higher potential energy region at the edge before it starts to move longitudinally in a steady state motion.

Very recently, the spin Hall induced motion of Néel skyrmions in synthetic antiferromagnets (SAF) due to damping-like spin–orbit torque was numerically demonstrated [98]. In the SAFs, since the two ferromagnetic layers are coupled by interlayer exchange coupling, it is shown that the skyrmion velocity of an order of magnitude larger could be expected. Recently, the current induced displacement of Néel skyrmions has been shown. It was demonstrated in Ta/CoFeB/TaO_x [89] multilayer systems and Pt/Co/Ta wires [77]. Even skyrmion Hall effects were observed by Jiang et al. [98] in Ta/CoFeB/TaOx multilayer, and Litzius et al. [82] Nature electronics, 3, 30 (2020) measured the skyrmion Hall effect in Pt/CoFeB/MgO. The skyrmion Hall effect is the phenomenon where the skyrmions move perpendicular to

the current. Analogous to the Hall effect in electrons. To actually use the skyrmions for applications there are still a number of open questions that need to be answered.

5 Conclusion

In the chapter, we have reviewed the fundamental physics for designing magnetic domain wall memories, especially domain wall racetrack memories. An overview of how the racetrack has been functionally improved and how the fundamental physics behind the operating mechanism has developed is shown. Material wise, the design of the racetrack has changed from using in-plane magnetic materials to out-of-plane magnetic materials. In the process of changing the material design resulted in new physics such as the SOTs and the DMI which resulted in domain wall motion with higher efficiency, and stability. The SOTs are the main mechanism in moving the domain walls efficiently by utilizing the SHE and the ISGE which have shown to be more efficient than the STT in current induced domain wall motion. The exact physics behind this SOT is still not well known, but it was well demonstrated that the SOT shows higher efficiency for DW motion. However, this SOT requires the DMI in order to act on the DWs. The DMI generates a certain chirality to the domain walls, especially forcing a Néel type DW. The Néel DW is required for the SOT to act as a driving force of the DWs. The different sections of the chapter have reviewed different physics and evidence of the SOT and DMI with the different experimental methods to quantify the SOT and DMI shown. Furthermore, as an outlook for the racetrack memory, we have reviewed the new exciting skyrmion racetrack memory which will be a possible future implementation of the racetrack memory.

References

1. S.S.P. Parkin, M. Hayashi, L. Thomas, Magnetic domain-wall racetrack memory. Science **320**, 190 (2008)
2. J.-S. Kim et al., Synchronous precessional motion of multiple domain walls in a ferromagnetic nanowire by perpendicular field pulses. Nat. Commun. **5**, 3429 (2014)
3. T.A. Moore et al., Magnetic-field-induced domain-wall motion in permalloy nanowires with modified Gilbert damping. Phys. Rev. B **82**, 094445 (2010)
4. J.C. Slonczewski, Current-driven excitation of magnetic multilayers. J. Magn. Magn. Mater. **159**, L1–L7 (1996)
5. L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current. Phys. Rev. B **54**, 9353–9358 (1996)
6. A. Brataas, A.D. Kent, H. Ohno, Current-induced torques in magnetic materials. Nat. Mater. **11**, 372–381 (2012)
7. M. Hayashi et al., Current driven domain wall velocities exceeding the spin angular momentum transfer rate in permalloy nanowires. Phys. Rev. Lett. **98**, 037204 (2007)
8. A. Thiaville, Y. Nakatani, J. Miltat, Y. Suzuki, Micromagnetic understanding of current-driven domain wall motion in patterned nanowires. EPL **69**, 990 (2005)

9. M. Kläui, Head-to-head domain walls in magnetic nanostructures. J. Phys. Condens. Matter **20**, 313001 (2008)
10. O. Boulle, G. Malinowski, M. Kläui, Current-induced domain wall motion in nanoscale ferromagnetic elements. Mater. Sci. Eng. R Reports **72**, 159–187 (2011)
11. O. Boulle et al., Nonadiabatic spin transfer torque in high anisotropy magnetic nanowires with narrow domain walls. Phys. Rev. Lett. **101**, 216601 (2008)
12. T. Koyama et al., Observation of the intrinsic pinning of a magnetic domain wall in a ferromagnetic nanowire. Nat. Mater. **10**, 194–197 (2011)
13. D. Ravelosona, D. Lacour, J.A. Katine, B.D. Terris, C. Chappert, Nanometer scale observation of high efficiency thermally assisted current-driven domain wall depinning. Phys. Rev. Lett. **95**, 117203 (2005)
14. T. Suzuki, S. Fukami, K. Nagahara, N. Ohshima, N. Ishiwata, Current-driven domain wall motion, nucleation, and propagation in a Co/Pt multi-layer strip with a stepped structure. IEEE Trans. Magn. **44**, 2535–2538 (2008)
15. T.A. Moore et al., High domain wall velocities induced by current in ultrathin Pt/Co/AlOx wires with perpendicular magnetic anisotropy. Appl. Phys. Lett. **93**, 262504–212404 (2008)
16. I.M. Miron et al., Domain wall spin torquemeter. Phys. Rev. Lett. **102**, 137202 (2009)
17. M. Cormier et al., Effect of electrical current pulses on domain walls in Pt/Co/Pt nanotracks with out-of-plane anisotropy: spin transfer torque versus Joule heating. Phys. Rev. B **81**, 024407 (2010)
18. R. Lavrijsen et al., Asymmetric Pt/Co/Pt-stack induced sign-control of current-induced magnetic domain-wall creep. Appl. Phys. Lett. **100**, 262408 (2012)
19. S. Parkin, S.-H. Yang, Memory on the racetrack. Nat. Nanotechnol. **10**, 195–198 (2015)
20. S. Emori, U. Bauer, S.-M. Ahn, E. Martinez, G.S.D. Beach, Current-driven dynamics of chiral ferromagnetic domain walls. Nat. Mater. **12**, 611–616 (2013)
21. S.D. Ganichev et al., Spin-galvanic effect. Nature **417**, 153–156 (2002)
22. Y.A. Bychkov, E.I. Rashba, Properties of a 2D electron gas with lifted spectral degeneracy. SovJEPT Lett. **39**, 78–81 (1984)
23. S. Zhang, P.M. Levy, A. Fert, Mechanisms of spin-polarized current-driven magnetization switching. Phys. Rev. Lett. **88**, 236601 (2002)
24. I. M. Miron et al., Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. Nat. Mater. **9**, 230 (2010)
25. I.M. Miron et al., Fast current-induced domain-wall motion controlled by the Rashba effect. Nat. Mater. **10**, 419–423 (2011)
26. M. Kavand et al, Quantitative inverse spin hall effect detection via precise control of the driving-field amplitude. Phys. Rev. B—Condens. Matter Mater. Phys. **95**, 161406(R) (2017)
27. L. Liu, T. Moriyama, D.C. Ralph, R.A. Buhrman, Spin-torque ferromagnetic resonance induced by the spin hall effect. Phys. Rev. Lett. **106**, 036601 (2011)
28. J. Sinova, S.O. Valenzuela, J. Wunderlich, C.H. Back, T. Jungwirth, Spin hall effects. Rev. Mod. Phys. **87**, 1213–1260 (2015)
29. M.I. Dyakonov, V.I. Perel, Current-induced spin orientation of electrons in semiconductors. Phys. Lett. A **35**, 459–460 (1971)
30. J.E. Hirsch, Spin hall effect. Phys. Rev. Lett. **83**, 1834–1837 (1999)
31. S. Murakami, N. Nagaosa, S.-C. Zhang, Dissipationless quantum spin current at room temperature. Science (80-.). **301**, 1348 LP–1351 (2003)
32. J. Sinova et al., Universal intrinsic spin hall effect. Phys. Rev. Lett. **92**, 126603 (2004)
33. N.F. Mott, On the interpretation of the relativity wave equation for two electrons. Proc. R. Soc. London. Ser. A **124**, 422LP–425 (1929)
34. N.F. Mott, The polarisation of electrons by double scattering. Proc. R. Soc. London. Ser. A **135**, 429LP–458 (1932)
35. A. Crépieux, P. Bruno, Relativistic corrections in magnetic systems. Phys. Rev. B **64**, 94434 (2001)
36. J.C. Sankey et al., Measurement of the spin-transfer-torque vector in magnetic tunnel junctions. Nat. Phys. **4**, 67–71 (2008)

37. W. Zhang, W. Han, X. Jiang, S.-H. Yang, S.S.P. Parkin, Role of transparency of platinum–ferromagnet interfaces in determining the intrinsic magnitude of the spin Hall effect. Nat. Phys. **11**, 496–502 (2015)
38. T. Nan et al., Comparison of spin-orbit torques and spin pumping across NiFe/Pt and NiFe/Cu/Pt interfaces. Phys. Rev. B **91**, 214416 (2015)
39. J. Kim et al., Spin-orbit torques associated with ferrimagnetic order in Pt/GdFeCo/MgO layers. Sci. Rep. **8**, 6017 (2018)
40. S. Woo, M. Mann, A.J. Tan, L. Caretta, G.S.D. Beach, Enhanced spin-orbit torques in Pt/Co/Ta heterostructures. Appl. Phys. Lett. **105**, 162507–122404 (2014)
41. J. Kim et al., Layer thickness dependence of the current-induced effective field vector in Ta|CoFeB|MgO. Nat. Mater. **12**, 240–245 (2012)
42. K. Garello et al., Symmetry and magnitude of spin–orbit torques in ferromagnetic heterostructures. Nat. Nanotechnol. **8**, 587–593 (2013)
43. Y. Fan et al., Magnetization switching through giant spin–orbit torque in a magnetically doped topological insulator heterostructure. Nat. Mater. **13**, 699–704 (2014)
44. A. Thiaville, S. Rohart, E. Jue, V. Cros, A. Fert, Dynamics of Dzyaloshinskii domain walls in ultrathin magnetic films. EPL **100**, 57002 (2012)
45. T. Schulz et al., Effective field analysis using the full angular spin-orbit torque magnetometry dependence. Phys. Rev. B **95**, 224409 (2017)
46. L. Liu, O.J. Lee, T.J. Gudmundsen, D.C. Ralph, R.A. Buhrman, Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from the spin hall effect. Phys. Rev. Lett. **109**, 096602 (2012)
47. J. Heinen et al., Determination of the spin torque non-adiabaticity in perpendicularly magnetized nanowires. J. Phys. Condens. Matter **24**, 24220 (2012)
48. T. Schulz et al., Spin-orbit torques for current parallel and perpendicular to a domain wall. Appl. Phys. Lett. **107**, 122405 (2015)
49. C. Zhang, S. Fukami, H. Sato, F. Matsukura, H. Ohno, Spin-orbit torque induced magnetization switching in nano-scale Ta/CoFeB/MgO. Appl. Phys. Lett. **107**, 122405–222401 (2015)
50. P.P.J. Haazen et al., Domain wall depinning governed by the spin Hall effect. Nat. Mater. **12**, 299–303 (2013)
51. R. Lo Conte et al., Role of B diffusion in the interfacial Dzyaloshinskii-Moriya interaction in Ta/Co20Fe60B20/MgO nanowires. Phys. Rev. B **91**, 14433 (2015)
52. I. Dzyaloshinsky, A thermodynamic theory of "weak" ferromagnetism of antiferromagnetics. J. Phys. Chem. Solids **4**, 241–255 (1958)
53. T. Moriya, Anisotropic superexchange interaction and weak ferromagnetism. Phys. Rev. **120**, 91–98 (1960)
54. S. Mühlbauer et al., Skyrmion lattice in a chiral magnet. Science (80-.) **323**, 915 (2009)
55. S.X. Huang, C.L. Chien, Extended skyrmion phase in epitaxial FeGe (111) Thin Films. Phys. Rev. Lett. **108**, 267201 (2012)
56. S. Rohart, A. Thiaville, Skyrmion confinement in ultrathin film nanostructures in the presence of Dzyaloshinskii-Moriya interaction. Phys. Rev. B **88**, 184422 (2013)
57. B. Krüger, M. Kläui topological defects in nanostructures---chiral domain walls and skyrmions. in *Topological Structures in Ferroic Materials: Domain Walls, Vortices and Skyrmions* ed. by J. Seidel (Springer International Publishing, 2016) pp. 199–218
58. F. Büttner, M. Kläui, Chapter 8 magnetic skyrmion dynamics. in *Skyrmions: Topological Structures, Properties, and Applications* (CRC Press, 2016), pp. 211–238
59. M. Bode et al., Chiral magnetic order at surfaces driven by inversion asymmetry. Nature **447**, 190–193 (2007)
60. G. Chen et al., Novel chiral magnetic domain wall structure in Fe/Ni / Cu (001) films. Phys. Rev. Lett. **110**, 177204 (2013)
61. S.G. Je et al., Asymmetric magnetic domain-wall motion by the Dzyaloshinskii-Moriya interaction. Phys. Rev. B **88**, 214401 (2013)
62. M. Vavnatka et al., Velocity asymmetry of Dzyaloshinskii domain walls in the creep and flow regimes. J. Phys. Condens. Matter **27**, 326002 (2015)

63. A. Hrabec et al., Measuring and tailoring the Dzyaloshinskii-Moriya interaction in perpendicularly magnetized thin films. Phys. Rev. B **90**, 020402 (2014)
64. K.-S. Ryu, L. Thomas, S.-H. Yang, S. Parkin, Chiral spin torque at magnetic domain walls. Nat. Nanotechnol. **8**, 527–533 (2013)
65. J. Torrejon et al., Interface control of the magnetic chirality in CoFeB/MgO heterostructures with heavy-metal underlayers. Nat. Commun. **5**, 144425 (2014)
66. D.-Y. Kim et al., Chirality-induced antisymmetry in magnetic domain wall speed. NPG Asia Mater. **10**, e464–e464 (2018)
67. D.-Y. Kim et al., Magnetic domain-wall tilting due to domain-wall speed asymmetry. Phys. Rev. B **97**, 134407 (2018)
68. E. Jué et al., Chiral damping of magnetic domain walls. Nat. Mater. **15**, 272 (2015)
69. J. Cho et al., Thickness dependence of the interfacial Dzyaloshinskii-Moriya interaction in inversion symmetry broken systems. Nat. Commun. **6**, 7635 (2015)
70. N.H. Kim et al., Interfacial Dzyaloshinskii-Moriya interaction, surface anisotropy energy, and spin pumping at spin orbit coupled Ir/Co interface. Appl. Phys. Lett. **108**, 142406–152403 (2016)
71. N.H. Kim et al., Improvement of the interfacial Dzyaloshinskii-Moriya interaction by introducing a Ta buffer layer. Appl. Phys. Lett. **107**, 142408–152403 (2015)
72. H.T. Nembach, J.M. Shaw, M. Weiler, E. Jué, T.J. Silva, Linear relation between heisenberg exchange and interfacial Dzyaloshinskii-Moriya interaction in metal films. Nat. Phys. **11**, 825–829 (2015)
73. H.S. Körner et al., Interfacial Dzyaloshinskii-Moriya interaction studied by time-resolved scanning kerr microscopy. Phys. Rev. B **92**, 220413 (2015)
74. K. Zakeri et al., Asymmetric spin-wave dispersion on Fe(110): direct evidence of the Dzyaloshinskii-Moriya interaction. Phys. Rev. Lett. **104**, 137203 (2010)
75. J.M. Lee et al., All-electrical measurement of interfacial dzyaloshinskii-moriya interaction using collective spin-wave dynamics. Nano Lett. **16**, 62–67 (2016)
76. J.R. Eshbach, R.W. Damon, Surface magnetostatic modes and surface spin waves. Phys. Rev. **118**, 1208–1210 (1960)
77. S. Woo et al., Observation of room-temperature magnetic skyrmions and their current-driven dynamics in ultrathin metallic ferromagnets. Nat. Mater. **15**, 501–506 (2016)
78. D.S. Han et al., Asymmetric hysteresis for probing Dzyaloshinskii-moriya interaction. Nano Lett. **16**, 4438–4446 (2016)
79. A. Fert, V. Cros, J. Sampaio, Skyrmions on the track. Nat. Nanotechnol. **8**, 152–156 (2013)
80. J. Müller, Magnetic skyrmions on a two-lane racetrack. New J. Phys. **19**, 25002 (2017)
81. F. Buttner et al., Dynamics and inertia of skyrmionic spin structures. Nat. Phys. **11**, 225–228 (2015)
82. K. Litzius et al., Skyrmion Hall effect revealed by direct time-resolved X-ray microscopy. Nat. Phys. **13**, 170–175 (2016)
83. J. Zázvorka et al., Thermal skyrmion diffusion used in a reshuffler device. Nat. Nanotechnol. **14**, 658–661 (2019)
84. J. Sampaio, V. Cros, S. Rohart, A. Thiaville, A. Fert, Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures. Nat. Nanotechnol. **8**, 839–844 (2013)
85. R. Tomasello et al., A strategy for the design of skyrmion racetrack memories. Sci. Rep. **4**, 6784 (2014)
86. S. Zhang et al., Topological computation based on direct magnetic logic communication. Sci. Rep. **5**, 15773 (2015)
87. X. Zhang et al., Skyrmion-skyrmion and skyrmion-edge repulsions in skyrmion-based racetrack memory. Sci. Rep. **5**, 7643 (2015)
88. G. Finocchio, F. Büttner, R. Tomasello, M. Carpentieri, M. Kläui, Magnetic skyrmions: from fundamental to applications. J. Phys. D. Appl. Phys. **49**, 423001 (2016)
89. W. Jiang et al, Blowing magnetic skyrmion bubbles. Science (80-.) **349**, 283 (2015)
90. C. Moreau-Luchaire et al., Additive interfacial chiral interaction in multilayers for stabilization of small individual skyrmions at room temperature. Nat. Nanotechnol. **11**, 444–448 (2016)

91. O. Boulle et al., Room temperature chiral magnetic skyrmion in ultrathin magnetic nanostructures. Nat. Nanotechnol. **11**, 449–454 (2016)
92. X.Z. Yu et al., Skyrmion flow near room temperature in an ultralow current density. Nat. Commun. **3**, 988 (2012)
93. T. Schulz et al., Emergent electrodynamics of skyrmions in a chiral magnet. Nat. Phys. **8**, 301–304 (2012)
94. K. Everschor, M. Garst, R.A. Duine, A. Rosch, Current-induced rotational torques in the skyrmion lattice phase of chiral magnets. Phys. Rev. B **84**, 064401 (2011)
95. J. Iwasaki, M. Mochizuki, N. Nagaosa, Universal current-velocity relation of skyrmion motion in chiral magnets. Nat. Commun. **4**, 1463 (2013)
96. J. Iwasaki, M. Mochizuki, N. Nagaosa, Current-induced skyrmion dynamics in constricted geometries. Nat. Nanotechnol. **8**, 742–747 (2013)
97. I. Makhfudz, B. Krüger, O. Tchernyshyov, Inertia and chiral edge modes of a skyrmion magnetic bubble. Phys. Rev. Lett. **109**, 217201 (2012)
98. X. Zhang, Y. Zhou, M. Ezawa, Magnetic bilayer-skyrmions without skyrmion Hall effect. Nat. Commun. **69**, 990 (2015)

Circuit Design for Non-volatile Magnetic Memory

Tony Tae-Hyoung Kim

Abstract High performance low power memory is a topmost requirement in advanced computing systems. Magnetic memory has been considered as a promising solution because of its performance and non-volatility. However, it has various design challenges such as small tunneling magnetoresistance (TMR) ratios and large variability that need to be tackled for reliable operation. This chapter will discuss those challenges and introduce state-of-the-art write and read techniques.

1 Introduction

The performance of most computing systems is generally limited by the characteristics of memory. Over the last few decades, charge-storage-based memories such as static random access memory (SRAM) and dynamic random access memory (DRAM) have been the mainstream memory solutions. However, these memory technologies are facing challenges in scaling, which also limits the implementable memory density and the system performance. In addition, many emerging systems with frequency idle states require to store data with extremely low power consumption. No conventional memory technology can satisfy the above requirements. To tackle these issues, various non-volatile memory devices such as magnetic memory (MRAM), phase-change memory (PCRAM), and resistive memory (RRAM) have been explored. Since these memory devices have two terminals, they demonstrate high scalability (<10 nm). In addition, the non-volatile characteristics allows these memory technologies applicable to many ultra-low power systems requiring almost-zero power during idle or standby modes. However, the above non-volatile memory candidates have various design issues to be overcome, which is the main focus of this chapter.

This chapter introduces the overview of magnetic memory cells in circuit designer's point of view. After that, basic memory architecture for magnetic memory

T. T.-H. Kim (✉)
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: THKIM@ntu.edu.sg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_6

is explained followed by design challenges in magnetic memory. Then, various circuit techniques for write and sensing for magnetic memory are discussed in the following section.

2 Overview of Magnetic Memory Cells

Magnetic memory cells are implemented with the magnetic tunnel junction (MTJ), which provides two different resistance values depending upon magnetization states of the top layer (i.e. free layer) and the bottom layer (i.e. pinned layer) as shown in Fig. 1a. The magnetization orientation is controlled by the bias voltage between the free layer and the pinned layer. When the layers have parallel magnetization, the magnetic memory device can conduct more current (Low Resistance State). Less current flows through the magnetic device when the magnetization orientations of the layers are opposite. The ratio between two different resistance values is called tunneling magnetoresistance (TMR) ratio. The resistance values can be controlled by the bias voltage as described in Fig. 1b. Applying higher bias voltage produces

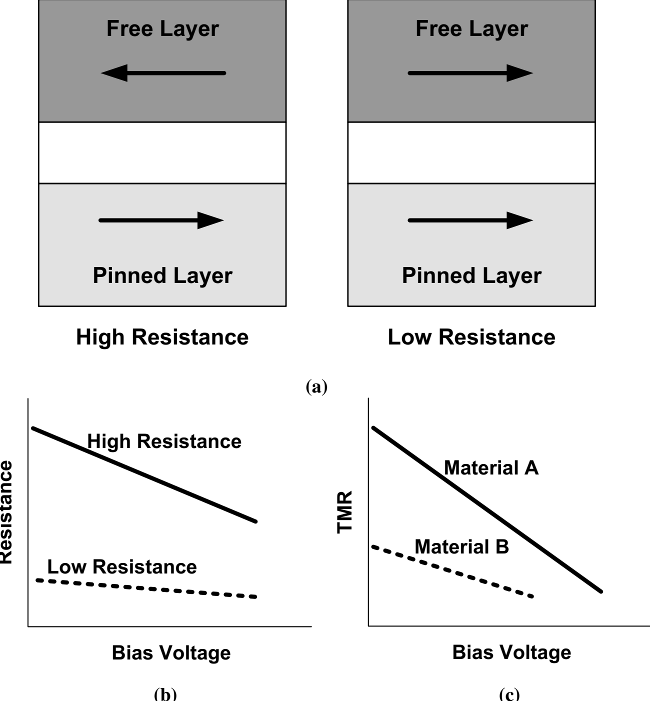


Fig. 1 **a** Schematic of magnetic memory cell structure, **b** effect of bias voltage on magnetic memory device resistance values, and **c** characteristics of TMR over bias voltage

lower resistance in both low and high resistance states. But, the high resistance state is affected more than the low resistance, which leads to a degradation in the overall TMR ratio as depicted in Fig. 1c. Many materials have also been explored for enhancing the TMR ratios (Fig. 1c). The TMR ratio is conceptually similar to on-current to off-current ratios in SRAM or DRAM cells. For reliable sensing, it is more desirable to have higher TMR ratios. However, in general, the TMR ratios of magnetic memory are much smaller than the on-current to off-current ratios of SRAMs or DRAMs. Therefore, reliable sensing in the magnetic memory is very challenging compared to SRAMs and DRAMs.

3 Magnetic Memory Array Architecture

Two commonly employed magnetic memory array architectures are the 1T1R array architecture shown in Fig. 2 and the crossbar array architecture in Fig. 3. In the 1T1R architecture, each magnetic memory cell consists of a selection transistor and a MTJ device connected in series. The selection transistor is employed to isolate a selected MTJ device from unselected MTJ devices during write and read operation. The 1T1R array architecture is similar to DRAM array architecture where one selection transistor and a capacitor connected in series form a cell and the selection transistor also isolate the selected capacitor from other unselected capacitors. The magnetic memory

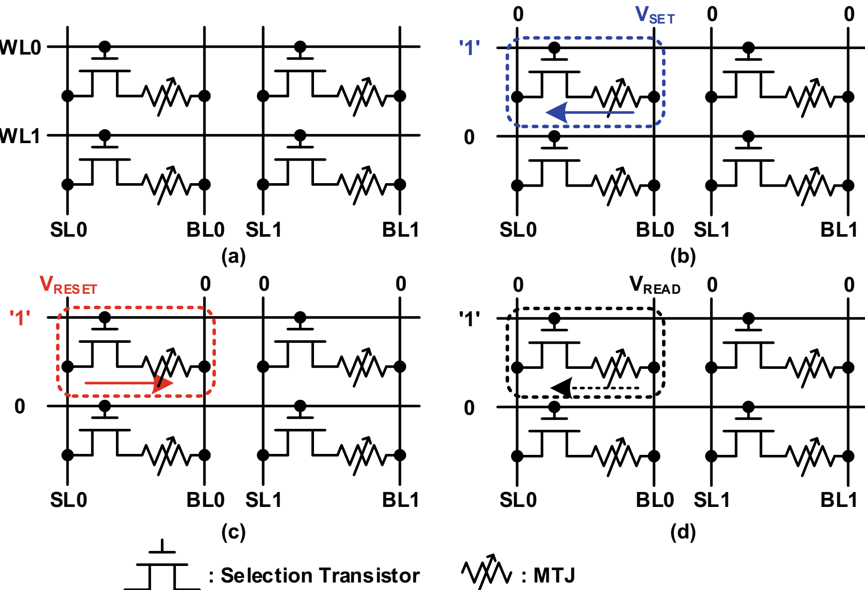


Fig. 2 **a** 1T1R array, **b** 1T1R in set operation, **c** 1T1R in reset operation, and **c** 1T1R in read operation

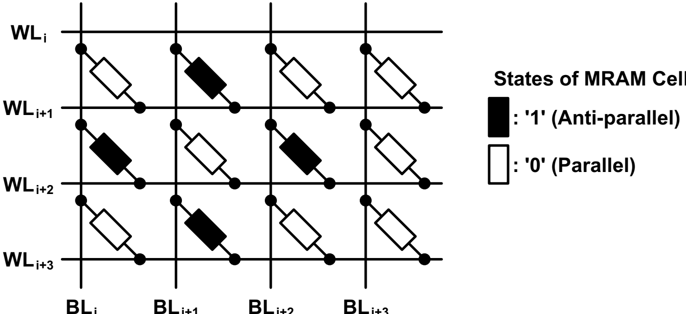


Fig. 3 Crossbar MRAM array architecture

cell size can be as small as the DRAM cell size since both require one transistor and a MTJ device and a capacitor are implemented above or below the transistor layer. However, the size of the selection transistor in the magnetic memory cell needs to be carefully chosen after considering the required write current. Therefore, the cell size is mostly determined by the area occupied by the selection transistor.

The operation of the 1T1R array architecture is as follows. The 1T1R array architecture employs three important signals. They are word lines (WLi), source lines (SLi), and bit lines (BLi). A word line selects a row to access. A source line and a bit line activate a column. The source lines are connected to the selection transistors while the bit lines are connected to the MTJs as described in Fig. 2a. SET (Writing '1') and RESET (Writing '0') operations are explained in Fig. 2b, c. For SET (Fig. 2b), a word line is selected by supplying high voltage, and SET voltage (V_{SET}) is applied to the bit line of a selected column. All the other signals are grounded. In the selected magnetic memory cell, current flows from BL0 to SL0 (Fig. 2b). Similarly, in RESET (Fig. 2c), RESET voltage (V_{RESET}) is applied to the source line of a selected column. Therefore, current flows from SL0 to BL0. This two different current directions and their values determine MTJ resistance values. Finally, for read operation (READ), read voltage (V_{READ}) is applied to the bit line of a selected column, which forms current flowing from BL0 to SL0. Smaller current will flow if the selected MTJ device is in the anti-parallel state. If the selected MTJ device is in the parallel state, relatively larger current will flow through the MTJ device. A sensing circuit tells the difference in the cell current for the anti-parallel state and the parallel state. The read operation is executed through a single bit line. Therefore, reference current or voltage is necessary for comparison in the sensing circuit. Another critical parameter to be considered is the level of V_{READ}. The resistance of the selected MTJ should not change during read operation. This requires V_{READ} to be low. However, if $\mathrm{V}_{\mathrm{READ}}$ is too low, the difference in the cell current deteriorates as well as the read speed. Unlike DRAMs, no interference occurs in unselected cells by biasing source lines and bit lines to ground.

Even though the 1T1R array architecture can provide the magnetic memory cell size as small as the DRAM cell size theoretically, the typical magnetic memory cell

size (12 F^2) for the 1T1R array architecture is larger than the DRAM cell size (6 F^2) due to two significant requirements. First, the 1T1R array architecture has two vertical lines (SLi and BLi) while the DRAM architecture has only one vertical line. In addition, the amount of required write current also limits the cell size minimization. In general, the amount of write current generated by a minimum size selection transistor is not large enough for reliable write operation (SET and RESET). Therefore, the transistor size should be increased, which also increase the magnetic memory cell size, accordingly. To tackle this issue, the crossbar array architecture has been introduced where no selection transistor is used as depicted in Fig. 3. In this architecture, word lines (WL_i) and bit lines (BL_i) are connected perpendicularly and magnetic memory cells are inserted between the word lines and the bit lines. The crossbar array architecture has been reported to have the cell area of 4 F^2 because of removing the selection transistor. Thus, the density of the magnetic memory in the crossbar array architecture is higher than that of the magnetic memory in the 1T1R array architecture and that of DRAMs. When compared to the 1T1R array architecture, the crossbar array architecture has no source lines. This prevents a selected magnetic memory cell from being isolated from other unselected cells. To address this issue, the crossbar array architecture typically employs magnetic memory cells consisting of one selector and one MTJ (1S1R).

Figure 4 compares the magnetic memory cell structures for the 1T1R array architecture and the crossbar array architecture. Unlike the 1T1R magnetic memory cell, a word line (WL) is connected to MTJ and a bit line (BL) is connected to the selector. The selector and the MTJ device are physically stacked without additional area overhead, achieving a higher integration density. However, the characteristics of the selector in 1S1R is worse than that of the selection transistor in 1T1R, which deteriorates the degree of isolation and produces relatively smaller margins for sensing.

Figure 5 illustrates read operation in the crossbar array architecture. During read operation, the word line of the selected row is grounded while all the other signals are biased at V_R. All the cells in the selected row will have bias voltage of V_R. This leads to current flow from the bit lines to the selected word line. The current in each selected cell will have two different values depending upon the magnetization state of the cell. This current is sensed by a current sensing circuit. An alternative sensing method is to use voltage-mode sensing. Here, constant current is supplied to each bit line and the bit line voltage is generated after multiplying the current and

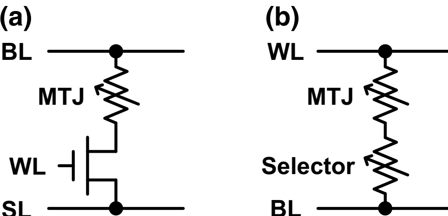


Fig. 4 Magnetic memory cell structures: **a** 1T1R and **b** 1S1R

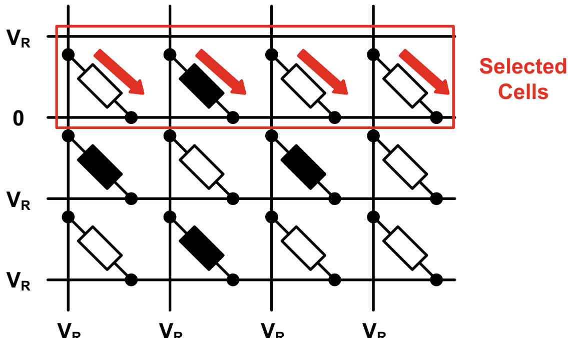


Fig. 5 Read operation in the crossbar array architecture

the resistance value of the cell. Two different cell resistance values will form two different voltage levels to be sensed by a voltage sense amplifier. Like SRAMs and DRAMs, all the cells in a selected row can be read out in parallel. This requires a sensing circuit in each bit line. Typically, the number of cells in a row is much larger than the required data width. Therefore, only a part of the read data is transferred to the output through multiplexers. All the other cells have no bias since both bit lines and word lines are biased at V_R. Therefore, no current flows through them.

In the typical crossbar array architecture, it is not possible to select one cell for write operation using the simple bias conditions from the 1T1R array architecture. Figure 6(a) is an example showing that multiple cells are selected for write operation when using only V_W and ground. Note that all the cells in the selected column are under the same bias condition. To address this issue, unselected word lines need to be biased at a different level. Sneak current is another significant concern in the crossbar array architecture. Various biasing schemes have been investigated to mitigate the aforementioned issues. Figure 6b introduces a write scheme utilizing $1/2V_W$ as the bias voltage for unselected signals. In this example, the selected word line is biased at V_W, the selected bit line is grounded, and all the other signals are bias at $1/2V_W$. The opposite current direction can be achieved by applying ground to the selected word line and V_W to the selected bit line. Regardless of the current direction, only the selected cell sees the voltage difference of V_W. The half-selected cells in the selected row see the voltage difference of $1/2V_W$ ($=V_W - 1/2V_W$), which is small enough to prevent unwanted write operation. Similarly, the half-selected cells in the selected row also see the voltage difference of $1/2V_W$ ($=1/2V_W - 0$). Note that this scheme is based upon the assumption that the magnetic memory cell resistance is not disturbed by $1/2V_W$. However, in the reliability point of view, it is more desirable to further lower the voltage difference seen by half-selected cells. Even though the unwanted write operation is eliminated, this scheme still suffers from sneak current caused by the half-selected cells. For reliable write operation, the required write current should be provided to the selected cell, which indicates that the actual amount of current provided by the word line driver (V_W) should be as large as the summation of the required write current and the sneak current. In addition, the amount of sneak current

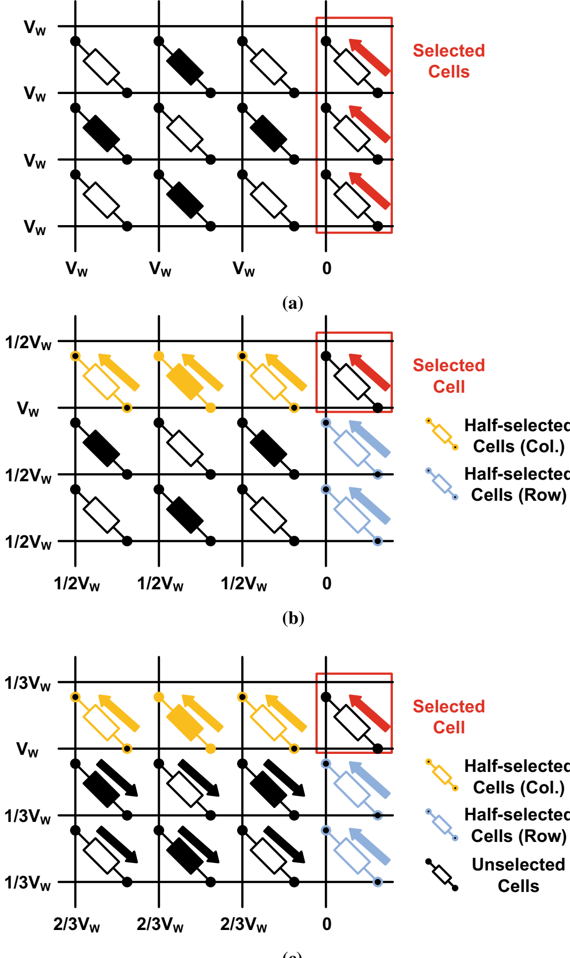


Fig. 6 Write operations in the crossbar array architecture: **a** conventional scheme, **b** 1/2V_W scheme, and **c** 1/3V_W scheme

is affected by the distribution of the anti-parallel state and the parallel state in the half-selected cells. In general, we need to design the word line driver in consideration of the worst case, which increases circuit area and power overheads. Figure 6c is another write scheme using different bias voltage levels in unselected word lines and bit lines, depending upon the write data. It uses the same bias condition as the one in Fig. 6b for the selected cell. The main difference comes from the bias voltage levels used in the un-selected rows and columns. In Fig. 6c, V_W, 1/3V_W, ground, and 2/3V_W are used in the selected row, the unselected rows, the selected column, and the unselected columns, respectively. Here, write current flows from the selected word line to the selected bit line. The opposite current direction can be implemented by applying ground, 2/3V_W, 1/3V_W, V_W, and 1/3V_W to the selected row, the unselected rows, the selected column, and the unselected columns, respectively. It can be observed that all

the cells in the unselected columns and the unselected rows see the voltage difference of $1/3V_W$. Therefore, it can be expected that the magnetic memory cells will have less disturbance in their resistance. However, all the cells except the selected cell see $1/3V_W$, generating more sneak current. Overall, this scheme is preferred when device reliability is a more serious concern than power consumption.

4 Design Challenges of Magnetic Memory

This section briefly summarizes general design challenges in magnetic memory. The issues include low TMR ratios, small sensing margins, large write current, IR drop in the interconnection, MTJ reliability, sneak current, etc. Many of the above issues are not independent but interrelated. One of the most critical issue in designing magnetic memory is the low TMR ratio. This is a similar to the on-current to off-current ratio of a transistor, conceptually. In general, the TMR ratio is <200% even though higher TMR ratios have been achieved in some recent research works. Other emerging memory devices such as PCRAM and RRAM usually show much higher on-to-off resistance ratios, which provides relatively large sensing margins. The low TMR ratio imposes more challenges in sensing circuit design. Write current is another significant issue that needs to be addressed in the magnetic memory design. As shown in Table 1, the write energy of magnetic memory is already much larger than that of SRAMs and DRAMs. The actual energy for write operation will become even higher when considering the sneak current in the crossbar architecture. IR drop in the interconnection also leads to addition power consumption since we have to supply higher voltage for biasing. In addition, the amount of IR drop is affected by the current, which is determined by the distribution of the anti-parallel state and the parallel state along the interconnection. MTJ reliability has been investigated and improved by properly setting the bias conditions. This is particularly critical

Table 1 Device characteristics of mainstream and emerging memory technologies [1]

	Mainstream memories				Emerging memories		
	SRAM	DRAM	NOR	NAND	MRAM	PCRAM	RRAM
Cell area	>100F^2	6F^2	10F^2	<4F^2 (3D)	6–50F^2	4–30F^2	4–12F^2
Multi Bit	1	1	2	3	1	2	2
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	3 V	3 V
Read time	~1 ns	~10 ns	~50 ns	~10 μs	<10 ns	<10 ns	<10 ns
Write time	~1 ns	~10 ns	10μs–1 ms	100μs–1 ms	<10 ns	~50 ns	<10 ns
Retention	NA	~64 ms	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6–1E12
Write energy (/bit)	~fJ	~100 fJ	~100 pJ	~10 fJ	~0.1 pJ	~10 pJ	~0.1 pJ

in the crossbar array architecture where multiple cells undergo certain bias voltage. The bias voltage is maintained below the threshold level so that no disturbance on the cell resistance occurs. However, the long term reliability of those magnetic memory cells need to be further investigated. Finally, sneak current is inevitable in the available magnetic memory array architectures. It affects, write current, IR drop, MTJ reliability, and overall power consumption. Two popular bias schemes are based upon the trade-off in the MTJ reliability and the power consumption.

5 Write Techniques for Magnetic Memory

Magnetic memory has been considered as a promising candidate for replacing SRAMs and DRAMs. However, one of the most critical issues is the large write current. In addition, variations in the characteristics of fabricated MTJ devices increase the write current even more to provide enough margins for write operation. It is important to control write current carefully for successful write operation without significant disturbance and power consumption. While simply increasing the size of the selection transistor is the easiest method for reducing write failures, it limits the scalability of the memory. In addition, the write pulse width also needs to be carefully selected to reduce power consumption without dramatic increase in write error probability. Figure 7 briefly shows the effects of write current and write pulse width on write error probability. At a given write current, the error probability decreases as the write pulse width increases. In addition, higher write current reduces the required write pulse width for the same error probability, which is desirable for high performance.

Other techniques for mitigating write failures are word line voltage boosting, write voltage boosting, body biasing, and additional external magnetic field. As explained in the comprehensive study in [2], the word line voltage boosting generally achieves the most reduction in the write power and the selection transistor size at iso-failure probability. However, many actual operating scenarios are not considered in [2]. Various design techniques have been reported to address these write-related issues in magnetic memory. This section introduces several state-of-the-art write techniques.

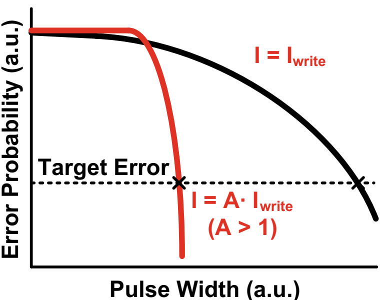


Fig. 7 Effects of write current and write pulse width on error probability

One of the most commonly used write technique is to control the strength of write drivers through current mirroring. A locally mirrored write driver is reported in [3]. Figure 8 depicts the simplified schematic of this write driver. When a bit line is selected for write operation through a multiplexer, it is connected to a shared write driver whose strength is controlled by a current source through current mirroring. The write timing and the write duration are controlled by "Pulse". When "Pulse" and "/Pre" are high, pull-down write current will be applied to the selected bit line for writing. The output node of the write driver is pre-charged to "VDD-Vtn" to prevent the potential current spike occurring at the beginning of the write operation. It is because the common node at the multiplexer output can have a voltage drop of "Vtn" due to the NMOS transistors in the multiplexer. By pre-charging it to "VDD-Vtn", current spikes can be avoided. In addition, it also improves write speed since the write current can flow in the selected bit line without delay. In this scheme, only unidirectional write current is implemented because of the cell structure. The pull-up write current can also be easily generated by mirroring the current source to a PMOS transistor in a similar way depicted in Fig. 8.

As discussed in the previous section, IR drop is a challenging issue in write operation due to the almost minimum interconnection width and high write current. This requires to supply even higher write current to compensate for the IR drop, which increases write driver devices and occupies significant chip area. One way to provide higher write current without increasing the device sizes and occupying large area is to use bootstrapped voltage. Figure 9 shows one example reported in [4]. Here, one thick oxide device is inserted between the access transistor and the column decoder. During pre-charge state, the decoder output ("Dec"), "Gate", and "Source" are grounded while "PreCh" is pre-charged to VDD. Once "Dec" becomes VDD, "Gate" rises through the thick oxide device. However, "Gate" cannot be VDD due to the voltage drop caused by the thick oxide NMOS device. After some delay,

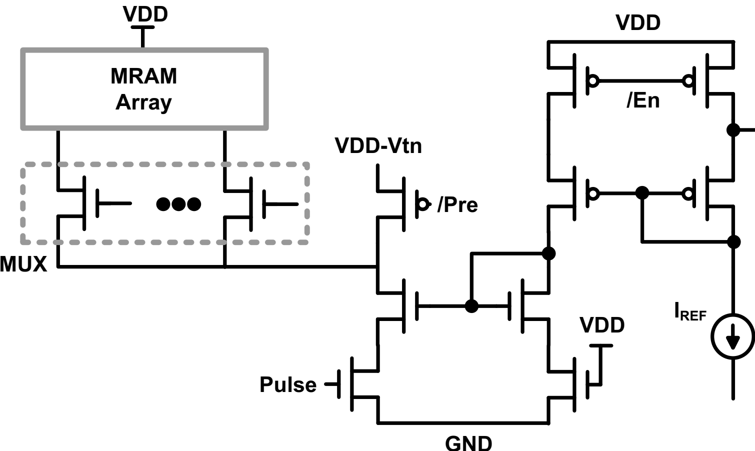


Fig. 8 Locally mirrored write driver scheme [3]

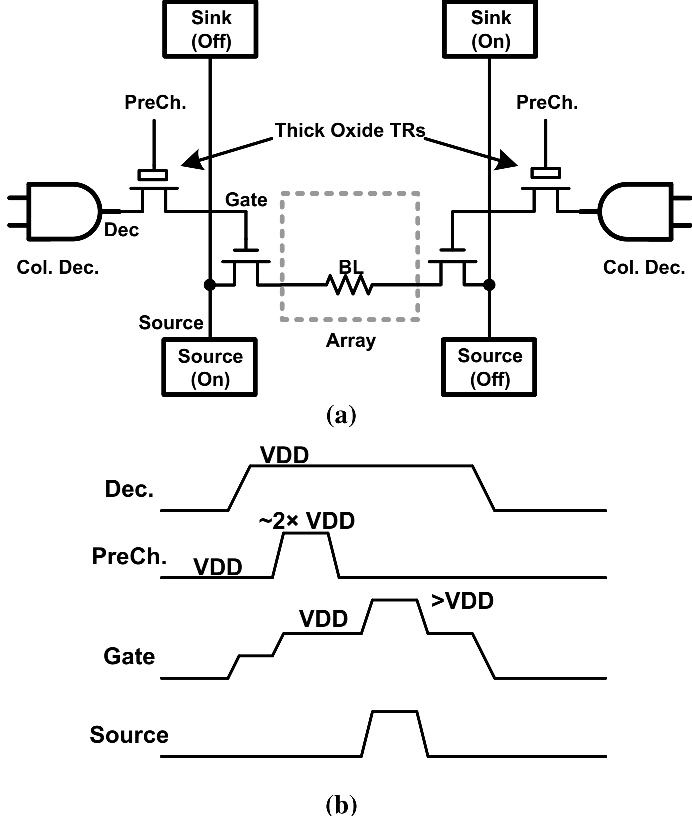


Fig. 9 Write driver using bootstrapped voltage: **a** schematic and **b** timing diagram [5]

"PreCh" is pulsed with ~2×VDD to raise "Gate" up to VDD. Finally, The "Source" node is raised, which raise "Gate" above VDD. The effective resistance of the write access transistor is reduced without creating any reliability issues. This technique is applied to a 16-Mb MRAM chip realizing the write current of 80 mA in 0.18 μm CMOS technology.

Programming magnetic memory is a stochastic process because of the random thermal fluctuations [5]. Therefore, it is challenging to remove write failures using a conventional fixed write current scheme without inducing significant power and area overheads. Probabilistic design techniques have been introduced to reduce write performance and write failures [5–7]. They include write-verify-rewrite with adaptive period (WRAP) in [5], verify-one-while-writing (VOW) in [5], variable energy write (VEW) in [6], and self-timed write operation in [7].

In the WRAP scheme, each write operation is followed by a read operation. Then, the read data is compared with the write data to decide whether to execute a rewrite operation. This iteration is repeated until the write operation is successful. Here, the

write pulse width can be adaptively changed to enhance the overall system performance. In [5], the optimal pulse width is based on Hamming weight. The optimal pulse width values for various Hamming weights at different temperatures are stored in a look-up table. The WRAP scheme demonstrated the write latency improvement of 40% and the energy saving of 26% when compared to the conventional fixed pulse write scheme. Figure 10 illustrates the system diagram of the WRAP scheme.

The VOW scheme is conceptually similar to the WRAP scheme. The main difference is that the VOW scheme executes write and verification operations in parallel. This technique is utilizing the write speed difference between "0" and "1". Since writing "0" is faster than writing "1", the VOW scheme verifies only writing "1". This is from the observation that the error probability of writing "0" is much lower than that of writing "1" with the same write pulse width. Figure 11 illustrates the VOW scheme. When writing "1" (WriteBit ⟨X⟩ = "1"), the corresponding done ⟨X⟩ becomes "1". If all the write data are successfully written, the completion flag signal

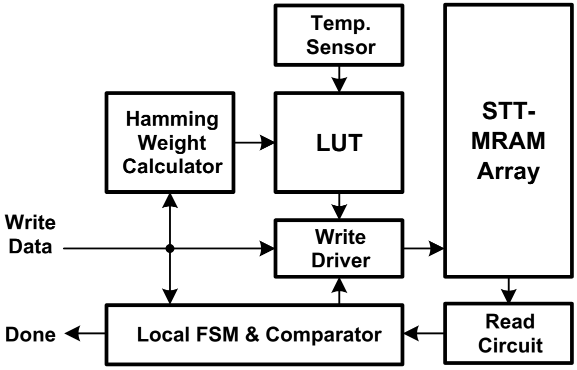


Fig. 10 System diagram of the WRAP scheme [5]

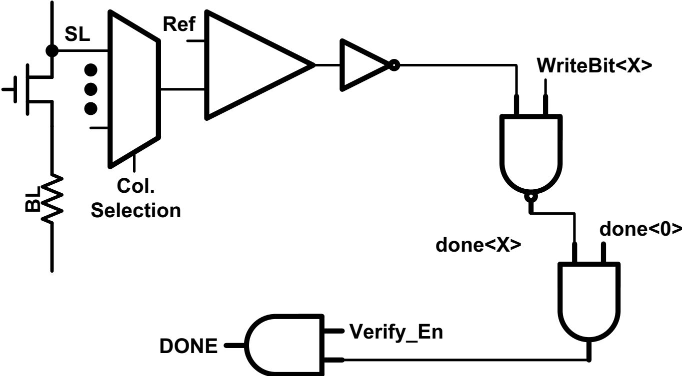
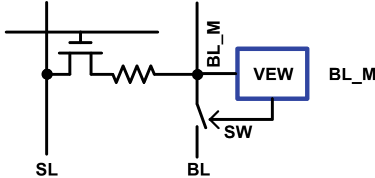


Fig. 11 Circuit diagram of the VOW scheme [5]

becomes high (i.e. DONE = "1") and the write operation ends. In [5], the VOW scheme achieved the write latency improvement of 52% and the energy saving of 29% compared to the conventional fixed pulse write scheme.

The VEW scheme is another probabilistic design technique for addressing the stochastic switching features of the STT MRAMs [6]. In this scheme, the write pulse ends when a successful write operation is achieved by monitoring the status of an STT MRAM cell during write operation. This scheme significantly reduces the write energy because of the reduced average write time. Figure 12 shows the circuit diagram of the VEW scheme [6]. During write operation, the BL and SL nodes are biased so that write current flows through the selected MTJ device. The voltage at BL_M can initially have two different levels depending on the state of the selected MTJ. Once the write current is large enough to change the resistance of the selected MTJ, the voltage at BL_M changes at a certain moment. The VEW circuit monitors this change and turns off the switch to remove the unnecessary current consumption. The VIEW circuit consists of two detection circuits, one for writing "1" and the other for writing "0" as shown in Fig. 12. When writing "1" (BL = "1" and SL = "0"), the voltage at BL_M changes abruptly, which is detected by a comparator (Comp.1) The flip-flops (FF) operate as delay elements for removing any possible feedback. The detected change makes SW "0" and cuts the current path from BL to BL_M. For writing "0" (BL = "0" and "SL = "1"), the voltage at BL_M changes in the opposite direction. This is detected by the second comparator



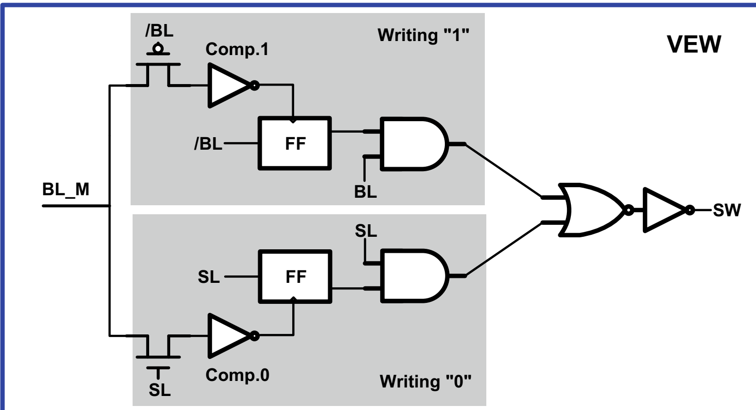


Fig. 12 Circuit diagram of the VEW scheme [6]

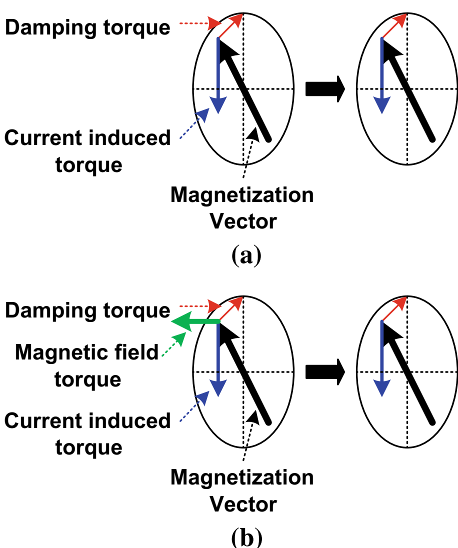


Fig. 13 Comparison of **a** typical STT and **b** field-assisted STT

(Comp.0), subsequently resetting SW to cut the current path. The VEW circuit needs to be implemented in each bit line, which is different from the VOW scheme. The VEW scheme is more efficient in saving write power at the cost of area overhead. In general, the VEW scheme provides more energy saving when the magnetic memory requires longer write time. If the write pulse is short, the energy saving can be easily nullified by the energy overhead from the VEW circuit itself. The self-time write operation in [7] is conceptually similar to the VOW and VEW schemes since it also detects write completion on the fly and immediately turns off current paths for energy saving.

Another write technique for magnetic memory is field assisted writing where additional magnetic field is applied to MTJ devices for easier switching with a larger spin-transfer torque [8]. Figure 13 depicts the switching concepts of the typical spin-torque transfer (STT) and the field-assisted STT. The additional magnetic field torque destabilizes the MTJ polarity toward the short axis, which reduces the switching latency [8]. Figure 14 shows the structure of the field-assisted STT MRAM cell. Similar works are also reported in [9–11]. In [9], an assisting magnetic field is utilized to set the MTJ devices to an initial reset state before writing operation. However, since each write operation requires to switch to a reset state, the write latency is doubled. In [10, 11], several structures and topologies of the field-assisted STT MRAM cells are described.

Write power reduction techniques are highly demanded since magnetic memory consumes much more power than other storage-based memories like SRAMs and DRAMs. One method of reducing power consumption is to selectively write data after reading and comparing [12]. If the read data is identical to the write data, no write operation is necessary. However, this scheme requires multiple cycles for write operation, which is not suitable for high performance.

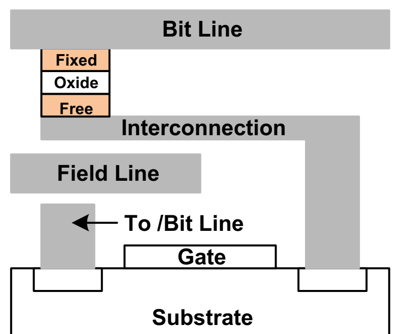


Fig. 14 Field-assisted STT MRAM cell structure

The production yield of magnetic memory is highly affected by large variations in the memory cells. One of the most critical yield limiting parameter is write current. For yield improvement, it is critical to characterize the memory cells and find proper write current, which is time consuming. In [13], an enhanced spiral search method is proposed to rapidly find required operating conditions. The enhanced spiral search method reduces the search time by 88–93% compared to the conventional scheme. This allows each chip to be configured with proper write current with the area overhead of 0.0254% in a 16-Mb MRAM test chip.

6 Sensing Techniques for Magnetic Memory

Small TMR ratios in magnetic memory impose a significant challenge in sensing data reliably. Various advanced sensing techniques have been developed. This section discusses various advanced sensing techniques.

Figure 15 illustrates the conventional sensing scheme in magnetic memory. A fixed amount of current is supplied to a selected memory cell through a switch controlled by a decoder. This will generate a voltage level at the bit line (BL), which is compared with a reference level by a sense amplifier. Since magnetic memory senses data through single bit lines, it is important to accurately generate reference in the middle of the anti-parallel state and the parallel state as illustrated in Fig. 16.

In [14], a midpoint reference generator is proposed by utilizing multiple MTJ devices that are connected in a series-parallel combination to generate a resistance value that is ½ (R_{high} + R_{low}). As depicted in Fig. 17, each V_{ref} generation unit consists of four MTJs and two selection transistors. The series connected MTJs gives R_{high} + R_{low} while the parallel connection sets the overall resistance ½ (R_{high} + R_{low}). Each V_{ref} generation unit is only shared by two rows, whose activation is controlled by word lines. In [14], the memory array has 1024 rows, which requires 512 V_{ref} generation units. A1-Mb MRAM in 0.6-μm CMOS technology demonstrated successful sensing operation at 3 V with a magnetoresistance ratio (MR) of ~45%.

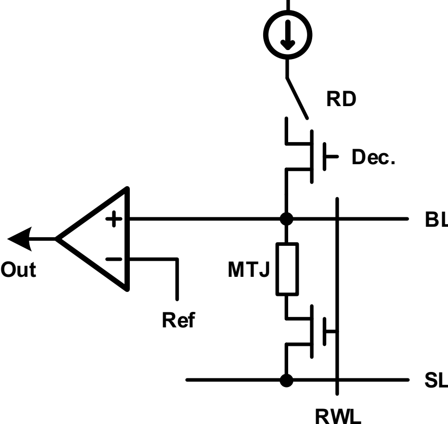


Fig. 15 Conventional sensing scheme

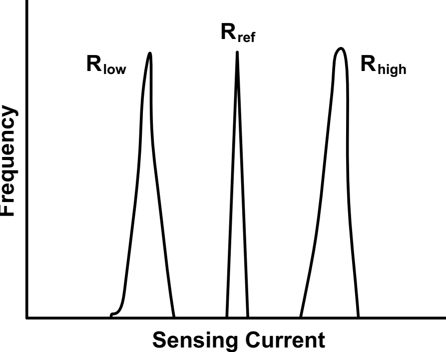


Fig. 16 Relationship between magnetic memory device resistance distribution and reference resistance

Jeong et al. proposed a sensing scheme that can generate reference using storage cells [15]. Since a storage cell itself generate its own reference voltage, it can remove sensing failures caused by variations in the tunneling oxide thickness. Figure 18 explains the operation of this self-reference generation scheme. In the first stage, current is applied to a selected MTJ device and converts the resistance to a voltage value. This voltage is stored in a capacitor. Depending upon the MTJ resistance, the stored voltage will have two different values. In the second stage, the selected MTJ is written to have the parallel state. After this, the MTJ resistance is again converted to voltage by applying current. Here, the current is slightly higher than the current used in the first stage. The converted voltage is also stored in a capacitor. Note that the stored voltage in the second stage is higher than that in the first stage if the original data is “0”. Similarly, the stored voltage in the second stage is lower than that in

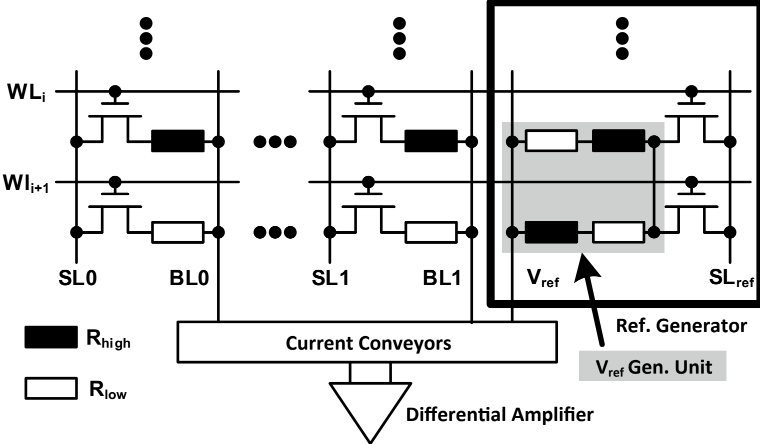


Fig. 17 Midpoint reference generator [14]

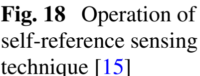
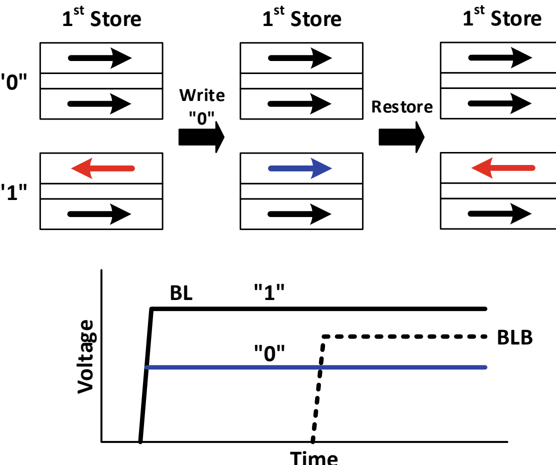


Fig. 18 Operation of self-reference sensing technique [15]



the first stage if the original data is “1”. Here, the voltage generated by the second stage is used as a reference. A sensing circuit will amplify the voltages stored in two capacitors. The voltage difference between the reference and the original data can be controlled by adjusting the amount of current used in the first stage and the second stage. The current in the second stage needs to be chosen carefully to have balanced sensing margins for both data “1” and “0”. In this scheme, the original data is deleted for generating reference voltage, which requires a restoring operation. The output of the sensing circuit should be used as write data for restoring the selected MTJ resistance. A similar self-referencing technique is also reported in [16]. Here, a noise shaping sense amplifier is proposed, reading data by a counter. Like the method

in [15], the selected cell is read first followed by write "1", read "1", write "0", and read "0". Comparing the counter values can tell the read data. After this, the selected cell is written with the original read data for restoring.

Then main limitation of the above self-reference sensing technique is requiring two write operations. Since write operations degrades the MTJ endurance significantly, an advanced sensing technique without destructing cell data is necessary. In [17], a nondestructive self-reference scheme is introduced. Figure 19 illustrates the schematic diagram of this nondestructive self-reference scheme. It looks similar to the above self-reference sensing technique [2] except that the negative node of the sense amplifier is connected to BL2 through a voltage divider using two resistors. The detailed operation of this scheme is as follows. During the first reading, a read current I_{R1} is applied to the selected memory cell to generate BL voltage. The BL voltage is stored in C1 through the NMOS transistor controlled by Sel1. Another read operation is executed using I_{R2}, which is larger than I_{R1}. This current will also generate BL voltage. The BL voltage is applied to the negative input of the sense amplifier through a voltage divider. The voltage at BL1 and the voltage generated by the voltage divider are compared by the sense amplifier and generate read data. If the voltage at BL1 is larger, the read data is "0" or vice versa. Here, the sensing margin for "1" and "0" are given as follows [17].

$$\Delta V_{BL,H} = I_{R1} \cdot (R_{H1} - R_{H2}) > 0 \tag{1}$$

$$\Delta V_{BL,L} = I_{R1} \cdot (R_{L1} - R_{L2}) \approx 0 \tag{2}$$

Here, R_{H1} and R_{H2} are the MTJ resistance values at high resistance state with I_{R1} and I_{R2}, respectively. Similarly, R_{L1} and R_{L2} are the MTJ resistance values at low resistance state with I_{R1} and I_{R2}, respectively. Here, I_{R1} and I_{R2} need to be carefully selected for reliable sensing. The detailed mathematical derivation can be found in

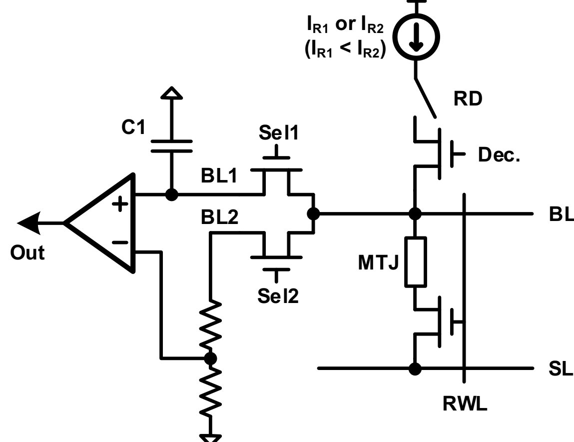


Fig. 19 Nondestructive self-reference sensing scheme [17]

[17]. Since this scheme has no write operation, the overall read latency and power consumption can be significantly improved compared to the previous self-reference scheme.

A balanced sensing architecture is preferred to achieve high-speed read. In [18], a symmetrical magnetic memory sensing architecture is introduced. In this scheme, two input nodes of the differential amplifiers have same amount of loading so that the read operation is not affected by the mismatches in the loading. Figure 20 shows the simplified architecture of this scheme. Reference voltage is generated by two reference bit lines, one programmed with "high" resistance and the other with "low" resistance. In the column selector block, two reference bit lines are merged to generate reference voltage and connected to two sense amplifiers. In this architecture two bit lines are accessed for read operation at the same time for balancing loading at the input of the sense amplifiers. One input of the sense amplifiers is connected to the selected bit lines while the other input of the sense amplifiers is connected to the merged reference bit lines. Since two reference bit lines are shared by two sense amplifiers, the effective loading is equal to one bit line.

Another sensing technique realizing balanced input capacitance to sense amplifiers is proposed in [3]. Here, the input capacitance of the sense amplifier is balanced through multiplexing and no balanced sensing architecture is necessary. Figure 21 illustrates the column multiplexing and the preamplifier of the technique in [3]. Like [18], two replica bit lines are employed to generate reference current. The two reference current components (I_H and I_L) are added and shared by two PMOS transistors (P2 and P3). Therefore, the current flowing in each P2 and P3 will be the average of I_H and I_L. The bias voltage generated by P2 is copied to P1, which also allows $(I_H + I_L)/2$ to flow through P1. The voltage at "OUTB" is used as reference voltage in the sense amplifier while the voltage at "OUT" is compared with "OUTB" by the second stage of the sense amplifier. Note that the voltage at "OUT" is formed by comparing $(I_H + I_L)/2$ with I_B. If I_B is higher than $(I_H + I_L)/2$, "OUT" will be lower than

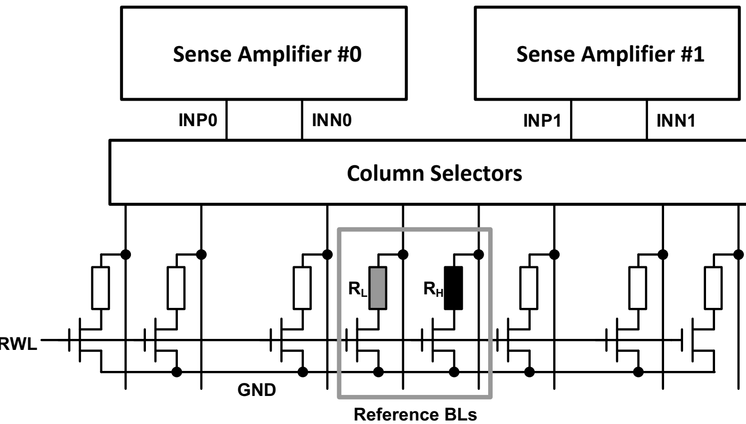


Fig. 20 Symmetrical sensing architecture [18]

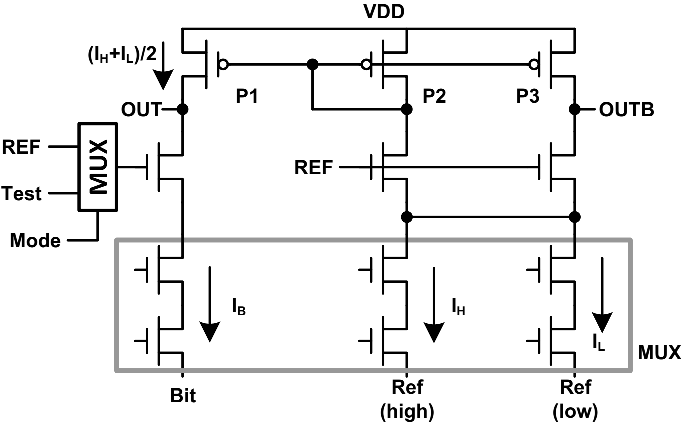


Fig. 21 3-input sensing for balanced capacitance [3]

"OUTB" and vice versa. In this sensing scheme, the capacitance values at "OUT" and "OUTB" are identical. Since "OUT" and "OUTB" are connected the second stage of the sense amplifier, it is obvious that the sense amplifier has no mismatch in the input capacitance. This scheme is also employed in the sensing scheme reported in [19, 20] with minor modifications such as adding degeneration PMOS devices and self-body-biasing. Another type of preamplifier is explained in [21]. Since the TMR ratio is not high in general, it is important to realize high sensitivity in the preamplifier. In [21], the cross-coupled NMOS load improves the sensitivity of the preamplifier to the small TMR ratio (e.g. 25%).

Offset voltage is another challenging issue in sensing. A dual-reference-voltage sensing scheme (DVSS) is introduced in [22] where an optimal reference level out of two reference levels is selected after reading data using them. Two reference voltage levels should be carefully selected so that at least one reference level can sense all data successfully. Figure 22 illustrates the concept of the DVSS. Ideally, any reference voltage (Ref+ or Ref−) can sense the read data. With positive offset voltage, only Ref+ can sense. Similarly, Ref− can sense data with negative offset voltage. However, this scheme requires accurate sense amplifiers since the reference levels

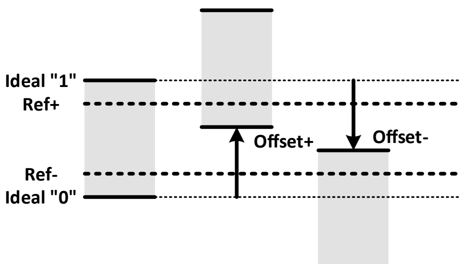


Fig. 22 Concept of dual-reference sensing scheme [22]

are relatively biased close to ideal "1" and "0" levels, which generates unbalanced sensing margins.

7 Summary

This chapter presents a brief overview of circuits design for magnetic memory. First, popular magnetic memory array architectures are explained. Two popular architectures are the 1T1R and crossbar architectures. While the crossbar architecture provides higher density, the 1T1R architecture is more prevalent in the magnetic memory. The relatively small TMR ratios of magnetic memory prohibit the proliferation of the crossbar architecture in magnetic memory. Magnetic memory has been considered as a promising next generation memory solution. However, compared to CMOS memories such as SRAMs and DRAMs, magnetic memories still need to overcome various challenges in circuits design. This chapter also briefly summarizes the general challenges in magnetic memory circuits design. After this, various state-of-the-art write techniques techniques are introduced. They include write driver design using a current source, bootstrapped word line voltage, field-assisted STT MRAM, and several probabilistic design techniques. In the probabilistic design techniques, this chapter explains the write-verify-rewrite with adaptive period (WRAP), verify-one-while-writing (VOW), and variable energy write (VEW) schemes. This chapter also discusses several sensing techniques such as midpoint reference generation, self-reference sensing, nondestructive self-reference scheme, symmetrical sensing architecture, dual-reference-voltage sensing scheme (DVSS), etc.

References

1. S. Yu, P.-Y. Chen, Emerging memory technologies. IEEE Solid-State Circuits Mag. **8**(2), 43–56 (2016)
2. X. Fong, Y. Kim, S.H. Choday, K. Roy, Failure mitigation techniques for 1T-1MTJ spin-transfer torque MRAM Bit-cells. IEEE Trans. VLSI Syst. **22**(2), 384–395 (2012) (Write techs. study)
3. T.W. Andre, J.J. Nahas, C.K. Subramanian, B.J. Garni, H.S. Lin, A. Omair, W.L. Martino, A 4-Mb 0.18-μm 1T1MTJ Toggle MRAM with balanced three input sensing scheme and locally mirrored unidirectional write drivers. IEEE J. Solid-State Circuits **40**(1), 301–309, (2005) (3-input sensing and unidirectional write drivers)
4. D. Gogl, C. Arndt, J.C. Barwin, A. Bette, J. DeBrosse, E. Gow, H. Hoenigschmid, S. Lammers, M. Lamorey, Y. Lu, T. Maffitt, K. Maloney, W. Obermaier, A. Sturm, H. Viehmann, D. Willmott, M. Wood, W.J. Gallagher, G. Mueller, A.R. Sitaram, A 16-Mb MRAM featuring bootstrapped. IEEE J. Solid-State Circuits **40**(4), 902–908 (2005) (Boostrapped write driver)
5. X. Bi, Z. Sun, H. Li, W. Wu, Probabilistic design methodology to improve run-time stability and performance of STT-RAM caches, in Proceedings of IEEE/ACM ICCAD, pp. 88–94 (2012) (VOW)
6. J. Park, T. Zheng, M. Erez, M. Orshansky, Variation-tolerant write completion circuit for variable-energy write STT-RAM architecture. IEEE Trans. VLSI Syst. **24**(4), 1351–1360 (2016) (write completion, VEW)

7. R. Bishnoi, F. Oboril, M. Ebrahimi, M.B. Tahoori, Self-timed read and write operations in STT-MRAM. IEEE Trans. VLSI Syst. **24**(5), 1783–1793 (2016) (Self-timed write)
8. R. Patel, X. Guo, Q. Guo, E. Ipek, E.G. Friedman, Reducing switching latency and energy in STT-MRAM caches with field-assisted writing. IEEE Trans. VLSI Syst. **24**(1), 129–138 (2016) (Field-assisted writing)
9. T. Andre, S. Tehrani, J. Slaughter, N. Rizzo, Structures and methods for a field-reset spin-torque MRAM, U.S. Patent 8 228 715, Jul. 24, 2012
10. Y. Ding, Method and system for using a pulsed field to assist spin transfer induced switching of magnetic memory elements, U.S. Patent 7 502 249, Mar. 10, 2009
11. X. Wang et al., Magnetic field assisted STRAM cells, U.S. Patent 8 400 825, Mar. 19, 2013
12. J.J. Nahas, T.W. Andre, B. Garni, C. Subramanian, H. Lin, S.M. Alam, K. Papworth, W.L. Martino, A 180 Kbit Embeddable MRAM Memory Module. IEEE J. Solid-State Circuits **43**(8), 1826–1834 (2007) (selective write)
13. C.-Y. Chen, S.-H. Wang, C.-W. Wu, Write current self-configuration scheme for MRAM yield improvement. IEEE Trans. VLSI Syst. **21**(1), 1260–1270 (2012) (Write current self-config.)
14. M. Durlam, P.J. Naji, A. Omair, M. De Herrera, J. Calder, J.M. Slaughter, B.N. Engel, N.D. Rizzo, G. Grynkewich, B. Butcher, C. Tracy, K. Smith, K.W. Kyler, J. Jack Ren, J.A. Molla, W.A. Feil, R.G. Williams, S. Tehrani, A 1-Mbit MRAM based on 1T1MTJ Bit cell integrated with copper interconnects. IEEE J. Solid-State Circuits **38**(5), 769–773 (2003) (Midpoint ref gen)
15. G. Jeong, W. Cho, S. Ahn, H. Jeong, G. Koh, Y. Hwang, K. Kim, A 0.24-μm 2.0-V 1T1MTJ 16-kb nonvolatile magnetoresistance RAM with self-reference sensing scheme. IEEE J. Solid-State Circuits **38**(11), 1906–1910 (2003) (Self-reference)
16. M.B. Leslie, R.J. Baker, Noise-shaping sense amplifier for MRAM cross-point arrays. IEEE J. Solid-State Circuits **41**(3), 699–704 (2006) (noise shaping and self-referencing)
17. Y. Chen, H. Li, X. Wang, W. Zhu, W. Xu, T. Zhang, A 130 nm 1.2 V/3.3 V 16 Kb spin-transfer torque random access memory with nondestructive self-reference sensing scheme. IEEE J. Solid-State Circuits **47**(2), 560–573 (2012) (Self-reference)
18. J. De Brosse, D. Gogl, A. Bette, H. Hoenigschmid, R. Robertazzi, C. Arndt, D. Braun, D. Casarotto, R. Havreluk, S. Lammers, W. Obermaier, W.R. Reohr, H. Viehmann, W.J. Gallagher, G. Müller, A high-speed 128-kb MRAM core for future universal memory applications. IEEE J. Solid-State Circuits **39**(4), 678–683 (2004) (Symmetrical archi)
19. J. Kim, K. Ryu, S.H. Kang, S.-O. Jung, A novel sensing circuit for deep submicron spin transfer torque MRAM (STT-MRAM). IEEE Trans. VLSI Syst. **20**(1), 181–186 (2012) (Iref-sensing)
20. J. Kim, K. Ryu, J.P. Kim, S.H. Kang, S.-O. Jung, STT-MRAM sensing circuit with self-body biasing in deep submicron technologies. IEEE Trans. VLSI Syst. **22**(7), 1630–1634 (2014). (sensing with body bias)
21. T. Sugibayashi, N. Sakimura, T. Honda, K. Nagahara, K. Tsuji, H. Numata, S. Miura, K. Shimura, Y. Kato, S. Saito, Y. Fukumoto, H. Honjo, T. Suzuki, K. Suemitsu, T. Mukai, K. Mori, R. Nebashi, S. Fukami, N. Ohshima, H. Hada, N. Ishiwata, N. Kasai, S. Tahara, A 16-Mb Toggle MRAM with burst modes. IEEE J. Solid-State Circuits **42**(11), 2378–2385 (2007) (preamp)
22. T. Na, J. Kim, B. Song, J.P. Kim, S.H. Kang, S.-O. Jung, An offset-tolerant dual-reference-voltage sensing scheme for deep submicrometer STT-RAM. IEEE Trans. VLSI Syst. **24**(4), 1361–1370 (2016) (Dual ref volt sensing)

Domain Wall Programmable Magnetic Logic

Sarjoosing Goolaup, Chandrasekhar Murapaka, and Wen Siang Lew

Abstract Traditional micro-electronics work by controlling the flow of electron charge through transistor switches. Spintronics, which exploits the spin degree of freedom of electron, can lead to devices that outstrip the performance of traditional semiconductor technology. The spin moment in magnetic nanostructures by virtue of their inherent non-volatility can potentially offer the opportunity of ultra-low power and high speed devices. This chapter describes proof-of-concept devices that utilize magnetic domain walls in nanowire network to perform logical operations. A magnetic reconfigurable logic device is discussed, whose operation is mediated via domain wall motion. The logical operation can be programmed at run-time and the whole booelan logic family, from simple NOT to complex operations such as NAND can be obtained. The second part of the chapter explores the feasibility of encoding logical bit within the internal state of domain wall, leading to mobile data bits. Simple single to double bits logical operations are demonstrated.

1 Introduction

Spintronics, which combines the spin degree of freedom with the charge transport characteristics of electrons, has the potential to change the basic principles of logic operations in microelectronics devices [1, 2]. By exploiting the binary state of electron spin in nanomagnets for logic operation, an all-magnetic logic computing architecture is possible. This approach has the advantages of non-volatility, zero

S. Goolaup · C. Murapaka · W. S. Lew (✉)
Division of Physics and Applied Physics, Nanyang Technological University, Singapore, Singapore
e-mail: WenSiang@ntu.edu.sg

S. Goolaup
School of Engineering and Innovation, The Open University, Milton Keynes, UK

C. Murapaka
Department of Materials Science and Metallurgical Engineering, Indian Institute of Technology Hyderabad, Kandi 502285, Telangana, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_7

quiescent power, high scalability and high speed. The manipulation of the magnetization direction of nanomagnets, via magnetic field [3–8] or spin transfer torque effect [9], has spawned numerous techniques for magnetic logic [3–12]. In these logic schemes, different physical designs are needed for full logic functionalities. However, a logic which can be reconfigurable at run-time shall simplify the device architecture by eliminating interconnects and can potentially increase the computing speed. The reconfigurability makes the device attractive as a single device can then be programmed for many applications [13]. Though conceptual proposals [14, 15] of programmable spin logic exist, experimental demonstration of an all-magnetic reconfigurable logic is still lacking. Semiconductor based reconfigurable logic schemes have been demonstrated combining the magnetism and semiconductor technologies [16]. The logic operations in the device are performed by controlling the motion of charge carriers in p-n junction diodes by the application of a magnetic field. Simple circuit combinations of diode structures are used to perform various logic operations.

The propagation of magnetic domain wall (DW) in magnetic nanowires has been proposed towards making high-density magnetic memories [17], spin logic devices [3, 7] and shift registers [18, 19]. The approach to domain wall magnetic logic has been to drive DW to switch the binary state of nanostructures. The success of these technologies will rely inevitably on the perfect control and understanding of magnetic DWs in nanowires. This has proven to be a challenging task, and much research has been devoted to understanding the DW dynamics in magnetic nanowires [20–25]. Changes in the DW structure are significant in device applications where DW motion is controlled via interaction with artificial defects, as the detailed spin distribution in the wall affects the nature and strength of the pinning potential [25–27].

The topological defects of Transverse DWs (TDW) are of paramount importance as regards to the deterministic pinning and movement of DW within complex networks of conduits. The fidelity of the data transmission may also depend on preserving the DW structure. Although much progress has been made, to date there has been a limited number of DW-based devices for industrial applications. In situ control of the DW topological defects in nanowire conduits may pave the way for novel DW logic applications [28].

2 Domain Wall Reconfigurable Logic

A DW based reconfigurable logic is proposed and demonstrated where-in controlled motion of a domain wall in a magnetic network structure results in binary logical operations [29]. The device is capable of performing all basic logic functionalities on a single structure. The selection of a particular logic functionality is achieved by using an in situ local Oersted field via a magnetic gate. The details of selective motion of domain wall in network structures and the concept of magnetic charge associated with DW are further elaborated in the chapter.

2.1 Domain Wall Selective Switching

In thin film narrow magnetic nanowires, the TDW is the stable configuration. TDWs are characterized by a magnetostatic charge; positive for head-to head (HH) DWs where the magnetizations are pointing towards each other and negative for tail to tail (TT) DWs when magnetizations are pointing away. In wider and thicker nanostructures, vortex DWs (VDW) are stable configurations and are characterized by the spins curling around a vortex core. In narrow and thin nanostructures, TDWs are stabilized, with spins converging or diverging along a transverse orientation orthogonal to the nanowire magnetization. Both VDW and TDW intrinsically possess a chirality which determine the orientation of the spins as the magnetization changes from one domain to another. The directionality (chirality) of the spin rotation can be either clockwise (CW) or anti-clockwise (ACW). The transverse component of the TDW gives the sense of rotation of the spins within the wall, leading the TDW chirality. The chirality of the wall can be defined as either "UP" or "DOWN", reflecting the orientation of the transverse component of the wall. The spins within the TDW (along y-direction) with 90° and 270° rotation with respect to the nanowire magnetization (along $+x$), are referred to as "UP" (U) and "DOWN" (D) chirality, respectively. Another approach to describing TDW is to view it as a composite object of elementary topological defects [30]. The edge defects have half-integer winding numbers, either $+½$ or $-½$, representing the spin configuration of the DW at the edges of the nanowire, as illustrated in Fig. 1.

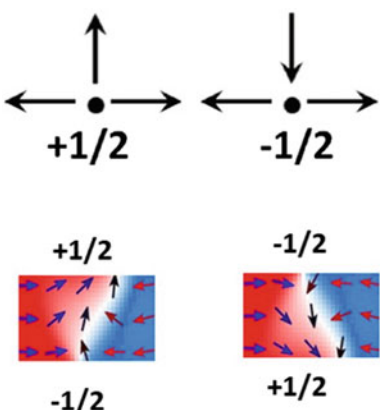


Fig. 1 Schematics to represent the winding numbers for the edge defects of transverse

2.2 Chirality Dependent Transverse Domain Wall Selective Switching

Transverse DW trajectory in a bifurcated nanowire based on its chirality is key to the development of the proposed magnetic network logic device. The field driven magnetization reversal process in a network structure was first investigated using OOMMF micromagnetic simulations [31]. The width and thickness of the bifurcated nanowire are chosen to be 100 and 10 nm, respectively to ensure TDWs are the only stable configurations. The branches of the bifurcated nanowire deviate at an angle $\phi = 70°$ from the nanowire long axis (x-axis). Material parameters for Permalloy, $Ni_{81}Fe_{19,}$ were considered for the simulation.

Figure 2 depicts the schematic of the magnetization configuration of the network structure as a TT TDW with an "UP" chirality (TT-U TDW) propagates along the nanowire conduit. A magnetic field is applied to drive the DW through the network structure and that selectively switches the magnetization of the lower branch as shown in Fig. 2. For reliable performance of the logic structure, a complete understanding of the DW dynamic behavior at the bifurcation is necessary. The spin state evolution of the DW at the bifurcation is extracted to better understand the DW selective movement along the lower branch structure. The corresponding spin states, as obtained from

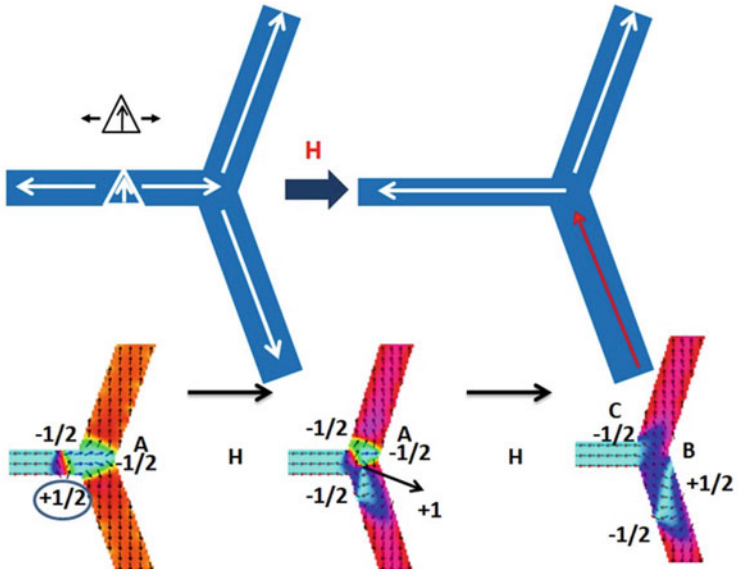


Fig. 2 Schematic of a dual branch structure with a tail to tail DW with up chirality before and after the application of magnetic field to drive the DW. Bottom insets show snapshots of the magnetization evolution at the bifurcation of dual branch

micromagnetic simulation, are shown in Fig. 2. The DW motion from the bifurcation to the lower branch is completed via a complex DW transformation.

Besides chirality, the DW can also be represented as the composite of topological edge defects [30]. For instance, a vortex wall is a combination of two −½ edge defects (vertices) and a +1 bulk defect (core). A transverse DW is a composite of two half integer edge defects with their positions dependent on its chirality. The magnetization switching in the bifurcated nanowire is governed by the conservation of these topological edge defects. Schematic representations of the elementary defects with +½ and −½ winding numbers are shown in Fig. 1. A defect in which all the spins are diverging is assigned a +½ winding number. Defect with two spins diverging and one spin converging is assigned with −½ winding number. A TT-U TDW has a, −½ ~ +½, winding number at the top and bottom edge of the DW respectively as seen in Fig. 1. At the bifurcation, there is vertex, which is characterized by a −1/2 winding number. When the DW reaches the bifurcation, it gets pinned before interacting with the vertex state. The TTU TDW has a CW spin orientation. When the DW reaches the bifurcation, the spins adopt a CW orientation which in turn pushes the vertex core towards the upper branch. The DW has higher energy at the +½ edge defect as compared to the −½ edge defect due to the transverse variation of the DW width [32]. With the increase of magnetic field strength, the bottom part of the DW depins and collides with the vertex at the bifurcation. The collision between the TDW and the edge defect at the vertex leads to the formation of a VDW with a CW orientation. Conservation of topological charge dictates that the total winding numbers of edge defects should be conserved during DW interaction or transformation [30]. The VDW is characterized by a +1 bulk defect at the core and two −½ defects at the edges. In this system, the transformation of the +½ defect from the TT-U TDW to a +1 defect leads to formation of the vortex configuration. To conserve the total winding number of the system, a −½ defect is nucleated along the same edge of the transformed +½ defect, as shown in the bottom inset of Fig. 2. Further increasing the magnetic field strength causes the core of the VDW to move towards point A, where the core eventually annihilates. To maintain the total topological charge at the bifurcation, the annihilation of the vortex core (+1 defect) leaves behind −½ edge defect from the TT-U TDW at the bifurcation and a new TDW is nucleated within the lower branch as shown in Fig. 2. The motion of the TDW through the lower branch changes the magnetization orientation of the branch. The DW motion through the network structure displaces the position of the edge defect from position A to B. It shows that the total topological winding number is always −½ before and after DW motion through the dual branch structure.

To investigate the effect of the DW chirality on the reversal process, a TT TDW with "DOWN" chirality (TT-D TDW) is relaxed in the nanowire as shown in Fig. 3. The magnetization switching of the TT-D TDW follows the same process as the TT-U TDW. However, when the DW is driven through the Y-shaped structure, the TT-D selectively travels through the upper branch and switches its magnetization, as shown in Fig. 3. A TT-D is composed of a +½ defect at the top edge and a −½ defect at the bottom edge, as shown in Fig. 2. Similar TDW pinning at the bifurcation occurs when a magnetic field is applied. The spins in the TT-D TDW always rotates

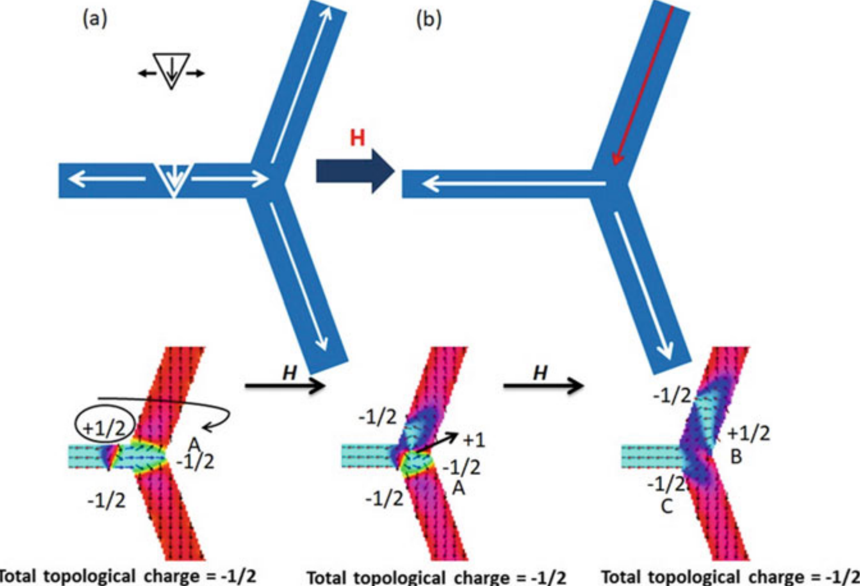


Fig. 3 Schematic of a dual branch structure with a tail to tail DW with down chirality before and after application of magnetic field to drive the DW. Bottom insets show snapshots of the magnetization evolution at the bifurcation of dual branch

in anti-clockwise (ACW) orientation. The spin relaxation at the bifurcation causes the edge defect at the vertex to be displaced towards the lower branch. The TDW depinning leads to the formation of an ACW VDW. The DW transformation occurs via the annihilation of the +½ defect and generation of a −½ defect in the upper branch. An increase in the magnetic field strength leads to the annihilation of the VDW, leaving behind the −½ defect at the vertex position C and nucleation of the TDW in the upper branch as shown in in the bottom inset of Fig. 3. The motion of this TDW towards the end of the upper branch switches the magnetization. The switching process displaces the position of the edge defect from A to C. Thus, the arrangement of edge defects in the TDWs affects the reversal process of the branch structure. The DW always moves towards the branch at which +½ defect (of the TDW) is facing. This implies that depending on the chirality of the injected DW, the magnetization in the branch structure can be selectively switched.

To validate the simulation results, we have experimentally investigated the DW dynamic at the bifurcation in a Y-shaped magnetic network structure. A scanning electron microscopy image of the fabricated network structure is shown in Fig. 4. The structure is a thin film stack of Ta(5 nm)/$Ni_{81}Fe_{19}$(10 nm)/Ta(5 nm). For DW injection and chirality selection mechanism, a diamond-shaped NiFe nucleation pad of area 2 μm × 2 μm, together with a transverse nanowire was attached to the left end of the nanowire.

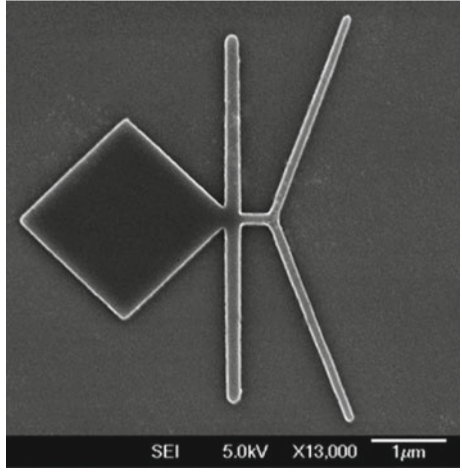


Fig. 4 Scanning electron microscopy image of the patterned dual branch structure

The transverse nanowire, which is 4 μm long and 100 nm wide, acts as a chirality selector which is positioned between the nucleation pad and the nanowire. In the presence of a transverse nanowire, the DW has been shown to preserve its spin structure beyond the walker breakdown field. The DW fidelity length, the region within which the DW retains its internal spin configuration, in an external field of 50 Oe, has been shown to be close to 500 nm [33]. As such, the length of the longitudinal nanowire in our structure is chosen to be 400 nm. The chirality of the injected DW can be set by fixing the magnetization orientation of the transverse nanowire [34].

Direct observation of the selective switching was achieved via magnetic force microscopy (MFM) imaging. The chirality selector is first magnetized with a 1 kOe field along the y direction, to set the magnetization direction of the selector. An injected DW can acquire an "UP" or "DOWN" chirality, via the magnetization orientation ($+y$ or $-y$) of the selector. To inject a TT TDW into the nanowire conduit a magnetic field is applied along the $-x$-direction. Under increasing fields, a DW is nucleated in the pad and subsequently moves into the nanowire conduit after passing the chirality selector. The magnetization direction of the chirality selector remains unchanged during the field application along the x-direction due to the strong shape induced anisotropy along the y direction.

Figure 5a shows the MFM image of the network structure when the selector and the nanowire are saturated with the 1 kOe field along the $+y$ and $+x$-directions, respectively. The dark magnetic contrast at the top end point of the selector, as well as at each end point of the two output branches, indicate that the magnetizations are aligned along the $+y$ and $+x$ directions, respectively. When a 50 Oe field is applied along the $-x$ direction, a TT TDW with "UP" chirality (TT-U TDW) is injected from the pad into the nanowire conduit after passing through the selector. The TT-U TDW propagates to the bifurcation of the network structure. The DW stops at the bifurcation due to pinning. When the field strength is increased to 100 Oe, the DW overcomes the pinning and moves into one of the two branches. MFM

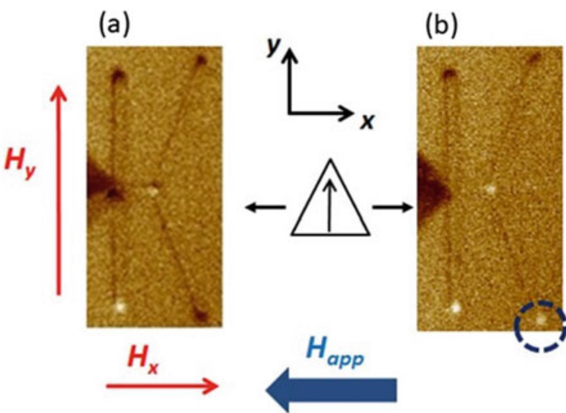


Fig. 5 **a** MFM image of the initial configuration of the dual branch structure when the nanowire and the chirality selector are saturated in +x and +y-directions, respectively. **b** MFM image of the final configuration when the TT DW with up chirality is driven through it

image in Fig. 5b shows the final magnetic configuration. The magnetic contrast at the end point of the lower branch changes to bright contrast whereas that at the end point of the upper branch remains as dark contrast. This result implies that only the magnetization of the lower branch has changed direction. Next, to verify the effect of the magnetization direction of the selector on the branch switching, the selector magnetization was reversed to the $-y$ direction.

Figure 6a shows the initial magnetization configuration of the network structure after the similar $+x$ saturation field application. The MFM image reveals that the top end point of the chirality selector shows a bright contrast, confirming the magnetization direction switch; while each end point of the two branches show dark contrast. Similarly, a DW injection field of 100 Oe was applied along the $-x$-direction to nucleate and drive the TDW, in this case a TT-D was driven along the nanowire conduit before pinning at the bifurcation. The final magnetic configuration is shown in the MFM image in Figure 6b. As opposed to the previous scenario, the magnetic contrast at the lower branch remains the same whereas that of the upper branch

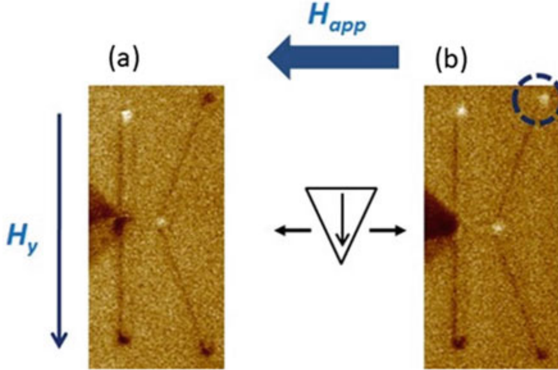


Fig. 6 **a** MFM image of the initial configuration of the dual branch structure when the the chirality selector is saturated along −y-direction. **b** MFM image of the final configuration when the TT DW with down chirality is driven through it

changed from dark to bright contrast, indicating a switch of the magnetization direction. This result clearly demonstrates that selective switching at either branch of the network structure can be obtained by controlling the chirality of the transverse DW.

2.3 Chirality Dependant Vortex Domain Wall Selective Switching

Similar investigations have been performed earlier by Pushp et al. [35] on VDW motion at a bifurcation to understand the effect of chirality on the trajectory of the DW. VDWs of opposite chirality moves along different branches giving rise to the sorting effect similar to TDWs. Here, we discuss briefly the chirality dependant VDW motion in magnetic branch structure. To ensure VDWs are stabilized, the wire width was chosen to be 300 nm while the $Ni_{81}Fe_{19}$ film thickness was 20 nm. An alternative method of DW injection was used. By generating a local Oersted field around a metallic strip placed orthogonal to the nanowire conduit, DW can be nucleated on demand. To generate the DW with a particular chirality, a notch was patterned beneath the strip line at the bottom edge of the nanowire. The notch presets the magnetization rotation either in CW or ACW orientation depending on the initial magnetization direction of the nanowire. For instance, when the nanowire is saturated along $-x$-direction, the magnetic moments curl around the notch in ACW fashion. When a voltage pulse is applied across the strip line, magnetization reversal occurs in the region beneath the injection line. In this particular case, two VDWs, HH and TT VDWs with ACW orientation are formed due to the influence of the initial magnetization around the notch. However, when the strength of the injection pulse is increased, it causes the oscillatory buckling of the magnetization along the nanowire, alternating between the top and bottom edges with a periodicity of 2 times the nanowire width. As the chirality of the VDW created at the bottom edge is always ACW, the VDW created at the top edge is of CW orientation. Thus by carefully controlling the strength of injection pulse one can generate a CW VDW that is created at the top edge of the nanowire. In short, the strength of the injection pulse can determine the chirality of the injected VDWs.

To demonstrate the chirality dependant VDW motion in branch structure, Y-shaped nanostructure with each branch deviating 30° from the nanowire are patterned, as shown in Fig. 7. When the HH VDW with an ACW chirality is injected and driven with a magnetic field of 75 Oe, the VDW selectively moves towards the lower branch, switching its magnetization, while the magnetization orientation of upper branch remains unaffected as shown in Fig. 8a. In contrast, when a HH VDW with CW chirality is injected (Fig. 8b), the DW selectively moves towards the upper branch. In addition to the chirality, VDW is also characterized by its polarity of the core which can be pointing in one of the two directions perpendicular plane of the nanowire magnetization. The polarity of the vortex core does also affect the VDW motion at the bifurcation. Certain combination of the vortex chirality and polarity needs less

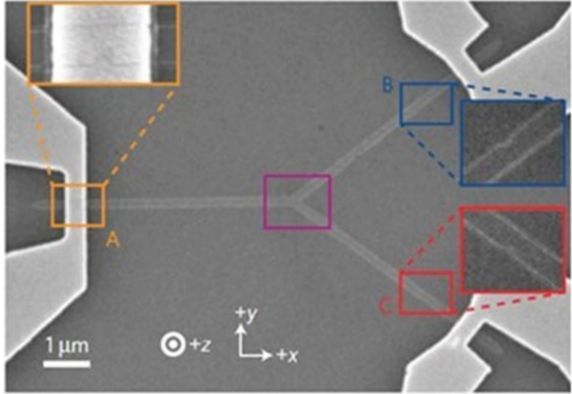


Fig. 7 Scanning electron micrograph of the dual branch structure along with contact pads to inject and drive the vortex DWs

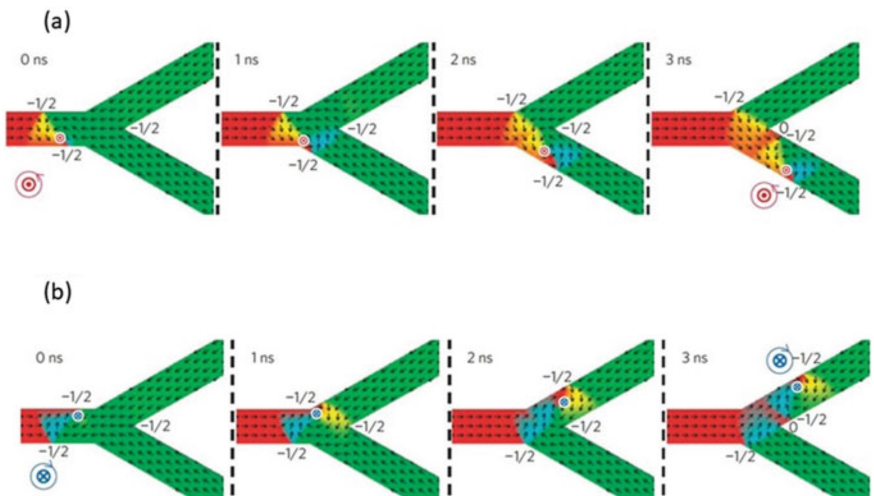


Fig. 8 Magnetization configurations acquired from micromagnetic simulations. **a** when a vortex DW with ACW chirality is injected and driven through branch structure, **b** when a vortex DW with CW chirality is injected and driven through branch structure

driving field to move beyond the bifurcation and the magnetization reversal is faster. However, when the combination of vortex chirality and core polarity is not suited, the polarity has to switch first before the DW moves further into one of the branches determined by its chirality making the switching process slow and requiring higher field. To summarize, the VDW follows selective switching depending on its chirality irrespective of the polarity of the vortex core.

The chirality dependant selective switching in in-plane nanowires forms the basis for logic functionality. Phung et al., demonstrated simple two-bit biplexer or demultiplexer that sorts the VDWs into one of the two branches of the Y-shaped branch

structure depending on its initial chirality [36]. This method is found to be robust that sorting mechanism is insensitive to the angle between the two branches to the strong topological protection coming from the topological edge defects constituted within the VDW. However, this structure sorts the VDWs based on their chirality at the bifurcation only.

If the VDW chirality is not stable and switches during its motion towards the bifurcation, the VDW sorter may direct the VDW towards a wrong branch irrespective of its initial chirality. The switching of the chirality during the DW driving can happen due to the continuous precession of the DW structure when the driving field is sufficiently higher due to Walker breakdown phenomenon.

To test the repeatability of the selective switching phenomenon, we have fabricated 16 branch structures close to each other. Figure 9a shows schematic of a branch structure, together with the corresponding SEM image, chosen to verify the chirality dependent DW trajectory. For a head-to-head (HH) DW, in which the spins point towards each other, a CW chirality VDW is expected to propagate along the upper branch whereas an ACW chirality VDW moves along the lower branch. The branch structure comprises of a circular pad of diameter 2 μm, a longitudinal nanowire of width 300 nm, and a 'U' shaped branch of width 300 nm. The circular pad is used for nucleating and injecting a DW into the longitudinal nanowire. A 6-μm-long and 100-nm-wide chirality selector is positioned to assign chirality to the injected DW at a distance of 1 μm from the output branch [37].

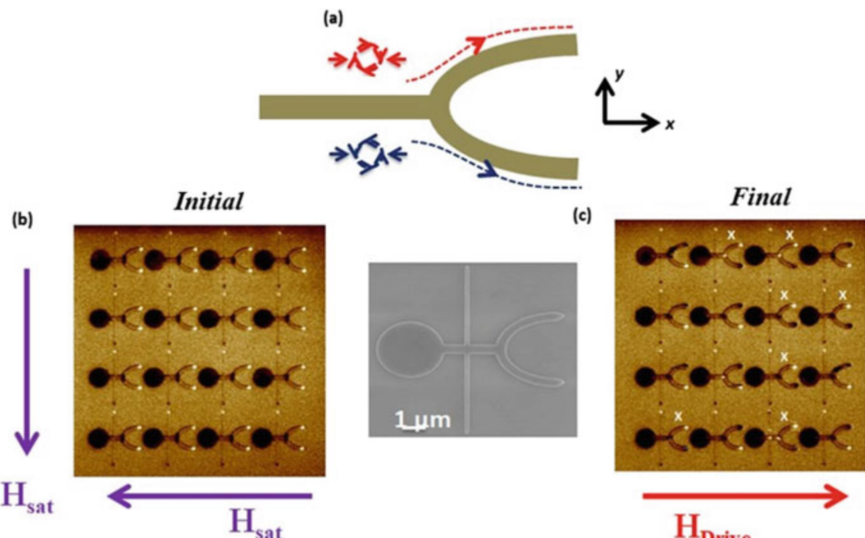


Fig. 9 Magnetization configurations acquired from micromagnetic simulations. **a** when a vortex DW with ACW chirality is injected and driven through branch structure, **b** when a vortex DW with CW chirality is injected and driven through branch structure

When the magnetization of the transverse nanowire is aligned along the $+y$ direction, the injected VDW attains an ACW chirality. Conversely, a CW chirality VDW is injected if the magnetization is aligned along the $-y$ direction. MFM imaging was carried out on an array of structures to determine the trajectory of field-driven VDW into the branch. Figure 9b, c shows MFM images of the initial and final magnetization states of the structures. In the initial state, the magnetization orientation of the transverse nanowire and the branch structure were set along the $-y$ and $-x$ directions, respectively. As the external magnetic field is gradually increased along the $+x$ direction, a HH VDW with a CW chirality is injected and driven in the longitudinal nanowire. According to the chirality dependent motion, the VDW is expected to move along the upper branch and switch its magnetization which is reflected by the change in magnetic contrast from bright to dark. However, MFM image of the final state of the structures shows that the VDW trajectory into the output branch was random. In some of the structures the DW propagated along the lower branch (shown by 'X' marks). This observation points to the possibility that the VDW chirality is not always preserved during VDW motion along the nanowire. We performed 30 MFM scans to estimate the number of successful and failed trials. We observed that in 70% of the trials the DW followed a deterministic trajectory governed by its initial chirality.

Figure 10 shows the relative distribution of successful and failed trials when a VDW propagates in the symmetric structure. The success is achieved in a trial when the VDW follows a selective trajectory governed by its initial chirality. We performed *t-test* to ascertain the 95% confidence level. The range of successful trials was found to be 65–75%. Thus the trajectory is not completely random however the success rate of selective trajectory based on initial chirality is not very high. To investigate the effect of Walker breakdown on the DW trajectory in a branch structure, micromagnetic simulations were carried out. Figure 11 depicts the simulated magnetization configurations of the DW as a function of the simulation time when driven by applied

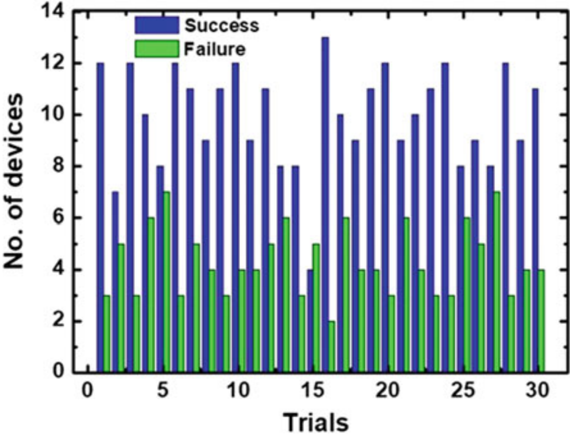


Fig. 10 Relative distribution of the success and failures of the chirality dependant vortex domain wall motion in dual branch structure

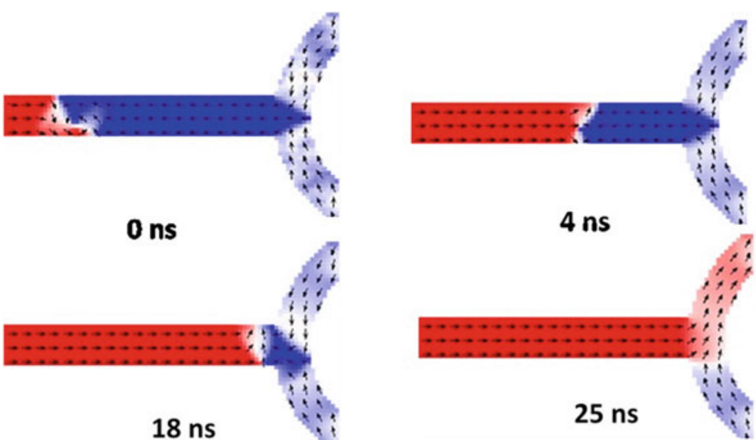


Fig. 11 Snapshot images of the magnetization configuration of the DW in a nanowire at various instances to represent the Walker breakdown where the DW internal structure transforms to different configurations when driven with a magnetic field of 65 Oe

field strength of 65 Oe. The initial chirality of the VDW is ACW. As the VDW propagates, it breaks and transforms into an anti-vortex configuration which eventually transforms to a TDW before entering the bifurcation. The TDW propagates along the upper branch contrary to what is expected. These results indicate that Walker breakdown [23, 38] leads to the VDW propagating into an arbitrary branch.

The above results show that the selective switching based on chirality seems promising but may not be reliable for logic operation when the nanowire is relatively longer as the output of a logic functionality needs to be deterministic rather than probabilistic. To overcome the issue of stochastic nature involved in the DW motion in Y-branch structure, an asymmetric branch structure is proposed, where the two branches are deviated at two different angles at the bifurcation. This is achieved by displacing the output branch at the bifurcation along y direction. Figure 12 shows schematic of the structure where the output branch has been displaced in the $-y$ direction, this configuration is labelled as 'pull-down' (PD). D_{offset} represents the distance by which the branch has been displaced from the center of the longitudinal nanowire. The simulation results for $D_{\text{offset}} = 200$ nm when CW & ACW VDWs are driven in this structure are shown in Fig. 12b(i–iv). Figure 12b-i depicts the initial magnetization state of the structure with an ACW VDW before driving. Figure 12b-ii represents the final magnetization state of the structure, showing that the magnetization of the upper branch switched from $-x$ to $+x$ direction. This implies that the ACW VDW propagates along the upper branch. Figure 12b-iii depicts the initial magnetization state of the structure with a CW VDW before driving. The CW VDW also moves in the upper branch as shown in Fig. 12b-iv. The results reveal that irrespective of the initial chirality, the DW propagates along the upper branch. Alternatively, the branch can be displaced in the $+y$ direction, this structure is labeled as 'pull up' (PU)

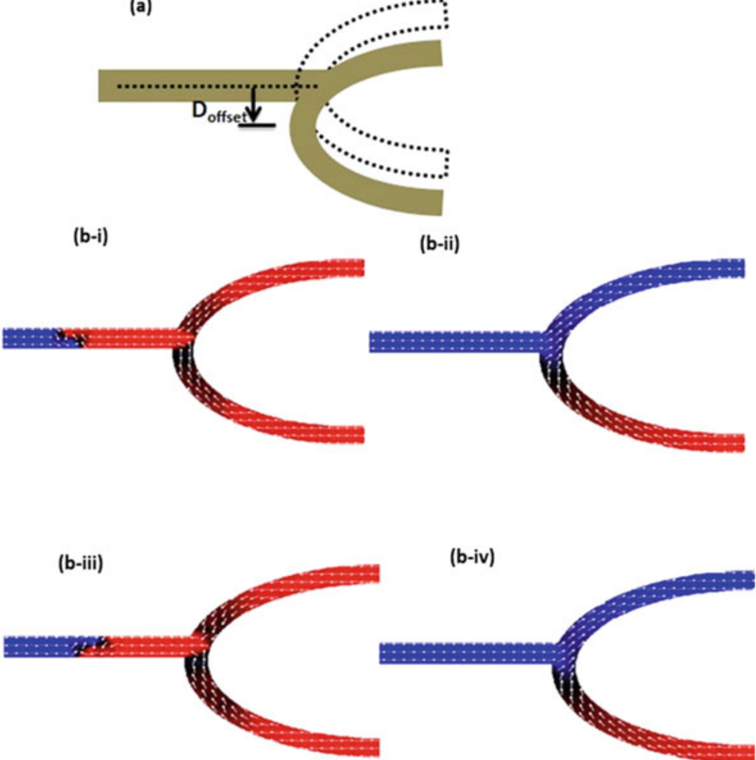


Fig. 12 **a** Schematic of the asymmetric branch with an offset at the bifurcation. **b**(i & ii) snap shot images of the initial and final magnetization configurations when a DW with ACW chirality is injected and driven. **b**(iii & iv) snapshot images of the initial and final magnetization configurations when a DW with CW chirality is injected and driven

structure. With the displacement along the $+y$ direction, the VDW moves along the lower branch irrespective of its chirality.

Experimental study was carried out on DW motion in the asymmetric branch structures using MFM imaging technique to validate the simulations. Figure 13(a) shows the SEM image of a PD structure with similar dimensions as the previous test structure. In this PD structure, the D_{offset} is 200 nm. Figure 13(b-i) shows the MFM image of the initial magnetization state when the transverse nanowire and the branch structure are magnetized in the $+y$ and $-x$ directions, respectively. As the magnetic field was increased along the $+x$ direction, HH-ACW VDW was injected and driven to the upper branch at a field strength of 65 Oe as shown in Fig. 13(b-ii).

This results in the change of the contrast of the upper branch from bright to dark. The final magnetization state for an array of PD structures as obtained from MFM imaging is shown in Fig. 13b-iii. Out of 16 structures in the array, 15 structures showed change in the magnetic contrast of the upper branch. Only one structure

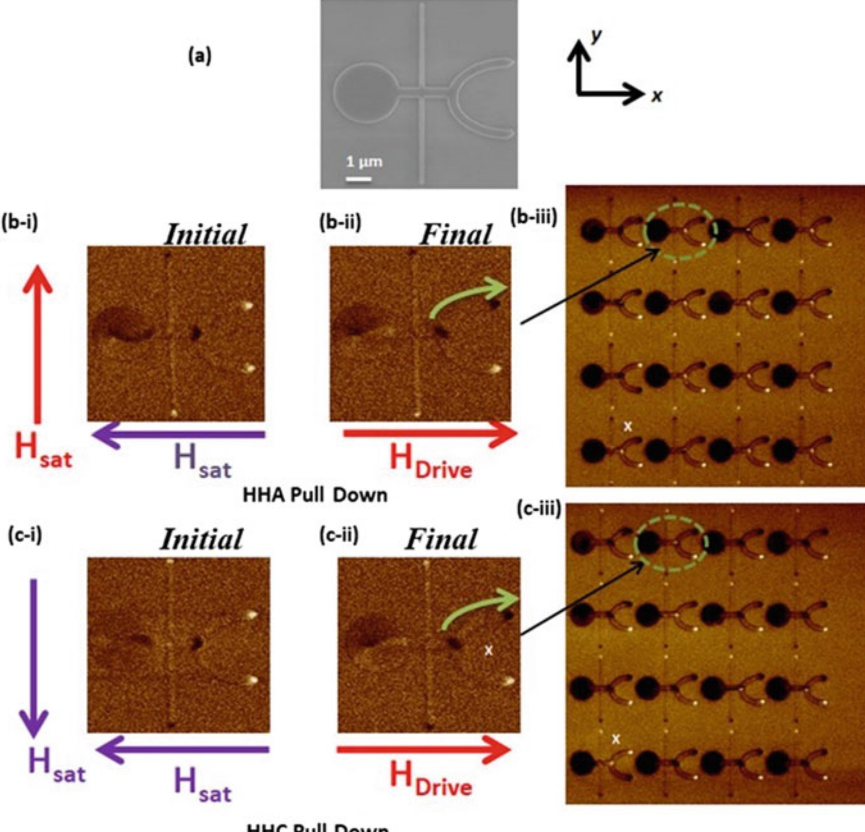


Fig. 13 **a** SEM image of the dual branch structure with a displacement along −y direction at the bifurcation (pull down). **b** (i & ii) MFM images of the initial and final magnetization configuration of the structure when an anti-clockwise vortex DW is injected and driven. **c** (i & ii) MFM images of the initial and final magnetization configuration of the structure when a clockwise vortex DW is injected and driven through the structure

resulted in opposite trend. However, repeated measurements on the devices showed that the same device does not fail every time. The variation in the observed output may be due to thermal instability. Figure 13c-i shows the MFM image of the initial state when the transverse nanowire and the branch structure are saturated in the $-y$ and $-x$ directions, respectively. As the magnetic field increased along the $+x$ direction, HH-CW VDW was injected and driven to the upper branch, as shown in Fig. 13c-ii. Similar observation was obtained for an array of such structures where barring one structure, the rest followed the same trend (Fig. 13c-iii). The above MFM measurements were repeated on a single structure 20 times for both HH-ACW and HH-CW configurations and it was found that the VDW always propagates along the upper branch in PD structure. The above experiments were carried out for the

PU structure as well. The results indicate that the DW moves to the lower branch independent of its assigned chirality. Thus, by introducing an asymmetry in the device geometry, the VDW can be deterministically propagated in a specific trajectory.

From these results we can understand that the VDW can be selectively driven towards one of the two branches in a branch structure irrespective of initial chirality using asymmetric branch structure. One cannot perform the logic operations if the VDW is always moving towards one branch. It brings in the necessity of additional degree of control to selectively drive the VDW trajectory in to the branch structure.

2.4 Exploring the Domain Wall Charge Distribution

We have explored the transverse magnetic charge distribution of a TDW to impose the additional control. We have calculated the magnetic volume pole along the TDW which is analogous to the magnetic charge. As shown in Fig. 14, HH TDW possesses a positive magnetic charge while TT TDW possesses a negative magnetic charge, respectively. The charge distribution of TDW depends on the orientation of the magnetic moment at the wall. The HH TDW is characterized by a positive magnetic charge due to convergence of the magnetization (inset) whereas the TT TDW is characterized by negative magnetic charge due to the divergence of the magnetization at

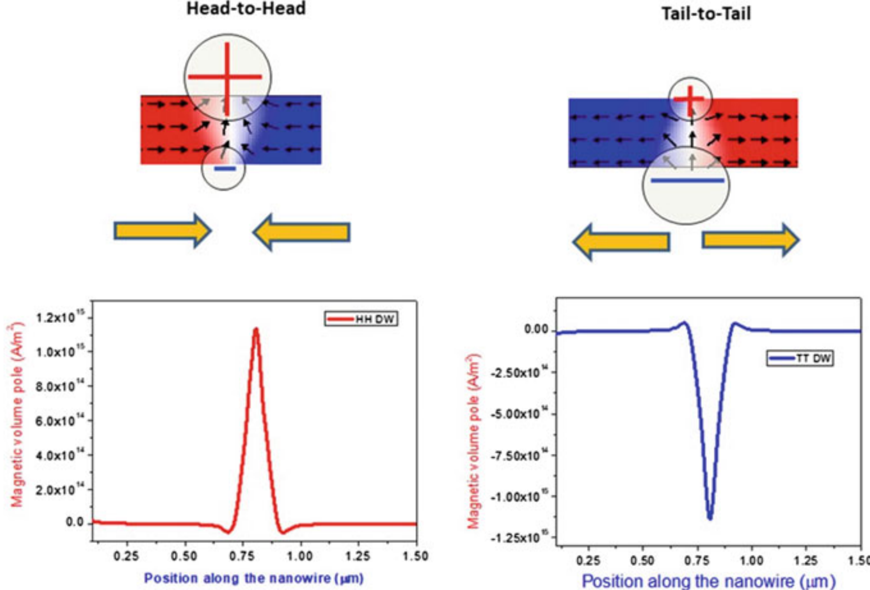


Fig. 14 Magnetic volume pole analogous to magnetic charge calculated by taking the divergence of the magnetization across the DW width for head-to-head and tail-to-tail configurations

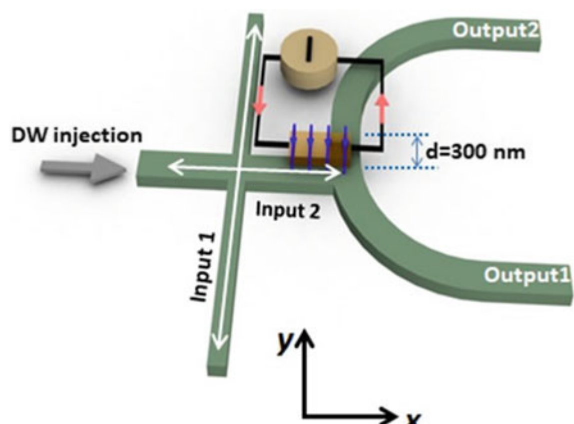


Fig. 15 Schematic image of the dual branch structure with metallic strip at the bifurcation to generate Oersted magnetic field

the wall (inset). Moreover, the TDW assumes a characteristic triangular shape which gives rise to a transversely varying magnetic charge concentration. The base of the triangular spin structure always has higher charge component as compared to the apex of the TDW.

To explore the charge distribution of the TDW for selective motion into the branch structure, we have patterned a metallic strip at the upper edge of the nanowire to generate Oersted field as shown in the schematic Fig. 15. When the current is applied through the metallic strip,

Oersted field is generated that can be used to repel or attract the TDW at the bifurcation. The TDW motion into the U-shaped half ring can be controlled depending on the polarity of the current and the charge distribution of TDW. The Oersted field distribution around the magnetic gate is estimated by COMSOL simulation.

Shown in Fig. 16(a) is simulated model in which the magnetic gate is overlapped only 25% of the nanowire width. The transverse magnetic field distribution in the NiFe nanowire due to the current is shown in Fig. 16c. It shows that the top edge of the NiFe which is directly beneath the Au pad experiences a higher magnetic field compared to the bottom edge of the NiFe which is away from the Au pad. The strength of magnetic field generated by current is plotted as a function of position along the NiFe, Fig. 16c. The field strength is estimated to be ~110 Gauss beneath the Au pad while it is ~10 Gauss at the bottom edge when the current density is 1×10^{11} A/m^2. The local Oersted field interacts with the magnetic charge of the TDW and influences the motion of TDW having higher charge concentration along the upper edge. This dual control of selective motion of the TDW allows the device to perform programmable logic operation within a single structure. By simply changing the direction of the local Oersted field, two universal logic gate operations can be obtained.

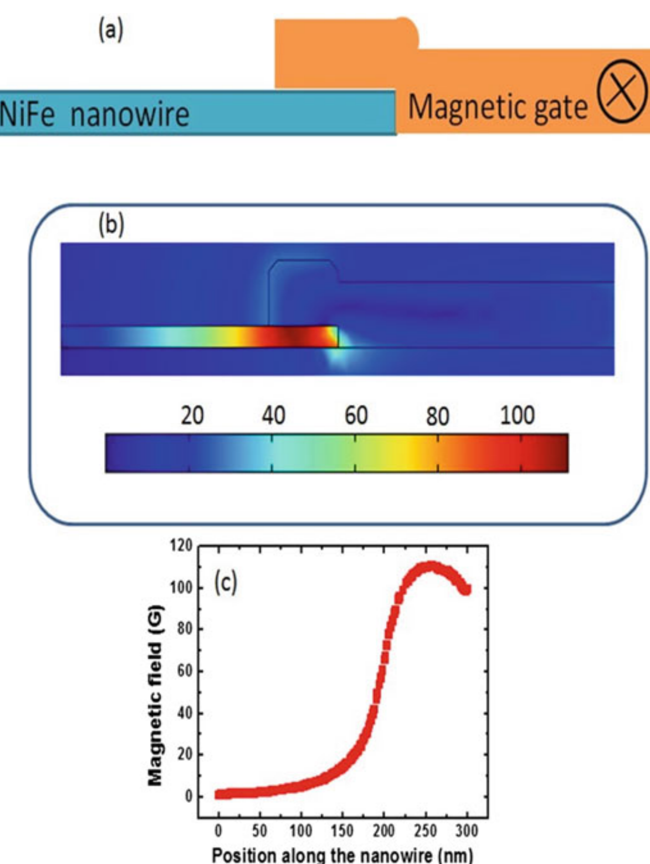


Fig. 16 **a** Schematic image of the magnetic nanowire overlapped with metallic strip to generate the Oersted field. Current is flowing perpendicular to plane, **b** Simulated Oersted magnetic field distribution along cross-section of the NiFe due to the current flowing along the Gate, **c** Magnetic field strength as a function of position along the NiFe nanowire

2.5 Reconfigurable Logic

By combining the above mentioned two effects: deterministic VDW motion in asymmetric branch and the magnetic gate influence on the TDW charge distribution we have demonstrated a reconfigurable logic on a single magnetic structure in which one can carry out all the basic logic functionalities. Figure 17 shows the SEM image of the device structure. The vertical nanowire (width of 120 nm and length of 6 μm) is placed 100 nm away from the bifurcation to transform the DW from vortex to transverse configuration with a specific chirality as mentioned in the earlier section. The half

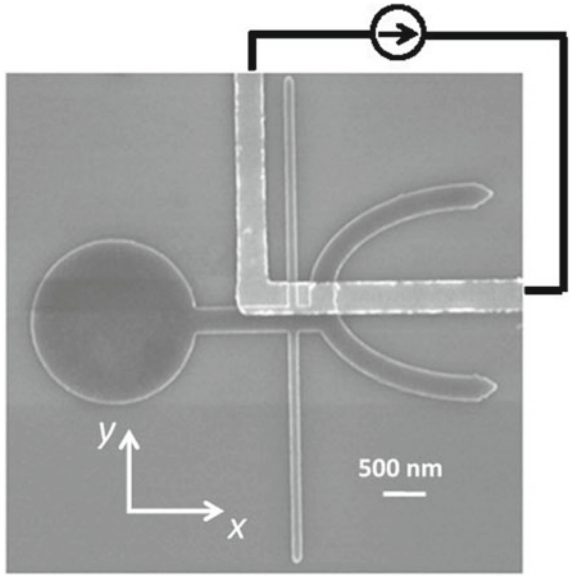


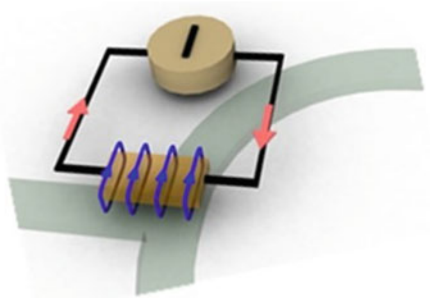
Fig. 17 Scanning electron microscopy (SEM) image of the fabricated device with schematic of the circuit for current injection

ring is displaced along $+y$ direction with an offset of 300 nm to introduce an asymmetry at the bifurcation. The metallic strip (magnetic gate) of Cr(3 nm)/Au(30 nm) is patterned so as to overlap the upper edge of the horizontal nanowire.

2.5.1 NAND and AND Boolean Logic Operations

Here we discuss in detail the experimental verification of the NAND and AND logic gate operations. A current of 6 mA is passed from B to A through the metallic strip as shown in the schematic of Fig. 18. The Oersted field generated curls around the metallic strip attracts negative magnetic charges and repels positive magnetic charges.

Inputs "0" and "0": The initial configuration with both input bits in logical "0" state is captured by using MFM as shown in Fig. 18-i. The magnetisation orientations of Input 1 and Input 2 are both pointing along the negative direction ($-y$ and $-x$, respectively). Both states of the UHR and LHR are in logical bit "0". Following the application of a linear magnetic field of 100 Oe along $+x$ direction, a HH TDW is injected into the nanowire. The transverse component of the HH TDW points in the $-y$ direction ("DOWN" chirality) as dictated by Input 1 (vertical nanowire). After the application of a linear magnetic field, the MFM image shows a change in magnetic contrast of the LHR from bright to dark. The effect of the Oersted field from the magnetic gate on the HH-D TDW is negligible as the higher charge concentration of the TDW is at the lower edge of the nanowire. Hence the TDW is influenced by the asymmetry resulting in the switching of the LHR magnetisation from $-x$ to +



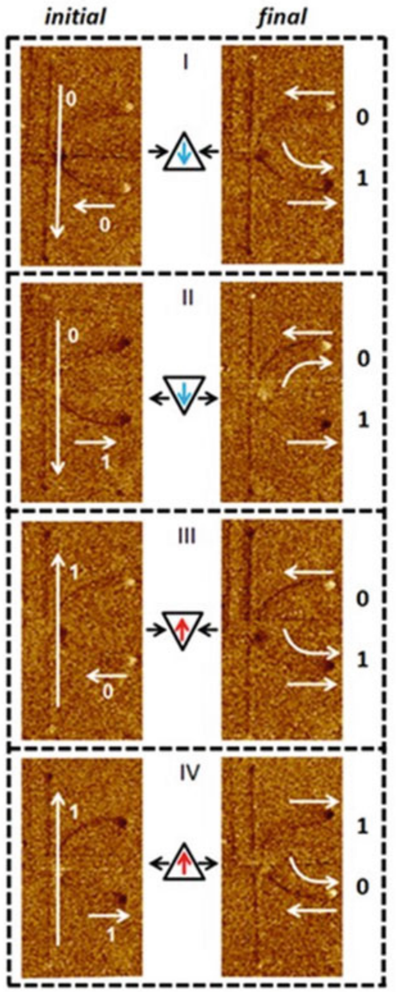


Fig. 18 Logical NAND, AND, NOT and COPY operations, **a** Schematic to show the device with the direction of current flowing from B to A through the metallic gate. The red arrow represents the current direction and the blue lines represent the direction of current induced Oersted field. **b** MFM images of the initial and final configurations of the structure for four different combinations of the magnetisation directions of chirality selector (y) and the nanowire (x)

x direction. Hence the output state at the LHR is logical bit “1”, while that for the UHR is logical bit “0”.

Inputs “0” and “1”: In this particular logic combination, the magnetisation orientation of the horizontal nanowire (Input 2) is aligned along the $+x$ direction. This configuration for Input 2 presets the outputs of the device (UHR and LHR) to logical “1”. Following the application of a linear magnetic field of 100 Oe along the $-x$ direction, a TT TDW propagates through the horizontal nanowire (Input 2). MFM imaging after the field application reveals a change in the magnetic contrast at the UHR from dark to bright, as seen in Fig. 18-ii. The transverse component of the injected TT-DW points to the $-y$ direction (“DOWN” chirality) as dictated by the vertical nanowire (Input 1) magnetic configuration. For a TT-D TDW, the higher charge concentration is at the upper edge of the horizontal nanowire. In this case, the attraction from the Oersted field overcomes the potential barrier created by the asymmetry at the bifurcation.

The Oersted field from the magnetic gate attracts and guides the TT-D TDW into the UHR. Consequently, the UHR magnetisation orientation is switched, representing a change from logical bit “1” to “0”. The magnetisation of the LHR remains unchanged, *i.e.* at logical bit “1”. Without the Oersted field from the magnetic gate, the TT-D TDW would move into the LHR as dictated by the asymmetry of the structure. For our design, the minimum current density required to generate the Oersted field to overcome the potential barrier created along the UHR was ~5×10^{11} A/m^2.

Inputs “1” and “0”: In this configuration, the magnetisation orientation of Input 1 is aligned along $+y$ direction (“1”) and the magnetisation orientation of Input 2 is aligned along the $-x$ direction (“0”). The MFM image of the final state, following the application of 100 Oe driving magnetic field along $+x$ direction. The magnetic contrast change at the LHR indicates the switching of magnetisation orientation of the LHR. A HH-U TDW is injected at the bifurcation as the magnetisation orientation in Input 1 is aligned along $+y$ direction. Similar to the case of “0” and “0”, the HH-U TDW is influenced by both the local Oersted field and asymmetry. As such, the HH-U TDW moves into the LHR resulting in the reversal of the magnetisation orientation to $+x$ direction, which corresponds to an output bit of “1”. The logical state of the UHR remains unchanged as logical bit “0”.

Inputs “1” and “1”: When the magnetisation orientations of the vertical and horizontal nanowires are aligned along the $+y$ and $+x$ directions, respectively, both logical input bits (Input 1 and Input 2) are “1”. Under the external magnetic field, a TT-U TDW is injected into the structure. For this TDW configuration, the higher charge concentration is along the lower edge of the nanowire. MFM image of the structure after 100 Oe field application along $-x$ direction shows that the magnetisation orientation of LHR has switched as seen in Fig. 18-IV. The effect of the Oersted field from the magnetic gate on the TT-U TDW is negligible as the higher charge concentration of the DW is at the lower edge of the nanowire. The effect of asymmetry overpowers the attraction from local Oersted field at the bifurcation. This leads to the TT-U TDW motion into the LHR which switches the magnetisation orientation

Table 1 NAND and AND Truth table formed based on the logic operations of the device showing NAND and AND functionalities

My (Input 1)	Mx (Input 2)	M (UHR) (Output 1) AND	M (LHR) (Output 2) NAND
↓ 0	← 0	← 0	→ 1
↓ 0	→ 1	← 0	→ 1
↑ 1	← 0	← 0	→ 1
↑ 1	→ 1	→ 1	← 0

corresponding to an output bit of "0" while that for the UHR remains unchanged as a logic bit of "1".

The logical output at the UHR and LHR for all the different input combinations are summarized in the truth table in Table 1. For all the logic operations, the resulting magnetic configuration of the UHR and LHR are complementary. For current flowing from B to A through the metallic strip, the outputs at the LHR reveal a NAND logic gate operation, while an AND logic gate operation is obtained at the UHR. The striking feature of this device is that in a single clock cycle, two complementary logic operations can be performed simultaneously.

2.5.2 NOT and COPY Boolean Logic Operations

Single Boolean logic bit operation was achieved by fixing Input 1 (vertical nanowire) to logical bit "1" (magnetisation orientation along $+y$ direction) and flowing current from B to A through the metallic strip. When the structure was saturated along the $-x$ direction, which corresponds to the bit "0" in input 2, the DW propagation leads to the switching of LHR, resulting in a logical "1" output as shown in Fig. 18. For logical input bit "1", the output at LHR reveals a logical bit "0" as presented in Fig. 18. The output bit at the UHR always follows the input bit. The truth table formed in Table 2 shows the structure operates as logical NOT gate when the output is read at the LHR while a COPY operation is obtained at the UHR in this configuration.

2.5.3 NOR and oR Boolean Logic Operations

By reversing the direction of the current flow through the metallic strip *i.e.* from A to B, the polarity of the local Oersted field can be switched, as shown in Fig. 19. The local Oersted field now attracts the DW with a positive charge but repels TDW with a negative charge. From the NAND operation discussion, the magnetic gate has no effect on the DW when the two input bits are the same, *i.e.* input 1 and input 2 are

Table 2 NOT and COPY Truth table describing the NOT and COPY functionalities for single bit logic operation by keeping the chirality selector fixed along +y direction

My	Mx (Input)	M (UHR) (Output) COPY	M (LHR) (Output) NOT
↑ 1	← 0	← 0	→ 1
↑ 1	→ 1	→ 1	← 0

"0"("1") and "0"("1"), respectively as the charge is concentrated at the lower edge of the nanowire. Therefore, the output remains the same for these configurations irrespective of the polarity of the Oersted field as seen in Fig. 19. Here we only discuss the other two input combinations. As shown in Fig. 19, when Input 1 is logic bit "0" and Input 2 is logic bit "1", TT-D TDW is injected at the bifurcation with the application of magnetic field. As the resulting TDW has a higher charge concentration (negative) along the upper edge of the nanowire, it is repelled by the field from the magnetic gate. The TDW is pushed into the LHR and switches the magnetisation orientation of the LHR leading to a logical output bit "0". This is clearly observed by the magnetic contrast change from dark to bright at the LHR, as shown in Fig. 19.

When Input 1 is logical bit "1" and Input 2 is logical bit "0", respectively, a HH-U TDW is formed at the bifurcation. This TDW is characterized by a higher (positive) charge concentration along the upper edge of the nanowire. The HH-U TDW is attracted by the Oersted field from the magnetic gate, and is guided into the UHR.

The magnetic contrast at the UHR switches from bright to dark after the logic operation, as seen from the MFM image in Fig. 19. The results for the logical operation, where the magnetic gate programmed with current flowing from A to B, are summarised in the truth table, Table 3. The logical output at the LHR shows a NOR gate operation. As the UHR is complimentary to the LHR, a logical OR gate operation is obtained at the UHR. The experimental results clearly show that the device can be programmed to perform both universal logic operations (NAND and NOR) by changing the direction of current flow through the metallic strip.

3 Transverse Domain Wall Profile for Logic

For logic operations, information can be encoded and processed within the internal states of magnetic domain wall, by leveraging on the degree of freedom associated with the chirality of TDW. This concept paves the way for mobile data bit within the system enabling more complicated logical operations. The bit representation for the proposed logic scheme is depicted in Fig. 20. The bits "0" and "1" correspond to

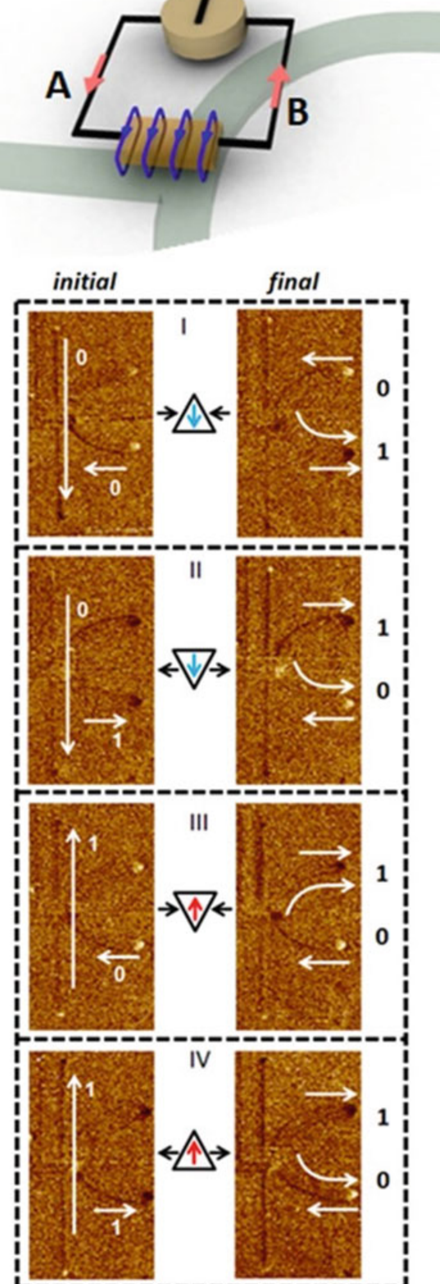


Fig. 19 Logical NOR & OR operations, **a** Schematic to show the device with the direction of current flowing from A to B through the metallic gate. **b** MFM images of the initial and final configurations of the structure for four different combinations of the magnetisation directions of chirality selector (y) and the nanowire (x)

Table 3 NOR and OR Truth table formed based on the logic operations of the device showing NOR and OR functionalities

My (Input 1)	Mx (Input 2)	M (UHR) (Output 1) OR	M (LHR) (Output 2) NOR
↓ 0	← 0	← 0	→ 1
↓ 0	→ 1	→ 1	← 0
↑ 1	← 0	→ 1	← 0
↑ 1	→ 1	→ 1	← 0

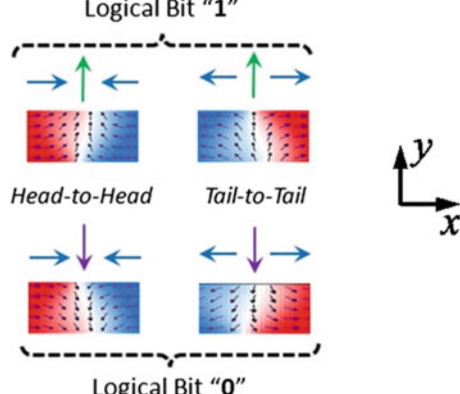


Fig. 20 Logical bit representation *via* transverse spin profile of domain wall, logical bit "1" ("0") is represented by transverse spin pointing along the +y (−y) orientation

the transverse profile pointing in the −y ("DOWN" profile) and +y ("UP" profile) orientations, respectively. In this scheme, irrespective of the type (HH or TT) of TDW, the device structure allows for repeatable logical operation via control of the TDW transverse profile. As discussed earlier, another approach to describing.

TDW is to view it as composite object of elementary topological defects [33]. The edge defects have half-integer winding numbers, either +½ or −½, representing the spin configuration of the DW at the edge of the nanowire, as illustrated in Fig. 20.

3.1 Topological Rectification of TDW

Figure 21 depicts the schematic representation of the topological manipulation of TDW via two devices; Rectifier and Inverter. Both the Rectifier and Inverter comprise of an angled rectangle making a +α° orientation with respect to the nanowire conduit. The main difference between the Rectifier and Inverter is the range of angle α°.

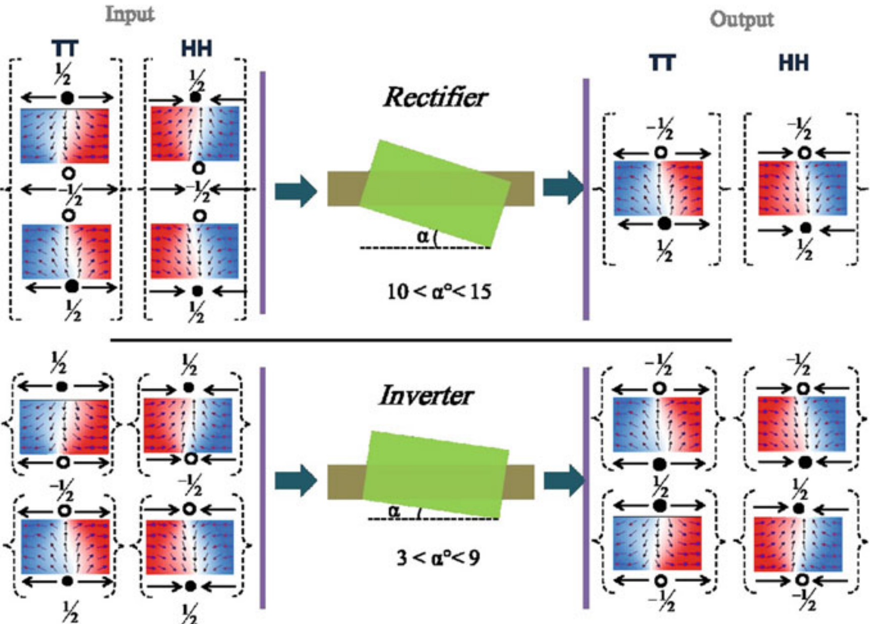


Fig. 21 Schematic representation of the operation of the *Rectifier* and *Inverter*, with all the possible input configurations and corresponding output

In the rectifier, irrespective of the type of TDW input, the resulting output TDW always has a −½ ~ +½ edge defects. For instance, independent of the type of TTDW flowing through rectifier, the output is always a TTDW with transverse component in the +*y* direction, which has a −½ ~ +½ edge defects. Conversely, for inverter, the output TDW will result in opposite edge defects as the input.

The spin state evolution obtained via micromagnetic simulation, as a TTDW is passes through the Rectifier is shown in Fig. 22. Rectification of the TDW occurs when α is in the range of 10 to 15. A TT TDW with transverse component along + *y* direction (TTU), has composite edge defects of −½ ~ +½, whereas for a TTDW with transverse component along −*y* direction (TTD) the composite defects are + ½ ~ −½. As seen in Fig. 22a, a TTU DW undergoes a transformation from TTU TDW to VDW and subsequently to TTD TDW as the DW propagates through the structure. At the entrance of rectifier, the spins along the left-hand edge of the angled rectangle are opposite to the incoming transverse component of the TTD DW. For TDW pinning is most effective when the higher energy component of the TDW encounters a potential barrier [39–41]. Given that the +½ edge defect has a higher energy component [39, 41], the TTD DW is pinned at the entrance of the Rectifier. Depinning result in the nucleation of a vortex core at the entrance of the Rectifier. The chirality of the vortex core is set by the transverse component of the TTD DW.

The transverse spins within TTD DW point along the −*y*-direction, resulting in the left-hand side of the vortex adopting this spin structure. As such, a vortex with

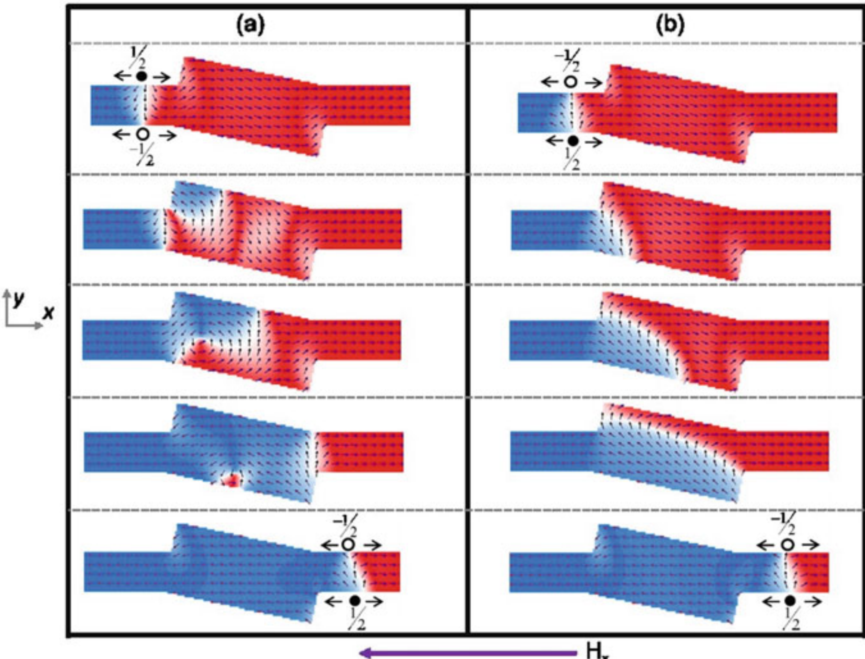


Fig. 22 Spin state evolution of Tail-to-Tail transverse domain wall flowing through the Rectifier structure. Irrespective of the chirality of the input domain wall, **a** Down-chirality, **b** Up-chirality, the output domain wall has an Up-chirality

ACW configuration is nucleated within the rectifier structure. The transformation of the TDW into a vortex wall, enables it to overcome the pinning potential and move along the angled rectangle. As the VDW propagates through the Rectifier, the vortex core moves transverse to the applied field direction to the lower edge of the structure and is subsequently annihilated. Since the VDW has ACW configuration, the spins on the right-hand side points along the $+y$ direction, "UP". The DW exiting the rectifier into the conduit is then a TTU TDW with the transverse component pointing along $+y$, with a topological edge defect of $-½ \sim +½$. The external in-plane field along the +x direction required for this particular operation is ~320 Oe. To mitigate the risk of Walker breakdown, the field can be reset to zero once the DW exits the rectifier, to ensure controllable DW motion and preserve the TDW fidelity.

For a TTU TDW flowing through the structure, as seen in Fig. 22b, the transverse component of the DW and the moments at the left-hand edge of the rectifier point in the same direction. In this case, the higher energy component of the TDW (+½ edge defect) is along the lower edge of the nanowire and hence, does not encounter any potential barrier while moving into the angled rectangle. The lower half (+½ edge defect) of the TDW thus propagates through the structure, following the contour of the lower edge of the structure. The top half of the TDW (−½ edge defect) is

pinned at the upper edge of the structure. This results in the spins within the structure to adopt a transverse alignment along the $+y$ direction. When the +½ edge defect reaches the end of the angled rectangle, a DW with same input spin configuration exits into the conduit. Hence, the rectifier, performs a topological rectification with the output always being −½ ~ +½. As topological edge defects are consistent for Head-to-Head and Tail-to-Tail DW, irrespective of the transverse component of the wall, the structure allows for the topological rectification of the TDW. By mirroring the angled rectangle along the nanowire length, such that the spins within the angled rectangle influences the lower edge of the nanowire, the structure can perform a topological rectification with the output always being +½ ~ −½.

Experimental Verification of Rectifier

Experimental verification of the rectifier is conducted by exploiting the chirality dependent selective movement of TDW in a branch structure. For a +½ ~ −½ TDW flowing through a branch structure, the conservation of the topological defect results in the TDW moving along the upper edge of a Y shape structure. Conversely, for a −½ ~ +½ winding number, the TDW will move to the lower branch of the Y shape structure.

We have experimentally verified this principle using the same nanowire conduit dimensions as the rectifier structure. The topological detector is shown in the scanning electron microscope image. A nucleation pad with a transverse nanowire, acting as a selector is used, so as to ensure proper control on the type of TDW being injected in the conduit. The selector sets the transverse component of the DW exiting the nucleation pad. The Y-shape detector comprises of two wire angled at ~ 70° from the horizontal. To ensure that the TDW is the stable configuration, the width of the nanowire conduit, DW detector and selector are kept at 120 nm.

To test the rectifier structure, the angled rectangle is patterned along the nanowire conduit, as seen in Fig. 23a. The rectifier has a width of ~240 nm (~$2 \times w$) and length ~480 nm (~$4 \times w$), angled with respect to the horizontal axis, at ~ +11°, for rectifier (Fig. 23a). The selector is initially set along the $+y$ direction, characterized by bright and dark contrasts, and the spins along the conduit are aligned the $+x$ direction as seen in Fig 23(b)I. By increasing the external field to 150 Oe, along the $-x$ direction, a TTU DW (−½ ~ +½ winding numbers) is injected into the conduit, as seen in the MFM image, Fig. 23b-II. The lower branch of the detector changes direction, as evident from the MFM image of Fig. 23b-II. This implies that the DW exiting the rectifier has as −½ ~ +½ winding numbers, implying that the TTU DW does not undergo any topological change as it moves through the rectifier.

The MFM configuration when the selector and conduit are magnetized along the $-y$ and $+x$ direction respectively, is shown in Fig. 23c-I. As expected, the selector has a bright and dark contrast on the upper and lower edges. Following the application of a field of 150 Oe along the $-x$ direction, the contrast of the lower branch of the detector changes to bright, as can be seen in Fig. 23c-II. This indicates that a DW with winding numbers of −½ ~ +½, reached the detector. As the nucleated DW has a topological edge defect of +½ ~ −½, this implies that structure A effectively, rectifies the output of any incoming DW to a −½ ~ +½ configuration.

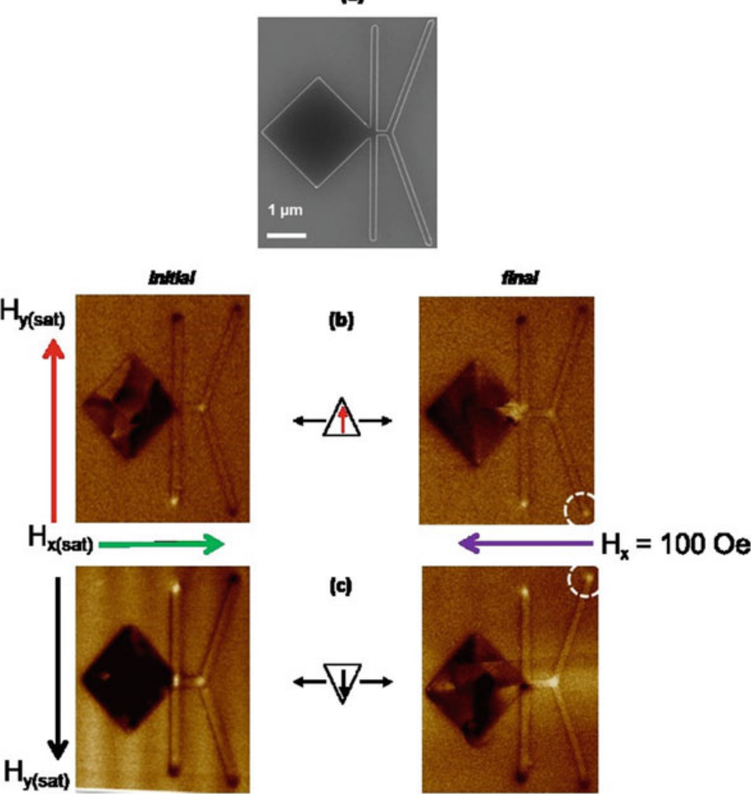


Fig. 23 **a** Scanning electron microscope image of a topological rectifier comprising of a nucleation pad and Y-shape structure. Magnetic force microscope image of the initial and final magnetic configuration of the topological rectifier for **b** Tail-to-Tail Up Domain wall and **c** Tail-to-Tail Down Domain wall

3.2 Topological Inverter of TDW

By tuning the value of α, the angled rectangle structure can be made to carry out topological inversion of the edge defects, as seen in Fig. 24. Simulations reveal that for inversion, α needs to be in the range of 3–8. The TTD DW (+½ ~ −½), undergoes a similar spin transformation as previously described for the rectifier structure. The output DW from the angled structure is a TTU DW with an edge defect of −½ ~ + ½. For a TTU DW flowing through the structure, a TTD DW exits into the nanowire conduit. As the higher energy component of the TTU DW (+½ winding number) is at the lower edge of the nanowire, the lower half of the DW moves into the structure. As the spins along the upper right edge follow the contour of the angle structure, the −½ edge defect which depins from the left edge cannot follow the +½ edge defect. Unlike the rectifier, the +½ edge defect transforming into a vortex core at the lower

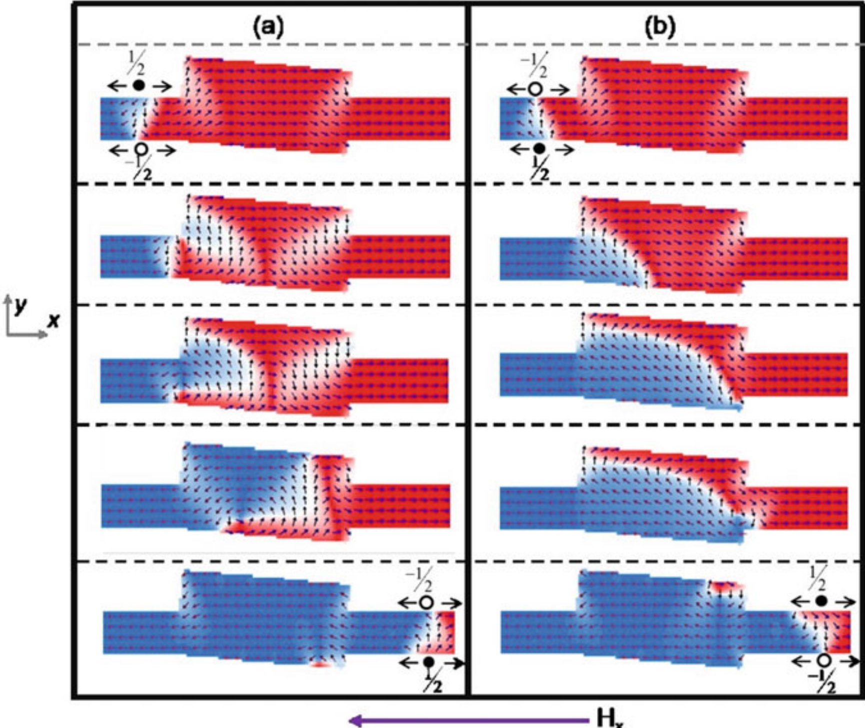


Fig. 24 Spin state evolution as a TDW goes through the rectifier for **a** TT TDW with DOWN chirality undergoes a chirality flip and a TDW with UP chirality exits the rectifier, **b** A TT TDW with UP undergoes reversal and is output as a TT TDW with DOWN chirality

right hand-edge of the angled rectangle. The vortex core adopts a CW configuration as the transverse spin's component was pointing along the $+y$ direction. The vortex core moves towards the upper right edge of the structure, injecting a TDW into the nanowire conduit in the process. The TDW exiting the structure has the transverse component along the $-y$ direction, i.e. a TTD DW with $+½ \sim -½$ edge defects. As such, this structure performs topological inversion of TDW. Similarly, for a HH DW, if the input is HHU (HHD) DW then the output will always be HHD (HHU) DW.

Experimental Verification of Inverter

The SEM image in Fig. 25a, shows the angled rectangle structure with ~ $+4°$ orientation with respect to the vertical axis. The initial configuration of the inverter with the selector set in the $-y$ direction and the spins in the conduit aligned along the $+x$ direction is shown in Fig. 25b-I. Increasing the magnetic field along the $-x$ direction results in the injection of a TTD DW into the conduit. The resulting MFM image after application of a field of 150 Oe along the $-x$ direction is shown in Fig. 25b-I. This results in the upper branch of the detector switching direction, as seen from the MFM

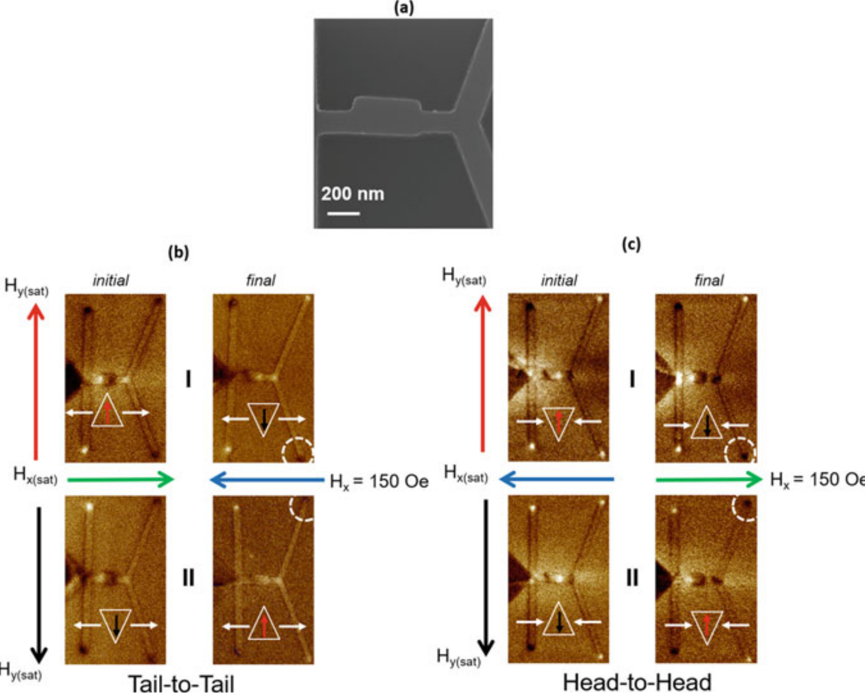


Fig. 25 (a) Scanning electron microscopy image of inverter patterned along the nanowire. Magnetic force microscopy images of the topological inverter with initial and final configuration for; (b)-I TTU DW, (b)-II TTD DW and (c)-I HHU DW and (c)-II HHD DW

image of final configuration Fig. 25b-I. This implies that the DW has undergone a topological inversion and the output is a TTU DW.

The MFM configuration when the selector and conduit are magnetized along the $+y$ and $+x$ direction respectively, is shown in Fig. 25b-II. Following the application of a field of 150 Oe along the $-x$ direction, the contrast of the lower branch of the detector changes to bright, as can be seen from the final configuration in Fig. 25b-II. As expected, the output DW into the conduit is a TTD DW.

In Fig. 25c, the inverter operation for HH TDWs is presented. The initial configuration of the inverter with the selector set in the $+y$ direction and the spins in the conduit aligned along the $-x$ direction is shown in Fig. 25c-I. Increasing the field in the $+x$ direction injects a HHU TDW, with $+½ \sim -½$ edge defects into the conduit. The final MFM configuration, in Fig. 25c-I, shows that the lower branch of the detector has switched. As discussed previously, a TDW with $+½ \sim -½$ would propagate in the upper branch of the Y-shape detector. Thus, for the lower branch to switch, a TDW with $-½ \sim +½$ edge defect should have moved to the bifurcation. This confirms that the inverter flips the edge defects of the input TDW to output an HHD TDW. Similarly, when a HHD TWD is injected into the conduit, the output

from the inverter switched the upper branch of the detector. This implies that the HHD (−½ ~ +½) TWD has been transformed into an HHU (+½ ~ −½) TWD.

The above results show that it is possible to control the winding numbers of TDW in magnetic nanowires by exploiting the relaxation process of a DW passing through an angled rectangle. This work paves the way for technique employing DW profile (winding numbers) for performing logical operation. The TDW can be considered as a mobile data bit and the information is encoded as the transverse profile of the DW. This concept can be exported to chiral DW in materials with perpendicular magnetic anisotropy.

Even though the fabricated structures have slight variations, as seen from the SEM image of Figs. 23 and 25, the topological rectification/inversion is still possible as evidence from the MFM measurements. The variation in the shape of the structure does not affect the rectification/inversion process, as the device exploits the inherent internal DW profile for pinning/depinning at the modulation. This shows that this technique is robust and can be applied experimentally for various technological applications.

The field needed for the rectification/inversion operation is lower in experiment as compared to micromagnetic simulation is attributed to thermally activated depinning [42]. A field of 150 Oe, for device operation, may inevitably lead to Walker breakdown, which will affect the integrity of the DW. The fidelity length of a TDW decreases asymptotically to around 350 nm at high fields [33]. Additionally, it has been reported that edge defects along the nanowire conduit may also lead to shift Walker breakdown to higher field [43]. For all our samples, the nanowire conduit length between the nucleation pad to the rectifier/inverter and subsequently the detector, is kept to 200 nm. This is well below the reported fidelity length of TDW and ensures that the integrity of our input and output DW are not compromised. However, for practical applications, the nanowire conduit may be structurally modulated hereby stabilising DW and preserving its chirality over large distances [44].

4 Summary

In this chapter, we have demonstrated the concept of manipulating information by engineering the internal structure of transverse domain wall. By exploiting the dynamics on domain wall within defined structures, different types of logical operations can be carried out. We show that by leveraging on the charge intrinsic to the transverse domain wall profile, a gate programmable reconfigurable logic structure can be designed. The structure is capable of carrying out all the logical operations and more importantly produce complementary output. Finally, we show that by modulation the propagation of the domain wall through modulated nanowires, the profile of the transverse domain wall can be deterministically controlled, enabling a novel paradigm in magnetic logic. The domain wall profile for logic operation can be further extended to domain walls in materials with perpendicular magnetic anisotropy.

References

1. G.A. Prinz, Magnetoelectronics. Science **282**, 1660 (1998)
2. S.A. Wolf et al., Spintronics: a spin-based electronic vision for the future. Science **294**, 1488 (2001)
3. A. Imre et al., Majority logic gate for magnetic quantum-dot cellular automata. Science **311**, 205 (2006)
4. I. Amlani et al., Digital logic gate using quantum-dot cellular automata. Science **284**, 289–291 (1999)
5. R.P. Cowburn, M.E. Welland, Room temperature magnetic quantum cellular automata. Science **287**, 1466–1468 (2000)
6. D.A. Allwood et al., Submicrometer ferromagnetic NOT gate and shift register. Science **296**, 2003–2006 (2002)
7. D.A. Allwood et al., Magnetic domain-wall logic. Science **309**, 1688–1692 (2005)
8. S. Goolaup, M. Ramu, C. Murapaka, W.S. Lew, Transverse domain wall profile for logic applications. Sci Rep **5**, 9603 (2015)
9. P. Xu, et al. An all-metallic logic gate based on current-induced domain wall motion. Nat. Nanotech. **3**, 97–100 (2008)
10. G. Reiss et al., New materials and applications for magnetic tunnel junctions. Phys. Stat. Sol. **201**, 1628–1634 (2004)
11. B. Behin-Aein, D. Datta, S. Salahuddin, S. Datta, Proposal for an all-spin logic device with built-in memory. Nat. Nanotech. **5**, 266–270 (2010)
12. S. Patil, A. Lyle, J. Harms, D. Lilja, J.-P. Wang, Spintronic logic gates for spintronic data using magnetic tunnel junctions, in Proceedings of IEEE International Conference Computer Design, pp. 125–131 (2010)
13. Z. Navabi, *Digital Design and Implementation with Field Programmable Devices* (Kulwer Academic Publishers, Dordrecht, Netherlands, 2005)
14. A. Ney, C. Pampuch, R. Koch, K.H. Ploog, Programmable computing with a single magnetoresistive element. Nature **425**, 485–487 (2003)
15. H. Dery, P. Dalal, L. Cywinski, L.J. Sham, Spin-based logic in semiconductors for reconfigurable large-scale circuits. Nature **447**, 573–576 (2007)
16. S. Joo et al., Magnetic-field-controlled reconfigurable semiconductor logic. Nature **494**, 72–76 (2013)
17. S.S.P. Parkin, M. Hayashi, L. Thomas, Magnetic domain wall racetrack memory. Science **320**, 190 (2008)
18. M. Hayashi, L. Thomas, R. Moriya, C. Rettner, S.S.P. Parkin, Current controlled magnetic domain wall nanowire Shift register. Science **320**, 209 (2008)
19. J.H. Franken, H.J.K. Swagten, B. Koopmans, Shift registers based on magnetic domain wall ratchets with perpendicular anisotropy. Nat. Nanotech. **7**, 499 (2012)
20. S. Glathe et al., Splitting of a moving transverse domain wall in a magnetic nanostripe in a transverse field. Phys. Rev. B **81**, 020412 (2010)
21. Tretiakov, O. A. et al. Dynamics of domain walls in magnetic nanostrips. Phys. Rev. Lett. **100**, 127204 (2008)
22. T. Ono et al., Propagation of a magnetic domain wall in a submicrometer magnetic wire. Science **284**, 468 (1999)
23. G.S.D. Beach et al., Dynamics of field driven domain wall propagation in ferromagnetic nanowires. Nat. Mater. **4**, 741 (2005)
24. X. Jiang et al., Enhanced stochasticity of domain wall in magnetic racetracks due to dynamic pinning. Nat. Comm. **1**, 25 (2010)
25. D. Atkinson et al., Magnetic domain wall dynamics in a submicrometer ferromagnetic structure. Nat. Mater. **2**, 85 (2003)
26. L.K. Bogart et al., Dependence of domain wall pinning potential landscapes on domain wall chirality and pinning site geometry in planar nanowires. Phys. Rev. B **79**, 054414 (2009)

27. S. Lepadatu, A. Vanhaverbeke, D. Atkinson, R. Allenspach, C.H. Marrows, Dependence of domain-wall depinning threshold current on pinning profile. Phys. Rev. Lett. **102**, 127203 (2009)
28. H.K. Teoh, S. Goolaup, W.S. Lew, Programmable logic operation via domain wall profile manipulation. IEEE Magn. Lett. https://doi.org/10.1109/lmag.2015.2434801 (2015)
29. C. Murapaka et al., Reconfigurable logic via gate-controlled domain wall trajectory in magnetic network structure. Sci. Rep. **6**, 20130 (2016)
30. O. Tchernyshyov, G.W. Chern, Fractional vortices and composite Domain walls in flat nanomagnets. Phys. Rev. Lett. **95**, 197204 (2005)
31. M.J. Donahue, D.G. Porter, *OOMMF User's Guide, Version 1.0 Interagency Report NISTIR 6376*. National Institute of Standards and Technology, Gaithersburg, MD (Sept 1999)
32. C.H. Marrows, Adv. Phys. **54**, 585 (2005)
33. E.R. Lewis, D. Petit, A.-V. Jausovec, L.O'Brien, D.E. Read, H.T. Zeng, R.P. Cowburn, Phys. Rev. Lett. **102**, 057209 (2009)
34. C. Murapaka et al., Appl. Phys. Express **7**, 113003 (2014)
35. A. Pushp, T. Phung, C. Rettner, B.P. Hughes, S.-H. Yang, L. Thomas, S.S.P. Parkin, Nat. Phys. **9**, 505 (2013)
36. T. Phung et al., Appl. Phys. Lett. **105**, 222404 (2014)
37. P. Sethi et al., Direct observation of deterministic domain wall trajectory in magnetic network structures. Sci. Rep. **6**, 19027 (2016)
38. N.L. Schryer, L.R. Walker, The motion of 180o domain walls in uniform dc magnetic fields. J. Appl. Phys. **45**, 5406 (1974)
39. D. Atkinson, D.S. Eastwood, L.K. Bogart, Controlling domain wall pinning in planar nanowires by selecting domain wall type and its application in a memory concept. Appl. Phys. Lett. **92**, 022510 (2008)
40. M. Chandra Sekhar, S. Goolaup, I. Purnama, W.S. Lew, Depinning assisted by domain wall deformation in cylindrical NiFe nanowires. J. Appl. Phys. **115**, 083913 (2014)
41. S. Goolaup, S.C. Low, M. Chandra Sekhar, W.S. Lew, Dependence of pinning on domain wall spin structure and notch geometry. J. Phys. Conf. Ser. **266**, 012079 (2011)
42. A. Himeno, T. Okuno, T. Ono, K. Mibu, S. Nasu, T. Shinjo, Temperature dependence of depinning fields in submicron magnetic wires with an artificial neck. J. Magn. Magn. Mat. **286**, 167 (2005)
43. Y. Nakatani, A. Thiaville, J. Miltat, Faster magnetic walls in rough wires. Nat. Mater. **2**, 251 (2003)
44. D. Burn, D. Atkinson, Suppression of Walker breakdown in magnetic domain wall propagation through structural control of spin wave emission. Appl. Phys. Lett. **102**, 242414 (2013)

3D Nanomagnetic Logic

Markus Becherer

Abstract Digital computation by magnetic ordering? That sounds useless or crazy having powerful CMOS technologies available everywhere at low cost. However, if it comes to massively parallel and pipelined digital operations with stringent power constraints, 3D Nanomagnetic Logic might pay off. This chapter gives an insight on an experimentally demonstrated complete set of logic devices, where the entities are not electrically connected but fully powered and operated by magnetic clocking fields. The computing elements are comprised of ferromagnetic islands of sub-micrometer size, whereas the binary "0" and "1" is encoded in magnetic north- and south-pole. This is accomplished by ferromagnetic thin-film materials that show magnetic anisotropy perpendicular to the plane of the chip. The anisotropy is locally reduced by focused Ga^+ ion radiation in order to program computation functionality into close-by magnetic islands. Digital signals are propagated by field-driven domain-walls in lateral direction and via field coupling in vertical direction, enabling monolithic 3D integration. Based on majority votes that can be re-programmed to the universal NAND or NOR function, a hybrid co-processor integrated in the back-end-of-line CMOS technology is envisioned. Still far from being ready for mass fabrication, but containing all basic elements for 3D integrated computation with nonvolatile magnetic states.

1 Introduction to the Concept of pNML

In order to pursue the ongoing scaling of CMOS circuits, beyond CMOS devices are heavily researched [1]. At the same time, it is a well known fact, that planar technology having been a story of great success for Integrated Circuits (ICs) over the last 50 years, will face a limit, latest when approaching atomistic length-scales. However, in order to pursue the path of the ITRS roadmap—namely increasing device density by an exponential law—there is a strong trend towards 3D integration [2].

M. Becherer (✉)
Chair of Nanoelectronics, Technical University of Munich, Theresienstraße 90, 80333 Munich, Germany
e-mail: markus.becherer@tum.de

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_8

The current key enabling technology for 3D integration are thinned wafers, stacked and connected with through-silicon vias [3, 4].

To give an example for 3D integration, NAND Flash technology for nonvolatile memory has already been economically successful [5, 6]. This is mainly due to the fact that fault tolerances for memory are less critical and memories are highly regular structures easier to integrate in a 3D manner by stacking dies [2]. However, true monolithic 3D integration is still in its beginnings especially for logic ICs and reasons for that are i.e. the very low acceptable error rates, non-regular layouts and GHz clocking frequencies. Easily speaking, logic ICs are far more demanding and for monolithic device integration, process technologies with reduced thermal budget and at the same time high quality conducting, semiconducting and insulating films are needed. Furthermore, interconnects of different length scales are a must: from very small, to medium and to very long distances and at the same time, they have to be provided in a strictly hierarchical manner. To be precise, through silicon vias are not an option for monolithic integration, as they connect dies only on a long range manner. It is worthwhile to mention that planar fabrication technology is not contradictory to 3D integration and by overcoming the wiring problem, namely electrical connection to every single switching element, it would tremendously reduce complexity of the fabrication technology.

There is a broad spectrum of assessment going on for logic devices that are discussed as beyond CMOS candidates. Following the Taxonomy of ITRS as reviewed and summarized in [7] there are

- Conventional CMOS FETS,
- Alternate Channel Materials and Structures,
- Charge-Based Devices Beyond Conventional FETs and
- Devices Using Information Carriers other than Electronic Charge,

whereas the latter class of devices with—Information Carriers other than Electronic Charge, are

- All-spin logic,
- Bilayer Pseudospin FET,
- Excitonic FET,
- Nanomagnetic Logic,
- Spin torque majority gate and
- Spine Wave devices.

The concept of perpendicular Nanomagnetic Logic (pNML) for digital computation—mainly developed at Technische Universität München and Notre Dame University—is a subclass of Nanomagnetic Logic, where multilayer stacks are utilized, showing a ferromagnetic magnetization perpendicular to the film plane. The pNML technology avoids *electronic switches* for computation by using *ferromagnetic bistable entities*, coupled by magnetic fields. We claim that it is highly beneficial to use such non-volatile computational states and combine logic functionality with inherent memory

ability. This has already been pursued by ferromagnetic logic device implementations of different flavors. In order to distinguish our approach from other magnetic device implementations, we restrict ourselves to the following preconditions:

- Boolean algebra with non-volatile binary states (digital information processing) from ferromagnetic entities. The binary states encode logic ‘0’ and ‘1’.
- Planar fabrication with standard microelectronic equipment (top-down approach), compatible with CMOS, beneficially as a back-end of line (BEOL) process with low temperature budget.
- No electrical current flow except for in-/output interfaces, latching/synchronization of signals (sequential logic) and clocking circuitry for power supply.

The main implementations fulfilling these conditions are:

1. Magnetic domain-wall (DW)-logic, e.g. [8].
2. In-plane Nanomagnetic Logic (iNML) [9–13].
3. Perpendicular Nanomagnetic Logic (pNML) [14–18].

For (1) *DW-logic* the reader is referred to detailed descriptions in e.g. [8], for (2) *In-plane Nanomagnetic Logic* e.g. [11, 13]. This book chapter will exclusively focus on (3) *pNML devices and circuits* composed of magnetic multilayers with *perpendicular-to-plane* magnetic anisotropy. The principle layout of such a pNML computing system is depicted in Fig. 1. It consists of thin planar ferromagnetic islands that are lithographically defined on a plane surface. The digital values ‘1’ or ‘0’ are assigned to the polarization of the magnetic islands, i.e. corresponding to the magnetization pointing ‘up’ or ‘down’ with respect to the film plane. This simplification of binary assignment holds, as long as the ferromagnets are in the so-called *single-domain* state. Magnetic spins are acting as an ensemble, stabilizing each other and forming a magnetic macrospin. It is worthwhile to mention that these ferromagnetic states are non-volatile, however, for logic implementation they will be switched on nanosecond scale. Digital computation is implemented by magnetic field-coupling between ferromagnetic islands, reaching a local low-energy state with anti-parallel alignment of the magnetization state. For signal routing the z-dimension is exploited

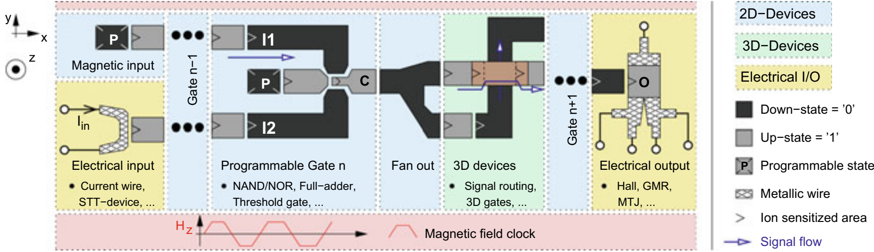


Fig. 1 Top view Layout of a pNML system including 2D and 3D pNML building blocks, electrical I/O and clocking circuitry (schematically drawn as alternating pulse-shaped waveform). Digital computation is performed by the field-coupled nanomagnets and signal propagation takes place in elongated islands, acting as domain-wall conduits. Reprinted with permission from [19]

by stacking the highly planar 2D structures. With that, 3D monolithic computing structures, interconnected by domain wall conduits in lateral direction (wire-shaped ferromagnetic islands) and *magnetic vias* (established by coupling fields) in vertical direction are achieved. As computation is taking place in a well defined volume—so-called *nucleation centers* of the islands—field-coupled vias and signal crossings are achievable without electrical (metallic) interconnects. In other words, computation is taking place in one part of a ferromagnetic island, whereas signal propagation takes place in another part, where domain-walls propagate under an external magnetic field. Electrical input is envisioned by e.g. current carrying (metallic) wires generating a magnetic field to switch a close-by magnet or by running electrical current through more complex magnetic arrangements like magnetic-tunnel-junctions (MTJs). Electrical output is implemented by Hall-elements, giant-magneto-resistance (GMR)- or (MTJ)-structures. The clocking apparatus, a spatially homogeneous alternating magnetic field generated by a ferromagnetic on-chip inductor, corresponds to the 'power supply' in CMOS circuits.

The book chapter is organized as follows: in Sect. 2, the concept of computing in pNML devices is explained and 2D-pNML devices are reviewed. Section 3 is highlighting the need for experimentally calibrated models to bridge the gap between the micromagnetic domain and circuit level simulations. Section 4 addresses the monolithic 3D integration of the presented 2D devices in order to keep up with integration density and to provide field-coupled interconnects for signal distribution. In Sect. 5, a co-processing unit as a potential BEOL CMOS process together with an on-chip inductor for clock-field generation is described.

2 The Planar Technology of pNML—Fabrication of Devices

In this section, the pNML fabrication technology is summarized and the basic operating principles in 2D pNML are discussed. Furthermore, state-of-the-art 2D pNML devices are introduced before the *five tenets* of digital computation are reviewed.

2.1 pNML Fabrication Technology

The basic materials for pNML are most commonly (but not limited to) sputter deposited Co/Pt or Co/Ni multilayer stacks. With the single Co layer chosen not thicker than in the order of 1.5 nm, the magnetization will align perpendicular to the film surface. There is a complex interplay between shape anisotropy (a 2D extended flat magnetic film) and crystalline/interfacial anisotropy [20, 21]. Multiple repetitions of non-magnet/ferromagnet (NM/FM) or ferromagnet/ferromagnet (FM/FM) bilayers, enhance the anisotropy in perpendicular direction. In order to identify prospective films for NML devices both high areal magnetization [mA] and

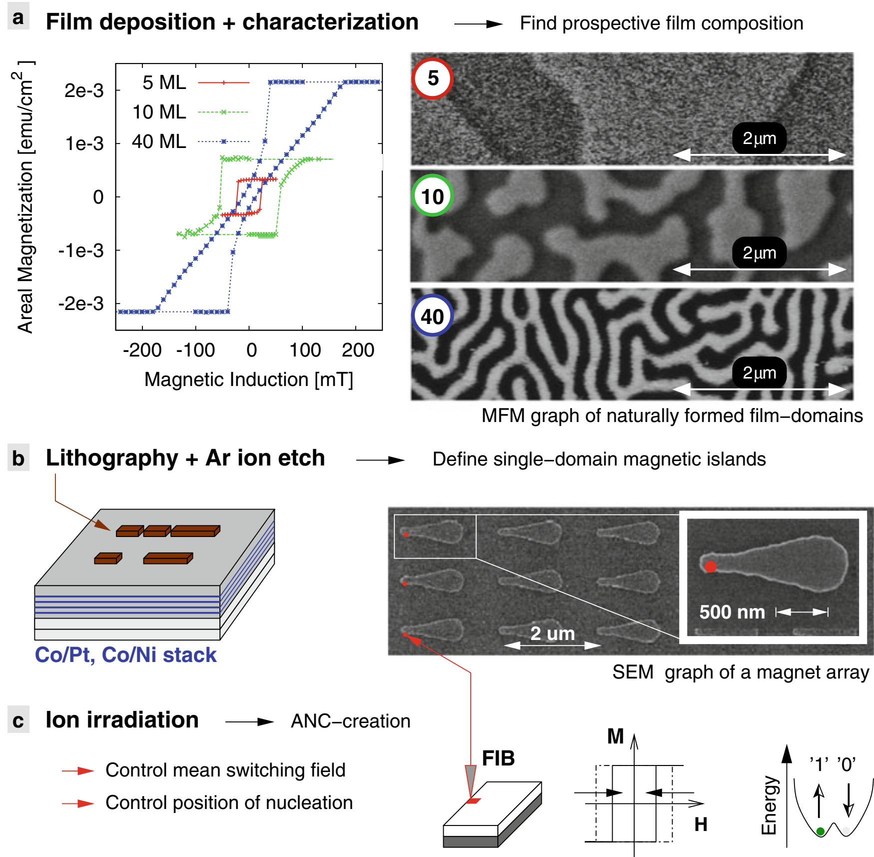


Fig. 2 Fabrication technology for 2D pNML, which can be subdivided in three major steps: **a** Film deposition and characterization, **b** magnetic island lithography and etching and **c** artificial nucleation center (ANC) formation with FIB irradiation

sufficient magnetic uniaxial anisotropy [$\mathrm{J\,m^{-3}}$] are searched for. [1] When demagnetized, the sputter deposited films show natural strip domains as depicted in Fig. 2a. With increasing number of bilayers, the magnetic moment is increased, the hysteresis loops become sheared and the natural domain width is significantly reduced [22]. A typical material composition optimized over the years that turned out to be the *workhorse* for pNML research in our group reads $\mathrm{Ta_{3.0\,nm}/Pt_{3.0\,nm}/N \times [Co_{0.8\,nm} + Pt_{1.0\,nm}]/Pt_{4.0\,nm}}$, where N is the number of bilayer repetitions.

In Fig. 2b, the second fabrication step is visualized. Co/Pt islands are defined by e-beam (EB) or focused ion beam (FIB) lithography, hard mask patterning and con-

[1]High amplitudes of areal magnetization are providing sufficient stray-fields for the computing operation whereas magnetic uniaxial anisotropy is keeping magnetization in perpendicular-to-film direction, a prerequisite for this type of computing devices.

secutive Ar ion etching. For the definition of the artificial nucleation center (ANC), two methods are applied, either in a one step process during FIB lithography with increased areal dose at the ANC spots, or in a second alignment step after island etching. The ANC itself is most commonly not visible in scanning electron microscopy imaging and hence, highlighted by red circles in the SEM image of Fig. 2b. ANC creation and its importance for field-coupling is summarized in Fig. 2c. Basically, the switching field (the magnetic field that has to be applied for magnetization reversal) is controlled by the dose, species and energy of ion irradiation. But also spatial extension and the position of the FIB ANC is vital for NML operation. It is indeed a complex process and still a matter of current research. Overall, fabrication of 2D pNML devices can be subdivided in three major steps:

1. Film deposition and characterization for viable film composition,
2. High resolution lithography and island definition,
3. ANC creation during lithography or in a consecutive, aligned FIB irradiation step.

2.2 pNML Operation Principles

The basic elements for both computation and signal propagation are lithographically patterned –in at least one dimension– sub-micron sized, ferromagnetic islands. As explained in the previous section, two magnetization directions are preferred, both perpendicular to the film plane. The binary logic states '*0*' and '*1*' are represented by the down- or up-magnetization direction of the islands, pointing either 'in' or 'out' of the surface as depicted in Fig. 3a. The magnetic anisotropy, which defines the direction of lowest energy for the magnetization vector, is governed by an interplay of crystalline and interfacial anisotropy. Hence, for thin film media with perpendicular magnetization, various shapes of magnetic islands are possible, ranging from squares or circles to elongated wires and stripes, the latter also referred to as domain-wall (DW) conductors. As a matter of fact, shape anisotropy is in a first order approximation constant over the 2D plane and hence, negligible, providing many degrees of freedom in the design of logic gates. However, it is worthwhile to mention, that the *position* of the ANC is vitally important for logic operation.

In order to reach a lowest energy state, magnetic coupling fields are acting strongly between next neighbor magnets and can be used for logic operations as shown in the energy diagram of Fig. 3a. For two neighboring magnets, there are 4 possible stable states, from which the 2 anti-parallel states are of lower energy. The principle mechanism behind nanomagnetic computation is to switch the magnets with externally applied field pulses. At the same time they relax to the lowest energy state by summing the external clocking field with the stray fields of the next neighbor magnets right within the ANC.

Coupling fields of magnetic islands decay with $H_c \propto 1/r^n$ and $1 \leq n \leq 3$, where H_c is the coupling field, r the spatial distance and n a rational number depending on the spatial geometry of the arrangement. Coupling is intrinsically symmetric, i.e.

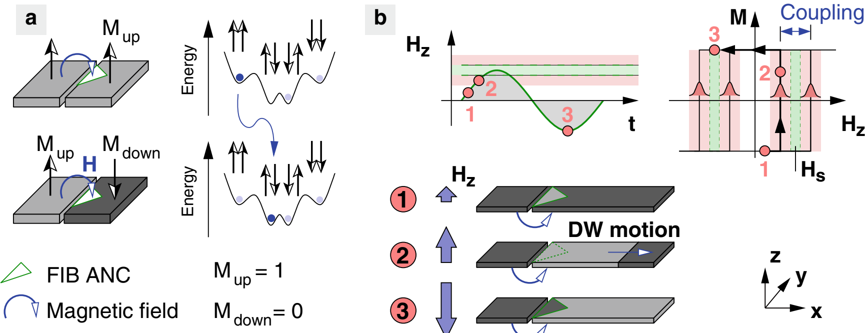


Fig. 3 The basic principle of pNML: Magnetic anisotropy is locally reduced by focused ion beam radiation, forming artificial nucleation centers (ANCs)—small soft-magnetic volumes—where coupling fields are acting strongly. Field-coupling between magnetic islands can be exploited for computation with stable ferromagnetic states as the lowest energy state is favored. **b** Ferromagnetic islands show distinct hysteretic behavior with full remanence during switching. Due to local ion irradiation on the left edge of a magnet, ANCs are formed and non-reciprocal signal flow (antiparallel direction of perpendicular magnetization) in an alternating magnetic clock-field can be achieved. Reprinted with permission from [19]

two neighboring magnets of equal shape are experiencing the same magnetic fields leading to an unwanted reciprocity. However, this can be overcome by (1) engineering the geometry or (2) controlling the local sensitivity of the magnets to the generated stray-fields. Method (2) was successfully employed by use of localized ion irradiation with Ga^+ ions [23]. The perpendicular magnetic anisotropy can be changed locally on length scales far below 100 nm. The focused ion radiation is intermixing the Co/Pt or Co/Ni interfaces such, that this magnetized volume is very likely to be reversed (switched) in an external magnetic field. It becomes soft-magnetic and by irradiating especially in close vicinity to neighboring islands where coupling (stray) fields are strong, control of directed signal propagation in complex pNML configurations is achieved. By incorporating method (1), i.e. maximizing the magnetic volume of close-by neighboring magnets, the coupling strength can be further enhanced. The invention of this method in 2011 [23] was the key for upcoming pNML device research at TUM.

The reversal mechanism of a ferromagnet, locally irradiated at one edge to form a so-called artificial nucleation center (ANC), is depicted in Fig. 3b. Three snapshots in time are given for the switching of a coupled pair of magnets from the parallel (high energy) state, to the anti-parallel (low energy) state. At the irradiated spot (ANC indicated by triangle), the magnetic anisotropy is reduced in a controlled way by focused ion beam (FIB) radiation. First, a negative field pulse in z-direction (not shown) saturated both magnets in 'down' direction for initialization. By applying a positive field pulse H_z, a domain is nucleated in the ANC and propagates until it is stuck at the edge of the irradiation spot as shown in Fig. 3b (1), where the DW experiences a step in magnetic anisotropy [24, 25]. In the irradiated (nucleation)

volume, all magnetic fields i.e. the stray field from the neighboring dot and the externally applied Zeeman clocking field H_z superpose.

Now, two scenarios are possible: (2) The stray field of the left magnet is *adding* to the Zeeman field, it supports to overcome the pinning barrier and the DW is immediately propagating through the magnet, fully reversing it. By contrast, the second scenario is shown in (3). The stray field of the left magnet is *opposing* the Zeeman field. Hence, the DW is not propagating as the energy barrier is *not* overcome. Consequently, the magnet is *not* switched back to the parallel state. In Fig. 3b, the effect of magnetic coupling is additionally visualized by means of a hysteresis loop of the switched magnet. By choosing a perpendicular field amplitude H_s which equals the switching field of a single magnet, a symmetric clocking window opens, where computation (=anti-parallel ordering) takes place. Additionally, a switching field distribution is sketched in the graph, which is due to thermally activated broadening of the switching field when overcoming the energy barrier (see e.g. [26, 27]). Please note, as long as field-coupling from the neighboring magnet is strong enough, reliable switching (from the parallel to the anti-parallel state) or suppression of switching (in case the states are already anti-parallel) is achieved. The unique behavior of unidirectional signal flow in an external Zeeman-field, is achieved by side-irradiation of magnetic islands with a focused ion beam. There have been related experiments, where a 'ratchet-like' behavior was achieved in an irradiated domain-wall conduit by Franken et al. [28].

In order to sum up the basic principle of NML operation, the ANC can be described as *summing* point for all generated stray fields in close vicinity and the externally applied Zeeman field. Depending on the leftmost neighbor, the magnet will switch into the anti-parrallel configuration, due to the ANC close to the edge of the island. Assuming that the ANC is infinitesimally small, magnetic islands can be described by lumped elements. Stray fields are calculated by the shape and the areal magnetic moment of the islands and those fields will act on the ANC, where the logic decision is made. The nonlinear switching behavior (highly non-linear hysteresis loop) will result in a sudden reversal of the magnetic domain, starting with a DW from the ANC and spreading over the whole magnetic island.

2.3 Major Building Blocks of Planar 2D pNML

By the introduced technique of localized ion irradiation for ANC creation and superposition of stray fields in the ANC, a 2D pNML family of logic devices has been demonstrated. The major building blocks are depicted as schematic layout and summarized in Fig. 4. The devices are briefly discussed in the following. The simplest structure is the binary state, an island with magnetization pointing parallel/antiparallel to the film normal. The logic values '0' and '1' are assigned to it. A chain of magnetic islands with unidirectional data flow in a spatially homogeneous clocking field (Zeeman field)—often referred to as inverter chain— has been demonstrated in Breitkreutz et al. in [23], Eichwald et al. in [29] and Kiermaier et

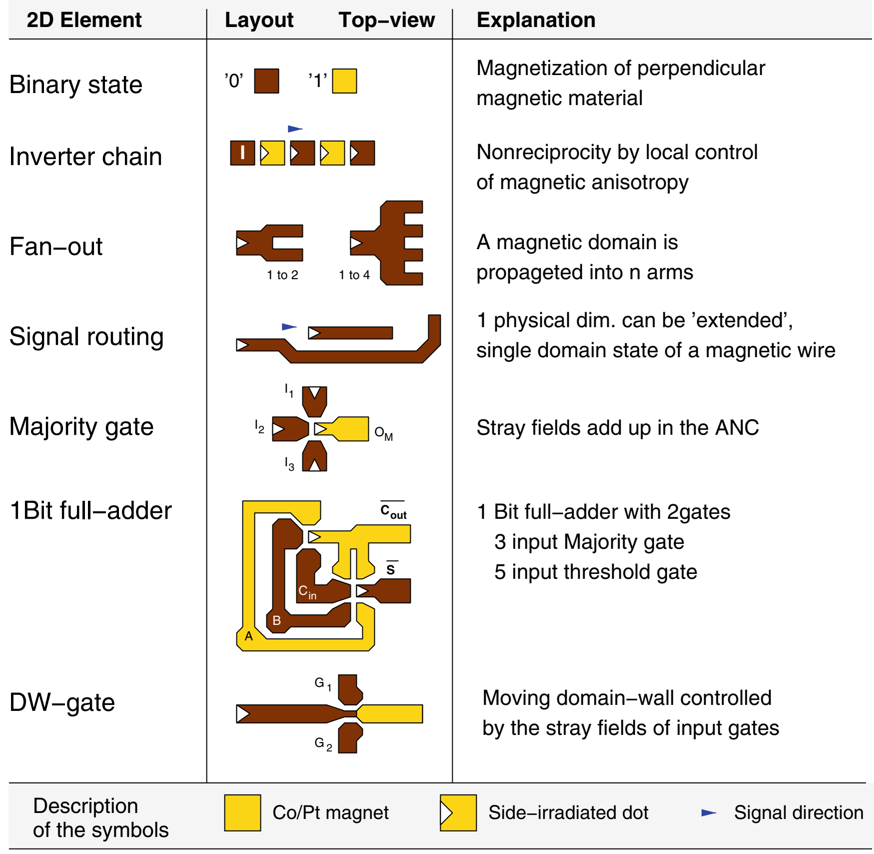


Fig. 4 Experimentally demonstrated elements for a 2D pNML logic family with typical sketch of the layout. Reprinted with permission from [19]

al. in [30]. Fan-out capability was experimentally demonstrated in Breitkreutz et al. in 2012 [25]. Inverter chain, fan-out and elongated magnetic islands—so called domain-wall conductors—are the basic elements for signal routing. It is worthwhile to mention, that for signal transmission in a pNML circuit, inverter chains are only used, when a signal *delay* has to be introduced, in other cases domain-wall conductors are preferable.

The *natural* gate of pNML is the *majority* decision with three input magnets and one output magnet, experimentally demonstrated in [16]. The output is aligned to the opposite of the majority of the inputs, therefore the name *Majority gate*. It is a universal gate which can be programmed to NAND or NOR functionality when *clamping* one input to a fixed state. The latter could be achieved by simply using hard-magnetic material (a non-irradiated island) or by programming one input with a current-wire or spin-valve structure. A full-adder employing a 5-input majority gate

(more generally a universal threshold gate) was theoretically predicted by Papp et al. in [31] and experimentally demonstrated by Breitkreutz et al. in [17]. It demonstrates the very efficient implementation of complex boolean functions in pNML, as in terms of magnet counts it is the most compact design: only 5 magnets for 3 inputs and 2 outputs are deployed.

In order to control the signal flow in DW conductors, two gate magnets close to a constriction are applied in the DW gate of Breitkreutz et al. in [32] and Breitkreutz et al. in [33]. By applying two DW gates in series, the structure is extended to a device similar to latches in conventional CMOS, which is an essential building block for data processing and synchronization. This is a very important finding, as it highlights that the nonvolatile computing state of a ferromagnet not automatically results in integrated memory devices.

2.4 Foundations for Digital Computation

The preceding section introduced the concept of digital computation in pNML together with experimental demonstration of the basic pNML building blocks. At this point it is adequate to raise the question, if the demonstrated pNML devices fulfill all prerequisites for digital computation. For that, the well known 5 *tenets* for digital computation systems [34] are adapted for pNML as follows:

1. *Nonlinear device characteristics.* Nonlinear hysteresis of the nanomagnetic switching, i.e. nucleation event, overcoming an energy barrier, rapid DW motion. The switching is first blocked by an energy barrier due to a step in magnetic anisotropy at the edges of the ANC. Suddenly, when overcoming the barrier, the magnetization of an island is reversed by a domain wall spreading out.
2. *Enable functionally complete Boolean logic set.* The majority gate [16] or threshold majority gate [17] fulfill this. With this device, NAND/NOR and hence inverters as well as all other functions can be implemented. However, we argue, that also latching behavior (integrated memory of non-volatile type) together with complex functions like the demonstrated full-adder shall be exploited for area, time and power efficient circuits and architectures.
3. *Power amplification (gain).* An external magnetic field together with nonlinear switching provides the energy. Ferromagnetism itself is an ensemble effect, refreshing the computing state due to exchange energy (magnetic spins tend to align in the same direction) [30]. Compared to electronic switches, magnetic states stabilize each other, and for switching, only tens of $k_B T$ are needed. By contrast, each electron in an electronic switch dissipates thousands of $k_B T$.
4. *Output of one device must drive another, fan-out.* This is realized by fork like branching of DW conductors (elongated magnets, stripes) [25]. Proper design of the fork-like structures are necessary in order not to pin a domain wall at the slits/notches, especially when switching at low fields and on short time-scales.

5. *Dataflow directionality.* The ANC is *breaking field coupling symmetry,* [23]. By this, properties of magnetic islands are locally changed in order to achieve signal directionality. As the ANC can be in the 10 nm-range when ultimately scaled, this method is prospective for high-density integration.

To the best of our knowledge, all fundamental properties for digital computation are fulfilled. But following the achievements in logic devices of the last decades it is obvious, that many more requirements are necessary to be listed as potential beyond CMOS candidate. Even worse, a single shortcoming might be sufficient not to be selected. To be more specific, due to [35] *further technological requirements* are: (1) room temperature or higher temperature operation, (2) low sensitivity to parameters (e.g. fabrication variations), (3) operational reliability, (4) CMOS architectural compatibility, (5) CMOS process compatibility and (6) comprehending intrinsic/extrinsic parasitic and their interface to interconnect.

For those requirements more detailed benchmarking in future is needed, in order to decide on pNML applicability in products. Our recent experimental investigations project, that all the requirements (1–6, see above) claimed in [35] can be fulfilled, but it is also clear, that they are hardly tackled by pure university research. A lot of engineering is needed, to achieve a mature computing systems for adding further functionality to CMOS circuits. At the bottom line, it becomes obvious that *further physical needs*, namely size (scalability), switching time (speed) and switching energy decide on whether pNML is competitive or not [35].

3 Device and Circuit Modeling on Different Levels of Abstraction

So far, the concept of 2D pNML was introduced and the experimental demonstration of the fundamental logic devices was summarized. In order to benchmark device performance and to extrapolate for larger circuits and systems, simulations on different levels of abstraction are needed. For that, a three step approach was found to be well suited. First, micro-magnetic finite difference simulations provide insight in the physics of magnetization reversal dynamics and domain-wall propagation of single pNML islands and devices. Second, together with magneto-optically characterized test-structures, switching distributions of magnets are extracted and simplified behavioral models are formulated. Third, the parameterized compact-models form the basis of Verilog-A simulations, utilizing the well-known strength of abstraction in system level simulation.

3.1 Micromagnetics with OOMMF Finite Difference LLG Solver

Commonly, small arrangements of nanomagnets are simulated in the *micromagnetic domain*, analyzing the dynamic behavior of macro-spins in applied fields by solving the Landau-Lifshitz-Gilbert equation (LLG). From such simulations—e.g. performed with the open-source LLG-solver OOMMF [36]—detailed knowledge on DW motion, reversal mechanism and stray field generation can be obtained. In the early work of Co/Pt research for NML application, it was found that a spatially varying anisotropy can be linked to focused ion radiation [22, 37, 38]. For that, experiments with Extraordinary-Hall effect sensors of Co/Pt bilayer stacks were compared with micromagnetic simulations in OOMMF [36]. Basically, the OOMMF simulation framework solves the 'Landau-Lifshitz-Gilbert' ordinary differential equation by means of finite difference methods. For NML films, the multilayer is treated as a 2-D effective media where the saturation magnetization M_s, anisotropy constant K_u, and exchange stiffness A_{exch} are kept constant in the direction perpendicular to the film plane. It was found, that in order to model FIB irradiation, the anisotropy *constant* is varied for the lateral dimensions and experimental data are taken to validate the model. The methodology is summarized in Fig. 5. A 5 bilayer Co/Pt film is investigated and compares (a) the polynomial fit of the measured coercivity, plotted over irradiation dose and (b) the linear fit of coercivity versus the anisotropy constant K_u as it was extracted from micromagnetic simulations. Both functions are monotonous and reversible, which allows to mathematically equate the coercivity from both fitted functions. This leads for a given technology exemplarily to a analytical mapping of

$$\frac{K_u}{[\text{J/m}^3]} = 1.957 \times 10^{-22} \cdot \frac{x_d^2}{[1/\text{cm}^4]} - 4.858 \times 10^{-9} \cdot \frac{x_d}{[1/\text{cm}^2]} + 3.388 \times 10^5, \quad (1)$$

with anisotropy constant K_u as a function of the experimentally measured areal irradiation dose x_d. Equation 1 is plotted in Fig. 5. It is valid in the boundaries (parameter space), where in this example both simulation and experiments where conducted. This equation can be adapted for different multilayer films and irradiation parameters as proposed in [39]. Therefore it is useful for mapping the ion dose to

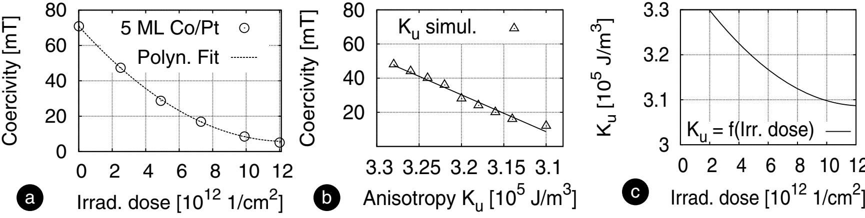


Fig. 5 **a** By measuring and **b** micromagnetically simulating the coercive field, **c** K_u can be described as a function of irradiation dose applied to a Co/Pt film. Results are adapted from [22]

the anisotropy constant for device and circuit simulation. However, this method has several shortcomings:

- In easy-axis hysteresis loops, defects in a film act as strong nucleation centers and the 'true' switching field cannot be attributed to the LLG-simulated values using K_u as 'fitting' parameter.
- The area of measurement is several 100 μm^2 whereas simulation area is a few μm^2 maximum.
- The time-scale of the experiment (several seconds in Hall-measurements) is not related to the 'virtual time-scales' in the simulation, where the LLG is solved by minimizing the energy term.
- The experiment is taken at room temperature, the simulations are at zero Kelvin, hence thermally distributed switching fields are not covered.

Soon after, it became clear in NML research that only a focused spot of ion radiation would solve both the non-reciprocity and effective field-coupling for a logic gate [23]. Further experiments showed, that different ion dose in a concentrated spot—called Artificial Nucleation Center (ANC) from now on—were significantly changing the switching field [25] . Franken et al. at the same time modeled and measured the behavior of a step-like anisotropy change generated by FIB patterning of Co/Pt strips [24] further confirming the findings of [25]. At this point it became obvious, that detailed ANC modeling would be necessary to map the distributions found in the experiment, and get reasonably low switching fields without using K_u as simple fitting parameter in the micromagnetic models.

3.2 ANC Modeling and Switching Field Distributions

Arrhenius-Type Models
In principle, the mean switching field of a nanomagnet can be described by the so-called Sharrock-formalism [40]. It basically says, that the mean switching field follows an Arrhenius law with increasing field amplitude for switching at short time-scales. The probability P_{sw} that the magnetization of a magnet is reversed in an external field is given by [26] to

$$P_{\mathrm{sw}}(t_p, B) = 1 - \exp\left(-t_p/\tau(B)\right) \tag{2}$$

$$\tau(B) = f_0^{-1} \exp\left(\frac{E_0 \cdot (1 - \frac{B}{B_{s0}})^2}{k_B T}\right) \tag{3}$$

with $\tau(B)$: inverse of the switching rate, $B = \mu_0 H$: magnetic induction, t_p: pulse time, $f_0 = 2 \times 10^9$ Hz: reversal attempt frequency, E_0: energy barrier at zero field,

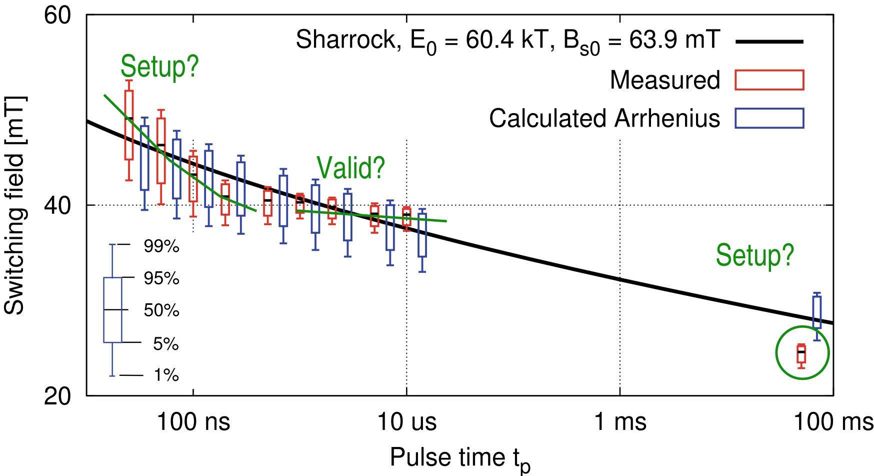


Fig. 6 Switching field distributions and Sharrock plot for a switching experiment measured by MOKE. Assumed inaccuracies of the setup marked in green color. Graph adapted from [27]

μ_0: magnetic vacuum permeability, B_{s0}: magnetic switching field[2] at zero temperature, k_B: Boltzmann constant and $T = 293\,\mathrm{K}$: ambient temperature.

Setting the switching probability in Eq. 2 to 0.5, Eqs. 2 and 3 can be reformulated to a switching *field* B_{sw} for various pulse times t_p as expressed by the Sharrock equation [41]

$$B_{\mathrm{sw}} = B_{s0}\left[1 - \left(\frac{k_B T}{E_0} ln\left(\frac{f_0 t_p}{ln(2)}\right)\right)^{1/2}\right] \tag{4}$$

with the parameters as given in Eqs. (2) and (3). The Sharrock equation is derived from the ideal case of a so-called *Stoner-Wohlfarth particle*. However, here the model is applied to a switching process, where a nucleation event triggers the reversal by domain wall motion. It describes the increase in mean switching field on shorter time-scales, ideally over several orders of magnitude.

Figure 6 summarizes measurements for pulsed switching probabilities on micron-sized nanomagnets with partial ion radiation as described in [27] and extended in [42]. In the experiment, off-chip generated field pulses with $t_p = 50\,\mathrm{ms}$ and on-chip generated pulse times ranging from $25\,\mathrm{ns} \le t_p \le 10\,\mu\mathrm{s}$ are applied. 100 switching experiments per field-pulse amplitude are carried out from which the probability for switching is extracted. The measured values are fitted by Eq. (2) with B_{s0} and E_0 as free parameters and the quantiles are plotted in red. Additionally, Fig. 6 reproduces

[2] For brevity, the magnetic flux density B is denoted by a *'field'* as it is easily converted to a magnetic field H with the equation $H = B/\mu_0$.

the Sharrock plot (black line) for the averaged B_{s0} and E_0 value. The blue quantiles show the numerical evaluation of the Arrhenius law of Eq. 3.

It is obvious, that for pulse times in the deep sub-micron range below $t_p \leq 200$ ns the Sharrock law does not fit the experimental findings. A rather steep increase of the mean switching field was found, highlighted by the green line as guide for the eye and not being in good agreement with a fixed parameter set of the Sharrock equation. However, at that time of the experiment it was attributed to the measurement setup which showed inaccuracies in the pulse width for short timescale and the Arrhenius model was in good agreement for medium pulse times, reproducing also the switching field distributions to a reasonable extent. Nevertheless, the results were not satisfying and after thorough refinement of the setup, the experiment was repeated by Ziemys et al. [43]. Figure 7 shows the investigated nanomagnets with FIB created ANC and field-pulses provided by on-chip inductors. The experimental time-scales were extended from $20\,\text{ns} \leq t_p \leq 200\,\mu\text{s}$ with improved on-chip coils (scaled footprint, better alignment of the structures) and from $30\,\text{ms} \leq t_p \leq 1$ s with an off-chip inductor. With that, eight orders of magnitude in pulse time for magnet switching were experimentally covered. For pulse times $t > 100$ ns the standard Arrhenius model (blue line in Fig. 8) with the following parameters was found to best fit the experimental data; $\text{E}_0 = 33.5 \cdot \text{k}_\text{b}\text{T}$, $\text{B}_0 = 30\text{mT}$ and $\text{f}_0 = 2$ GHz.

But, the experiment manifested the previous finding that at time-scales $t < 100$ ns, the standard Arrhenius model with thermally excited switching does not apply. For that time-scales, the theoretical predictions in [45] are able to explain the sudden steep increase in switching field and are modeled with a *dynamic* temperature independent time constant τ.

Even though the artificial nucleation center is far from matching the theoretical idealized model of small ferromagnetic ellipsoidal grains that were used to derive the dynamic time constant in [45], the analytical model is in good agreement with the experimental data of [44]. This might be due to the fact, that the weakest *grain* in the ANC acts as a nucleation volume from which a domain wall is starting to propagate and fully reversing the magnetic island.

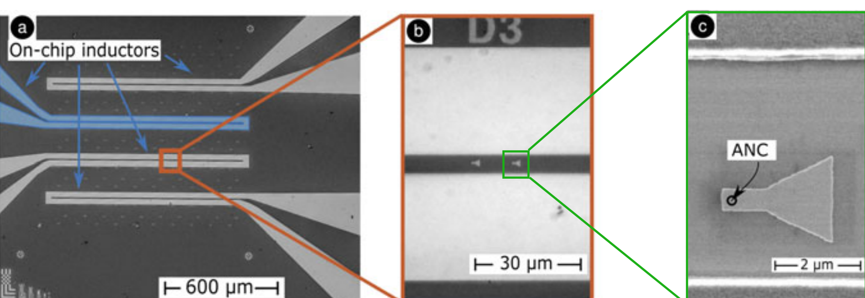


Fig. 7 **a** SEM image of on-chip coils for pulsed experiments. **b** Zoom-in with the position of the investigated magnetic islands and **c** high-resolution image of one magnetic island with ANC position, overlayed by black circle. The graphs are adapted from AIP [44]

But another important aspect became obvious: the measured time-scales, where dynamic switching is starting to dominate, is still somewhat surprising. One would expect dynamic effects for ANC reversal to take place far below 10 ns and one can speculate, that either damping in the investigated Co/Pt stacks is very high and/or domain-wall-speed in the device-under-test (DUT) is 'hiding' the dynamics of ANC reversal. Recapitulating, that the MOKE laser beam is integrating over the whole magnet area, DW-speed indeed becomes important when the magnet is pulsed at short-time scales. Special care was taken to apply a magnetic *propagation field* slightly lower than the effective switching field of the DUT in order to fully propagate the DW throughout the magnet after the short field pulse.

However, this method turned out to be insufficient, most likely due to unexpectedly large domain-wall pinning. Hence, the results for dynamic reversal depicted in Fig. 8 were revisited in consecutive measurements and are discussed in the following. Assuming, that 50% of magnet reversal was detected as switching event in the measurement, the DW would have to travel more than 1 μm as estimated from Fig. 7. For the pulse width of 100 ns, at which the steep increase of switching field starts, a DW-speed of $v_{\mathrm{DW}} = 10\,\mathrm{m\,s^{-1}}$—which is a typical speed for the Co/Pt films—must be assumed. As a consequence, the dynamic-switching model *at this time-scale* was an artifact from measurements, not stemming from the ANC reversal but rather from limited DW propagation-speed and it could not be detected in Laser-MOKE mea-

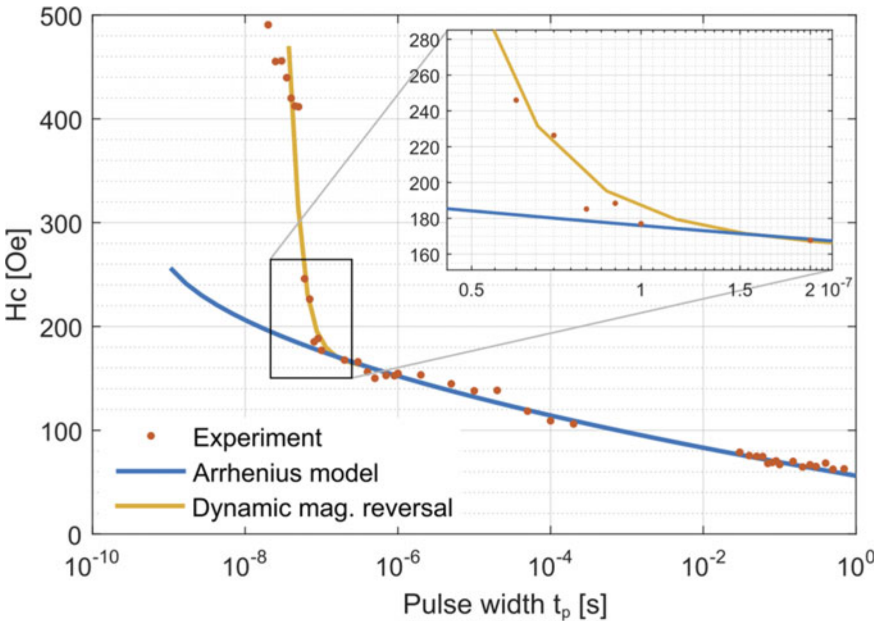


Fig. 8 Increase of switching field at small timescales as measured in pulsing experiments of a magnetic island with ANC. Fit-curve of two applied models for longer (blue) and shorter (orange) timescales. Reprinted with permission from [44]

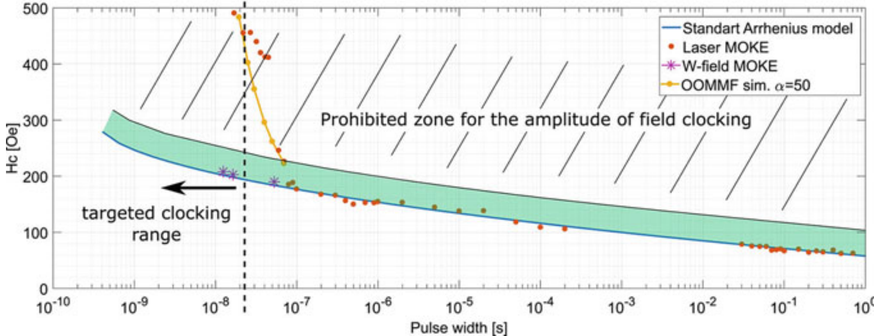


Fig. 9 Switching field measured with wide-field MOKE at small timescales (violet). Fit-curve of the Arrhenius model (blue fit) over 8 order of magnitude. The *dynamic extension* as suspected in earlier experiments turned out to be not applicable (orange fit). Reprinted with permission from [46]

surements. Instead, wide-field MOKE measurements with higher spatial resolution were performed and the results replotted in violet color ($*$) in Fig. 9. The switching amplitude for shorter time-scales between 20 ns $\leq t_p \leq$ 50 ns still follow strictly the Arrhenius-type reversal, hence the ANC reversal *is* Arrhenius-like, whereas DW-motion becomes the bottleneck for the switching speed of a Co/Pt island of this size. From this discussion, it is clear that on such short time-scales special care has to be taken to explain the switching of ANC modified Co/Pt devices as recently reported in [46]. As optical methods are limited in terms of spatial resolution, magnetic force microscopy techniques could complement switching experiments, even though optical techniques are indispensable for switching field statistics, due to their measurement speed.

To conclude, we do not expect dynamic reversal beyond the Arrhenius-type models to come into play when aiming higher clocking frequencies up to $\approx$100 MHz which might play a role when exploring materials with faster domain-wall speed like e.g. in CoNi and CoFeB. For very high frequencies above $\approx$1 GHz, the modified Arrhenius (with dynamic contribution) should be reconsidered and can provide a simple behavioral model when extracting the experimental data for a given ANC, or more general, an investigated fabrication technology.

Thermal Micromagnetic Simulations

The presented results of pulsing experiments and their analytical modeling showed good agreement for the mean switching field on nanosecond timescales. However, the measurements are using the non-volatility of the magnets rather than resolving the dynamics of magnetization reversal. For that, far more sensitive sensing methods would be necessary. But both, spatial and time resolution are not easily available, especially when talking about statistically relevant distributions in deep-submicron ANCs—those experiments are simply to costly.

In order to look at the underlying time evolution of magnetization during the pulsing experiments, micromagnetic simulations are the method of choice. By using the thermal module of [36] implemented by the Wiesendanger group [47], which adds a *stochastic differential equation of the Langevin type*, extensive simulations were performed, mapping the pulsing experiment in reasonable good agreement. For that, both the time-scale of up to hundreds of nano-seconds and the area of the ANC under investigation, are to be modeled. In order to extract statistically relevant data, at least 100 repeated simulations per pulsing amplitude are performed. The *computing cost* for one simulation run are several hours on a single computing core, hence the *parallel-script* [48] was beneficially used to distribute the large amount of stochastical simulations on far more than 100 cores in parallel.

In the following, a typical procedure for combined in-plane and out-of-plane pulses for thermal ANC switching simulations is reviewed [49]. Figure 10a shows the ANC-model that consists of a spatially varying perpendicular anisotropy K_u. Combined out-of-plane (B_z)- and in-plane (B_x)-pulses are 'applied' as depicted in Fig. 10c whereas the in-plane (B_x-field) is kept at 30% of the out-of-plane B_z-field amplitude. A time-constant of $\tau = 2$ ns is chosen to reproduce the rise time of the pulsing current in the experiment. The z-pulse is effective over 100 ns, whereas the in-plane pulse is starting after 60 ns with a pulse-width of 20 ns.

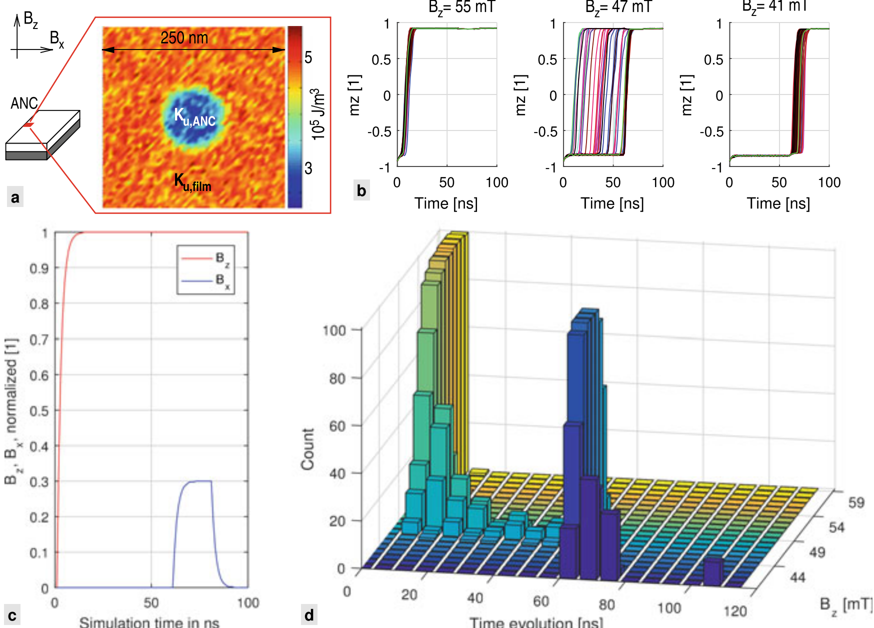


Fig. 10 **a** Spatially varying K_u for ANC modeling. **b** Time-evolution of the normalized magnetization in z-direction for distinct B_z amplitudes *applied* in the simulation. **c** Timing of the B_x-B_z-pulses in the simulation. **d** Switching counts over time for the interval $40\,\text{mT} \leq B_z \leq 59\,\text{mT}$

Typical results of OOMMF thermal micromagnetic simulations are depicted in Fig. 10b. The magnetization is recorded for each simulation result and each magnet reversal is evaluated in terms of switching time. For high B_z pulses (55 mT), the switching immediately occurs during the rising z-field pulse. Medium B_z pulses (47 mT) have a broadly distributed switching time from the beginning of the z-pulse up until the x-field is applied. For low B_z pulses (41 mT), the in-plane field pulse is most effective and the switching time can be as narrow as the in-plane pulse in x-direction. This means, that in-plane pulses can very effectively reduce the distribution and can be used for synchronous switching of the ANC. Figure 10d is summarizing the results for increasing B_z-pulse amplitude ($40\,\mathrm{mT} \leq B_z \leq 59\,\mathrm{mT}$) with narrowest switching field distribution for $42\,\mathrm{mT} \leq B_z \leq 45\,\mathrm{mT}$. This is of great importance for ANC operation: For memory applications, one could of course use pulses with highest B_z amplitude, as only highest switching probability is needed. However, for logic applications in field-coupled devices, the range of medium or low B_z amplitudes are of great importance, as in those ranges field coupling can be applied most effectively. Please note, the absolute values discussed do not present a general physical rule but rather a specific technology that was calibrated with a certain type of Co/Pt bilayer stack. It demonstrated the effectiveness of in-plane pulses for ANC reversal with spatially varying perpendicular anisotropy as experimentally demonstrated in [50]. The vector nature of the applied field-pulses are of vital importance for device optimization, as neither the local coupling-fields nor the externally applied Zeeman-fields can be reduced to simple out-of-plane fields. Instead the switching amplitudes can be influenced by in-plane fields and there is research on-the-way which refines the ANC switching models by a combination of in-plane and out-of-plane pulses with good control over pulse-timing. To summarize the state-of-the-art in ANC modeling, the following procedure turned out to be practicable:

1. Measure film anisotropy with common methods like the Extraordinary Hall-effect and rotating magnetic field [51]. This gives an upper limit for the anisotropy of a scaled NML device.
2. Fabricate small islands and create an ANC, measure the switching field on small time-scales for field-pulses, possibly in in-plane and and out-of-plane-direction.
3. Match the switching field by adopting A_{exch} and K_u-profile in the ANC and simulate at the same time-scales. ANC dimensions are estimated from the software collection Stopping and Range of Ions in Matter (SRIM) simulations or SEM-images.
4. Thermal OOMMF-simulations [36] with a Langevin term provide distributions and comparison to the experiment on the nanosecond time-scale.
5. Compare both experiment and micromagnetic simulation.
6. Parameterize possible analytical descriptions of the nucleation and reversal process.

3.3 Verilog-A Models

With micromagnetic solvers, one is able to understand switching behavior in quite good detail. However, such solvers are not capable for simulations of large pNML systems due to the time-intensive evaluation of the LLG-equation. For that reason, we developed physical-based compact models (as these are commonly used in circuit design) in order to mimic the behavior of larger pNML arrangements [52, 53].

By micromagnetic modeling and accurate parameter extraction from sample structures in the experiment, e.g. magnetic layer composition, island geometries, local ion irradiation, etc. (see e.g. [54] for measurement and modeling of the nucleation probability), the full design space for pNML can be explored and a basis for benchmarking against other emerging computing devices is feasible. In [55] we proposed a compact model for pNML devices as depicted in Fig. 11. The pNML inverter and M-gate are split into four parts, the field-sum is evaluated by adding the external clocking field H_{ext} with the input coupling fields H_{in}. These fields are acting in the ANC, which is modeled by the Arrhenius-law including the switching field distributions as explained in the preceding paragraphs. After nucleation, the signal is propagated via a domain-wall, excited by the external clocking field H_{ext}, whereas the domain-wall speed is modeled for the regimes Creep—Depinning—Flow—Saturation as a piecewise defined function of field-amplitude [56]. As a last component, an output field is generated by the so-called field-generator, which mainly depends on the geometry of ferromagnetic material surrounding the consecutive ANC. In order to simplify stray-field generation, a simple 2D torus segment is chosen for field-calculation, giving a reasonably good approximation [55].

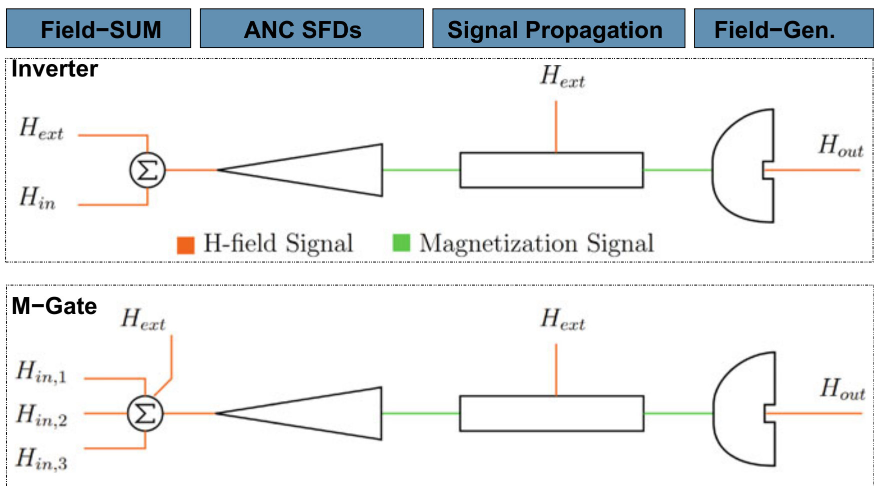


Fig. 11 Behavioral model for pNML Inverter and Majority-gate. The device is split into four separate parts: field-sum, ANC switching field distributions (ANC SFDs), signal propagation and coupling field generator

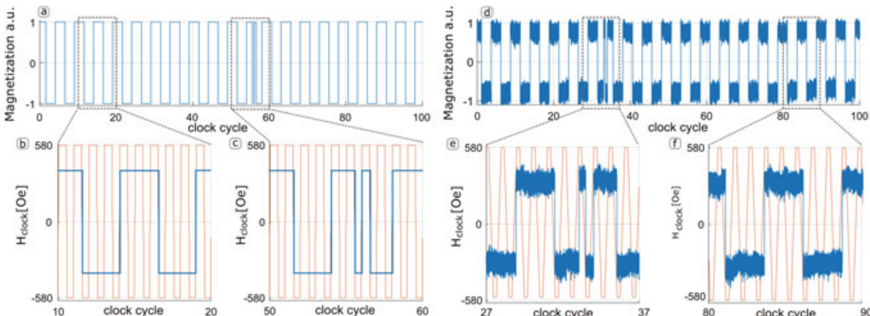


Fig. 12 Cadence Virtuose compact modeling **a**–**c** and experimental data **d**–**f** of a 5:1 pNML clock divider. The time signal of the magnetic clocking field (orange) and the binary state of one magnet in the loop (blue) are compared. Reprinted with permission from [44]

By using the compact models presented in this chapter, pNML devices and circuits can be simulated as shown in [57] for a 3D programmable ALU. Even more, they are suitable for benchmarking purposes against other emerging research devices [58]. A comparison of an experimentally demonstrated 1:5 pNML clock-divider circuit [30] and the compact modeling as performed in [44, 55] is shown in Fig. 12. A magnetic field-clock (orange) is applied to a loop of 5 in series connected pNML inverters. At one inverter, the magnetic binary state is recorded (blue). Both, the Verilog-A simulation in Fig. 12a, b and the experimental data in Fig. 12d, f show correct operation by dividing the clocking frequency by a factor of 5. However, there are also switching errors Fig. 12c, e which are occurring both in simulation and experiment.

There is further need for compact modeling of pNML devices, as they are the foundation for system and architectural level simulation. We claim, that it is mandatory to investigate appropriate architectures for pNML. For example, it is known, that systolic architectures are very attractive coming along with much less wiring and at the same time profiting from non-volatile computing states [57, 59–62]. Furthermore, stochastic computing circuits in NML are investigated for NML [63] and architectural explorations are employed [64]. This has to be kept in mind when working on emerging technologies: from the physical device up to the circuit, architecture and application level, a tailored *environment* for efficient operation has to be provided.

4 A Monolithically Integrated 3D Computing Circuit in pNML

In this section, pNML as back-end-of-line (BEOL) 3D technology is proposed and a summary on demonstrated 3D devices in pNML is given. Implementations and challenges of electrical I/O and signal routing in 3D structures are addressed. In order to operate a pNML circuit, the need of a magnetic field-clock as power supply is discussed.

4.1 Complicated Signal Routing in 2D Arrangements

One of the most important aspect in 3D integration is signal routing. In principle, electrical interconnects are superfluous in pNML as computation is executed by field-coupling. The concept of field-coupling for computation is very much related to cellular automata where only *next neighbors* are interacting for computing operations. However, as DW conductors are used for signal routing, a 2D pNML layout is very much complicating signal distribution, as line crossings are not possible. By contrast, in CMOS, numerous metal layers allow for high flexibility and there are interconnects on short-, medium- and long-range scale, organized in a strict hierarchical manner. This is simply not feasible in 2D nanomagnetic logic circuits. But in principal, as magnetic coupling fields are acting in all 3 dimensions of space, it is straight forward to extend the 2D pNML system to logic gates and signals propagating in 3D arrangements.

The basic idea to boost pNML technology is true monolithic 3D integration. It means that 3D arrangements of pNML devices are interacting in a dense configuration but at the same time keeping the benefits of a top down planar technology. The question arises, if coupling fields could be used in both directions, namely x-y (in-plane) and in z (out-of-plane.) Fig. 13b shows the basic idea for 3D integration of pNML devices by exploiting the *vertical* field coupling of magnetic islands. By sketching the cross-section of three magnetic islands, assumed to be separated by a dielectric material, ANCs can be placed in such a way, that the stray field of magnet #1 is acting on the ANC of magnet #2 such, that magnetization will align in *parallel* with magnet #1. Of course, there is a time delay, as first magnet #1 will reverse the nucleation volume of the ANC, followed by a domain-wall propagation through magnet #1. Afterwards—but in the same field-clocking pulse—the ANC of magnet #2 is reversed followed by domain-wall motion through magnet #2. Next, ANC volume magnetization of magnet #3 reverses followed by domain-wall motion through magnet #3. Overall, by a single external field-pulse of sufficient amplitude, the magnetization of all three magnets is aligned in the same direction meaning magnetic

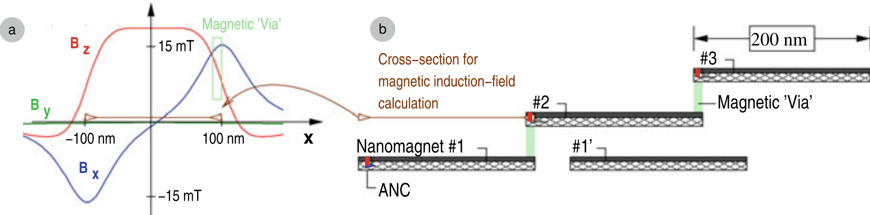


Fig. 13 The idea of vertical field-coupling in pNML layers separated by a planarizing spin-on dielectric. **a** Calculated induction field values at the section of graph b), 40 nm above the surface of nanomagnet #1. **b** Cross-section of the arrangement of NML devices leading to field-coupled vias and vertical signal transmission from magnet #1 via magnet #2 to magnet #3. Magnet #1' has no ANC and is not influenced by the neighboring islands

information is propagating in z-direction without having vertically conductive channels nor magnetic material deposited for the via. Simply speaking, the field coupling is generating a so called (virtual) *magnetic via*, transferring information from layer 1 to layer 2 and to layer 3.

It is important to mention, that magnet #1' is not affected, as long as there is no ANC fabricated. Please note the difference of magnets in the same layer to magnets coupled with a magnetic via in vertical direction. The magnetization of magnets in the same layer couples in anti-parallel manner and alignment happens in the following field-pulse (opposite direction), whereas the magnetization of vertically stacked magnets are reversed within the same clocking pulse. Hence, for signal propagation in vertical direction, a single clock-pulse is sufficient, but for logic operation with neighboring magnets, a consecutive (opposite) field pulse is mandatory.

In order to estimate the field amplitudes of vertical coupling, Fig. 13a exemplarily shows the calculated magnetic induction fields 40 nm above the crossection of a magnetic island. At the position of the ANC (highlighted in light green), there are coupling fields of strong amplitude acting in both z- and x-direction. They are oriented in a way, that an ANC above the magnet can be easily reversed as described above. There is another important aspect visible in this simple calculations, namely that a precise (and optimized) position of the ANC is vital for correct operation, as combined B_x and B_z fraction of amplitude massively influence the reversal field of magnetization. At this point it becomes obvious, that the demands for fabrication technology are high and a wisely chosen ANC position in the direction closer to the origin, where the z-field is more constant turns out to be more robust.

4.2 Details on 3D NML Fabrication

A closer look at an employed fabrication technique for 3D stacked NML devices is presented in Fig. 14. In principle, fabrication technology seems not to strongly differ from the standard 2D layout of magnetic islands, however there are some details to be considered:

- The seed layer is very important for magnetic thin-film growth, hence both magnetic layer stacks should be grown on identical dielectric layers. Hydrogen silsesquioxane (HSQ, [65]) has been found to be a promising material, as it can be spun at low temperatures and at the same time planarizes the surface after consecutive processing steps.
- Standard ANC fabrication is done at 50 kV acceleration voltage, depth control of the straggling ions has to be considered in order to only pattern the top-most magnetic layer.
- Alignment of four consecutive processing steps—(1) Co/Pt island patterning layer 1, (2) ANC formation layer 1, (3) Co/Pt island patterning layer 2, (4) ANC formation layer 2—is vital for device operation. It is hence beneficial to perform all

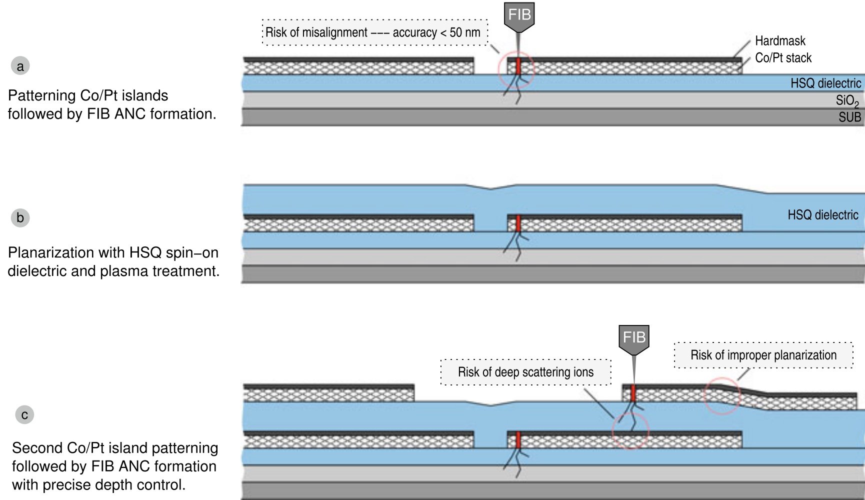


Fig. 14 Manufacturing process of 3D NML devices: **a** First pNML functional layer, followed by **b** HSQ planarization and **c** second NML functional layer fabrication. Especially the risk of deep scattering ions and insufficient planarization are to be considered

four alignment steps for FIB lithography/direct patterning in the same fabrication tool with integrated imaging capability.

In Fig. 14a, first a HSQ dielectric layer is spun on an oxidized silicon substrate for the formation of Co/Pt islands by FIB lithography followed by a mild Ar-ion physical etching process. In a second step, the FIB ANCs are patterned with standard focused ion radiation. After the first NML device layer, the magnetic islands are planarized with HSQ as shown in Fig. 14b. The HSQ is baked at a hotplate at $T = 225\,°\mathrm{C}$ or alternatively exposed to an O_2 plasma for *hardening*. As the trenches between the magnetic islands are in the range of 10 nm–15 nm, an optional chemical-mechanical polishing (CMP) step is avoided still giving sufficient planarity for the fabrication of the second NML device layer. Figure 14c is showing two risks that might deteriorate device operation. First, deep scattering ions that are not sufficiently *stopped* in vertical direction and second, improper planarization of HSQ leading to rough surfaces or wedge-like surfaces above edges at the buried islands.

Whereas the planarization was not found to be critical, the first risk of straggling ions was intensely studied for 3D pNML research. The well-established simulation tool Stopping and Range of Ions in Matter (SRIM) [66, 67] was used to identify film configurations, that could serve both needs, namely a thin planarization layer for strong vertical magnetic field-coupling and sufficient stopping of ions to secure buried magnetic layers from being irradiated. Figure 15a sketches the cross-section of 2 Co/Pt NML layers separated by a HSQ dielectric. For simulation, the thickness of the Titanium hard-mask is varied between (b) $t_{Ti} = 3\,\mathrm{nm}$ and (c) $t_{Ti} = 8\,\mathrm{nm}$. By comparing the results for 50 ions of Ga^+ at 50 kV (typical values for FIB ANC

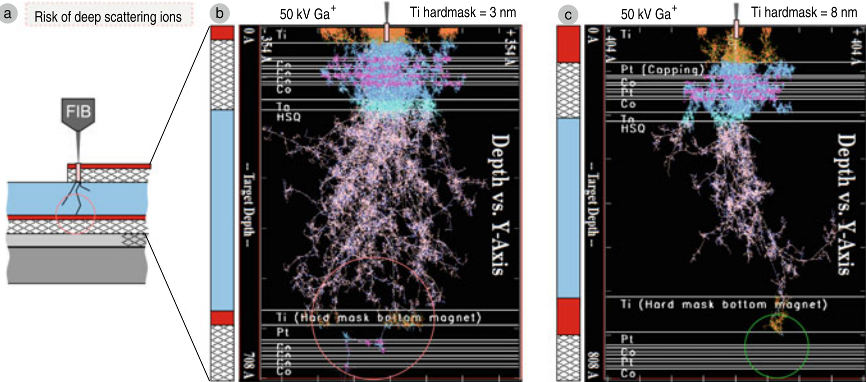


Fig. 15 **a** Cross-section of 3D Co/Pt layer stacks as simulated in SRIM. By changing the thickness of the Ti-hard mask from **b** t_{Ti} =3 nm to **c** t_{Ti} =8 nm, the amount of deep scattering ions can be significantly reduced. Adapted from [68]

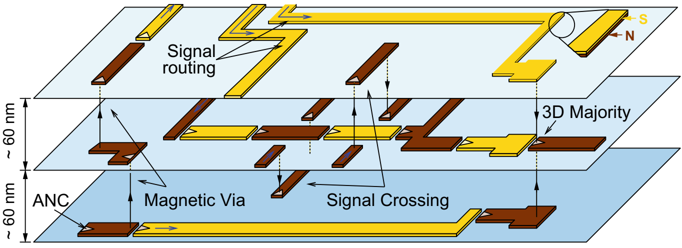


Fig. 16 Layout of a pNML system arranged in 3D. Ferromagnetic islands and stripes, incorporated into planar, dielectric films of ≈ 60 nm distance. Magnetic vias, signal crossing and 3D majority gates can be implemented. The typical layer separation in vertical direction is in the 60 nm range

creation in those stacks), there are few ions penetrating the lower Co/Pt stack and possibly changing the magnetic properties (Fig. 15b). By contrast, for the 8 nm Ti-layer, the ions are fully stopped in the thicker hard-mask of both layers (Fig. 15b), preventing deep scattering ions in the lower magnetic film stack as analyzed in detail [68]. From such simulations it becomes clear, that ion irradiation has to be adapted whenever changing the magnetic layer composition, even more in 3D arrangements. The simple change of hard-mask thickness already leads to significant changes of the magnetic patterning parameters in pNML device fabrication.

In order to give an idea of 3D integrated pNML devices, Fig. 16 sketches three stacked NML layers incorporating magnetic islands for signal routing, stacked field-coupled vias, signal crossing elements and a 3D majority gate. Please note, that the structure is highly planar but at the same time allows for monolithic 3D integration.

4.3 Major Building Blocks of 3D pNML

In the following, the state-of-the-art in 3D pNML devices is briefly reviewed. Figure 17 gives an overview on the recently demonstrated 3D devices. All demonstrated 3D devices have in common that two planar stacked layers of pNML are separated by a thin dielectric planarizing layer with thicknesses in the tens of nanometer range. As explained before, the fabrication is subdivided in 3 major steps, (1) the first layer of pNML is fabricated, (2) the planarizing dielectric is deposited, (3) the second layer of pNML is deposited and aligned with respect to the first layer.

A magnetic via consists of two stacked and laterally shifted magnets. Depending on the position of coupling ANC either in the top or bottom magnet, the signal propagates in vertical up- or down-direction. Both magnets couple by magnetic fields in a way that the magnets are switched in consecutive manner, but within the same clocking pulse [69]. This is very important for signal distribution on chip, as only the delay of nucleation and domain-wall motion is summing up, but switching occurs in the same clocking pulse. There is no additional clock-delay introduced, as e.g. in 2D inverter structures. With this method, signals can in principle travel over several layers up or down 3D NML circuits. Incorporating two magnetic vias

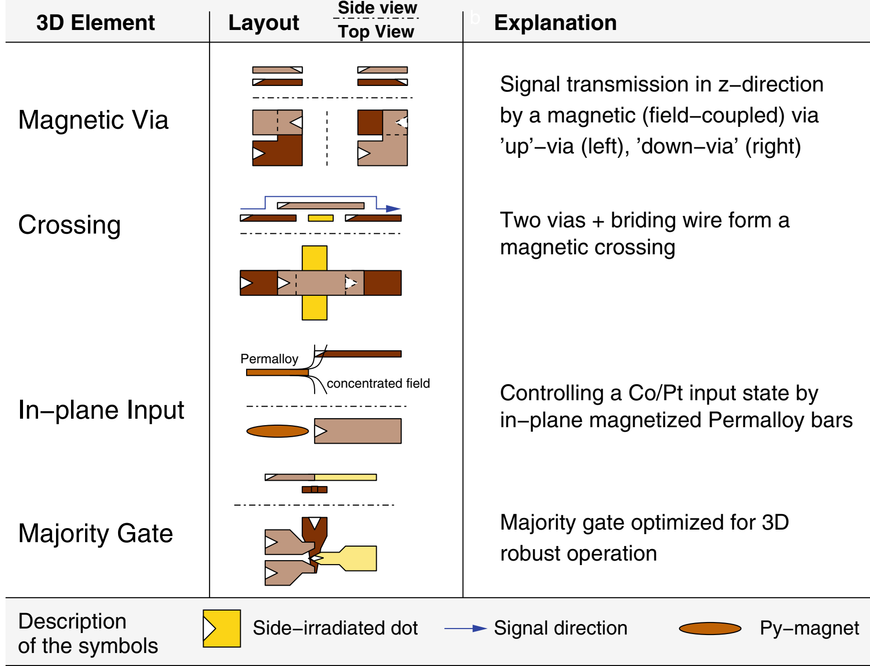


Fig. 17 Experimentally demonstrated elements of a 3D pNML logic family with typical sketch of the layout and cross-section

of opposite directionality (down-up, up-down) together with a stripe shaped magnet, a magnetic signal crossing is achieved [70]. With crossing devices in 3D, the layout of circuits can be very much simplified compared to 2D layouts. In summary, binary states can be propagated within one clocking pulse (due to ferromagnetic coupling in perpendicular direction) over several layers of pNML in positive or negative z-direction within one clocking pulse. A similar functionality as provided by electrical interconnects in CMOS devices providing high flexibility in circuit design.

Another viable device is the so-called soft-magnetic input, where a Permalloy magnet of ellipsoidal shape is setting (programming) a Co/Pt input magnet [71]. It is able to transform and at the same time *focus* or *concentrate* globally applied in-plane fields to out-of-plane fields, acting in a local manner and reprogramming Co/Pt devices i.e. to switch a M-Gate from NAND to NOR behavior and vice versa during runtime. Further use of the Permalloy structures is envisioned as field-concentrator for electrical input wires or as cladding material of those.

The 2D majority gate is transformed to a real 3D gate by positioning one input into a second pNML plane [72]. The most robust M-gate would use 3 active layers with maximum coupling-field area, with the drawback of a more demanding fabrication of three active layers. In principle, the demonstrated 2D 5-input majority gate could be further optimized for multiple active layers in 3D. With that, robust operation and a more compact design is achieved. As an example the reader is referred to a 3D pNML Arithmetic-Logic-Unit (ALU) optimized for compact design in a 3D implementation [57].

5 A Co-processing Unit as Back-End of Line Technology for pNML

In this section, the vision of a 3D pNML co-processing unit is discussed including electrical I/O and the integrated magnetic power clock as *power supply* for pNML circuits.

5.1 pNML as a Technology for the BEOL

For a monolithically integrated 3D NML system, the 2D pNML planar circuits are stacked in the BEOL of a standard CMOS process as sketched in Fig. 18. For pNML, there is no semiconducting crystalline substrate needed, only a highly planarized and defect free underlayer for seed layer and magnetic thin film growth is mandatory. This can be easily implemented in a process that already incorporates magnetic RAM (MRAM) for non-volatile on-chip memory heavily researched in these days. A co-processor design is targeted, where the main processor is a general purpose CMOS IC with electrical to magnetic interfaces (e.g. giant-magneto-resistance (GMR) devices,

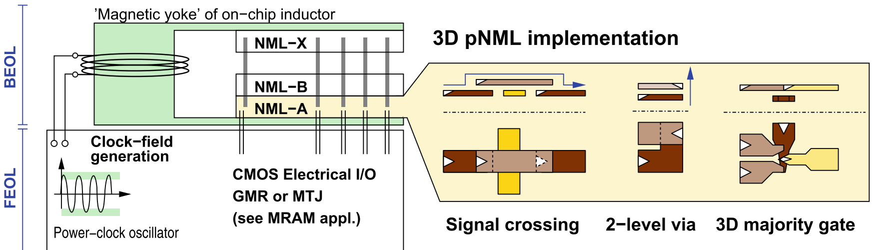


Fig. 18 The co-processor design of a 3D pNML computing system. On top of a CMOS main processor chip (front-end-of-line (FEOL)), pNML is fabricated in 3 dimensions in the BEOL. As the pNML co-processor is *power-supplied* by a magnetic clocking field, electrical interconnects are needed for I/O and field-clock generation circuitry (ferromagnetic 'yoke' of on-chip-inductor) only. The co-processing unit in pNML can be tailored for specialized tasks, e.g. pattern-matching, parallel and massively pipelined data processing. Reprinted with permission from [19]

magnetic-tunnel-junction (MTJ) devices as known from magnetic RAM technology). Electrical wiring is reduced to a minimum as computation takes place via field-coupling and magnetic signals are moved along extended DW conductors.

Further, a clock-field generation circuit for logic operation in the pNML circuits has to be implemented. The magnetic fields are generated by harmonically driven on-chip inductors. A ferromagnetic yoke (on-chip inductor with soft-magnetic cladding) is exciting and power-supplying the 3D pNML system in the BEOL with an alternating field amplitude. The 3D stacking of functional computing layers (up to 10 or more layers are expected to be a reasonable stacking height) with non-volatile computing states provides signal crossings and vertical signal flow in an easy way without need of metallic wiring.

5.2 Electrical I/O

For interfacing the pNML chip with CMOS circuitry, different electrical I/O concepts have already been investigated.

For electrical input we proposed and demonstrated a current carrying wire close to the ANC of the input magnet [73]. As the distance of the wire can be chosen to the tens of nanometer range, small field pulses are sufficient to support switching of the input (or suppression of switching). Please note, this is different to toggle-MRAM, where every storage cell has to be addressed in a cross-bar structure. For high data-throughput of pNML devices, repetitive pulses on the tens of nanosecond time-scale have to be generated and synchronized with the applied field-clock for electrical input. As in pNML not more than several hundreds of electrical inputs are envisioned, circuit complexity for addressing is not a critical issue. It is worth to mention that other switching mechanisms for electrical input (e.g. the STT-effect,

Spin-Hall-effect or switching with multiferroic materials) could also be suited for electrical input, but have not yet been in the scope of our work.

Regarding the electrical output, the Extraordinary-Hall effect (EHE) has been investigated with a special sensing geometry [74–76]. In the so-called split-current arrangement, three current contacts are needed to directly contact the output magnet of a pNML circuit. The side of the magnetic strip is free for incorporating an ANC for field-coupling to its next neighbor. An advantage is, that small volume magnets can be sensed and the output signal is proportional to the magnetization, but the EHE-effect is very likely too small for industrial applications with high data-throughput. Instead, MTJ or GMR sensors (see e.g. [77, 78]), either as remote stray-field sensors, or integrated with the free layer as pNML output magnet, could be applied. The latter concept, a path towards a GMR stack for pNML output sensing was envisioned in [79] and experimentally demonstrated in [68, 80].

5.3 A New Sensing Concept: Split-Stack Magneto-Resistance (MR) Device

As described above, there are highly optimized and industry-ripe sensors for MRAM and hard-drive technology. In principle, one could apply those sensors for pNML technology. However, there are operating conditions, that have to be addressed and tailored for pNML integration. One could make e.g. use of the stray field emanating from the pNML output magnet to flip the free-magnet (FM) of a GMR stack for sensing but fabrication is complicated by putting complex GMR stacks as new devices above the pNML output magnets. In our point of view, it is most efficient to tailor a perpendicular GMR sensor where the free-layer of the magneto-resistance (MR) device is at the same time the pNML output device and the permanent magnet is stacked or buried below. Hence, summarizing the results of [68, 80], a sensor was developed, which has a split structure as depicted in the cross-sectional views of Fig. 19a–c.

First, a pNML stack is deposited—ending with an inert Pt-layer—followed by hard-mask patterning for later Co/Pt island etching through a lift-off process. After that, the device is coated with a Co/Cu/Co GMR interface (Fig. 19b), followed by a Co/Pt stack forming the latter permanent magnet (PM). In a further step the second hard-mask is deposited, overlapping with the buried first hardmask. As a last step (Fig. 19c), the split-stack MR sensor is physically etched by Ar ion etching. The ANC creation can take place in step (a) or after (c). The fabrication steps result in a split-structure forming a standard pNML output device with ANC and domain-wall conduit (left) and an MR-stack with Co/Cu/Co GMR interface. The pNML stack serves as free-magnet (FM) and will be switched by a domain wall moving from the ANC throughout the lower stack. In Fig. 19d an SEM image shows the fabricated structure, where the hard-mask overlap and the ANC formation are clearly visible.

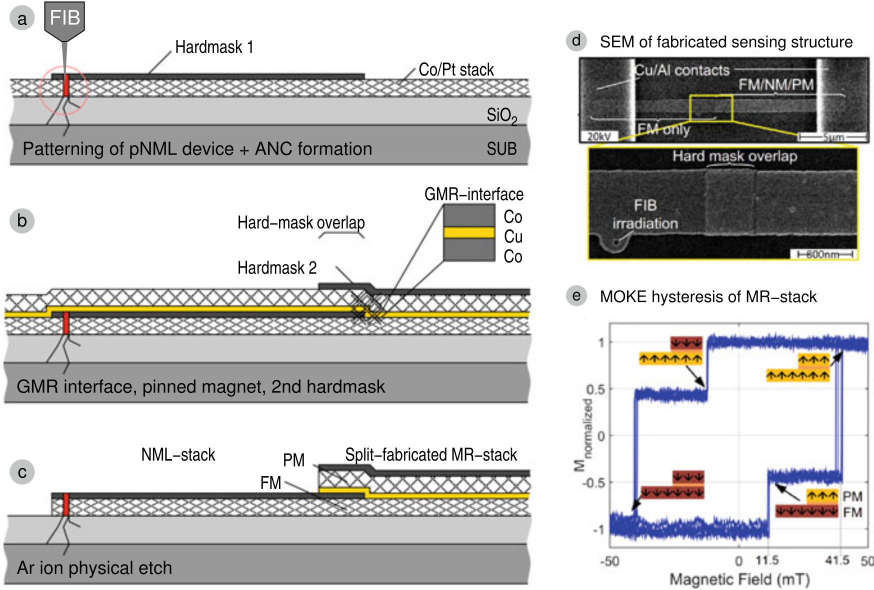


Fig. 19 **a**–**c** Fabrication steps for a pNML output sensor with a split MR-stack, employing buried hard-masks for etching. **d** SEM image of the fabricated current-perpendicular-to-plane (cpp) MR sensor. **e** MOKE hysteresis loop of the PM/FM stack. Images are adapted from [68]

Magneto-optical Kerr-effect (MOKE) measurements as shown in Fig. 19e reveal a decoupled switching of the free and pinned magnet with distinct square hysteresis. However, the device is not fully functional, as the designed pinned magnet has lower switching field than the pNML layer. It would be possible to use the sensor in a dynamic mode, where the pinned magnet is following the external clocking amplitude and the NML output magnet either switches or not, depending on the logic operation of the pNML circuit. But conceptually, one would prefer a real pinned layer, that could be achieved by an artificial antiferromagnet (AAF) which exchange couples the pinned magnet (PM) and compensates the magnetic moment of the PM. Even though the fabricated sensor showed the expected behavior in terms of decoupled switching (prohibited exchange coupling through the Copper layer), the measured change in resistance is very small. By cycling through the major hysteresis loops, the typical changes in resistance for the MR structure are detected via a matched Wheatstone-bridge. For bias currents of $I_{bias} = 5\,\mu A$ differential voltages of $U_d \approx 200\,nV$ are detected corresponding to a an MR-ratio of 0.34‰ which is far too small for device application and further optimization of the layer stacks are needed. To summarize the results for the integrated read-out sensor, the following improvements are proposed: First, the PM should be pinned by an artificial anti-ferromagnet to achieve higher switching fields than the ANC as common in GMR device fabrication. Second, a higher MR-ratio has to be targeted, in order to readout the sensor at high clocking frequencies. However, the measurements demonstrated,

that a split-stack MR structure can be fabricated in an easy way for direct integration into pNML. For that, the pNML output magnet serves as the free-layer of the sensor and we found the concept prospective for future pNML research.

5.4 A Field-Clock as Power Supply

In order to operate the 3D pNML devices, a bipolar magnetic field-clock with constant amplitude over the pNML device-space has to be provided. Clocking amplitudes are equal to the switching field of the ANC, and range from $\approx 10\,\text{mT}$ to $100\,\text{mT}$ for typical Co/Pt films. There might be applications, where an external magnetic field is conceivable but for pNML computing, the field clock has to be generated on-chip. As already depicted in Fig. 18, magnetic fields are most efficiently generated by inductors with a ferromagnetic core incorporating a thin *slit*. By contrast, simple air-coils can hardly provide the needed clocking amplitude even though lower switching fields are feasible by using low PMA magnetic stacks [68]. Our concept of pNML clocking is based on a planar design of a current carrying copper wire, sandwiched between two ferromagnetic films. The idea is depicted in Fig. 20a–c. A bulk ferromagnetic inductor with N windings carries a current I and generates a magnetic induction field B in the slit width w of

$$B = \mu_0 H = \frac{\mu_0 N I}{w + l/\mu_r}, \quad \text{and for } w \gg l/\mu_r \; : \; B = \mu_0 N \frac{I}{w} \tag{5}$$

with the free-space permeability μ_0, the relative *medium* permeability μ_r and the effective length l of the magnetic core (Fig. 20b).

In order to *squeeze* the inductor to chip-size, N is set to 1 and the current wire is placed on the chip-plane in a meander-like manner. This leads to *clocking zones* of opposite field-clock polarity in which the pNML devices can be placed (Fig. 20c). The ferromagnetic *core* is designed as an underlayer/overlayer film, possibly made of highly permeable soft-magnetic material with low electrical conductivity (to avoid eddy-currents) and minimal hysteresis (to eliminate hysteresis losses). The material should be proven up to the clocking frequency of pNML in the 100 MHz regime as e.g. reviewed in [81]. A rule-of-thumb calculation, adapted to the dimensions of our

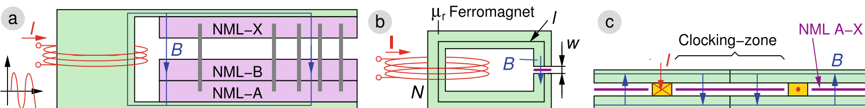


Fig. 20 *Evolution* of the pNML inductor design. **a** The 3D pNML layers are stacked in the slit of a ferromagnetic inductor core. **b** The slit-distance is reduced to a minimum, as pNML layers are very thin. **c** On-chip version, where the N wire turns are reduced to a single copper wire, sandwiched between the ferromagnetic *core*

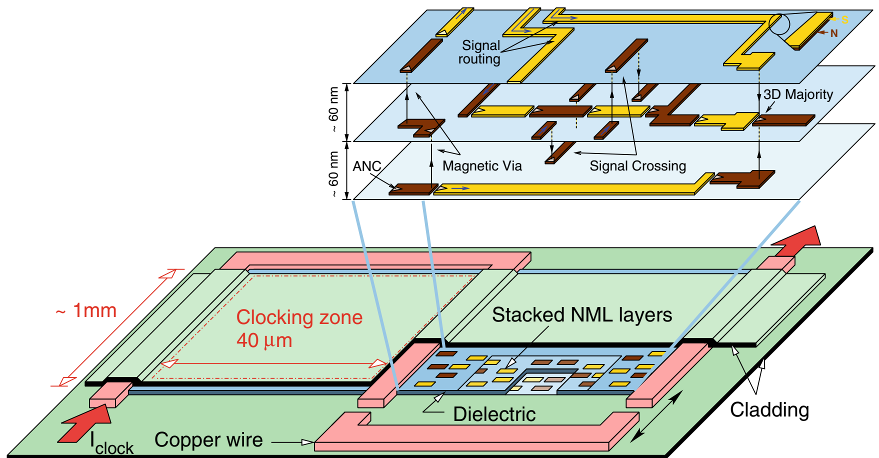


Fig. 21 On-chip inductor for clocking: Copper wires in a meander-like layout are sandwiched between soft-magnetic films. Magnetic fields perpendicular to the NML layer stack are generated. The pNML system is embedded in a dielectric matrix. Typical clocking zones are 40 µm wide and 1 mm long. Graph adapted and reprinted with permission from [19]

envisioned pNML on-chip inductor will be given in the following. For the inductor estimated by $w = 1\,\mu\text{m}$, $l = 80\,\mu\text{m}$ (40 µm-wide clocking zone) and $\mu_r = 1000$, the simplification in Eqn. 5 holds, as $1\,\mu\text{m} \gg 80\,\mu\text{m}/1000 = 80\,\text{nm}$. With that, the magnetic induction field can be calculated with $I = 30\,\text{mA}$ $(j = 3 \times 10^9\,\text{A}\,\text{m}^{-2}$ for a wire with cross-section area of $10\,\mu\text{m}^2)$ to $B \approx \mu_0 I/2w = 19\text{mT}$, which is a reasonable high value for pNML on-chip clocking circuits.

This on-chip inductor design was proposed in [82] and studied by finite-element simulations in greater detail in [27, 42]. The main results are briefly summarized in the following: The investigated on-chip inductor with soft magnetic cladding is depicted in Fig. 21.

The ferromagnetic inductor is highly planar with Cu wires in a meander geometry, whereas the pNML system is positioned in the '*slit*' of the ferromagnetic yoke. The vertical distance between pNML computing layers is in the 60 nm-range. Hence, several layers (we estimated 10 layers for the given geometry) of pNML can be stacked while still be penetrated by a spatially homogeneous magnetic field amplitude over large rectangular clocking zones of $\approx 40\,\mu\text{m} \times 1\,\text{mm}$. First calculations for this inductor geometry are based on materials that are discussed in [83] and predict, that pNML can be clocked with an on-chip inductor of $L = 10\,\text{nH}$ in the 50 MHz frequency range dissipating only $\approx 3\text{W}\,\text{cm}^{-2}$ [27, 42]. But there is plenty of room for engineering the inductor material and geometry for NML power-clock operation.

However, with the given device, it can be calculated, that only a few aJ per NAND operation are dissipated when operating the inductor at 50 MHz [42]. A first benchmarking against CMOS [84] gives a 35 times power saving potential at comparable binary throughput for 10 vertically stacked functional layers of NAND (or majority gate) based circuits in pNML.

6 Conclusion and Outlook

Using the local ion irradiation technique for pNML combined with the 3D geometrical arrangements, directed signal flow and logic operation in monolithically integrated three dimensional computing systems is experimentally demonstrated. A complete logic family was developed and is now ready for optimization in terms of integration density, operational speed and robustness against variations. In contrast to CMOS microelectronics, pNML does not require signal and supply wiring to each gate, and it has no leakage currents. Furthermore, pNML is non-volatile, and provides an easy way for global clocking, preventing the complex wiring needed in CMOS monolithic integration. Even though, computation in the ferromagnetic domain is mainly unidentified in the domain of IC design, the presented results may provide a promising technology for future 3D integrated circuits and systems especially when implemented in a Co-processor or Systolic architecture. But there are limitations and shortcomings that have to be addressed in future NML research:

Materials: The investigated Co/Pt bilayer stacks are possibly not the best choice when it comes to bigger circuits, as switching fields are rather high and domain-wall-speed is low. Softer material stacks like Co/Ni or CoFeB are potentially better suited. Recent work in our group also considers asymmetric films stacks with DMI for enhanced domain wall speeds [85].

ANC: Even though focused ion beam processing is well suited for lab experiments, IC fabrication would need a parallel process step as e.g. provided through ion-implanter technologies combined with high resolution hard-mask fabrication on chip. As ANC is a *defect-based* method to form nucleation sites, one should also consider local doping or alloying as gentle alternative for ANC creation.

Distributions: Thermally induced switching field distributions can be overcome by strong field coupling. But distributions stemming from fabrication and ANC creation are much more difficult to tackle. It is beneficial, that a global and continuous field clock is fixing the time-scale on which switching occurs (Arrhenius-type variations of switching fields are the same for all magnets) however, due to grains and pinning in the so-far used perpendicular media, most complex *circuits* were limited to a full adder and few ANCs.

On-chip clock: Experimental demonstration of an on-chip field clock as given in Fig. 20 is still lacking. But we see no fundamental obstacles, that one could efficiently generate fast and low-loss clocking fields on chip. Field generated by currents should be more straight-forward than using e.g. multi-ferroic materials

or even current induced switching techniques like Spin-transfer torque (STT) or Spin-orbit torque (SOT) by electrically accessing each element on device level.

Switching speed: We think, that switching speed (clocking speed) in the MHz-regime is not a show-stopper for NML technology, as due to nonvolatility and high integration density data throughput is competitive with CMOS. However, CMOS is a circuit implementation, optimized over decades and an emerging technology has to promise (and at some point has to deliver) improvements of orders in magnitude to be accepted in an industrial environment.

The given list is far from being comprehensive, and many aspects of NML pros and cons are not taken into account. But micro-, nanomagnetism and spintronics are flourishing fields and new effects and novel material are around the corner. The given pNML platform is ripe to be improved by scientists and engineers: in any case, pNML 2.0 (if the presented work could be summarized as pNML 1.0) will have to be faster, lower-energy, more robust and fault-tolerant, thus closer to a real disruptive pNML BEOL technology for logic circuits. Potentially—by overcoming the given challenges—it could be well suited for digital but non-charge based beyond CMOS co-processor architectures.

Acknowledgements There are numerous people who greatly contributed to this work, too many to name them all personally. I would like to thank the Bachelor and Master students of the TUM nanomagnetic group, the staff of the former TUM Chair for Technical Electronics, the Researchers Xueming Yu, Josef Kiermaier, Stephan Breitkreutz-v. Gamm, Irina Eichwald and Grazvidas Žiemys, and my mentors Gyorgy Csaba, Wolfgang Porod and Doris Schmitt-Landsiedel. Financial support by the German Research Foundation DFG under Grant SCHM 1478/9-1, CS 62/2-1, SCHM 1478/9-2 and SCHM 1478/11-1 and the Technical University of Munich—Institute for Advanced Study, funded by the German Excellence Initiative, is gratefully acknowledged.

References

1. The International Technology Roadmap for Semiconductors (ITRS): Emerging Research Devices. http://www.itrs2.net (2015) http://www.itrs2.net/itrs-reports.html
2. The International Technology Roadmap for Semiconductors (ITRS) 2.0 Executive Summary. http://www.itrs2.net/itrs-reports.html
3. M. Koyanagi, T. Fukushima, T. Tanaka, Proc. IEEE **97**(1), 49 (2009). https://doi.org/10.1109/JPROC.2008.2007463
4. P.A.T. adn, X. Gu, R. Alapati, M.S. Bakir, IEEE Trans. Components, Packaging Manuf. Technolo. **6**(7), 1007 (2016). https://doi.org/10.1109/TCPMT.2016.2524691
5. T. Tanaka, M. Helm, T. Vali, R. Ghodsi, K. Kawai, J.K. Park, S. Yamada, F. Pan, Y. Einaga, A. Ghalam, T. Tanzawa, J. Guo, T. Ichikawa, E. Yu, S. Tamada, T. Manabe, J. Kishimoto, Y. Oikawa, Y. Takashima, H. Kuge, M. Morooka, A. Mohammadzadeh, J. Kang, J. Tsai, E. Sirizotti, E. Lee, L. Vu, Y. Liu, H. Choi, K. Cheon, D. Song, D. Shin, J.H. Yun, M. Piccardi, K.F. Chan, Y. Luthra, D. Srinivasan, S. Deshmukh, K. Kavalipurapu, D. Nguyen, G. Gallo, S. Ramprasad, M. Luo, Q. Tang, M. Incarnati, A. Macerola, L. Pilolli, L.D. Santis, M. Rossini, V. Moschiano, G. Santin, B. Tronca, H. Lee, V. Patel, T. Pekny, A. Yip, N. Prabhu, P. Sule, T. Bemalkhedkar, K. Upadhyayula, C. Jaramillo, in *2016 IEEE International Solid-State Circuits Conference (ISSCC)* (2016), pp. 142–144. https://doi.org/10.1109/ISSCC.2016.7417947

6. D. Kang, W. Jeong, C. Kim, D.H. Kim, Y.S. Cho, K.T. Kang, J. Ryu, K.M. Kang, S. Lee, W. Kim, H. Lee, J. Yu, N. Choi, D.S. Jang, J.D. Ihm, D. Kim, Y.S. Min, M.S. Kim, A.S. Park, J.I. Son, I.M. Kim, P. Kwak, B.K. Jung, D.S. Lee, H. Kim, H.J. Yang, D.S. Byeon, K.T. Park, K.H. Kyung, J.H. Choi, in *2016 IEEE International Solid-State Circuits Conference (ISSCC)* (2016), pp. 130–131. https://doi.org/10.1109/ISSCC.2016.7417941
7. S. Das, *ITRS Assessment and Benchmarking of Emerging Logic Devices* (Wiley, Chichester, 2014). https://doi.org/10.1002/9781118958254.ch20
8. D.A. Allwood, G. Xiong, C.C. Faulkner, D. Atkinson, D. Petit, R.P. Cowburn, Science **309**, 1688 (2005). https://doi.org/10.1126/science.1108813
9. G. Csaba, W. Porod, A.I. Csurgay, Int. J. Circuit Theory Appl. **31**, 67 (2003). https://doi.org/10.1002/cta.226
10. A. Imre, G. Csaba, L. Ji, A. Orlov, G. Bernstein, W. Porod, Science **311**, 205 (2006). https://doi.org/10.1126/science.1120506
11. M.T. Niemier, G. Bernstein, G. Csaba, A. Dingler, X. Hu, S. Kurtz, S. Liu, J. Nahas, W. Porod, M. Siddiq, E. Varga, J. Phys.: Cond. Matter **23**, 493202 (2011). https://doi.org/10.1088/0953-8984/23/49/493202
12. E. Varga, M. Niemier, G. Csaba, G. Bernstein, W. Porod, IEEE Trans. Magnetics **49**(7), 4452 (2013). https://doi.org/10.1109/TMAG.2013.2249576
13. G. Csaba, G.H. Bernstein, W. Porod, A. Orlov, M.T. Niemier, X.S. Hu, M. Becherer, S. Breitkreutz, J. Kiermaier, I. Eichwald, D. Schmitt-Landsiedel, X. Ju, P. Lugli, *Nanomagnet Logic* (Wiley, 2015), pp. 1–18. https://doi.org/10.1002/047134608X.W8242
14. G. Csaba, P. Lugli, M. Becherer, D. Schmitt-Landsiedel, W. Porod, J. Comput. Electron. **7**, 454 (2008). https://doi.org/10.1007/s10825-007-0157-3
15. M. Becherer, G. Csaba, R. Emling, W. Porod, P. Lugli, D. Schmitt-Landsiedel, in *Digest Technical Papers IEEE International Solid-State Circuits Conference. ISSCC 474–475* (2009). https://doi.org/10.1109/ISSCC.2009.4977514
16. S. Breitkreutz, J. Kiermaier, I. Eichwald, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, IEEE Trans. Magn. **48**(11), 4336 (2012). https://doi.org/10.1109/TMAG.2012.2197184
17. S. Breitkreutz, I. Eichwald, J. Kiermaier, A. Papp, G. Csaba, M. Niemier, W. Porod, D. Schmitt-Landsiedel, M. Becherer, EPJ Web Conf. **75**, 05001 (2014). https://doi.org/10.1051/epjconf/20147505001
18. S. Breitkreutz, R.L. Stamps, J. Åkerman, A.V. Chumak, Y.C. Otani, G.E.W. Bauer, J.U. Thiele, M. Bowen, S.A. Majetich, M. Kläui, I.L. Prejbeanu, B. Dieny, N.M. Dempsey, B. Hillebrands, J. Phys. D: Appl. Phys. **47**(33), 333001 (2014). https://doi.org/10.1088/0022-3727/47/33/333001
19. M. Becherer, S. Breitkreutz-v. Gamm, I. Eichwald, G. Žiemys, J. Kiermaier, G. Csaba, D. Schmitt-Landsiedel, Solid-State Electron. **115**, Part B, 74 (2016). https://doi.org/10.1016/j.sse.2015.08.004
20. O. Hellwig, A. Berger, J.B. Kortright, E.E. Fullerton, J. Magn. Magn. Mater. **319**(1–2), 13 (2007). https://doi.org/10.1016/j.jmmm.2007.04.035
21. M. Johnson, P. Bloemen, F. Den Broeder, J. De Vries, Rep. Prog. Phys. **59**(11), 1409 (1996). https://doi.org/10.1088/0034-4885/59/11/002
22. M. Becherer, *Nanomagnetic Logic in Focused Ion Beam Engineered Co/Pt Multilayer Films*, *Selected Topics of Electronics and Micromechatronics*, vol. 38 (Shaker Verlag, 2011)
23. S. Breitkreutz, J. Kiermaier, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, in *IEEE Proceedings of the 41st European Solid-State Device Research Conference ESSDERC* (2011), pp. 323–326. https://doi.org/10.1109/ESSDERC.2011.6044169
24. J.H. Franken, M. Hoeijmakers, R. Lavrijsen, H.J.M. Swagten, J. Phys.: Condens. Matter **24**, 024216 (2012). https://doi.org/10.1088/0953-8984/24/2/024216
25. S. Breitkreutz, J. Kiermaier, S.V. Karthik, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **111**(7), 07A715 (2012). https://doi.org/10.1063/1.3675171
26. W. Wernsdorfer, E.B. Orozco, K. Hasselbach, A. Benoit, B. Barbara, N. Demoncy, A. Loiseau, H. Pascard, D. Mailly, Phys. Rev. Lett. **78**, 1791 (1997). https://doi.org/10.1103/PhysRevLett.78.1791

27. M. Becherer, J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Csaba, D. Schmitt-Landsiedel, in *IEEE Proceedings of the 43rd European Solid-State Device Research Conference (ESSDERC)* (2013), pp. 276–279. https://doi.org/10.1109/ESSDERC.2013.6818872
28. J.H. Franken, H.J.M. Swagten, B. Koopmans, Nat. Nanotechnol **7**(8), 499 (2012). https://doi.org/10.1038/nnano.2012.111
29. I. Eichwald, A. Bartel, J. Kiermaier, S. Breitkreutz, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, IEEE Trans. Magn. **48**(11), 4332 (2012). https://doi.org/10.1109/TMAG.2012.2196030
30. J. Kiermaier, S. Breitkreutz, I. Eichwald, M. Engelstädter, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **113**(17), 17B902 (2013). https://doi.org/10.1063/1.4794184
31. A. Papp, M.T. Niemier, A. Csurgay, M. Becherer, S. Breitkreutz, J. Kiermaier, I. Eichwald, X. Hu, X. Ju, W. Porod, G. Csaba, IEEE Trans. Nanotechnol **13**(5), 990 (2014). https://doi.org/10.1109/TNANO.2014.2342659
32. S. Breitkreutz, G. Žiemys, I. Eichwald, J. Kiermaier, G. Csaba, W. Porod, D. Schmitt-Landsiedel, M. Becherer, in *Proceedings of the IEEE International Electron Devices Meeting (IEDM)* (2013). https://doi.org/10.1109/IEDM.2013.6724684
33. S. Breitkreutz, I. Eichwald, J. Kiermaier, G. Hiblot, G. Csaba, W. Porod, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **115**, 17D506 (2014). https://doi.org/10.1063/1.4864737
34. R. Waser (ed.), *Nanoelectronics and Information Technology*, 3rd edn. Advanced Electronic Materials and Novel Devices (Wiley-VCH, 2012)
35. D.E. Nikonov, I.A. Young, Proc. IEEE **101**(12), 2498 (2013). https://doi.org/10.1109/JPROC.2013.2252317
36. OOMMF - Object Oriented Micro Magnetic Framework. http://math.nist.gov/oommf/
37. J.M. Shaw, S.E. Russek, T. Thomson, M.J. Donahue, B.D. Terris, O. Hellwig, E. Dobisz, M.L. Schneider, Phys. Rev. B (Cond. Matter Mater. Phys.) **78**(024414), 024414 (2008). https://doi.org/10.1103/physrevb.78.024414
38. M. Becherer, G. Csaba, W. Porod, R. Emling, P. Lugli, D. Schmitt-Landsiedel, IEEE Trans. Nanotechnol **7**(3), 316 (2008). https://doi.org/10.1109/tnano.2008.917796
39. M. Becherer, J. Kiermaier, G. Csaba, J. Rezgani, C. Yilmaz, P. Osswald, P. Lugli, D. Schmitt-Landsiedel, in *IEEE Proceedings of the 39th European Solid-State Device Research Conference ESSDERC* (Athens, 2009), pp. 105–108
40. M.P. Sharrock, J. Appl. Phys. **76**, 6413 (1994). https://doi.org/10.1063/1.358282
41. M.P. Sharrock, J.T. McKinney, IEEE Trans. Magn. **17**(6), 3020 (1981). https://doi.org/10.1109/TMAG.1981.1061755
42. M. Becherer, J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Žiemys, G. Csaba, D. Schmitt-Landsiedel, Solid-State Electron. **2013**(102), 46 (2014). https://doi.org/10.1016/j.sse.2014.06.012 (Selected papers from ESSDERC)
43. G. Žiemys., C. Trummer, S. Breitkreutz-v. Gamm, I. Eichwald, D. Schmitt-Landsiedel, M. Becherer, AIP Adv. **6**(5), 056404 (2016). https://doi.org/10.1063/1.4944336
44. G. Žiemys, A. Giebfried, M. Becherer, I. Eichwald, D. Schmitt-Landsiedel, S.B. v. Gamm, Solid-State Electron. **125**, 247 (2016). https://doi.org/10.1016/j.sse.2016.05.013. Extended papers selected from ESSDERC 2015
45. V.L. Safonov, H.N. Bertram, J. Appl. Phys. **87**(9), 5681 (2000). https://doi.org/10.1063/1.372488
46. G. Žiemys, G. Csaba, M. Becherer, AIP Adv. **8**(5), 056003 (2018). https://doi.org/10.1063/1.5006935
47. O. Lemcke. Theta evolve, oommf extension module (2002). http://www.nanoscience.de/group_r/stm-spstm/projects/temperature/theory.shtml
48. O. Tange, ;login: The USENIX Magazine **36**(1), 42 (2011). http://www.gnu.org/s/parallel
49. M. Becherer, G. Ziemys, I. Eichwald, D. Schmitt-Landsiedel, G. Csaba, S. Breitkreutz-v. Gamm, in *13*textsuperscriptth *Joint MMM-Intermag Conference, San Diego* (2016)
50. J. Goertz, G. Ziemys, I. Eichwald, M. Becherer, H. Swagten, S. Breitkreutz-V, Gamm, AIP Advances **6**, 5 (2016). https://doi.org/10.1063/1.4944698

51. K.W. Moon, J.C. Lee, S.B. Choe, K.H. Shin, Rev Sci. Instrum **80**(113904), 1 (2009). https://doi.org/10.1063/1.3262635
52. S. Breitkreutz, J. Kiermaier, I. Eichwald, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, in *Nanoelectronic Device Applications Handbook*, Devices, Circuits, and Systems (CRC Press, 2013), chap. 62, pp. 779–790
53. S. Breitkreutz, I. Eichwald, J. Kiermaier, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **115**(17), 17D104 (2014). https://doi.org/10.1063/1.4857555
54. S. Breitkreutz, A. Fischer, S. Kaffah, S. Weigl, I. Eichwald, G. Žiemys, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **117**, 17B503/1 (2015). https://doi.org/10.1063/1.4906440
55. G. Žiemys, A. Giebfried, I. Eichwald, D. Schmitt-Landsiedel, M. Becherer, S. Breitkreutz-v. Gamm, in *45th IEEE European Solid-State Device Research Conference (ESSDERC)* (2015), pp. 97–100. https://doi.org/10.1109/essderc.2015.7324722
56. S. Breitkreutz, I. Eichwald, G. Žiemys, D. Schmitt-Landsiedel, M. Becherer, in *Proceedings of the 14th IEEE International Conference on Nanotechnology (IEEE-NANO)* (2014), pp. 104–107. https://doi.org/10.1109/NANO.2014.6968096
57. S. Breitkreutz-v. Gamm, I. Eichwald, G. Žiemys, G. Csaba, G.H. Bernstein, M. Niemier, W. Porod, M. Graziano, D. Schmitt-Landsiedel, M. Becherer, in *IEEE Nanotechnology Materials and Devices Conference (NMDC)* (Anchorage, USA, 2015), pp. 125–126. https://doi.org/10.1109/NMDC.2015.7439269
58. S. Breitkreutz-v. Gamm, G. Žiemys, I. Eichwald, D. Schmitt-Landsiedel, M. Becherer, G. Csaba, H. Bernstein, W. Porod, in *15th IEEE International Conference on Nanotechnology (IEEE-NANO)* (2015). https://doi.org/10.1109/NANO.2015.7388796
59. M. Niemier, X. Ju, M. Becherer, G. Csaba, X. Hu, D. Schmitt-Landsiedel, P. Lugli, W. Porod, in *Silicon Nanoelectronics Workshop (SNW), 2012 IEEE* (2012), pp. 1–2. https://doi.org/10.1109/SNW.2012.6243329
60. X. Ju, M. Becherer, P. Lugli, M. Niemier, W. Porod, G. Csaba, in *15th International Workshop on Computational Electronics (IWCE)* (2012), pp. 1–3. https://doi.org/10.1109/IWCE.2012.6242837
61. G. Causapruno, F. Riente, G. Turvani, M. Vacca, M.R. Roch, M. Zamboni, M. Graziano, IEEE Trans. Very Large Scale Integration (VLSI) Syst. **24**(11), 3208 (2016). https://doi.org/10.1109/TVLSI.2016.2547422
62. G. Causapruno, U. Garlando, F. Cairo, M. Zamboni, M. Graziano, in *Proceedings of the IEEE Computer Society Annual Symposium VLSI (ISVLSI)* (2016), pp. 99–104. https://doi.org/10.1109/ISVLSI.2016.36
63. R. Perricone, Y. Liu, A. Dingler, X.S. Hu, M. Niemier, IEEE Trans. Nanotechnol. **15**(2), 179 (2016). https://doi.org/10.1109/TNANO.2015.2511072
64. F. Cairo, G. Turvani, F. Riente, M. Vacca, S. Breitkreutz-v. Gamm, M. Becherer, G. M., M. Zamboni, in *15th IEEE International Conference on Nanotechnology (IEEE-NANO)* (Rome, Italy, 2015). https://doi.org/10.1109/NANO.2015.7388798
65. DowCorning. Hydrogen silsesquioxane (HSQ), DOW CORNING ® XR-1541 E-beam resistin MIBK
66. J.F. Ziegler, J.P. Biersack, *The Stopping and Range of Ions in Solids, Stopping and Ranges of Ions in Matter*, vol. 1 (Pergamon Press, New York, 1984)
67. J.F. Ziegler. SRIM: The stopping and range of ions in matter (2006). http://www.srim.org/
68. I. Eichwald, *Perpendicular nanomagnetic logic: Three-dimensional devices for non-volatile field-coupled computing* (Technical University Munich, Doktorarbeit, 2015)
69. I. Eichwald, J. Kiermaier, S. Breitkreutz, J. Wu, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, IEEE Trans. Magn. **49**(7), 4468 (2013). https://doi.org/10.1109/TMAG.2013.2238898
70. I. Eichwald, S. Breitkreutz, J. Kiermaier, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **115**(17), 17E510 (2014). https://doi.org/10.1063/1.4863810
71. J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, EPJ Web Conf. **40**, 16007 (2013). https://doi.org/10.1051/epjconf/20134016007
72. I. Eichwald, S. Breitkreutz, G. Žiemys, G. Csaba, W. Porod, M. Becherer, Nanotechnology **25**(33), 335202 (2014). https://doi.org/10.1088/0957-4484/25/33/335202

73. J. Kiermaier, S. Breitkreutz, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, J. Appl. Phys. **111**(7), 07E341 (2012). https://doi.org/10.1063/1.3678584
74. M. Becherer, J. Kiermaier, S. Breitkreutz, G. Csaba, X. Ju, J. Rezgani, T. Kiessling, C. Yilmaz, P. Osswald, P. Lugli, D. Schmitt-Landsiedel, Solid-State Electron. **54**(9), 1027 (2010). https://doi.org/10.1016/j.sse.2010.04.011
75. J. Kiermaier, S. Breitkreutz, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, in *IEEE Proceedings of the 40th European Solid-State Device Research Conference ESSDERC* (Sevillia, Spain, 2010), pp. 214–217. https://doi.org/10.1109/ESSDERC.2010.5618385
76. J. Kiermaier, S. Breitkreutz, X. Ju, G. Csaba, D. Schmitt-Landsiedel, M. Becherer, Solid-State Electronics, Selected papers of ESSDERC 2010 **65-66**(1), 240 (2011). https://doi.org/10.1016/j.sse.2011.06.027
77. S. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, M. Samant, Proc. EEE **91**(5), 661 (2003). https://doi.org/10.1109/JPROC.2003.811807
78. Z.R. Tadisina, A. Natarajarathinam, B.D. Clark, A.L. Highsmith, T. Mewes, S. Gupta, E. Chen, S. Wang, J. Appl. Phys. **107**, 09C703 (2010). https://doi.org/10.1063/1.3358242
79. J. Kiermaier, *Integrated Nanomagnetic Logic System in Perpendicular Magnetic Media*, *Selected Topics of Electronics and Micromechatronics*, vol. 45 (Shaker Verlag, 2014)
80. I. Eichwald, S. Breitkreutz-v. Gamm, G. Žiemys, M. Becherer, D. Schmitt-Landsiedel, in *20th International Conference on Magnetism (ICM)* (Barcelona, Spain, 2015)
81. D. Gardner, G. Schrom, P. Hazucha, F. Paillet, T. Karnik, S. Borkar, R. Hallstein, T. Dambrauskas, C. Hill, C. Linde, W. Worwag, R. Baresel, S. Muthukumar, J. Appl. Phys. **103**, 07E927:1 (2008)
82. M. Becherer, J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Csaba, D. Schmitt-Landsiedel, in *Frontiers in Electronic Materials: Correlation Effects, Spintronics, and Memristive Phenomena - Fundamentals and Applications*, ed. by J. Heber, D. Schlom, Y. Tokura, R. Waser, M. Wuttig (Wiley-VCH, 2012), pp. 194–195
83. D.S. Gardner, G. Schrom, F. Paillet, B. Jamieson, T. Karnik, S. Borkar, IEEE Trans. Magn. **45**(10), 4760 (2009). https://doi.org/10.1109/TMAG.2009.2030590
84. The International Technology Roadmap for Semiconductors (ITRS): Emerging Research Devices. http://www.itrs.net (2013). http://www.itrs.net/Links/2013ITRS/Home2013.htm
85. G. Žiemys, V. Ahrens, S. Mendisch, G. Csaba, M. Becherer, AIP Adv. **8**(5), 056310 (2018). https://doi.org/10.1063/1.5007308

Spintronics for Neuromorphic Engineering

Gerard Joseph Lim, Calvin Ching Ian Ang, and Wen Siang Lew

Abstract Today's machine learning and artificial neural networks rely heavily on conventional electronic circuits. Progress in machine learning models and algorithms will eventually be limited by issues such as high power dissipation and scaling challenges posed by CMOS, and it is necessary to resolve these through a bottom-up approach. Here, we discuss how spintronic devices can overcome energy efficiency and scalability, and serve as artificial synaptic and neuronal devices using materials with large spin-orbit coupling, as well as magnetic textures such as chiral domain walls and skyrmions. We also explore the how these spintronic devices can mimic the biological brain-inspired neuronal and synaptic behaviours, to develop beyond-CMOS neuromorphic hardware for more efficient computation.

1 Introduction

Computers have come a long way since the invention of the first mechanical calculators. Today's computers use solid state electronic transistors and are designed to perform complex arithmetic and logical operations with performance that far exceeds human capability. However, the human brain still outperforms digital computers when it comes to classification tasks such as recognizing complex patterns, while utilizing only a fraction of energy that computers require for performing similar tasks. This is largely due to a completely different architecture between the human brain and conventional computers. Modern computers are typically built around the von Neumann architecture that has its advantages in flexibility, simplicity, and economy. A key limiting factor of this architecture, however, is that a shared communications bus prevents instruction fetch and data operations from being executed simultaneously, as depicted in Fig. 1a. Therefore, the processor tends to idle while memory is being accessed, resulting in a limitation known as the von Neumann bottleneck, which is detrimental to performance and efficiency.

G. J. Lim · C. C. I. Ang · W. S. Lew (✉)
Nanyang Technological University, Singapore, Singapore
e-mail: WenSiang@ntu.edu.sg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_9

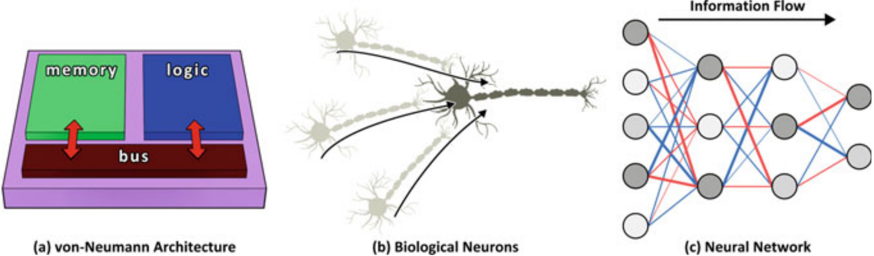


Fig. 1 **a** In the von Neumann architecture, memory and logic are physically separate. They communicate through a shared communications bus, and as a result, prohibits simultaneous bidirectional data transfer, resulting in the von Neumann bottleneck. **b** The biological brain and nervous system of most animals consist of specialised cells called neurons communicating with each other via electrical impulses. The arrows indicate the flow of electrical impulses along pre-synaptic neurons to the a post-synaptic neuron. **c** A neural network built upon certain characteristics of the biological brain, where neurons (grey circles) are connected to each other through synapses with varying weights (red and blue lines). The arrow indicates the general flow of information in this example of a neural network

As computers take on more complex tasks, efforts to mitigate the impact as a result of the bottleneck include increasing the memory bandwidth in and out from the processor, multithreading, and improving the overall performance of the memory and processor by increasing transistor density. Over the years, memory density has steadily increased at a rate described by Moore's law. Moore's law is an observation of the doubling of transistor count approximately every eighteen months. However, this is accompanied by a plateauing of processor clock rates. The trend of shrinking transistor size cannot go on indefinitely, due to the approaching physical limits which will restrict further progress in scaling, and aggravated by issues such as current leakage and limitations to material choices for ultra-thin gates and channels [1, 2]. Innovations to sustain Moore's law include building 3D transistors such as FinFETs, as well as 3D integrated circuit (IC) stacking to increase device density, making use of the 'vertical real estate' [3]. The continued reliance on CMOS transistors also means that achieving lower power consumption continues to pose a challenge. Brute processing power through von Neumann computers will not be able to keep up with ever increasing complexity inherent in computational problems such as image recognition, machine vision, and other classification tasks. Consequently, there is increasing demand for more efficient computing by taking inspiration from the biological brain.

The human brain is an extremely complex structure that exhibits self-awareness, expresses consciousness and creativity, and is the only organ attempting to understand itself. The brain employs massively parallel processing [4, 5], is fault tolerant [6], adaptive (plastic) and capable of self-learning. It manages and coordinates between sensory inputs, cognition, and motor control via electrical impulses along specialised cells called neurons as illustrated in Fig. 1b. While the brain consumes a substantial portion of the body's total energy, it is remarkably energy efficient when compared

with modern day computational algorithms and hardware [7]. As a result, brain-inspired computing using artificial neural networks such as depicted in Fig. 1c have become the de facto approach for classification tasks such as computer vision and speech recognition.

For instance, DeepMind's AlphaZero is an artificial intelligence (AI) designed to master the games of chess, shogi, and go [8]. Rather than performing computation using consumer desktop processors, AlphaZero relies on optimized and purpose-built application-specific integrated circuits (ASIC). Over 5000 units of Google's proprietary tensor processing units (TPU) are used in AlphaZero, with each TPU estimated to consume around 200 W [9]. Each TPU alone far exceeds the power utilization of the biological brain of 20 W. Neuromorphic chips have been developed to function as artificial synapses and neurons using conventional CMOS transistors. For instance, IBM's TrueNorth utilizes 5.4 billion transistors across 1 million programmable neurons and 256 million configurable synapses, and operates at comparatively lower power than general purpose microprocessors [10, 11]. These brain-inspired neuromorphic engineering and computing models are still implemented using conventional transistor circuits. Having a deeper understanding of the biological brain and emulating brain processes and attributes allow for the development of highly efficient brain-inspired computing technologies. The prospects of beyond CMOS technologies are also promising.

Emerging memory technologies such as spintronics offer a promising prospect for fulfilling such needs and applications. In utilizing spintronic device physics, one can engineer neuromorphic primitives that mimic the functionality of biological neurons, synapses, and processes in a top-down approach. Spintronic devices are inherently non-volatile, high-speed, radiation-hard, have zero static power requirements, practically infinite endurance, and can thus overcome the drawbacks inherent in transistor-based devices. Spintronic devices have also rapidly gained technological maturity in the form of magnetic random access memory (MRAM) that use magnetic tunnel junctions (MTJ) for data storage applications. In these MRAM devices, MTJs are deterministically switched to store binary information. In research, spintronic devices that rely on the motion of domain walls (DW) to achieve multi-state capabilities as well as in devices that operate in the stochastic regime have been demonstrated. These characteristics have opened opportunities and possibilities in developing brain-inspired functional primitives such as artificial neurons and synapses.

2 Fundamentals of Spintronic Device and Physical Phenomena

In addition to the fundamental charge property of electrons, spintronic devices also make use of the intrinsic electron spin and its associated magnetic moment. In the absence of external influence such as applied magnetic fields and spin currents, a

magnetic moment tends to orient itself along the direction of lowest energy. This preference in magnetic alignment is known as magnetic anisotropy and is determined by several factors such as the crystal structure and shape of the magnetic material. In the general case of thin film magnets, the magnetization can be in-plane (IP) or out-of-plane (OOP) of the film surface, as a result of in-plane magnetic anisotropy (IMA) or perpendicular magnetic anisotropy (PMA), respectively. Figure 1a shows arrows of opposite magnetizations corresponding to blue and red regions for IMA and PMA materials. Materials with IMA commonly form in bulk magnetic thin films. For PMA materials, the dominant magnetocrystalline anisotropy forces the magnetization direction to be along the normal of the film plane [12, 13]. The axis of anisotropy is also known as the easy axis and can be identified through magnetic hysteresis measurements. PMA materials are favoured for high-density memory applications due to much smaller demagnetization field as compared to IMA materials. Aside from memory density, the stability of the memory is also an important factor. The stability of a magnet is determined by the tendency of switching between magnetization states due to thermal noise. It is largely dependent on the magnetic material volume and magnetic anisotropy, and its retention time is given by:

$$\tau = \tau_0 \exp\left(\frac{K_u V}{k_B T}\right)$$

where τ_0 is the characteristic attempt time of 1 ns, K_u is the anisotropy per unit volume, V is the volume of the magnetic material, $k\text{-}_B$ is the Boltzmann constant, and T is the temperature. For a magnetic element suited for long term memory storage stability with $\tau = 10$ years, the energy barrier $E_B = K_u V$ should be $> 40k_B T$. Such magnetic elements can maintain a stable up or down state due to the large E_B, but can be deterministically flipped between states by overcoming E_B with additional spin torques as shown in Fig. 2b. On the other hand, magnetic elements with low E_B are susceptible to environmental perturbation such as thermal noise that introduce stochastic and random magnetization switching, and remain relevant in probabilistic computing and solving neural network problems [14, 15].

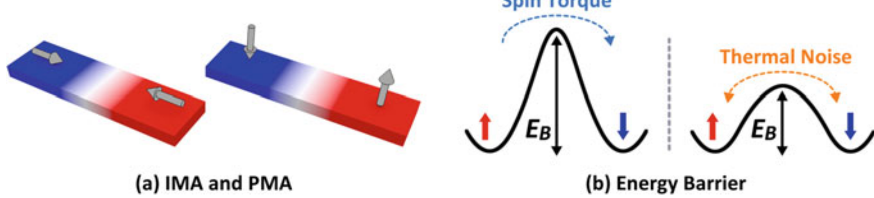


Fig. 2 **a** IMA and PMA materials form magnetic domains with magnetization in-plane and out-of-plane of the material **b** long term memory storage dictates that a large enough E_B is necessary. Energy is required to overcome E_B in order to switch magnetization states, and can be achieved through current-induced spin torques. Materials with low E_B are susceptible to random switching due to thermal noise, but remain relevant for applications in probabilistic computing and solving neural network problems

Fig. 3 **a** The dynamics of a magnetic moment subjected to damping, precessional, and spin torques, due to an effective magnetic field and spin current I_s. **b** A vertical spin valve in which the spacer layer is a non-magnetic conductive material. If the spacer layer is replaced with a dielectric, the device is known as a magnetic tunnel junction

Magnetization dynamics as illustrated in Fig. 3a can be described by the Landau-Lifshitz-Gilbert (LLG) equation:

$$\frac{d\hat{\mathbf{m}}}{dt} = -\gamma\left(\hat{\mathbf{m}} \times \mathbf{H}_{eff}\right) + \alpha\left(\hat{\mathbf{m}} \times \frac{d\hat{\mathbf{m}}}{dt}\right)$$

where $\hat{\mathbf{m}}$ is the unit vector of the magnetization, γ is the electron gyromagnetic ratio, α is the Gilbert damping constant, $\mathbf{H}_{eff}$ is the effective magnetic field. The first term describes the precessional motion of the magnetization in the presence of an effective field which includes the external magnetic field, the uniaxial anisotropy field, and other field contributions. The phenomenological damping torque, which is an intrinsic property of the material and described in the second term, tends to align the magnetization along the direction of the effective field.

Experimentally, the device magnetization orientation can be measured electrically using spin valve (SV) structures as depicted in Fig. 3b. By sandwiching a normal metal (NM) spacer such as Cu between ferromagnetic (FM) layers, the relative states between magnetic layers can be measured based on the magnetoresistance across both layers. The two magnetic layers differ such that the 'free layer' (FL) has a lower anisotropy than the 'pinned layer' (PL). Such a device can store a single bit of information in the form of high or low resistance states, measured electrically by sending a small non-perturbing current through the device. When the FL and PL magnetization are in the parallel (P) configuration, the resistance R_P is low, while in anti-parallel (AP) configuration, the resistance R_{AP} is high.

If the conductive spacer layer is substituted with a thin dielectric layer such as MgO, it behaves as a tunnel barrier between the FM layers, and the device is then identified as an MTJ. In MTJs, a spin-dependent tunnelling process occurs across the tunnel barrier, and the tunnelling magnetoresistance (TMR) far larger than GMR can be achieved [16]. A typical MTJ TMR curve plotted across a sweeping external magnetic field is shown in Fig. 4a. By switching both PL and FL across the external magnetic field range, the resistances at AP and P configurations can be measured.

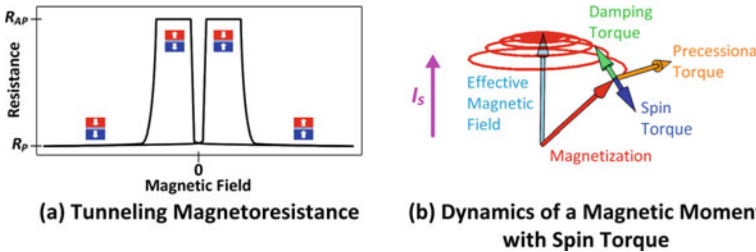


Fig. 4 **a** The magnetization configuration of an MTJ can be switched between parallel and anti-parallel, resulting in resistance states R_P and R_{AP}, respectively. **b** A vertical spin valve in which the spacer layer is a non-magnetic conductive material. If the spacer layer is replaced with a dielectric, the device is known as a magnetic tunnel junction

The TMR ratio which describes the change in magnetoresistance between AP and P states is given by:

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} \times 100\%$$

MTJs are mature technologies that have been commercialised in the form of toggle and spin-transfer torque (STT) MRAM. In a toggle MRAM, a magnetic field is used to switch the FL. To further simplify the circuitry and increase device density, two-terminal MTJs utilise STT to switch the FL. In the STT technique, polarized spin current from the PL exerts a torque on the FL magnetization. With a large enough injected spin current, the FL can be switched to align in the same orientation as the PL, resulting in an MTJ in P state. Switching the MTJ to AP state can be achieved by sending current in the opposite direction. Due to spin filtering, spins not aligned to the PL get reflected back to the FL, causing the FL to switch orientation, resulting in an MTJ in AP state. The STT can be described by a Slonczewski-like torque acting in the plane of the damping torque as depicted in Fig. 4b [17]. To account for the effects of STT, the LLG equation can be modified to include an additional STT term:

$$\frac{d\hat{\mathbf{m}}}{dt} = -\gamma\left(\hat{\mathbf{m}} \times \mathbf{H}_{eff}\right) + \alpha\left(\hat{\mathrm{m}} \times \frac{d\hat{\mathbf{m}}}{dt}\right) + \frac{\mu_B\left(\hat{\mathbf{m}} \times \mathbf{I}_s \times \hat{\mathbf{m}}\right)}{qM_sV}$$

where μ_B is the Bohr magneton, I_s is the input spin current, q is the electronic charge, M_s is the saturation magnetization, and V is the FL volume.

2.1 Domain Walls

In a magnetic material, individual moments can align to form various magnetic textures. In general, magnetic domains form when regions of magnetic moments are

oriented along the same direction. A single domain can exist in a sufficiently small region of magnetic material with strongly coupled magnetic moments, effectively forming monodomain magnets. The magnetization orientation transitions across the boundaries of domains to form a DW. DWs can form in magnetic materials with IMA and PMA. The types of DW formation differs between IMA and PMA nanowires. In IMA nanowires, the magnetization can form transverse DWs in narrow nanowires, or vortex DWs in wider nanowires as shown in Fig. 5a, b. In PMA nanowires, Néel and Bloch DWs form in narrow and wide nanowires, respectively, as shown in Fig. 5c, d. While monodomain magnets can function as binary bits, certain applications require multi-state or analog-like behaviour which can be achieved with multiple magnetic domains. DWs in these devices can be translated along the nanowire through the application of external fields or current-induced spin-torque. The translating or shifting of these DWs corresponds to the expansion of one domain, and the contraction of another. The DW position therefore determines the total magnetization of the magnetic material. More importantly, the DWs can be translated along a magnetic material through excitations such as external magnetic fields or current-induced spin torques, and is the basis of various device applications. One notable example is in DW racetrack memory in which data is stored at DWs in a magnetic nanowire. The position of DWs are electrically translated to the desired position for write and read operations.

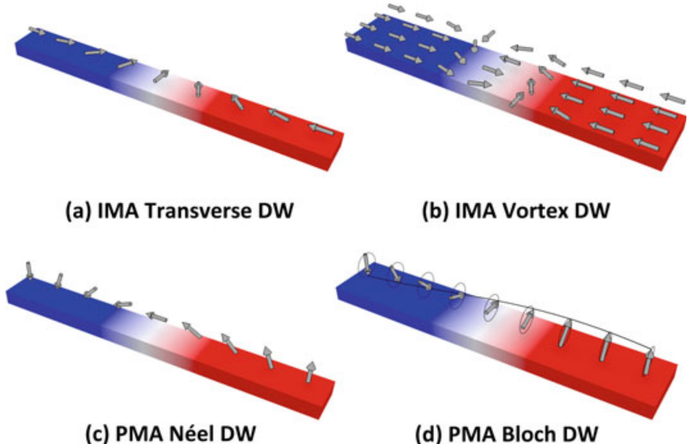


Fig. 5 **a** Tranverse and **b** vortex DW in narrow and wide IMA nanowires, respectively. **c** Neel and **d** Bloch DW in narrow and wide PMA nanowires, respectively. Red and blue regions indicate oppositely magnetized magnetic moments along their anisotropy axis. The transition along the white region is the DW, where the magnetic moment is twisted away from the anisotropy axis

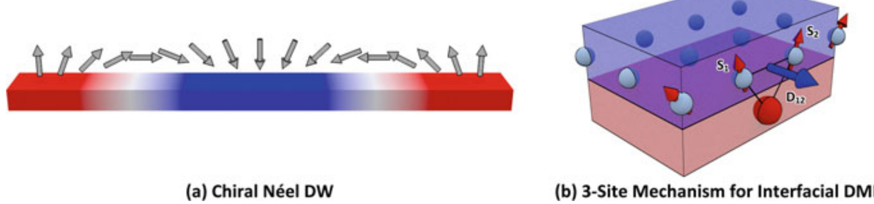


Fig. 6 **a** A Néel DW with right-handed chirality due to the DMI. **b** Dzyaloshinskii-Moriya interaction at the interface of a HM with large SOC and a FM layer

2.2 Dzyalonshinskii-Moriya Interaction

Dzyalonshinskii-Moriya interaction (DMI) manifests in materials with inversion asymmetry, such as at the interfaces formed by HM and FM layers [18]. This interaction stabilizes certain DW chirality as shown in Fig. 6a, which in turn gives rise to unique magnetic textures such as skyrmions [19–22], as well as for efficient magnetization switching and DW propagation [23–25].

First proposed by Dzyaloshinskii in 1958, he attributed the weak ferromagnetism found in antiferromagnetic crystals such as Fe_2O_3 and $MnCO_3$ to antisymmetric spin-coupling, showed that the weak ferromagnetism was caused by the relativistic spin-lattice and the magnetic dipole interactions, and derived the energy term [26]. In the following year, Anderson developed the quantum formalism for superexchange interaction, explaining the magnetic property of oxide and fluoride antiferromagnets, which subsequently provided Moriya with the formalism for describing a localised magnetic system in a microscopic model [27]. Subsequently, Moriya derived the energy term by including the spin-orbit interaction to Anderson's formalism [28].

DMI arises from spin-orbit coupling as well as the lack of inversion symmetry. While DMI in bulk materials is typically weak, it can manifest at the interface of dissimilar materials such as that formed between a magnetic material and a material with large SOC as shown in Fig. 6b. Interfacial DMI is a result of a 3-site indirect exchange mechanism between two atomic spins S_1 and S_2, and the atom with large SOC. The DMI interaction $H_{DMI} = \mathbf{D}_{12}(\mathbf{S}_1 \times \mathbf{S}_2)$ favours canted alignment between neighbouring spins $\mathbf{S}_1$ and $\mathbf{S}_2$, where $\mathbf{D}_{12}$ is the DMI vector [27].

2.3 Magnetic Skyrmions

A skyrmion refers to a topological vector field configuration first proposed by Tony Hilton Royle Skyrme in 1961 in the field of particle physics [29]. The skyrmion configuration experiences topological protection and stability that arises from its topological charge; The difference in topological charge of a skyrmion from a saturated and uniform configuration prevents the continuous deformation from one

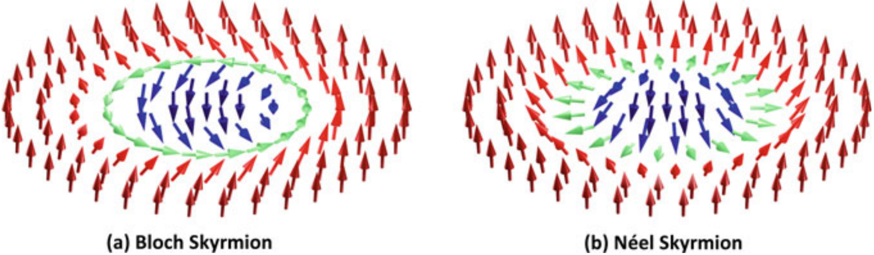


Fig. 7 Magnetization configuration of a **a** Bloch and **b** Néel magnetic skyrmion

state to the other. The actualization of this skyrmion state in a magnetic material was successfully observed in 2009 [30] and immense interest on this topological and magnetic particles was spurred thereafter for its novel physics and possible applications in spintronic devices.

Magnetic skyrmions can be visualized as a point of reversed magnetization in a ferromagnetic state with a finite radius for gradual reversal of magnetization as shown in Fig. 7a, b. They exist in 2 general forms, namely Bloch skyrmions and Néel skyrmions, where the 2 configurations differ in the direction the rotation of the magnetization vector from its centre to the outer uniform state. A Bloch skyrmion also known as the 'vortex' configuration has its magnetization rotate perpendicular to the plane between the opposite magnetizations. For a Néel skyrmion also referred to as the 'hedgehog' configuration, its magnetization rotates parallel to the plane between the opposite magnetizations [31].

Many desirable properties are found in magnetic skyrmions for its development in memory applications. The packing density of a magnetic skyrmion storage at the size of nanometres is highly competitive and comparable to that of the latest hard disk storage densities at the order of Tb/in^2. The non-volatility of magnetic skyrmions is in sharp contrast to dynamic random-access memory (DRAM) which requires continual refreshing of its capacitor state due to charge leakage. Skyrmion are particle-like and can be driven by electrical current, with reports of 110 nm skyrmions travelling up to velocities of 100 ms^{-1} [32]. This corresponds to read speeds in the order of nanoseconds. Hence, skyrmionic memory systems are promising in filling the niche between random access memories (RAM) and magnetic hard disk drives (HDDs), where it can provide reliable and non-volatile memory at high speeds of RAM but at the low cost of HDDs.

Skyrmions have been shown to be a superior alternative magnetic texture to DWs, and have been investigated for memory applications. Both magnetic skyrmions and DWs experience pinning which refers to the reluctance of the magnetic structure to propagate below a threshold driving current. Skyrmions have significantly lower pinning current at 10^6 Am^{-2} in comparison with DWs that is approximately 10^5 times larger at 10^{11} Am^{-2}. This supports skyrmions as an even lower power consuming memory device. The magnetic skyrmions' ability to deform without annihilation

allows it to traverse past minor defects, in contrast to DWs that are sensitive to defects which causes pinning.

Beyond storage and memory devices, skyrmions have been explored for applications in other fields such as microwave detector and harvester [33], probabilistic computing [34, 35], and artificial neuromorphic devices [36, 37]. With the abundant proposed uses of magnetic skyrmion illustrated in micromagnetic simulation, the experimental study of magnetic skyrmions on magnetic thin films becomes increasingly important and desired.

2.4 Spin Orbit Torque

STT-MTJ devices rely on high spin polarization factor offered by the PL. The high voltage and drive current required to effectively switch the FL magnetization can gradually damage the tunnel barrier, limiting the reliability and lifespan of MTJs. In order to reduce the required drive current, tunnel barriers under 1 nm are necessary to achieve a low resistance-area (RA) product. Furthermore, read-disturb can affect MTJ device reliability due to the shared read and write current paths, causing the device magnetization to unintentionally switch [38, 39]. An alternative technique to switch a FM layer magnetization is via spin-orbit torque (SOT) switching, allowing for separate read and write current paths, and improving the reliability and longevity of the MTJ. While the SOT has two possible origins—the spin Hall effect (SHE) and the Rashba effect – we will focus mainly on the SHE as the dominant contribution to the SOT in subsequent discussions. In SOT switching, an in-plane write current along a heavy metal (HM) such as Ta, Pt, and W, causes polarized spins to accumulate along the interfaces. The spin polarization is orthogonal to the charge current flow and spin current directions, and comes about due to the spin-orbit interaction and scattering of electrons in the HM as shown in Fig. 8a. Spins accumulating at the HM-FM interface

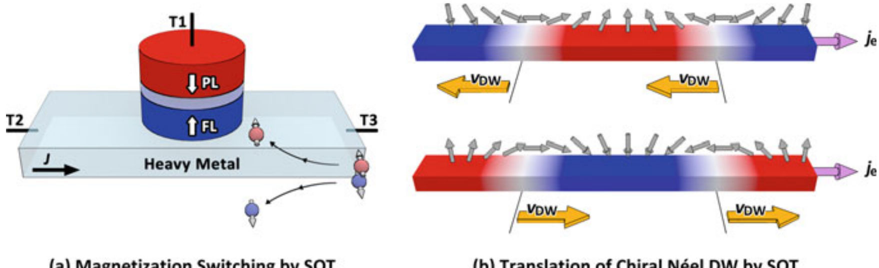


Fig. 8 **a** Spin-orbit torque acts on a nanomagnet due to charge current flow through a HM with large SOC, resulting in spin polarized electrons on the lateral surfaces of the HM. A spin torque acts on the FL magnetization due to spin accumulation at the HM-FL interface. **b** Chiral Néel DW motion in the direction of v_{DW} due to j_e. Opposite chirality or handedness results in opposite directions of v_{DW} for the same j_e

then acts on the magnetization of the FM, thereby switching it [40]. In materials with PMA, an externally applied field along the direction of the input charge current is necessary in order to break the switching symmetry [40, 41]. The SOT switching efficiency is dependent on the degree of spin polarization and spin current density J_s, described by the spin Hall angle θ_{SH} and input charge current density J_e [42]. SOT has also enabled high speed movement of chiral DWs along nanowires as illustrated in Fig. 8b [23]. The direction of DW propagation v_{DW} relative to injected current j_e is dependent on the chirality of the DW.

3 Biological Neurons and Synapses

Unlike conventional von Neumann computers, the biological brain consumes much less power, operates through massively parallel processing, is plastic and reconfigurable, and co-locates processing with memory. Brain-inspired models attempt to closely describe the various mechanisms to the likeness of the biological brain. The degree of likeness of the model to the biological brain is termed bio-fidelity. The nervous system of the biological brain is primarily composed of neurons, specialised cells that manage the transmission of nerve impulses as depicted in Fig. 9. Communication between neurons is achieved by sending neurotransmitters across a 20 nm gap known as the synaptic cleft [43]. Information is sent between different types of neurons that oversee the conveyance of information pertaining to sensory input, cognition, and motor control. A typical neuron comprises of several parts: (i) The cell body (or soma) which is the metabolic centre of the neuron. (ii) Dendrites which extend from the cell body and receive information impulses from pre-synaptic neurons. (iii) The axon which conducts electrical impulses away from the soma. (iv) The axon terminal which contains synaptic vesicles that hold neurotransmitters. Neurotransmitters are released across the synapse to post-synaptic neurons.

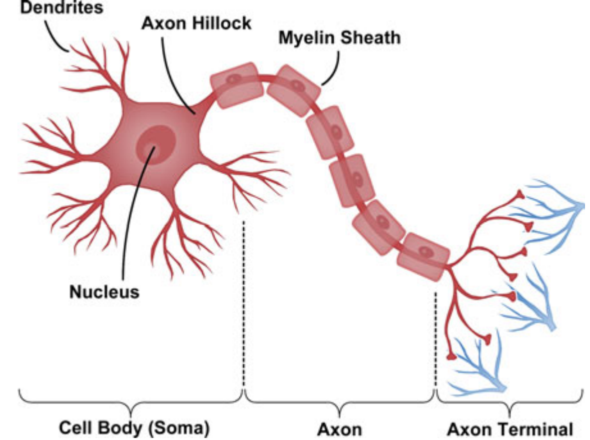


Fig. 9 A biological neuron comprises of a cell body containing the nucleus, an axon, and the axon terminal. An action potential sends an electrical impulse along the axon toward the axon terminal when the cumulative inputs from the dendrites exceed the threshold potential

Input received at the dendrites come from many other neurons, and can either be excitatory or inhibitory. The inputs contribute to an analog voltage signal called the membrane potential. The axon hillock operates on an all-or-none law that sums all the input signals and triggers an action potential when a threshold potential is breached. When sufficient excitatory input exceeds the potential threshold, an action potential will propagate along the axon to the axon terminals. Aside from the analog and digital mechanisms found within the brain, there are also dynamic, time-dependent features. For example, in the spike-timing-dependent plasticity (STDP) process, the synaptic strength between neurons are adjusted based on the relative timing between pre- and post-synaptic activity. Neuron behaviour vary vastly depending on their type and function [44]. In designing an artificial neuron/synapse, we trade-off between device complexity, computational power efficiency, and bio-fidelity.

4 Spintronics for Neuromorphic Engineering

Most of the ongoing research work in artificial neural networks and machine learning is through a top-down approach of developing algorithms to mimic biological neural processes while still relying on hardware based on mature CMOS technologies. Several takes on artificial neurons, synapses, and neural network hardware using conventional trasistors include the IBM TrueNorth and Intel Loihi chips. Such chips attempt to emulate biological neuronal and synaptic behaviour using transistors, but not necessarily capture the same bio-fidelity of actual biological neurons and synapses. However, properties and characteristics unique to emerging technologies such as spintronics enable a bottom-up approach in the development of brain-inspired computational hardware with greater bio-fidelity. In this section, we look at various spintronic devices and discuss how they can be used to mimic behaviours, mechanisms, and features of the biological brain, as well as in building functional features, efficient computational primitives, and circuits for real world applications.

4.1 Spintronic Synapses

The biological neuron receives excitatory and inhibitory post-synaptic potentials along dendrites as shown in Fig. 10a. The summation of the input potentials can ultimately lead towards generating an action potential. In an artificial neural network, synapses with synaptic weights $w_{m,n}$ determine the connection strengths between the pre- and post-synaptic neurons. The post-synaptic neuron N_n receives inputs I_m from pre-synaptic neurons as shown in Fig. 10b, such that $N_n = \sum I_m \cdot w_{m,n}$. An equivalent circuit in the form of a synaptic crossbar array is shown in Fig. 10c. In the crossbar array, a programmable device with memristive or analog-like behaviour is used to encode the synaptic weight at each crosspoint. The output at each neuron is then the sum of dot products between voltage inputs and synaptic conductance.

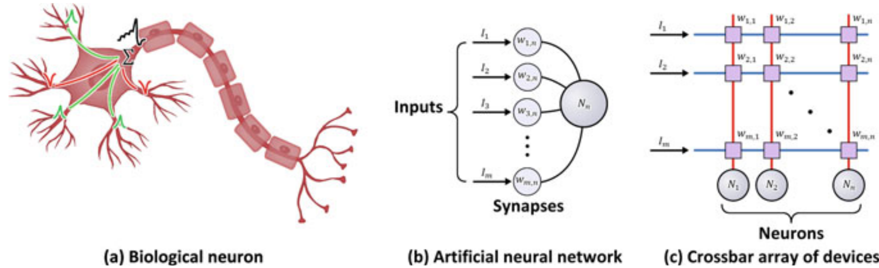


Fig. 10 **a** A biological neuron receives pre-synaptic input of excitatory and inhibitory nature. The signals are integrated at the axon hillock. Upon exceeding a threshold potential, an action potential is triggered. Electrical impulses travel down the axon and toward the axon terminal to other neurons. **b** A artificial neuron N_n mimics the biological counterpart by summing the product of inputs from multiple pre-synaptic inputs I_m and artificial synapses $w_{m,n}$. **c** A synaptic crossbar array where programmable resistive devices represent artificial synapses at each crosspoint

Synapses have plasticity, meaning that the strength between neurons can be adjusted to reflect the correlation between pre- and post-synaptic neurons. Several approaches using spintronic devices to function as artificial synapses are possible, generally revolving around the manipulation of relative sizes of magnetic domains in the device, while differing mainly in the sensing technique of the magnetization state.

MTJs have been available commercially in the form of MRAM, where a single bit is deterministically switched to store binary data. The reliability of such a memory element is only possible using precise fabrication processes. However, it has also been shown that "misbehaving" MTJs that do not switch deterministically, but rather, stochastically may also find purpose in neuromorphic engineering. An MTJ operating in the stochastic regime may switch up or down 50% of the time. Using this switching distribution, multiple MTJs with this quality can jointly function as a single synapse and demonstrate analog-like behaviour [45].

In a DW MTJ, the TMR corresponds with DW position, and the DW position can be manipulated by current-induced spin torques [46, 47]. Alternatively, synaptic weights can be represented by anomalous Hall voltage states in a Hall cross device with memristive or analog-like behaviour due to the magnetization polarization of the material as well as spin-orbit coupling [48, 49]. These spintronic synapses contain the weights in the form of non-volatile magnetization states, and can be adjusted through an iterative learning process by updating the device magnetization states.

4.2 *Spintronic Neurons*

Spintronic neurons can function and exist in several forms, ranging in behavioural complexity. The output of spintronic neurons can be described by their likeness to their biological counterpart, or bio-fidelity. In its simplest form, step-wise neurons

with binary output can be achieved with commercially available MTJs. However, one may require non-step transfer functions in order to handle non-binary inputs and outputs. For both step and non-step neurons, the magnetization dynamics and time-domain are not considered. We know that the biological brain does indeed have time-dependence behaviour, therefore, spintronic neurons with features such as in leaky-integrate-and-fire (LIF), or simply IF if no leaky feature is present, as well as stochasticity would emulate biological neurons more closely. The various spintronic neuronal behaviours are discussed in increasing bio-fidelity.

4.2.1 Stepwise Neurons

The simplest neuronal response is that of a perceptron with a step transfer function as shown in Fig. 11a. The most common spintronic device with such a response is the MTJ as shown in Fig. 11b, c, which switches between high and low resistance states via STT or SOT when the input $I > I_c$. The resistance states can be deterministically switched when the critical switching current I_c is exceeded, aligning the magnetization in the free and reference layers between parallel and anti-parallel orientations. For STT-MTJs, this requires a voltage bias through the tunnel junction. Such a switching mechanism has several drawbacks: (1) As the read and write current paths are shared, there is a tendency for unintentional switching due to read disturb. (2) The high voltage bias required affects the device endurance and is not energy efficient as compared to other switching mechanisms. As such, SOT is a more energy-efficient means of magnetization switching, at the expense of an additional terminal that results in separate read and write current paths. For such spintronic step-transfer function neurons, the MTJ can be extremely small such that only a mono-domain magnetization in the FL exists. The magnetization dynamics are neglected for a more simplistic binary output behaviour for such an artificial neuron.

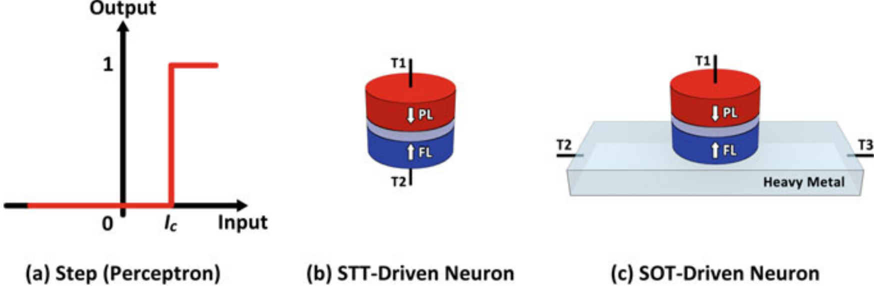


Fig. 11 **a** A step transfer function (perceptron) triggers at some threshold input I_c. **b** and **c** The step transfer function behaviour can be mimicked with STT or SOT MTJs. When $I > I_c$, the MTJ switches deterministically

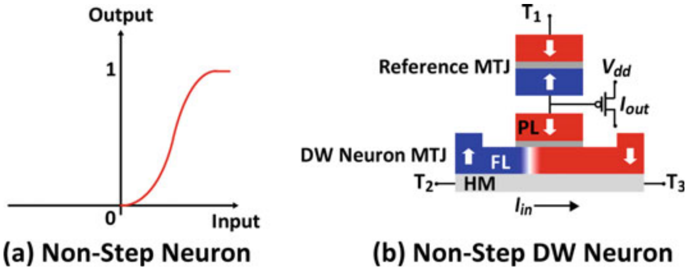


Fig. 12 **a** A non-step transfer function. **b** The resistance state of the DW neuron MTJ scales with DW position

4.2.2 Non-step Neurons

The transfer function that models a neuronal output is typically non-stepwise such as in Fig. 12a. In artificial neural networks, common transfer or activation functions include the sigmoid, hyperbolic tangent, or more popularly, the rectified linear unit (ReLU). To achieve beyond binary states in spintronic neuron devices, one would require multi-domain formation in the free layer. Rather than existing as either one of two possible magnetization states, a multi-domain free layer may form two oppositely magnetized domains separated by a DW. The relative proportions between domains can be adjusted by manipulating the DW position allowing for multiple resistance states to be exhibited in a DW neuron MTJ [47, 50, 51]. The DW position can be manipulated by current-induced SOT, and the resistance state of the device can be read out using an MTJ [52]. The divider circuit in Fig. 12b operates such that an increase in parallel orientation of the FL domain leads to lower resistance of the DW neuron MTJ. This drives the output transistor to approximately match I_{out} and I_{in} linearly.

4.2.3 (Leaky-)Integrate-and-Fire Neurons

The biological neuron receives multiple inputs from multiple unsynchronised pre-synaptic neurons. The excitatory and inhibitory inputs may not be received all at the same time, but across a span of time. These multiple inputs trigger the neuron when a threshold is reached and describes the IF functionality. Consider a similar structure to the non-step neuron discussed earlier, with multiple short injected current pulses to move the DW across the device, as depicted in Fig. 12a. At a certain point, the DW would have crossed a threshold position such that the resistance state of the neuron MTJ would cause the output inverter to be triggered as shown in Fig. 12b. The IF functionality therefore describes the biological brain's ability to summate all inputs over time and trigger the neuron when the threshold potential is reached. However, it does not address the potential decay as observed in biological systems, which further adds time-dependence to the input stimulus.

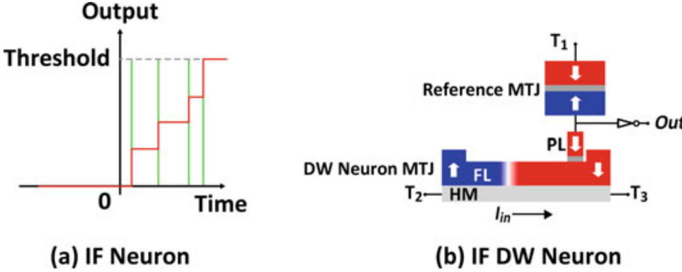


Fig. 13 **a** DW position and displacement due to input spikes. Green lines indicate input signal spikes that incrementally drive the DW, while the red line indicates the position of the DW. **b** The resistance state of the DW neuron MTJ switches only when the domain under the PL changes between up and down states

In the LIF functionality, consecutive inputs are not only summed over time, but the relative interval between inputs is also considered. Sparsely timed excitatory inputs may fail to trigger the neuron as the individually-triggered potential spikes decay over time, while closely timed excitatory inputs allow for successive integrations to exceed the threshold. Figure 13a shows input spikes (green lines) spaced close enough in time such that the DW is successfully displaced towards the threshold. If the input spikes were timed more sparsely, the DW may require more input spikes or even fail to cross the threshold. The leaky feature can be achieved by the simple implementation of a clocked current pulse in the opposite direction that drives the DW backwards. Alternatively, the DW can be allowed to naturally drift to leak without external stimuli by patterning a trapezoidal length of magnetic material, in which the DW would tend towards the narrower edge with lower DW energy [53]. In a similar energy-driven approach, the magnetic thin film can be grown such that the anisotropy is graded along the length within which the DW propagates as shown in Fig. 13b [54]. The DW will drift towards the region of lower anisotropy in the absence of external stimuli. Similarly, skyrmions have been proposed to deliver LIF neuron functionality [55] (Fig. 14).

4.2.4 Stochastic Neurons

In the previous discussions, achieving multi-state spintronic devices require the formation of multiple domains. The resolution of the multi-states is reduced as devices scale down, resulting in fewer domains and reducing the number of useable states. Therefore, an approach to encode the necessary information would be to use the time domain of a stochastic binary MTJ. When the energy barrier height E_B between magnetization states is reduced, such that thermal noise is able to randomly flip the free layer magnetization in an MTJ between parallel and anti-parallel states, its behaviour becomes stochastic [56]. Recently, voltage-controlled

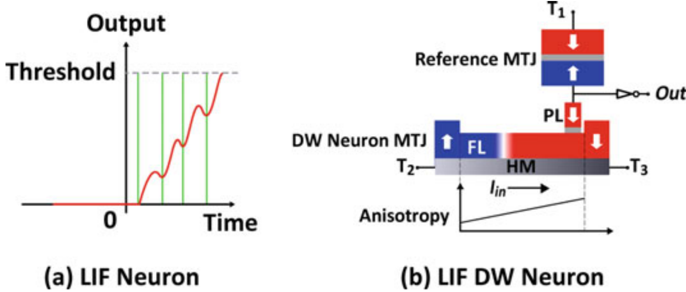


Fig. 14 **a** DW position and displacement due to input spikes. The DW moves in the opposite direction away from the threshold as part of the leaky feature. Green lines indicate input signal spikes that incrementally drive the DW, while the red line indicates the position of the DW. **b** An anisotropy gradient causes the DW to drift away from the MTJ towards the region of lower anisotropy when no input current spikes are injected

stochastic MTJs were demonstrated to output a sigmoidal probability activation function for handwritten digit recognition [57].

5 Summary and Outlook

Spintronic devices have been demonstrated to possess bio-plausible behaviours and characteristics, and can be used to develop a complete set of neuromorphic hardware primitives. Spintronic devices begin to take on a bigger role in overcoming challenges posed as transistor-based devices continue to shrink. Several proposals and proof-of-concepts for neuromorphic engineering and computing using spintronic devices show that it is progressively evolving, and continues to open new possibilities to developing brain-inspired neuromorphic hardware for more efficient computation. Presently, many of these devices serve niche roles, and more research works are necessary to broaden their functionalities. Spintronic neuromorphic devices are an encouraging approach to advancing brain-inspired computing, whether it is to mimic brain functionality for better understanding of the inner workings of the biological brain, or to develop better neuromorphic technologies for practical applications.

References

1. K.J. Kuhn, U. Avci, A. Cappellani, M.D. Giles, M. Haverty, S. Kim, R. Kotlyar, S. Manipatruni, D. Nikonov, C. Pawashe, M. Radosavljevic, R. Rios, S. Shankar, R. Vedula, R. Chau, I. Young, in *2012 International Electron Devices Meeting* (2012)
2. L.B. Kish, Phys. Lett. A **305**, 144 (2002)
3. W. Kuzmicz, in *2017 MIXDES - 24th International Conference Mixed Design of Integrated Circuits and Systems* (2017)

4. H. Wässle, Nat. Rev. Neurosci. **5**, 747 (2004)
5. H.V. Georgiou, Estimating the intrinsic dimension in fMRI space via dataset fractal analysis - Counting the 'cpu cores' of the human brain. arXiv e-prints (2014)
6. B.M. Yu, Nature **532**, 449 (2016)
7. M.E. Raichle, D.A. Gusnard, Proc. Natl. Acad. Sci. **99**, 10237 (2002)
8. D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Nature **529**, 484 (2016)
9. D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, D. Hassabis, Science **362**, 1140 (2018)
10. P.A. Merolla, J.V. Arthur, R. Alvarez-Icaza, A.S. Cassidy, J. Sawada, F. Akopyan, B.L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S.K. Esser, R. Appuswamy, B. Taba, A. Amir, M.D. Flickner, W.P. Risk, R. Manohar, D.S. Modha, Science **345**, 668 (2014)
11. M. Davies, N. Srinivasa, T. Lin, G. Chinya, Y. Cao, S.H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. Weng, A. Wild, Y. Yang, H. Wang, IEEE Micro **38**, 82 (2018)
12. M. Gajek, J.J. Nowak, J.Z. Sun, P.L. Trouilloud, E.J. O'Sullivan, D.W. Abraham, M.C. Gaidis, G. Hu, S. Brown, Y. Zhu, R.P. Robertazzi, W.J. Gallagher, D.C. Worledge, Appl. Phys. Lett. **100**, 132408 (2012)
13. S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H.D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, H. Ohno, Nat. Mater. **9**, 721 (2010)
14. S. Ganguly, K.Y. Camsari, A.W. Ghosh *Analog Signal Processing Using Stochastic Magnets.* arXiv e-prints (2018)
15. A. Graves, A.-r. Mohamed, G. Hinton *Speech Recognition with Deep Recurrent Neural Networks.* arXiv e-prints (2013)
16. S. Ikeda, J. Hayakawa, Y. Ashizawa, Y.M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, H. Ohno, Appl. Phys. Lett. **93**, 082508 (2008)
17. J.C. Slonczewski, Phys. Rev. B **39**, 6995 (1989)
18. A. Soumyanarayanan, M. Raju, A.L. Gonzalez Oyarce, A.K.C. Tan, M.-Y. Im, A.P. Petrović, P. Ho, K.H. Khoo, M. Tran, C.K. Gan, F. Ernult, C. Panagopoulos, Nat. Mater. **16**, 898 (2017)
19. S. Heinze, K. von Bergmann, M. Menzel, J. Brede, A. Kubetzka, R. Wiesendanger, G. Bihlmayer, S. Blügel, Nat. Phys. **7**, 713 (2011)
20. X.Z. Yu, Y. Onose, N. Kanazawa, J.H. Park, J.H. Han, Y. Matsui, N. Nagaosa, Y. Tokura, Nature **465**, 901 (2010)
21. I. Purnama, W.L. Gan, D.W. Wong, W.S. Lew, Sci. Rep. **5**, 10620 (2015)
22. J. Sampaio, V. Cros, S. Rohart, A. Thiaville, A. Fert, Nat. Nanotechnol. **8**, 839 (2013)
23. S. Emori, U. Bauer, S.-M. Ahn, E. Martinez, G.S.D. Beach, Nat. Mater. **12**, 611 (2013)
24. M. Cubukcu, O. Boulle, M. Drouard, K. Garello, C. Onur Avci, I. Mihai Miron, J. Langer, B. Ocker, P. Gambardella, G. Gaudin, Appl. Phys. Lett. **104**, 042406 (**2014**)
25. S. Emori, D.C. Bono, G.S.D. Beach, Appl. Phys. Lett. **101**, 042405 (2012)
26. I. Dzyaloshinsky, J. Phys. Chem. Solids **4**, 241 (1958)
27. P.W. Anderson, Phys. Rev. **115**, 2 (1959)
28. T. Moriya, Phys. Rev. **120**, 91 (1960)
29. T.H.R. Skyrme, B.F.J. Schonland, Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci. **260**, 127 (1961)
30. S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, P. Böni, Science **323**, 915 (2009)
31. A. Fert, V. Cros, J. Sampaio, Nat. Nanotechnol. **8**, 152 (2013)
32. K. Litzius, I. Lemesh, B. Krüger, P. Bassirian, L. Caretta, K. Richter, F. Büttner, K. Sato, O.A. Tretiakov, J. Förster, R.M. Reeve, M. Weigand, I. Bykova, H. Stoll, G. Schütz, G.S.D. Beach, M. Kläui, Nat. Phys. **13**, 170 (2017)
33. G. Finocchio, M. Ricci, R. Tomasello, A. Giordano, M. Lanuzza, V. Puliafito, P. Burrascano, B. Azzerboni, M. Carpentieri, Appl. Phys. Lett. **107**, 262401 (2015)

34. D. Pinna, F. Abreu Araujo, J.V. Kim, V. Cros, D. Querlioz, P. Bessiere, J. Droulez, J. Grollier, Phys. Rev. Appl. **9**, 064018 (2018)
35. J. Zázvorka, F. Jakobs, D. Heinze, N. Keil, S. Kromin, S. Jaiswal, K. Litzius, G. Jakob, P. Virnau, D. Pinna, K. Everschor-Sitte, A. Donges, U. Nowak, M. Kläui Thermal skyrmion diffusion applied in probabilistic computing. arXiv e-prints (2018)
36. S. Li, W. Kang, Y. Huang, X. Zhang, Y. Zhou, W. Zhao, Nanotechnology **28**, 31LT01 (**2017**)
37. Z. He, D. Fan Developing All-Skyrmion Spiking Neural Network. arXiv e-prints (2017)
38. X. Fong, Y. Kim, S.H. Choday, K. Roy, IEEE Trans. Very Large Scale Integration (VLSI) Syst. **22**, 384 (2014)
39. H. Naeimi, C. Augustine, A. Raychowdhury, L. Shih-Lien, J. Tschanz, Intel Technol. J. **17**, 54 (2013)
40. S. Fukami, T. Anekawa, C. Zhang, H. Ohno, Nat. Nanotechnol. **11**, 621 (2016)
41. L. Liu, C.-F. Pai, Y. Li, H.W. Tseng, D.C. Ralph, R.A. Buhrman, Science **336**, 555 (2012)
42. J.E. Hirsch, Phys. Rev. Lett. **1999**, 83 (1834)
43. B. Widrow, Y. Kim, D. Park, and J.K. Perin, *Chapter 1 - Nature's Learning Rule: The Hebbian-LMS Algorithm*, in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, ed. by R. Kozma, C. Alippi, Y. Choe, F.C. Morabito (Academic Press, 2019), p. 1
44. E.M. Izhikevich, IEEE Trans. Neural Networks **15**, 1063 (2004)
45. D. Zhang, L. Zeng, Y. Zhang, W. Zhao, J.O. Klein, in *2016 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)* (2016)
46. D. Bhowmik, U. Saxena, A. Dankar, A. Verma, D. Kaushik, S. Chatterjee, U. Singh, J. Magn. Magn. Mater. **489**, 165434 (2019)
47. S. Lequeux, J. Sampaio, V. Cros, K. Yakushiji, A. Fukushima, R. Matsumoto, H. Kubota, S. Yuasa, J. Grollier, Sci. Rep. **6**, 31510 (2016)
48. W.A. Borders, H. Akima, S. Fukami, S. Moriya, S. Kurihara, Y. Horio, S. Sato, H. Ohno, Appl. Phys. Express **10**, 013007 (2016)
49. A. Kurenkov, C. Zhang, S. DuttaGupta, S. Fukami, H. Ohno, Appl. Phys. Lett. **110**, 092410 (2017)
50. A. Chanthbouala, R. Matsumoto, J. Grollier, V. Cros, A. Anane, A. Fert, A.V. Khvalkovskiy, K.A. Zvezdin, K. Nishimura, Y. Nagamine, H. Maehara, K. Tsunekawa, A. Fukushima, S. Yuasa, Nat. Phys. **7**, 626 (2011)
51. D. Fan, Y. Shim, A. Raghunathan, K. Roy, IEEE Trans. Nanotechnol. **14**, 1013 (2015)
52. A. Sengupta, Y. Shim, K. Roy, IEEE Trans. Biomed. Circuits Syst. **10**, 1152 (2016)
53. W.H. Brigner, N. Hassan, L. Jiang-Wei, X. Hu, D. Saha, C.H. Bennett, M.J. Marinella, J.A.C. Incorvia, F. Garcia-Sanchez, J.S. Friedman, IEEE Trans. Electron Devices **66**, 4970 (2019)
54. W.H. Brigner, X. Hu, N. Hassan, C.H. Bennett, J.A.C. Incorvia, F. Garcia-Sanchez, J.S. Friedman, *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, **2019** 1
55. X. Chen, W. Kang, D. Zhu, X. Zhang, N. Lei, Y. Zhang, Y. Zhou, W. Zhao, Nanoscale **10**, 6139 (2018)
56. C.M. Liyanagedera, A. Sengupta, A. Jaiswal, K. Roy, Phys. Rev. Appl. **8**, 064017 (2017)
57. J. Cai, B. Fang, L. Zhang, W. Lv, B. Zhang, T. Zhou, G. Finocchio, Z. Zeng, Phys. Rev. Appl. **11**, 034015 (2019)

Resistive Random-Access Memory

Resistive Random Access Memory Device Physics and Array Architectures

Victor Yiqian Zhuo, Zhixian Chen, and King Jien Chui

Abstract Resistive random-access memories (RRAM) has garnered much interest in recent decades as a strong candidate to replace conventional memories like NAND flash, SRAM and DRAM. In contrast to the electrical charge changes in flash memories to define memory states, RRAM devices rely on non-volatile, reversible resistance changes within the device, hence its name. Apart from its superior performance: low power, high speed, high endurance, its simple two-terminal metal-insulator-metal (MIM) structure allows for a more scalable design and simpler fabrication processes. However, the RRAM is not without its problems and challenges. The resistive switching property of RRAM is inherently stochastic, resulting in variations between memory states, which could result in bit errors if unaccounted for. Also, the two-terminal RRAM requires a select device to prevent wrong selection of devices in an array. This chapter focuses on the redox-RRAM, where resistance changes take place through redox reactions within the insulator layer of the MIM, and will describe the basic operating principles of RRAM as well as various RRAM architectures.

1 Broad Perspective on Nonvolatile Memory

Nonvolatile memory has become an indispensable part of our everyday lives, powering a wide gamut of applications, from massive data storage such as memory cards and solid state drives, to consumer electronics like mobile phones, digital cameras, and music players. Since its invention by Fujio Masuoka [1] and its commercialization by Intel Corporation in 1988 [2], flash memory is the most prominent nonvolatile memory technology that drives the proliferation of consumer electronics of increasing functionality and storage density. In less than three decades, flash memory has grown into a US$20 Billion per year giant in the semiconductor industry

V. Y. Zhuo (✉) · Z. Chen · K. J. Chui
Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
e-mail: victor-zhuo@ime.a-star.edu.sg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_10

[3]. Despite its dominance, the long-term scalability of flash memory is very uncertain due to inherent physical limits and device reliability [4].

Consequently, there is a need for a new nonvolatile memory that can keep up with the scalability and performance presented by future technology nodes. Many alternative memory technologies have emerged and are extensively studied and developed. Amongst them, the resistive random-access memory (RRAM), which retains information in the form of distinct resistance states [5], is widely regarded as the most promising for massive data storage.

Historically, the resistive switching phenomenon dates back more than two centuries and was first demonstrated on the electric arc by Sir Humphry Davy in 1802 [6]. In the 1960s, advances in thin film technology enabled very high electric fields in tri-layered structures which led to observable resistive switching behaviour in ultrathin metal/oxide/metal films [7, 8]. However, research efforts in these devices diminished after a decade of intensive study [8] due to the development of silicon integrated circuits. In 1971, Leon Chua hypothesized the existence of memristors, the fourth basic circuit element [9]. The recent intense surge of RRAM research activities was renewed in the late 1990s by Asamitsu et al. [10], Beck et al. [11], and Kozicki et al. [12] due to the declining progress of silicon technology. Then in 2008, Hewlett Packard Labs realized and termed analog switching RRAM devices as memristors [13], thereby validating the fourth basic circuit element hypothesized by Leon Chua.

In 2013, Panasonic produced the first commercialized embedded RRAM chip [14]. Subsequently, Toshiba and Sandisk announced a 32 Gb high-density RRAM chip [15]. In 2017, Taiwan Semiconductor Manufacturing Company (TSMC) announced the production of embedded RRAM chips in 2019 with their 22 nm technology process [16]. In addition, to achieve ultrahigh storage density, the two-terminal RRAM devices are implemented in a crosspoint array structure. In 2007, Samsung demonstrated the first three-dimensional (3D) RRAM crosspoint array [17]. Notably, a 3D XPoint memory technology was released in 2015 by Intel Corporation and Micron Technology Inc. Likewise, Crossbar Inc. also released 3D RRAM products with high scalability beyond the 10 nm node, as well as better read latency, energy efficiency and write performance than the current NAND flash memory products.

Other than data storage, RRAM technology can also be used in the development of human brain-like computing systems with very high energy efficiency, computing capability and density scalability. For instance, matrix-vector multiplication or dot product can be realized with RRAM crosspoint architecture [18, 19]. Functional neuromorphic chips using binary or analog RRAM devices has been successfully demonstrated [20–23] for online training.

Despite the advancements in RRAM technology, there is still insufficient breakthrough in the complete understanding of the RRAM physical switching mechanism [24–26]. The main area of contention on RRAM mechanisms revolves around its conductive filament (CF), especially on its composition and on how it connects and ruptures.

In this chapter, we will first review the proposed RRAM mechanisms, followed by the resistive switching materials. Thereafter, the RRAM crosspoint array architecture

will be discussed with respect to the selector device and the self-rectifying RRAM implementations. It will then end with a conclusion and future outlook.

2 RRAM Device and Operation

A typical RRAM device has a simple metal-insulator-metal (MIM) structure, whereby the sandwiched insulator is the active layer that stores the data. In general, a RRAM device is able to show two distinct resistance levels, namely the high resistance state (HRS) and the low resistance state (LRS), which represent the logic '0' and logic '1', respectively. More than two resistance levels can also be obtained via careful tuning of the programming voltage pulse width and amplitude, which leads to high data storage density.

The two basic RRAM operations are programming and erasing. The transition from HRS to LRS is known as the SET or programming operation. Conversely, the RESET or erasing operation denotes the transition from LRS to HRS. Usually, a pristine RRAM device will have an initial high resistance which requires an applied voltage that is higher than the SET voltage to activate the resistive switching capability. This event is called the electroforming or forming process.

The RRAM switching mode can either be unipolar or bipolar as illustrated in Fig. 1. For unipolar operation, the resistive switching is independent of the voltage polarity and only depends on the magnitude of the voltage and its duration. Nonpolar switching is used to describe a unipolar switching that occurs in both voltage polarities. Conversely, for bipolar operation, the voltage polarity is important as the SET and RESET operations occur in opposite voltage polarities. The switching mode usually depends on the device structure, especially in the asymmetry obtained via fabrication or electrical methods.

There are many ways to classify RRAM devices based on their active material, switching mechanism or the switching phenomena. Here, we will focus on the redox-RRAM and its two subcategories, anion and cation devices, to simplify the discussion on their mechanisms.

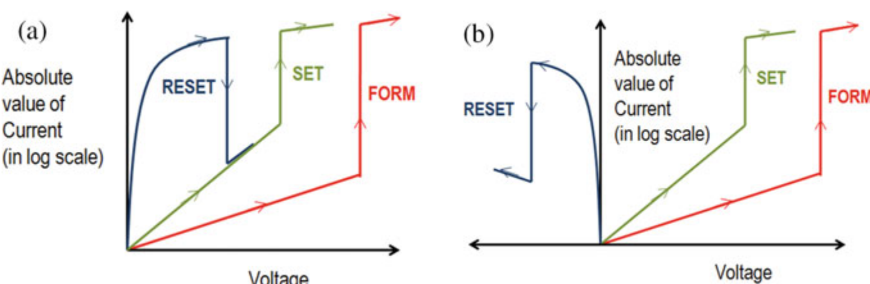


Fig. 1 Schematic of RRAM I-V curves showing two modes of operation **a** unipolar, and **b** bipolar

2.1 RRAM Conduction Mechanisms

In 2006, Szot et al. observed the CF inside insulating oxide using electron microscopy and proposed that the resistive switching was due to the reversible formation and rupture of the CF [27]. Ever since, the CF mechanism has been widely acknowledged. On the other hand, the non-filamentary switching phenomenon, also known as the interfacial switching mode, triggered by Schottky barrier changes at the dielectric-electrode interface, was also reported [5, 28, 29]. However, Schottky barrier changes are not exclusive to interfacial oxides and are also reported for filamentary switching oxides [5, 30–32]. The main difference between filamentary and interfacial switching modes is the dependence of the resistance to the device size, whereby for the former, the resistance is independent.

The CF within the dielectric is the localized conduction channel that is usually tens or hundreds of nanometers in diameter [33]. When the CF is formed between the electrodes, the RRAM device exhibits LRS. Conversely when the CF is ruptured, the device exhibits HRS. For cation RRAM devices, the resistive mechanisms are clear due to the ease at which the metal cations can be observed using microscopy techniques and thus verify the kinetic processes of the CF [34, 35]. As for anion RRAM, conduction depends on conductive channels created by oxygen vacancies, which is difficult to detect and to accurately fit with conventional current models. Though many attempts have been made to explain the conduction mechanism with Schottky emission, Poole-Frenkel emission, trap-assisted tunneling and other models, there is no single model that can accurately describe the conduction mechanism [31, 36].

In fact, there are many potential conduction mechanisms in RRAM as illustrated in Fig. 2. These conduction mechanisms can be grouped into two categories: (1) electrode-limited processes which consists of Schottky emission, Fowler-Nordheim tunneling, and direct tunneling as well as (2) bulk-limited processes, such as Poole-Frenkel emission, ohmic conduction, space-charge-limited conduction, hopping and trap assisted tunneling. For electrode-limited conduction mechanisms, the key factor

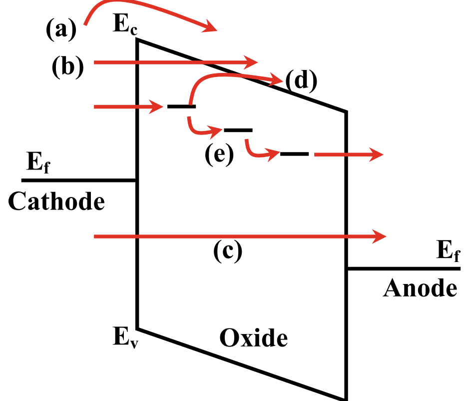


Fig. 2 Band diagram illustration of the possible conduction mechanisms in RRAM: electrode-limited processes which consists of **a** Schottky emission, **b** Fowler-Nordheim tunneling, and **c** direct tunneling, as well as bulk-limited processes, such as **d** Poole-Frenkel emission, and **e** trap–to–trap hopping

is the barrier height at the electrode-oxide interface whereas for bulk-limited conduction mechanisms, the electrical properties of the oxide play a critical role. Typically, RRAM devices exhibit ohmic or hopping conduction at the LRS [31]. However, during HRS, the dominant conduction mechanisms vary even for similar MIM stacks [37, 38] or could depend on the applied voltage regime [31].

2.2 *Anion Devices*

Switching Mechanisms

The active switching materials of anion RRAM devices include oxide dielectrics such as transition metal oxides, complex metal oxides, large bandgap dielectrics, as well as non-oxide dielectrics like chalcogenides and nitrides. In most metal oxides, the mobile species are widely believed to be the oxygen anions which is equivalent to the positively-charged oxygen vacancies. Hence, these RRAM devices are known as oxygen vacancy based RRAM (OxRRAM). Additionally, the anion migration leads to valence changes of the metal cations which leads to resistance change of the metal oxide and thus these devices are also termed as valence change memories (VCM) [5]. Since the resistance switching is caused by defects that modify electronic transport that is not limited to a specific electronic structure, almost all insulating oxides should show resistance switching behavior. In theory, the resistive switching phenomenon should also be observable in other insulating compounds like halides, borides, carbides and phosphides.

Since the earliest report of resistance switching in oxides by Hickmott in 1962 [7], oxides have been widely studied as anion-based switching materials, ranging from simple binary transition metal oxides, e.g. TiO_x [13, 32, 39], TaO_x [30, 40, 41], HfO_x [42–45], ZrO_x [46, 47], to rare-earth metal oxides, e.g. CeO_x [48], EuO_x [49], to perovskite-type complex oxides such as $SrTiO_3$ and others [50–52]. In the even larger set of non-oxide switching materials, the resistive switching phenomenon has been shown in nitrides, e.g. AlN [53], selenides, e.g. ZnSe [54], tellurides, e.g. ZnTe [55], and polymers [56, 57].

Insulating oxides can be regarded as semiconductors with native dopants. They can be either n-type or p-type semiconductors if they are oxygen deficient or excessive, respectively. Under applied high electric field and/or Joule heating, the electrically and/or thermally driven motion of these native dopants leads to chemical changes which results in resistance change. As the mobility and concentration of oxygen vacancies or cation interstitials are adequately high in transition metal oxides [5, 58], they are widely acknowledged as the mobile species responsible for the resistance switching and this is supported by experimental evidence [27, 33, 59–61]. However, more direct evidence is needed to confirm the actual mobile species even though it is very difficult to detect and track the migration of oxygen vacancies in RRAM devices [62–68].

The CF in OxRRAM is usually created by an electroforming process where the pristine device is subjected to a higher than normal voltage or current to activate the resistive switching capability [69–71]. Once activated, OxRRAM typically show either unipolar or bipolar resistive switching and have different switching dynamics and dominant driving forces.

In order to clarify the switching dynamics on their unipolar Pt/ZnO/Pt RRAM device, Chen et al. traced the CF evolution using in situ transmission electron microscopy (TEM) [72]. As depicted in Fig. 3a–c, during the electroforming process via applied voltage sweep, a darker contrast near the top electrode (TE) indicated the growth of the CF which eventually formed into a complete CF with the thinnest area in the middle. During the RESET process shown in Fig. 3d–f, the CF ruptured in the middle due to Joule heating, which ended in the dissolution of the CF towards the bottom electrode (BE). This in situ observation of CF formation confirms the important role of thermal effects in the CF rupture process, which suggests that the resistive switching in unipolar OxRRAM systems is mainly dominated by thermochemical effects.

Other plausible mechanisms in unipolar OxRRAM systems such as thermophoresis and diffusion were also explored by Strukov et al. [73], where they proposed a resistive switching model based on radial Soret-Fick diffusion equations. This model was experimentally verified with pulse length dependent tests to observe the response of switching dynamics in OxRRAM systems to the temperature gradient induced by Joule heating. As illustrated in Fig. 4a, in the case of neutral oxygen vacancies, a symmetrical CF is formed similar to that in Fig. 3b. However, for positively charged oxygen vacancies, an asymmetric CF is formed due to drift

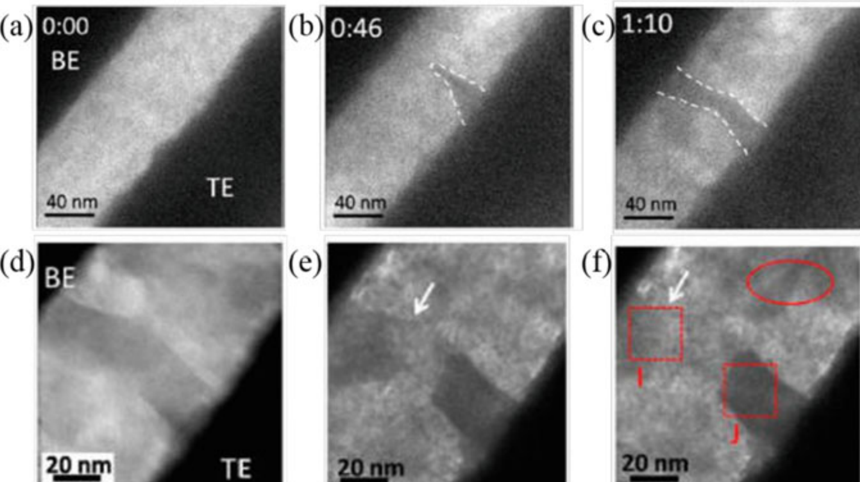


Fig. 3 In situ TEM of **a**–**c** electroforming process and **d**–**f** RESET process taken from video in Pt/ZnO/Pt system. Reprinted (adapted) with permission from [72]. Copyright (2013) American Chemical Society

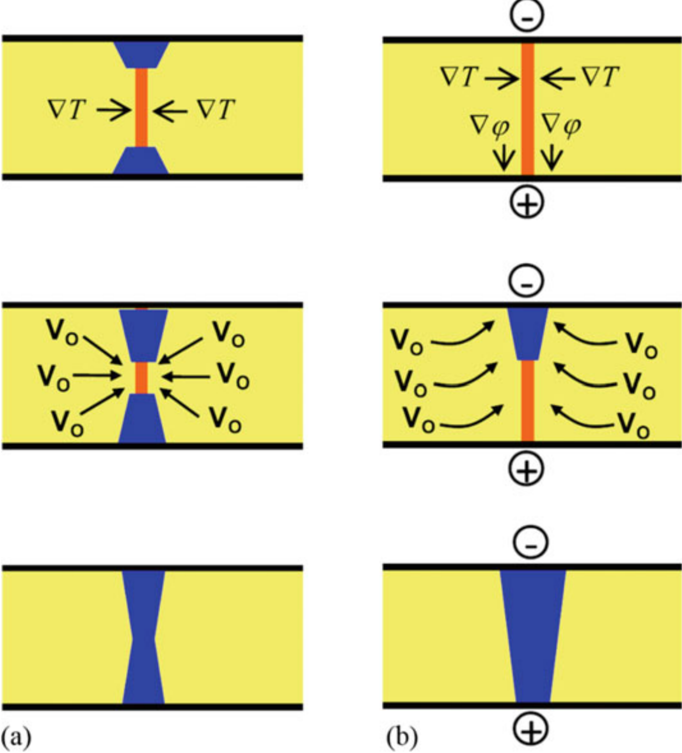


Fig. 4 Side view schematics of unipolar SET switching in RRAM device with **a** neutral oxygen vacancies and **b** positively charged oxygen vacancies. Reprinted by permission from Springer Nature, Applied Physics A: Materials Science & Processing [73], Copyright (2012)

(Fig. 4b) and this is observed in TiO_2 OxRRAM by Kwon et al. [59] using in situ TEM. The electroforming process creates a radial temperature gradient due to Joule heating within the CF. The Soret force then attracts the vacancies towards the CF and since the vacancies are positively charged, they are attracted towards the negative electrode.

By considering both **thermophoresis** and the Fick diffusion of oxygen ions under the influence of Joule heating, the ion dynamics in OxRRAM can be described by the following Soret-Fick continuity Eq. (1) [73, 74]:

$$\frac{\partial \eta_v}{\partial t} = \nabla J_{Fick} + \nabla J_{Soret} \tag{1}$$

where η_v is the vacancy density, J_{Fick} and J_{Soret} are the fluxes induced by Fick diffusion and thermophoresis, respectively.

Unlike unipolar OxRRAM systems, it is the electric field effects that dominate in bipolar OxRRAM systems. One of more widely researched bipolar OxRRAM

materials is TiO_2. Kwon et al. reported in situ studies on TiO_2 RRAM devices where the CF was made of a conductive Ti-O phase, termed Magnéli phase, with a Ti_nO_{2n-1} stoichiometry [59]. Similar oxygen-deficient CF with Magnéli phase was reported in WO_3 based RRAM systems by Tan et al. [75]. However, these observations rarely yielded any valid switching dynamics in bipolar OxRRAM systems that do not have stable intermediate phases such as HfO_2 and Ta_2O_5. This could be due to sample surface contaminants and the reliance of TEM observations on just the ion mass attribute. To circumvent this, a new technique was developed by Yang et al. in 2017 to characterize the ion transport and CF growth dynamics in bipolar TaO_x and HfO_x based OxRRAM using electrostatic force microscopy (EFM) [76]. Specifically, EFM is sensitive to the ion charge accumulation, has sub-10 nm spatial resolution with no strict requirements on sample thickness nor vacuum level [77, 78]. By merging EFM with systematic atomic force (AFM) and conductive atomic force microscopy (C-AFM) characterizations [79], this approach is able to provide a clearer representation of the dynamic oxygen ion transport during the resistive switching of OxRRAM. As depicted in Fig. 5, this approach is able to detect the migration and accumulation of oxygen ions to the interface and the following redox reactions as well as oxygen gas formation, which led to oxygen-deficient CF and structural distortions in the memory film. The formation of these CF in HfO_2 was directly identified using spherical-aberration (C_s)-corrected TEM and reversible ion migration was proposed to be responsible for the bipolar switching in HfO_2 [76].

Although bipolar switching in OxRRAM tends to be due to ionic motion and electrochemical reactions caused by electric field effects, thermal effects still play a crucial role during the RESET process [80, 81]. Panda et al. combined two models, an analytical temperature model [82] and a filament dissolution model [83, 84] and found that the temperature varies in a parabolic path within the CF, and is highest at the middle of the CF and lowest at the electrodes [85], which is in agreement with hypothesis that Joule heating assists the RESET mechanism in OxRRAM devices.

Material Selection Considerations

Since Joule heating is inevitable in the OxRRAM, it has a major influence on the material selection. In order for the OxRRAM to have stable and reliable resistive switching, both insulating and conductive phases are necessary. These two phases should not chemically react with each other to form a new phase, even at elevated temperatures caused by Joule heating. It should be a simple system with only two thermodynamically stable solid-state phases, one insulating phase and one relatively conductive phase for the conduction channel. Ideally, the conduction channel should have a high oxygen solubility to serve as a reliable oxygen reservoir to conserve the oxygen ions during repeated switching cycles [26]. TaO_x and HfO_x are the prime candidates that fulfil these requirements, exhibiting switching endurance over 10^{12} [86] and 10^{10} cycles [87], respectively.

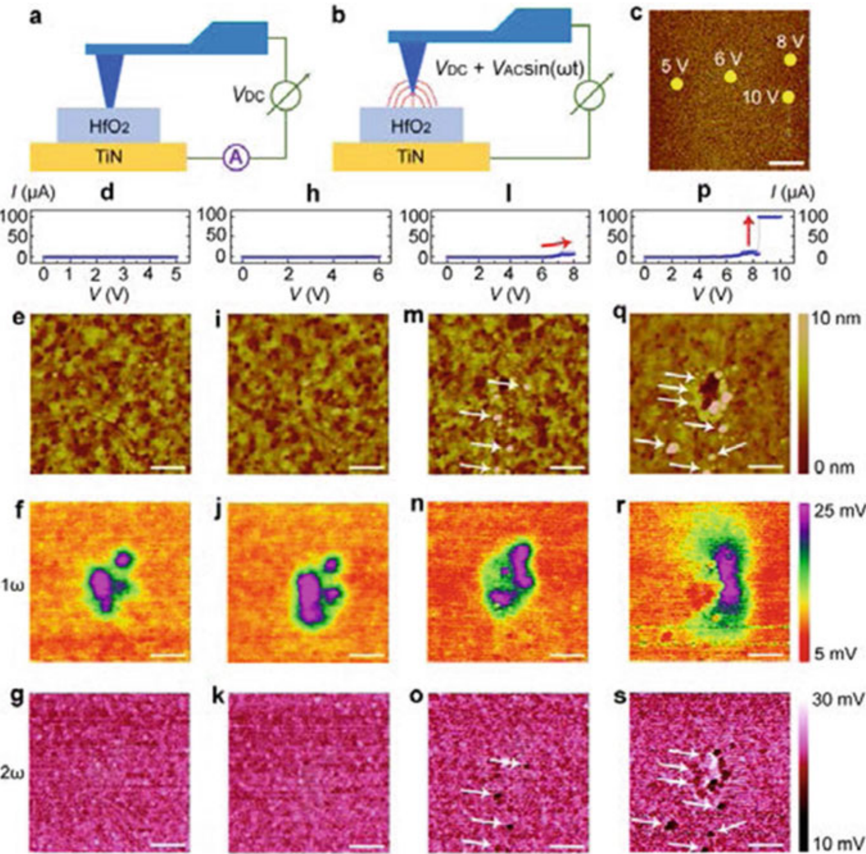


Fig. 5 Schematics of the **a** conductive atomic force microscopy (C-AFM) and **b** electrostatic force microscopy (EFM) measurements performed on HfO_2/TiN samples. **c** Topographic image showing locations where voltage sweeps with different amplitudes were performed. Electrical **d**, topographical **e**, 1ω (**f**), and 2ω **g** measurements on the region stimulated by voltage sweep up to 5 V during preceding C-AFM measurements. **h**–**k**, **l**–**o**, and **p**–**s** are the corresponding electrical and EFM results on the region stimulated by voltage sweep up to 6 V, 8 V, and 10 V during C-AFM measurements. Reprinted from [76] under a Creative Commons Attribution 4.0 International License. Full license terms at http://creativecommons.org/licenses/by/4.0/

2.3 Cation Devices

Switching Mechanisms

Cation-based RRAM devices are also known as metal ion based RRAM, electrochemical metallization memory, atomic switch or conductive bridge RRAM (CBRAM). First reported in the 1970s by Hirose et al. [88], and developed by Kozicki et al., in the late 1990s for data storage [89], the mobile species in CBRAM devices are metallic cations. Typically, the CBRAM MIM stack consists of a solid electrolyte

(SE) sandwiched between an electrochemically active electrode (AE) and an electrochemically inert electrode (CE). Candidates for the AE are limited to some metals that are mobile in the SE, and are usually Cu [90], Ag [91] or an alloy such as CuTe [35]. CE materials are less limited and include Pt, Ru, Ir, Au, W, Mo, Co, Cr, TiW, TaN or poly-Si [34]. Conventionally, the SE materials have been amorphous Si, C, doped organic semiconductors, iodides, methylsilesquioxane, selenides, sulphides, tellurides, ternary chalcogenides, or even non-solid electrolytes like water and even vacuum gaps [92–95]. Recently, more and more metal oxides and nitrides have been used, including Al_2O_3, CuO_x, HfO_2, GdO_2, MoO_x, SiO_2, Ta_2O_5, TiO_2, WO_3, ZnO, ZrO_2, and AlN [96–98]. The transition from traditional electrolytes to the complementary metal-oxide-semiconductor (CMOS) compatible and inexpensive oxides have enhanced the retention and operation voltages of CBRAM, making them more suitable for certain applications such as switches in large-scale integrated circuits [96]. More details on the material systems available for CBRAM can be obtained from the comprehensive review by Valov et al. [34].

Switching mechanisms in CBRAM are similar to that in OxRRAM. Electroforming in CBRAM also causes structural alterations to the SE and creates nanoscale CFs to accommodate the electrochemically active metal ions for the subsequent resistive switching [35]. Most CBRAM devices exhibit bipolar switching where by the electric field is the dominant driving force. However, there are also reports of unipolar CBRAM, which suggests the possibility of Joule heating effects [99, 100]. Unlike OxRRAM, the switching mechanisms are better understood since the metal ions are more easily observed using microscopy techniques, thus the kinetic processes of the CF can be more easily verified [101]. Since ion migration and associated redox processes during resistive switching occur at the nanoscale, the TEM, specifically the in situ TEM, is currently one of the best techniques used to understand the dynamic processes of resistive switching as it can yield details on the chemical state, composition, morphology, size and trace the evolution of the CF growth/dissolution down to individual metal nanoclusters. As summarized by Yang et al. [102], there are mainly four different electrochemical metallization processes in CBRAM devices depending on the cation mobility and the redox reaction rate as shown in Fig. 6. The cation mobility influences the nucleation site of the CF and the direction of the CF growth whereas the redox reaction rate determines the ion supply which decides the CF morphology [102].

As schematically shown in Fig. 6a, when cation mobility and redox reaction rate are high in the SE, it leads to inverted cone-shaped CFs that initiate from the inert CF. This has been directly verified using in situ TEM for Ag/H_2O/Pt (see Fig. 6e) [101], Cu/Al_2O_3/Pt [103] and Cu/Cu-GeTe/Pt–Ir [104] devices. On the other hand, when both cation mobility and redox reaction rate are low, the cations traverse a short distance from the AE to attain critical nucleation conditions within the dielectric which leads to metal cluster formation. These metal clusters then form a CF that grows from the AE to the CE as shown schematically in Fig. 6b and has been reported in Ag/SiO_2/W (see Fig. 6f) [105], Ag/ZrO_2/Pt [106], Ag/a-Si/Pt [107], and TiN/Al_2O_3/Cu [108].

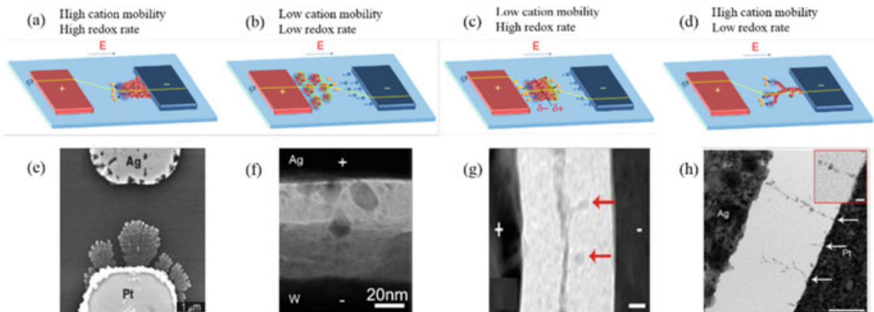


Fig. 6 **a**–**d** Schematic diagrams of filament growth and **e**–**h** TEM images of corresponding exemplary filament for when **a**, **e** both cation mobility and redox reaction rate are high **b**, **f** both cation mobility and redox rate are low **c**, **g** low cation mobility and high redox reaction rate, and **d**, **h** high cation mobility and low redox reaction rate. Adapted by permission from Springer Nature Journal of Electroceramics [102], Copyright (2017)

If the SE layer facilitates low cation mobility and high redox reaction rate as shown in Fig. 6c, nucleation of the CFs occurs inside the SE analogous to the case when both cation mobility and redox reaction rate are low. The exception lies in the high reaction rate which causes persistent reduction of the metal cations at the sites of nucleation, resulting in the backward CF growth from the nucleation site to the AE. This has been observed using in situ TEM by Yang et al. for Ni CF growth in evaporated SiO_2 thin films, as depicted in Fig. 6g [109]. For the last case shown in Fig. 6d where the cation mobility is high and redox reaction rate is low, CF nucleation occurs from the CE. The low redox reaction rate limits the ion supply such that ion reduction occurs at the edges of existing CFs, resulting in the formation of dendrite-like CFs towards the AE. This can be clearly observed in lateral Ag/SiO_2/Pt (see Fig. 6h) [107] and Ag/Ag-PEO/Pt devices [110].

In general, the formation and dissolution of nanoscale CFs has been widely accepted as the resistive switching mechanism for CBRAM devices. However, other mechanisms such as phase transitions has also been reported for specific material systems. One such example is the Ag/Ag_2S/W device that was reported by Xu et al. using in situ TEM where the resistive switching was induced by a phase transition in the Ag_2S between the conductive argentite phase and the insulating acanthite phase [111]. Unlike conventional CBRAM devices, these Ag/Ag_2S/W devices have an abundance of Ag cations which allows intrinsic resistive switching in the SE without the need of cation injection from the electrodes.

Although there has been much progress in the understanding of the resistive switching mechanisms in both CBRAM and OxRRAM, more research is still needed to confirm the microscopic details, especially for the OxRRAM devices. Linking the structural changes in the memory layer to the device variation, degradation and performance would be significant and will pave the way to increased RRAM commercialization.

3 RRAM Crosspoint Array

RRAM's simple MIM structure allows the implementation of the highly dense crosspoint memory array. As illustrated in Fig. 7, a crosspoint array has parallel metal lines, known as word and bit lines, at the top and bottom planes, perpendicular to one another. At each crosspoint or intersection of these metal lines, a two terminal memory device is integrated. The word and bit lines are used to select a memory cell and write/read data, respectively [29]. Assuming that the width of both metal lines and spaces is equal to F, where F is the minimum technology feature size, the effective cell area will be $4F^2$, which is the smallest single layer or two-dimensional (2D) footprint [112].

However, it is not easy to directly use this passive resistive network since accessing a designated cell induces sneak path currents from adjacent memory cells. As depicted in Fig. 8, the sneak path problem occurs when the selected memory cell is at HRS (red) and all its adjacent cells are in LRS (green). During the read operation, the current flow through the selected memory cell should be low ($I_{element}$). However, as the adjacent cells are all in LRS, significant sneak currents (I_{sneak}) will flow through all the cells, thus contaminating the actual information of the target cell. Moreover, the leakage currents will also increase the power consumption. These problems worsen

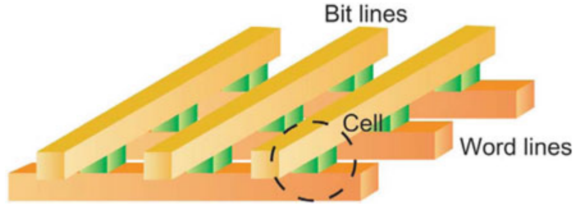


Fig. 7 Schematic diagram of a crosspoint memory array showing the word and bit lines. Reprinted from [29], Copyright (2008), with permission from Elsevier

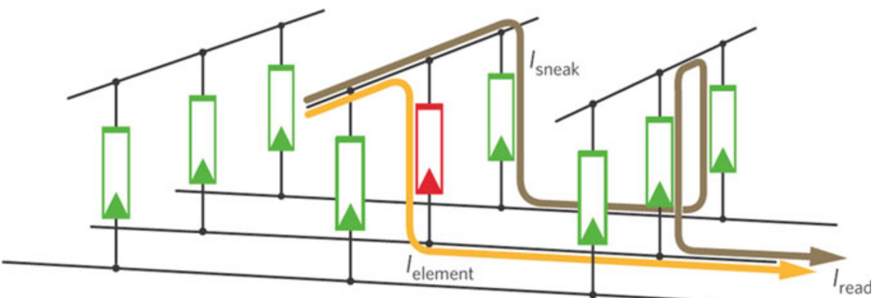


Fig. 8 Illustration of the sneak path issue in a crosspoint memory array. Reprinted by permission from Springer Nature Materials [112], Copyright (2010)

with increasing array size due to more current leakage paths, causing memory performance degradation. To suppress this sneak path problem, a prevalent solution at the device level is to incorporate a selection device with each RRAM device such as a transistor or a nonlinear selector, to get a 1T1R or 1S1R configuration, respectively.

Another key requirement for crosspoint memory arrays is high scalability such that the storage capacity can meet the specifications driven by the growth of internet data, data centers and mobile computers. A seemingly straightforward solution is to decrease the individual cell size such as reducing the RRAM device diameter. However, this raises concerns regarding increased spatial and temporal variability in the RRAM devices contained in the crosspoint array. Also, the interconnect and peripheral circuit area should decrease with downscaling which will increase series resistance due to geometry scaling and increased surface scattering [113], resulting in operational problems at high operating currents. Several methods have been reported to reduce interconnect resistance such as using graphene and carbon nanotubes as interconnect materials [114].

In order to circumvent the problems in downscaling, unique three-dimensional (3D) array architectures have been proposed such as the horizontal stacked 3D and vertical 3D structures. By stacking multiple 2D layers of memory crosspoint (Fig. 7) into a 3D structure, the minimal feature size is further reduced to $4F^2/n$, where n is the number of stacked layers. Between these two 3D structures, the vertical configuration has higher processing yield and cost effectiveness since only one critical lithography etch step is needed after the sequential deposition of multiple stacks. As depicted in Fig. 9, the memory cells in a vertical 3D crosspoint architecture, are formed at the sidewalls between the horizontal electrode and the vertical pillar electrode [115].

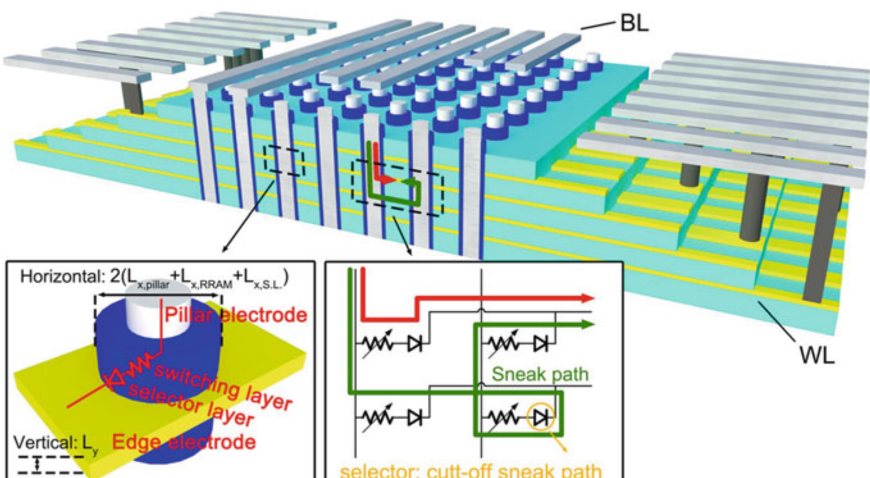


Fig. 9 Vertical 3D RRAM crosspoint architecture. Reprinted from [115] under a Creative Commons Attribution 4.0 International License. Full license terms at http://creativecommons.org/licenses/by/4.0/

Instead of reducing the cell size, high density in a 3D array is achieved by increasing the number of stacked layers.

3.1 RRAM Design Without Selector Device

The key advantages of a crosspoint array with no selector devices are the simple fabrication process and high array density. However, it will require the RRAM device to be self-rectifying in order to overcome the sneak path current problem.

One of the several self-rectifying concepts is the complementary resistive switching (CRS) devices first reported by Linn et al. in [112]. As shown in Fig. 10e, the CRS device is made up of two bipolar Cu/SE/Pt devices connected back to back with a shared Cu electrode in the middle. LRS occurs when a Cu filament is created across the SE by a positive voltage applied to the Cu electrode whereas a negative voltage will induce HRS via dissolution of the Cu filament. This CRS device is able to derive a unique I-V characteristic and four CRS states as shown in Fig. 10e–g. At both logic states (0 and 1), the leakage current is suppressed since one of the two Cu/SE/Pt cells within the CRS device is in HRS. A key disadvantage of this concept is that the read '1' operation is destructive; thus a write back process is required to revert the LRS (top cell)/LRS (bottom cell) to the original LRS/HRS state via an applied negative pulse. This increases the complexity of the peripheral circuitry and the power consumption [116]. Hitherto, the CRS phenomenon has been reported for several CBRAM devices [112, 117] as well as OxRRAM devices [86, 118, 119], using the anti-serial, back to back configuration. However, several challenges still

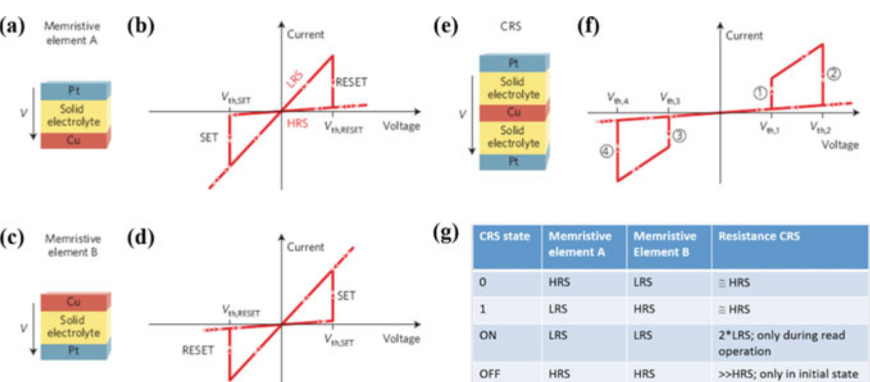


Fig. 10 **a** Bipolar memristive element A with a Pt/solid electrolyte/Cu stack. **b** I-V behavior of memristive element A shows that transition between HRS and LRS can be performed by exceeding $V_{th,RESET}$ and $V_{th,SET}$. **c** Bipolar memristive element A with a Cu/solid electrolyte/Pt stack. **d** I-V behavior of memristive element B. **e** Combining memristive elements A and B creates a CRS. **f** I-V behavior of a CRS. **g** All possible states of a CRS. Adapted by permission from Springer Nature Nature Materials [1124], Copyright (2010)

remain for practical CRS implementation such as the shared inner metal electrode and the inherent variability of each CBRAM in a CRS cell. It is also difficult to reduce the operation current in the CRS since the variability increases with the current reduction [120–122].

Another self-rectifying concept is to have a hybrid RRAM-selector cell such as the 3D double layer vertical structure demonstrated by Hou et al. [123–125]. By utilizing a Ta/TaO_x/TiO_x/Ti MIIM stack, stable bipolar resistive switching with over 10^3 self-rectifying ratio as well as superior endurance over 10^{10} cycles for both top and bottom MIIM cells. Homogeneous interfacial switching mechanism was proposed to elucidate the bipolar resistive switching mode and self-rectification [124]. The key merit of this structure is the absence of the additional inner middle electrode that separates the RRAM from the selector which makes it very attractive for the high density vertical 3D crosspoint architecture.

In 2017, Luo et al. demonstrated a bit cost scalable, 8-layer vertical 3D RRAM using a hybrid RRAM-selector cell [126]. With a TiN/HfO_2/TaO_x/Ti MIIM stack, the devices exhibited >100 non-linearity and high endurance (>10^7). In addition, scalability down to 5 nm device size and 4 nm vertical pitch in a 3D RRAM array was also demonstrated [126]. Despite its high performance and scalability, the high LRS and HRS resistances in the GΩ range might require innovative circuit design.

3.2 RRAM Design with Selection Device

Since most RRAM devices operate in the bipolar switching mode, we will focus on two-terminal selection devices compatible with bipolar RRAM. Based on I-V characteristics, Yu et al. broadly categorized these selection devices into two groups: Type I (Exponential I-V selectors) and Type II (Threshold I-V selectors) [127]. Figure 11 shows the typical I-V characteristics of a bipolar RRAM device with Type I and Type II selectors. The two key selector performance metrics are (1) non-linearity, which is the current ratio between V_w and $V_w/2$, and (2) drive current density, which ideally has to be higher than 10 MA/cm^2 if the device scales below the 10 nm technology node. The selector non-linearity ensures limited sneak current from unselected memory devices during write and read operations. Other selector performance metrics such as speed, cycling endurance, variability should ideally be as good as or better than the memory cell that it is paired with. In addition, the selector material and fabrication process should also be CMOS compatible.

As seen in Fig. 11a, Type I selectors employ an exponential I-V curve to switch on the selector with a current increase of several orders of magnitude, resulting in high non-linearity. Reported Type I selectors usually involve engineering the oxide/electrode interface to form a Schottky barrier [128–130] or MIM structures with a tunnelling oxide barrier [131–134].

Ideally, Type II selectors are more preferred due to their sudden turn-on property with steep slope. Type II selectors usually show a hysteresis in their I-V characteristics as depicted in Fig. 11b, where they turn on above a threshold voltage and turns off

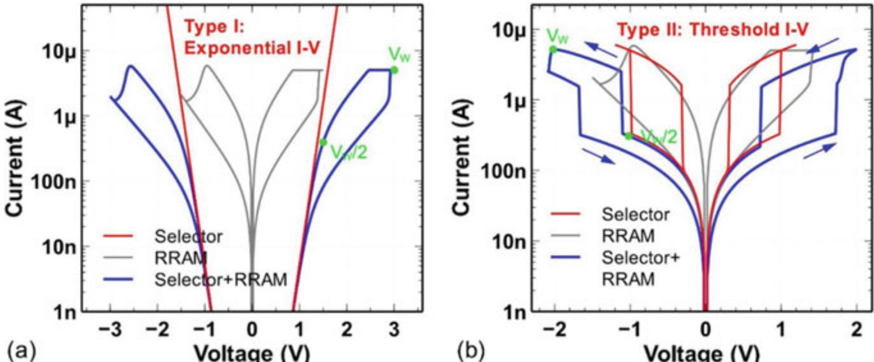


Fig. 11 Typical I-V characteristics of a bipolar RRAM with **a** Type I Exponential I-V selector and **b** Type II Threshold I-V selector, respectively. Reprinted by permission from Springer Nature Journal of Computational Electronics [127], Copyright (2017)

below a hold voltage. Such threshold switching is exhibited in metal-insulator transition (MIT) Mott oxide materials such as VO_2 [135–138] and NbO_2 [139–141]. MIT is an electrically or thermoelectrically-triggered, rapid, reversible transition between a high resistive insulating oxide state and a low resistive metallic state. Son et al. have demonstrated fast switching speed (<20 ns), large drive current density (>1MA/cm^2) and a relatively small non-linearity (~50) in their VO_2-based MIT selector devices [135]. Interestingly, only the nano-scale devices exhibited MIT I-V characteristics while the micro-scale devices showed ohmic behaviour, as shown in the inset of Fig. 14. In addition to its small non-linearity, the transition temperature of VO_2 is low at about 67 °C, which limits its practical applications since standard operating temperature is at 85 °C. Such limitations have led to new material considerations such as NbO_2, where higher non-linearity (~10^4) and higher drive current density (>30 MA/cm^2) has been demonstrated [139, 140]. This could be due to NbO_2 having a larger insulating bandgap and higher transition temperature as compared to VO_2 [142]. In addition, Park et al. developed a multi-layered $NiO_y/NbO_x/NiO_y$ structure capable of suppressing the leakage current such that relatively higher non-linearity (>5400), fast speed (<2 ns) and drift-free performance is achieved [143].

Other than Mott materials, Type II threshold selectors based on chalcogenide materials have shown excellent selector performance. Several of these selectors are based on the ovonic threshold switching (OTS) phenomenon, which was first reported by Ovshinsky in 1968 [144]. OTS is an electronic transition between high and low resistance states via a region of negative differential resistance. Such OTS selection devices are different from chalcogenide-based memory devices, where the latter must be actively switched between states [144]. Most OTS selectors reported high non-linearity (>10^4), which provides the full cell selectivity at read and write operations, and high on-state current (>30 MA/cm^2), which promotes scalability [145–147]. Several research groups have successfully demonstrated 1S1R integration using OTS selectors. In 2017, a joint project between IBM and Macronix demonstrated a 128 Gb

PCRAM crosspoint array using OTS selector that exhibited very low write (<300 ns) and read (<100 ns) latencies, good write high endurance (>10^7) and excellent thermal retention (>10^4 h at 85 °C) [148]. In terms of integration between OTS selector and RRAM, Lee et al. from Samsung reported in 2012, a 1S1R memory cell using a TaO_x-based RRAM and a AsGeTeSiN-based OTS select device [149]. The nanoscale selector showed high selectivity (10^3), endurance (10^8) and on-state current density (>10MA/cm^2).

Similarly, in 2017, Alayan et al. reported a 1S1R memory where HfO_x-based RRAM was stacked with a GeSe OTS selector [150]. A stable and novel reading strategy was proposed and up to 10^6 read cycles have been achieved in their 1S1R memory cell.

In 2014, Jo et al. from Crossbar Inc. reported a Field Assisted Superlinear Threshold (FAST) Selector based on a superlinear threshold layer (SLT) whereby a conduction channel is created at the threshold electric field [151]. By themselves, these Type II threshold selectors exhibited excellent selector performance such as very high non-linearity (>10^{10}), large drive current density (>5 MA/cm^2), high endurance (>10^8) and fast operation speeds (<50 ns). When the FAST selectors are integrated with low-current, forming-free RRAM cells in a 4 Mb 1S1R crosspoint array, the integrated 1S1R device demonstrated >10^2 memory on/off ratio and >10^6 non-linearity over 10^5 cycles [151].

Despite intense development of selector research in recent years, there is still no consensus on the best selector candidate as it must match the characteristics of the non-volatile memory device. Thus, the selector device still remains as a crucial challenge in the implementation of ultra-high density crosspoint memory architectures. Ultimately, a self-rectifying RRAM device would be ideal as it eliminates the requirement of an additional selector element.

4 Conclusions and Future Outlook

In the past decade, RRAM has seen much progress from single device up to 3D crosspoint array demonstrations. Nevertheless, challenges still persist such as detailed clarification of the resistive switching dynamics, the perfect matching of RRAM with selector or a reliable self-rectifying RRAM for true 3D vertical crosspoint implementation. Furthermore, new computing architectures are emerging, facilitated by the unique physical properties of RRAM and the perfect pairing between RRAM crosspoint arrays and the matrix-vector computation.

In the near future, RRAM technology is expected to have great potential in both data storage and computing. For instance, RRAM can be leveraged as a storage-class memory, which fills the gap between main memory and storage memory. In addition, the high-density, low-power and multi-level capabilities of RRAM has also motivated research into neuromorphic computing systems as a synaptic device, which would induce a paradigm shift in computing technology and a revolution in future electronic devices.

Acknowledgements The authors would like to acknowledge the Programmatic grant no. A1687b0033 from the Singapore government's Research, Innovation and Enterprise 2020 plan (Advanced Manufacturing and Engineering domain).

References

1. F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, S. Tanaka, A new flash E2PROM cell using triple polysilicon technology, in *Proceedings of the 1984 International Electron Devices Meeting*, pp. 464–467 (1984)
2. V.N. Kynett, A. Baker, M.L. Fandrich, G.P. Hoekstra, O. Jungroth, An in-system reprogrammable 32 K x 8 CMOS flash memory. IEEE J. Solid-State Circuits **23**(5), 1157–1163 (1988)
3. S. Lai, Non-volatile memory technologies: the quest for ever lower cost, in *Proceedings of the 2008 International Electron Devices Meeting* (2008)
4. H. Akinaga, H. Shima, Resistive random access memory (ReRAM) based on metal oxide. Proc. IEEE **98**(12), 2237–2251 (2010)
5. R. Waser, R. Dittmann, G. Staikov, K. Szot, Redox-based resistive switching memories—nanoionic mechanisms, prospects, and challenges. Adv. Mater. **21**(21), 2632–2663 (2009)
6. T. Prodromakis, C. Toumazou, L. Chua, Two centuries of memristors. Nat. Mater. **11**(6), 478–481 (2012)
7. T.W. Hickmott, Low-frequency negative resistance in thin anodic oxide films. J. Appl. Phys. **33**(9), 2669 (1962)
8. G. Dearnaley, A. Stoneham, D. Morgan, Electrical phenomena in amorphous oxide films. Reports Prog. Phys. **33**(3), 1129 (1970)
9. L.O. Chua, Memristor—The missing circuit element. IEEE Trans. Circuit Theory **18**, 507–519 (1971)
10. A. Asamitsu, Y. Tomioka, H. Kuwahara, Y. Tokura, Current switching of resistive states in magnetoresistive manganites. Nature **388**, 50–52 (1997)
11. A. Beck, J.G. Bednorz, C. Gerber, C. Rossel, D. Widmer, Reproducible switching effect in thin oxide films for memory applications. Appl. Phys. Lett. **77**(1), 139 (2000)
12. M.N. Kozicki, M. Yun, L. Hilt, A. Singh, Applications of programmable resistance changes in metal-doped chalcogenides, in *Proceedings of the Symposium Solid State Ion Devices,* no. 1, p. 849 (1999)
13. D.B. Strukov, G.S. Snider, D.R. Stewart, R.S. Williams, The missing memristor found. Nature **453**(12), 28–35 (2008)
14. Panasonic Starts World's First Mass Production of ReRAM Mounted Microcomputers, Panasonic Corporation, August 2013. [Online]. Available: https://news.panasonic.com/global/press/data/2013/07/en130730-2/en130730-2.html. Accessed 18 Nov 2018
15. T.-Y. Liu, T.H. Yan, R. Scheuerlein, Y. Chen, J.K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A 130.7-mm2 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology. IEEE J. Solid State Circuits **49**, 140–153 (2014)
16. R. Mertens, Digitimes, 2017. [Online]. Available: http://www.digitimes.com/newregister/join.asp?view=Article&DATEPUBLISH=2017/06/05&PAGES=PB&SEQ=200. Accessed 18 Nov 2018
17. M.J. Lee, Y. Park, B.S. Kang, S.E. Ahn, C. Lee, K. Kim, W.X. Xianyu, G. Stefanovich, J.H. Lee, S.J. Chung, Y.H. Kim, C.S. Lee, 2-Stack 1D-1R crosspoint structure with oxide diodes as switching elements for high density resistance RAM applications, in *2007 International Electron Devices Meeting* (2007)
18. C. Xu, X. Dong, N.P. Jouppi, Y. Xie, Design implications of memristor-based RRAM, in *DATE* (IEEE, New York, 2011)

19. M. Hu, H. Li, Q. Wu, G. S. Rose, Hardware realization of BSB recall function using memristor crossbar arrays, *Proceedings of the Design Automation Conference (DAC),* pp. 498–503 (2012)
20. M. Prezioso, F. Merrikh-Bayat, B.D. Hoskins, G.C. Adam, K.K. Likharev, D.B. Strukov, Training and operation of an integrated neuromorphic network based on metal-oxide memristors. Nature **521**, 61–64 (2015)
21. P.M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang, W.D. Lu, Sparse coding with memristor networks. Nature Nanotechnol. **12**, 784–789 (2017)
22. P. Yao, H. Wu, B. Gao, S.B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S.P. Wong, H. Qian, Face classification using electronic synapses. Nature Commun. **8**, 15199 (2017)
23. S. Yu, Binary neural network with 16 Mb RRAM macro chip for classification and online training, in *IEDM Technical Digest* (2016)
24. H.-S.P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F.T. Chen, M.-J. Tsai, Metal–Oxide RRAM, in *Proceedings of the IEEE,* pp. 1951–1970 (2012)
25. D.S. Jeong, R. Thomas, R. Katiyar, J. Scott, H. Kohlstedt, A. Petraru, C.S. Hwang, Emerging memories: resistive switching mechanisms and current status. Rep. Prog. Phys. **75**, 076502 (2012)
26. J.J. Yang, D.B. Strukov, D.R. Stewart, Memristive devices for computing. Nat. Nanotechnol. **8**, 13–24 (2013)
27. K. Szot, W. Speier, G. Bihlmayer, R. Waser, Switching the electrical resistance of individual dislocations in single-crystalline SrTiO3. Nat. Mater. **5**, 312–320 (2006)
28. K. Shibuya, R. Dittmann, S. Mi, R. Waser, Impact of defect distribution on resistive switching characteristics of Sr_2TiO_4 thin films. Adv. Mater. **22**, 411–414 (2010)
29. A. Sawa, Resistive switching in transition metal oxides. Mater. Today **11**, 28–36 (2008)
30. V.Y.-Q. Zhuo, Y. Jiang, M.H. Li, E.K. Chua, Z. Zhang, J.S. Pan, R. Zhao, L.P. Shi, Y. Yang, T.C. Chong, J. Robertson, Band alignment between Ta_2O_5 and metals for resistive random access memory. Appl. Phys. Lett. **102**, 062106 (2013)
31. V.Y.-Q. Zhuo, M.-H. Li, Y. Jiang, Electrode effects on the current conduction mechanisms in TaO_x-based RRAM, in *IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)* (2015)
32. J.J. Yang, M.D. Pickett, X. Li, D.A.A. Ohlberg, D.R. Stewart, R.S. Williams, Memristive switching mechanism for metal/oxide/metal nanodevices. Nat. Nanotechnol. **3**, 429 (2008)
33. J.P. Strachan, M.D. Pickett, J.J. Yang, S. Aloni, A.L.D. Kilcoyne, G. Medeiros-Ribeiro, R.S. Williams, Direct identification of the conducting channels in a functioning memristive device. Adv. Mater. **22**(32), 3573–3577 (2010)
34. I. Valov, R. Waser, J.R. Jameson, M.N. Kozicki, Electrochemical metallization memories—fundamentals, applications, prospects. Nanotechnology **22**(28), 289502 (2011)
35. W. Lu, D.S. Jeong, M. Kozicki, R. Waser, Electrochemical metallization cells—blending nanoionics into nanoelectronics?, *MRS Bull.* **37**(2), 124–130 (2012)
36. S. Yu, X. Guan, H.-S.P. Wong, Conduction mechanism of $TiN/HfO_x/Pt$ resistive switching memory: a trap-assisted-tunneling model. Appl. Phys. Lett. **99**, 063507 (2011)
37. Y.-M. Kim, J.-S. Lee, Reproducible resistance switching characteristics of hafnium oxide-based nonvolatile memory devices. J. Appl. Phys. **104**, 114115 (2008)
38. W.-Y. Chang, Y.-C. Lai, T.-B. Wu, S.-F. Wang, F. Chen, M.-J. Tsai, Unipolar resistive switching characteristics of ZnO thin films for nonvolatile memory applications. Appl. Phys. Lett. **92**, 022110 (2008)
39. B.J. Choi, D.S. Jeong, S.K. Kim, C. Rohde, S. Choi, J.H. Oh, H.J. Kim, C.S. Hwang, K. Szot, R. Waser, B. Reichenberg, S. Tiedke, Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition. J. Appl. Phys. **98**(3), 033715 (2005)
40. F. Miao, J.P. Strachan, J.J. Yang, M.X. Zhang, I. Goldfarb, A.C. Torrezan, P. Eschbach, R.D. Kelley, G. Medeiros-Ribeiro, R.S. Williams, Anatomy of a nanoscale conduction channel reveals the mechanism of a high-performance memristor. Adv. Mater. **23**, 5633 (2011)

41. V.Y.-Q. Zhuo, M. Li, Y. Guo, W. Wang, Y. Yang, Y. Jiang, J. Robertson, CMOS compatible electrode materials selection in oxide-based memory devices. J. Appl. Phys. **120**, 024504 (2016)
42. C.H. Wang, Y.-H. Tsai, K.-C. Lin, M.-F. Chang, Y.-C. King, C.-J. Lin, S.-S. Sheu, Y.-S. Chen, H.-Y. Lee, F.T. Chen, M.-J. Tsai, Three-dimensional 4F2 ReRAM cell with CMOS logic compatible process, in *IEDM Tech Digest,* pp. 664–667 (2010)
43. S. Privitera, G. Bersuker, S. Lombardo, C. Bongiorno, D.C. Gilmer, Conductive filament structure in HfO_2 resistive switching memory devices. Solid-State Electron. **111**, 161 (2015)
44. Z.X. Chen, Z. Fang, Y. Wang, Y. Yang, A. Kamath, X.P. Wang, N. Singh, G.-Q. Lo, D.-L. Kwong, Y.H. Wu, Impact of Ni concentration on the performance of Ni-Silicide/HfO_2/TiN resistive RAM (RRAM) cells. J. Electron. Mater. **43**(11), 4193–4198 (2014)
45. X.A. Tran, B. Gao, J.F. Kang, X. Wu, L. Wu, Z. Fang, Z.R. Wang, K.L. Pey, Y.C. Yeo, A.Y. Du, M. Liu, B.Y. Nguyen, M.F. Li, H.Y. Yu, Self-Rectifying and forming-free unipolar HfO_x based-high performance RRAM built by fab-available material, in *IEEE IEDM Technical Digest,* pp. 713–716 (2011)
46. D. Lee, H. Choi, H. Sim, D. Choi, Resistance switching of the nonstoichiometric zirconium oxide for nonvolatile memory applications. Electron Device Lett. **26**(10), 719–721 (2005)
47. H. Zhang, B. Gao, B. Sun, G. Chen, L. Zeng, L. Liu, X. Liu, J. Lu, R. Han, J. Kang, B. Yu, Ionic doping effect in ZrO_2 resistive switching memory. Appl. Phys. Lett. **96**, 123502 (2010)
48. X. Sun, B. Sun, L. Liu, N. Xu, X. Liu, R. Han, J. Kang, G. Xiong, T.P. Ma, Resistive switching in CeO_x films for nonvolatile memory application. IEEE Electron Device Lett. **30**(4), 334–336 (2009)
49. T.-M. Pan, C.-H. Lu, Switching behavior in rare-earth films fabricated in full room temperature. IEEE Trans Electr. Devices **59**(4), 956–961 (2012)
50. W. Shen, R. Dittmann, U. Breuer, R. Waser, Improved endurance behavior of resistive switching in (Ba, Sr)TiO_3 thin films with W top electrode. Appl. Phys. Lett. **93**, 222102 (2008)
51. X. Chen, H. Zhang, K. Ruan, W. Shi, Annealing effect on the bipolar resistive switching behaviors of $BiFeO_3$ thin films on $LaNiO_3$-buffered Si substrates. J. Alloys Compd. **529**, 108–112 (2012)
52. S.Q. Liu, N.J. Wu, A. Ignatiev, Electric-pulse-induced reversible resistance change effect in magnetoresistive films. Appl. Phys. Lett. **76**(19), 2749–2751 (2000)
53. X.P. Wang, Z. Fang, Z.X. Chen, A.R. Kamath, L.J. Tang, G.-Q. Lo, D.-L. Kwong, Ni-Containing Electrodes for Compact Integration of Resistive Random Access Memory With CMOS. IEEE Electron Device Lett. **34**(4), 508–510 (2013)
54. H.J. Hovel, J.J. Urgell, Switching and memory characteristics of ZnSe-Ge heterojunctions. J. Appl. Phys. **42**(12), 5076–5083 (1971)
55. N.G. Patel, Some observations on the switching and memory phenomena in ZnTe-Si. J. Mater. Sci. **21**(6), 2097–2099 (1986)
56. S. Gao, X. Yi, J. Shang, G. Liu, R.-W. Li, Organic and hybrid resistive switching materials and devices. Chem. Soc. Rev. pp. 1–35 (2018)
57. B.-H. Lee, H. Bae, H. Seong, D.-I. Lee, H. Park, Y.J. Choi, S.-G. Im, S.O. Kim, Y.-K. Choi, Direct observation of a carbon filament in water-resistant organic memory. ACS Nano **9**(7), 7306–7313 (2015)
58. R. Waser, M. Aono, Nanoionics-based resistive switching memories. Nat. Mater. **6**(11), 833–840 (2007)
59. D.-H. Kwon, K.M. Kim, J.H. Jang, J.M. Jeon, M.H. Lee, G.H. Kim, L.X.-S., G.-S. Park, B. Lee, S. Han, M. Kim, C.S. Hwang, Atomic structure of conducting nanofilaments in TiO_2 resistive switching memory. Nat. Nanotechnol. **5**(2), 148–153 (2010)
60. M. Janousch, G.I. Meijer, U. Staub, B. Delley, S.F. Karg, B.P. Andreasson, Role of oxygen vacancies in Cr-doped $SrTiO_3$ for resistance-change memory. Adv. Mater. **19**(17), 2232–2235 (2007)
61. Y.B. Nian, J. Strozier, N.J. Wu, X. Chen, A. Ignatiev, Evidence for an oxygen diffusion model for the electric pulse induced resistance change effect in transition-metal oxides. Phys. Rev. Lett. **98**(14), 146403 (2007)

62. C. Baeumer, C. Schmitz, A. Marchewka, D.N. Mueller, R. Valenta, J. Hackl, N. Raab, S.P. Rogers, M.I. Khan, S. Nemsak, Quantifying redox-induced Schottky barrier variations in memristive devices via in operando spectromicroscopy with graphene electrodes. Nat. Commun. **7**, 12398 (2016)
63. H. Du, C.-L. Jia, A. Koehl, J. Barthel, R. Dittmann, R. Waser, J. Mayer, Nanosized conducting filaments formed by atomic-scale defects in redox-based resistive switching memories. Chem. Mater. **29**, 3164–3173 (2017)
64. M. Sowinska, T. Bertaud, D. Walczyk, S. Thiess, M. Schubert, M. Lukosius, W. Drube, C. Walczyk, T. Schroeder, In-operando and non-destructive analysis of the resistive switching in the Ti/HfO_2/TiN-based system by hard x-ray photoelectron spectroscopy. Appl. Phys. Lett. **101**, 143501 (2012)
65. U. Celano, J.O.D. Beeck, S. Clima, M. Luebben, P.M. Koenraad, L. Goux, I. Valov, W. Vandervorst, Direct probing of the dielectric scavenging-layer interface in oxide filamentary-based valence change memory. ACS Appl. Mater. Interfaces. **9**, 10820–10824 (2017)
66. G. Niu, M. Schubert, S. Sharath, P. Zaumseil, S. Vogel, C. Wenger, E. Hildebrandt, S. Bhupathi, E. Perez, L. Alff, Direct probing of the dielectric scavenging-layer interface in oxide filamentary-based valence change memory. Nanotechnology **28**, 215702 (2017)
67. D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. La Torre, R.E. Dunin-Borkowski, S. Menzel, R. Waser, R. Dittmann, Anomalous resistance hysteresis in oxide ReRAM: oxygen evolution and reincorporation revealed by in situ TEM. Adv. Mater. **29**, 1–8 (2017)
68. J. Kwon, A.A. Sharma, C.-Y. Chen, A. Fantini, M. Jurczak, A.A. Herzing, J.A. Bain, Y.N. Picard, M. Skowronski, Transient thermometry and high-resolution transmission electron microscopy analysis of filamentary resistive switches. ACS Appl. Mater. Interfaces **8**, 20176–20184 (2016)
69. K.M. Kim, D.S. Jeong, C.S. Hwang, Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook. Nanotechnology **22**(25), 254002 (2011)
70. D.S. Jeong, H. Schroeder, U. Breuer, R. Waser, Characteristic electroforming behavior in Pt/TiO_2/Pt resistive switching cells depending on atmosphere. J. Appl. Phys. **104**(12), 123716 (2008)
71. J.J. Yang, F. Miao, M.D. Pickett, D.A. Ohlberg, D.R. Stewart, C.N. Lau, R.S. Williams, The mechanism of electroforming of metal oxide memristive switches. Nanotechnology **20**(21), 215201 (2009)
72. J.-Y. Chen, C.-L. Hsin, C.-H.C.Y.-T.H.S.-J.L.-W.L.-J.C.C.-W. Huang, Dynamic evolution of conducting nanofilament in resistive switching memories. Nano Lett. **13**, 3671–3677 (2013)
73. D.B. Strukov, F. Alibart, R.S. Williams, Thermophoresis/diffusion as a plausible mechanism for unipolar resistive switching in metal–oxide–metal memristors. Appl. Phys. A **107**, 509–518 (2012)
74. S. Kumar, Z. Wang, X. Huang, N. Kumari, N. Davila, J.P. Strachan, D. Vine, A.D. Kilcoyne, Y. Nishi, R.S. Williams, Conduction channel formation and dissolution due to oxygen thermophoresis/diffusion in hafnium oxide memristors. ACS Nano **10**, 11205–11210 (2016)
75. Z.H. Tan, R. Yang, K. Terabe, X.B. Yin, X.D. Zhang, X. Guo, Synaptic metaplasticity realized in oxide memristive devices. Adv. Mater. **28** (2016)
76. Y. Yang, X. Zhang, L. Qin, Q. Zeng, X. Qiu, R. Huang, Probing nanoscale oxygen ion motion in memristive systems. Nat. Commun. **8**, 15173 (2017)
77. Y. Jiang, Q. Qi, R. Wang, J. Zhang, Q. Xue, C. Wang, C. Jiang, X. Qiu, Direct observation and measurement of mobile charge carriers in a monolayer organic semiconductor on a dielectric substrate. ACS Nano **5**, 6195–6201 (2011)
78. S. Wang, R. Wang, X. Wang, D. Zhang, X. Qiu, Nanoscale charge distribution and energy band modification in defect-patterned graphene. Nanoscale **4**, 2651–2657 (2012)
79. J. Zhang, P. Chen, B. Yuan, W. Ji, Z. Cheng, X. Qiu, Real-space identification of intermolecular bonding with atomic force microscopy. Science **342**, 611–614 (2013)
80. S. Menzel, M. Waters, A. Marchewka, U. Böttger, R. Dittmann, R. Waser, Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. Adv. Funct. Mater. **21**, 4487–4492 (2011)

81. S. Yu, Y. Wu, H.-S.P. Wong, Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory. Appl. Phys. Lett. **98**, 103514 (2011)
82. U. Russo, D. Ielmini, C. Cagli, A.L. Lacaita, Filament Conduction and Reset Mechanism in NiO-Based Resistive-Switching Memory (RRAM) Devices. IEEE Trans. Electron Devices **56**, 186 (2009)
83. U. Russo, D. Ielmini, C. Cagli, A.L. Lacaita, S. Spiga, M. Perego, M. Fanciulli, Conductive-filament switching analysis and self-accelerated thermal dissolution model for reset in NiO-based RRAM. IEDM Tech. Dig. **2**, 775 (2007)
84. U. Russo, D. Ielmini, C. Cagli, A.L. Lacaita, Self-accelerated thermal dissolution model for reset programming in unipolar resistive-switching memory (RRAM) devices. IEEE Trans. Electron Devices **56**, 193 (2009)
85. D. Panda, P.P. Sahu, Thermal assisted reset modelling in nickel oxide based unipolar resistive switching. J. Appl. Phys. **121**, 204504 (2017)
86. M.-J. Lee, C.B. Lee, D. Lee, S.R. Lee, M. Chang, J.H. Hur, Y.-B. Kim, C.-J. Kim, D.H. Seo, S. Seo, U.I. Chung, I.-K. Yoo, A fast, high-endurance and scalable non-volatile memory device made from asymmetric $Ta_2O_{(5-x)}/TaO_{(2-x)}$ bilayer structures. Nat. Mater. **10**, 625 (2011)
87. 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, M.-J. Tsai, Evidence and solution of over-RESET problem for HfO_x based resistive memory with sub-ns switching speed and high endurance (2010)
88. Y. Hirose, H. Hirose, Polarity-dependent memory switching and behavior of Ag dendrite in Ag-photodoped amorphous As2S3 films. J. Appl. Phys. **47**, 2767–2772 (1976)
89. W.C. West, K. Sieradzki, B. Kardynal, M.N. Kozicki, Equivalent circuit modeling of the Ag vertical bar $As_{0.24}S_{0.36}Ag_{0.40}$ vertical bar Ag system prepared by photodissolution of Ag. J. Electrochem. Soc. **145**, 2971–2974 (1998)
90. T. Hasegawa, K. Terabe, T. Tsuruoka, M. Aono, Atomic switch: Atom/ ion movement controlled devices for beyond von-Neumann computers. Adv. Mater. **24**, 252–267 (2012)
91. S.H. Jo, K.H. Kim, W. Lu, Programmable resistance switching in nanoscale two-terminal devices. Nano Lett. **9**, 496–500 (2009)
92. U. Russo, D. Kamalanathan, D. Ielmini, A.L. Lacaita, M.N. Kozicki, Study of multilevel programming in programmable metallization cell (PMC) memory. Electron Dev. IEEE Trans. on **56**, 1040–1047 (2009)
93. N. Banno, T. Sakamoto, T. Hasegawa, K. Terabe, M. Aono, Effect of ion diffusion on switching voltage of solid-electrolyte nanometer switch. Jpn. J. Appl. Phys. **45**, 3666–3668 (2006)
94. I. Valov, I. Sapezanskaia, A. Nayak, T. Tsuruoka, T. Bredow, T. Hasegawa, G. Staikov, M. Aono, R. Waser, Atomically controlled electrochemical nucleation at superionic solid electrolyte surfaces. Nat. Mater. **11**, 530–535 (2012)
95. K. Terabe, T. Hasegawa, T. Nakayama, M. Aono, Quantized conductance atomic switch. Nature **433**, 47–50 (2005)
96. T. Sakamoto, K. Lister, N. Banno, T. Hasegawa, K. Terabe, M. Aono, Electronic transport in Ta_2O_5 resistive switch. Appl. Phys. Lett. **91**, 092110 (2007)
97. T. Kever, U. Bottger, C. Schindler, R. Waser, On the origin of bistable resistive switching in metal organic charge transfer complex memory cells. Appl. Phys. Lett. **91**, 083506 (2007)
98. C. Chen, Y.C. Yang, F. Zeng, F. Pan, Bipolar resistive switching in Cu/AlN/ Pt nonvolatile memory device. Appl. Phys. Lett. **97**, 083502–083503 (2010)
99. W.H. Guan, M. Liu, S.B. Long, Q. Liu, W. Wang, On the resistive switching mechanisms of Cu/ZrO_2:Cu/Pt. Appl. Phys. Lett. **93**, 223506 (2008)
100. T. Liu, M. Verma, Y. Kang, M.K. Orlowski, Coexistence of Bipolar and Unipolar Switching of Cu and Oxygen Vacancy Nanofilaments in $Cu/TaO_x/Pt$ Resistive Devices. ECS Solid State Lett. **1**(1), Q11–Q13 (2012)
101. X. Guo, C. Schindler, S. Menzel, R. Waser, Understanding the switching-off mechanism in Ag+ migration based resistively switching model systems. Appl. Phys. Lett. **91**, 133513 (2007)

102. Y. Yang, Y. Takahashi, A. Tsurumaki-Fukuchi, M. Arita, M. Moors, M. Buckwell, A. Mehonic, A.J.J. Kenyon, Probing electrochemistry at the nanoscale: in situ TEM and STM characterizations of conducting filaments in memristive devices. J. Electroceram. **39**(1–4), 73–93 (2017)
103. W.A. Hubbard, A. Kerelsky, G. Jasmin, E.R. White, J. Lodico, M. Mecklenburg, B.C. Regan, Nanofilament formation and regeneration during Cu/Al_2O_3 resistive memory switching. Nano Lett. **15**(6), 3983–3987 (2015)
104. S.-J. Choi, G.-S. Park, K.-H. Kim, S. Cho, W.-Y. Yang, X.-S. Li, J.-H. Moon, K.-J. Lee, K. Kim, In situ observation of voltage-induced multilevel resistive switching in solid electrolyte memory. Adv. Mater. **23**, 3273 (2011)
105. X.Z. Tian, S.Z. Yang, M. Zeng, L.F. Wang, J.K. Wei, Z. Xu, W.L. Wang, X.D. Bai, Bipolar electrochemical mechanism for mass transfer in nanoionic resistive memories. Adv. Mater. **26**, 3649 (2014)
106. Q. Liu, J. Sun, H. Lv, S. Long, K. Yin, N. Wan, Y. Li, L. Sun, M. Liu, Real-time observation on dynamic growth/dissolution of conductive filaments in oxide-electrolyte-based ReRAM. Adv. Mater. **24**(14), 1844–9 (2012)
107. Y. Yang, P. Gao, S. Gaba, T. Chang, X. Pan, W. Lu, Observation of conducting filament growth in nanoscale resistive memories. Nat. Commun. **3**, 732 (2012)
108. U. Celano, L. Goux, A. Belmonte, K. Opsomer, A. Franquet, A. Schulze, C. Detavernier, O. Richard, H. Bender, M. Jurczak, Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices. Nano Lett. **14**, 2401 (2014)
109. Y. Yang, P. Gao, L. Li, X. Pan, S. Tappertzhofen, S. Choi, R. Waser, I. Valov, W.D. Lu, Electrochemical dynamics of nanoscale metallic inclusions in dielectrics. Nat Commun. **5**, 4232 (2014)
110. C.P. Hsiung, H.W. Liao, J.Y. Gan, T.B. Wu, J.C. Hwang, F. Chen, M.J. Tsai, Formation and instability of silver nanofilament in Ag-based programmable metallization cells. ACS Nano **4**, 5414 (2010)
111. Z. Xu, Y. Bando, W. Wang, X. Bai, D. Golberg, Real-time in situ HRTEM-resolved resistance switching of Ag_2S nanoscale ionic conductor. ACS Nano **4**, 2515–2522 (2010)
112. E. Linn, R. Rosezin, C. Kügeler, R. Waser, Complementary resistive switches for passive nanocrossbar memories. Nat. Mater. **9**(5), 403–406 (2010)
113. P. Kapur, J.P. McVittie, K.C. Saraswat, Technology and reliability constrained future copper interconnects—Part I: Resistance modeling. IEEE Trans. Electron Devices **49**, 590–597 (2002)
114. A.K. Geim, K.S. Novoselov, The rise of graphene. Nat. Mater. **6**, 183–191 (2007)
115. S. Yu, Y. Deng, B. Gao, P. Huang, B. Chen, X. Liu, J. Kang, H.-Y. Chen, Z. Jiang, H.-S.P. Wong, Design Guidelines for 3D RRAM Cross-Point Architecture, in *2014 IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 421–424 (2014)
116. G.W. Burr, R.S. Shenoy, K. Virwani, P. Narayanan, A. Padilla, B. Kurdi, Access devices for 3D crosspoint memory. J. Vac. Sci. Technol. **32**(4), 040802 (2014)
117. R. Rosezin, E. Linn, L. Nielen, C. Kugeler, R. Bruchhaus, R. Waser, Integrated complementary resistive switches for passive high-density nanocrossbar arrays. IEEE Electron Device Lett. **32**(2), 191–193 (2011)
118. J. Lee, J. Shin, D. Lee, W. Lee, S. Jung, M. Jo, J. Park, K.P. Biju, S. Kim, S. Park, H. Hwang, Diode-less nano-scale ZrO_x/HfO_x RRAM device with excellent switching uniformity and reliability for high-density cross-point memory applications, in *Proceedings of the IEEE Electron Devices Meeting (IEDM)* (2010)
119. D.J. Wouters, L. Zhang, A. Fantini, R. Degraeve, L. Goux, Y.Y. Chen, B. Govoreanu, G.S. Kar, G.V. Groeseneken, M. Jurczak, Analysis of complementary RRAM switching. IEEE Electron Device Lett. **33**(8), 1186–1188 (2012)
120. A. Fantini, L. Goux, R. Degraeve, D.J. Wouters, N. Raghavan, G. Kar, A. Belmonte, Y.-Y. Chen, B. Govoreanu and M. Jurczak, Intrinsic switching variability in HfO_2 RRAM, in *Proceedings of the 5th IEEE International Memory Workshop (IMW)* (2013)

121. N. Raghavan, R. Degraeve, A. Fantini, L. Goux, D.J. Wouters, G. Groeseneken, M. Jurczak, Stochastic variability of vacancy filament configuration in ultra-thin dielectric RRAM and its impact on OFF-state reliability, in *Proceedings of the IEEE International Electron Devices Meeting (IEDM)*, pp. 21.1.1–21.1.4 (2013)
122. S. Yu, X. Guan, H.-S.P. Wong, On the stochastic nature of resistive switching in metal oxide RRAM: Physical modeling Monte Carlo simulation and experimental characterization, in *Proceedings of the IEEE International Electron Devices Meeting (IEDM)*, pp. 17.3.1–17.3.4 (2011)
123. C.-W. Hsu, I.-T. Wang, C.-L. Lo, M.-C. Chiang, W.-Y. Jang, C.-H. Lin, T.-H. Hou, Self-rectifying bipolar TaO_x/TiO_2RRAM with superior endurance over 10e12cycles for 3D high-density storage-class memory, in *Technical Digest VLSI Symposium*, pp. 166–167 (2013)
124. C.-W. Hsu, Y.-F. Wang, C.-C. Wan, I.-T. Wang, C.-T. Chou, W.-L. Lai, Y.-J. Lee, T.-H. Hou, Homogeneous barrier modulation of TaO_x/TiO_2 bilayers for ultra-high endurance three-dimensional storage-class memory. Nanotechnology **25**(16), 165202 (2014)
125. C.-W. Hsu, C.-C. Wan, I.-T. Wang, M.-C. Chen, C.-L. Lo, Y.-J. Lee, W.-Y. Jang, C.-H. Lin, T.-H. Hou, 3D vertical TaO_x/TiO_2 RRAM with over 10e3 self-rectifying ratio and sub-μA operating current, in *Proceedings of the IEEE International Electron Devices Meeting (IEDM)*, pp. 10.4.1–10.4.4 (2013)
126. Q. Luo, X. Xu, T. Gong, H. Lv, D. Dong, H. Ma, P. Yuan, J. Gao, J. Liu, Z. Yu, J. Li, S. Long, Q. Liu, M. Liu, 8-Layers 3D vertical RRAM with excellent scalability towards storage class memory applications, in *IEEE International Electron Devices Meeting (IEDM),* pp. 2.7.1–2.7.4 (2017)
127. X. Peng, R. Madler, P.-Y. Chen, S. Yu, Cross-point memory design challenges and survey of selector device characteristics. J. Comput. Electron. **16**(4), 1167–1174 (2017)
128. J.-J. Huang, Y.-M. Tseng, W.-C. Luo, C.-W. Hsu, T.-H. Hou, One selector-one resistor (1s1r) crossbar array for high-density flexible memory applications, *Proceedings of the IEEE International Electron Devices Meeting (IEDM)*, pp. 31.7.1–31.7.4 (2011)
129. J.-J. Huang, Y.-M. Tseng, C.-W. Hsu, T.-H. Hou, Bipolar nonlinear $Ni/TiO_2/Ni$ selector for 1S1R crossbar array applications. IEEE Electron Device Lett. **32**(10), 1427–1429 (2011)
130. J. Shin, I. Kim, K.P. Biju, M. Jo, J. Park, J. Lee, S. Jung, W. Lee, S. Kim, S. Park, H. Hwang, TiO_2-based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application. J. Appl. Phys. **109**, 033712 (2011)
131. B. Govoreanu, C. Adelmann, A. Redolfi, L. Zhang, S. Clima, M. Jurczak, High-performance metal-insulator-metal tunnel diode selectors. IEEE Electron Device Lett. **35**(1), 63–65 (2014)
132. J. Woo, J. Song, K. Moon, J.H. Lee, E. Cha, A. Prakash, D. Lee, S. Lee, S. Park, Y. Koo, C.G. Park, H. Hwang, Electrical and reliability characteristics of a scaled (~30 nm) tunnel barrier selector ($W/Ta_2O_5/TaO_x/TiO_2/TiN$) with excellent performance (JMAX > 107A/cm2), in *Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers,* pp. 1–2 (2014)
133. U. Chand, K.C. Huang, C.Y. Huang, T.Y. Tseng, Mechanism of Nonlinear Switching in HfO_2-Based Crossbar RRAM With Inserting Large Bandgap Tunneling Barrier Layer. IEEE Trans. Electron Devices **62**(11), 3665–3670 (2015)
134. B.J. Choi, J. Zhang, K. Norris, G. Gibson, K.M. Kim, W. Jackson, M.M. Zhang, Z. Li, J.J. Yang, R.S. Williams, Trilayer tunnel selectors for memristor memory cells. Adv. Mater. **28**, 356–362 (2016)
135. M. Son, J. Lee, J. Park, J. Shin, G. Choi, S. Jung, W. Lee, S. Kim, S. Park, H. Hwang, Excellent selector characteristics of nanoscale VO2 for high-density bipolar ReRAM applications. IEEE Electron Device Letter **32**(11), 1579–1581 (2011)
136. K. Zhang, B. Wang, F. Wang, Y. Han, X. Jian, H. Zhang, H.S.P. Wong, VO2-based selection device for passive resistive random access memory application. IEEE Electron Device Lett. **37**(8), 978–981 (2016)
137. I.P. Radu, B. Govoreanu, K. Marten, M. Toeller, A.P. Peter, M.R. Ikram, L.Q. Zhang, H. Hody, W. Kim, P. Favia, T. Conard, H.Y. Chou, B. Put, V.V. Afanasiev, A. Stesmans, A. Heyns, S.D. Gendt, M. Jurczak, Vanadium dioxide for selector applications. ECS Trans. **58**(7), 249–258 (2013)

138. K. Martens, I.P. Radu, S. Mertens, X. Shi, L. Nyns, S. Cosemans, P. Favia, H. Bender, T. Conard, M. Schaekers, S. De Gendt, V. Afanas'ev, J.A. Kittl, M. Heyns, M. Jurczak, The VO2 interface, the metal-insulator transition tunnel junction, and the metal-insulator transition switch on-off resistance. J. Appl. Phys. **112**, 124501 (2012)
139. E. Cha, J. Woo, D. Lee, S. Lee, J. Song, Y. Koo, J.H. Lee, C.G. Park, M.Y. Yang, K. Kamiya, K. Shiraishi, B. Magyari-Köpe, Y. Nishi, H. Hwang, Nanoscale (~10 nm) 3D vertical ReRAM and NbO_2 threshold selector with TiN electrode, *IEEE International Electron Devices Meeting,* pp. 10.5.1–10.5.4 (2013)
140. S. Kim, T.J. Ha, S. Kim, J.Y. Lee, K.K.W., J.H. Shin, Y.T. Park, S.P. Song, B.Y. Kim, W.G. Kim, J.C. Lee, H.S. Lee, J.H. Song, E.R. Hwang, S.H. Cho, J.C. Ku, J.I. Kim, K.S. Kim, J.H. Yoo, H.J. Kim, H.G. Jung, Improvement of characteristics of NbO_2 selector and full integration of 4F2 2x-nm tech 1S1R ReRAM, in *IEEE International Electron Devices Meeting (IEDM),* pp. 10.3.1–10.3.4 (2015)
141. X. Liu, S.K. Nandi, D.K. Venkatachalam, K. Belay, S. Song, R.G. Elliman, Reduced threshold current in NbO_2 selector by engineering device structure. IEEE Electron Device Lett. **35**(10), 1055–1057 (2014)
142. T. Sakata, K. Sakata, G. Höfer, T. Horiuchi, Preparation of NbO_2 single crystals by chemical transport reaction. J. Cryst. Growth **12**(2), 88–92 (1972)
143. J. Park, T. Hadamek, A.B. Posadas, E. Cha, A.A. Demkov, H. Hwang, Multi-layered $NiO_y/NbO_x/NiO_y$ fast drift-free threshold switch with high I_{on}/I_{off} ratio for selector application. Sci. Rep. **7**, 4068 (2017)
144. S.R. Ovshinsky, Reversible electrical switching phenomena in disordered structures. Phys. Rev. Lett. **21**(20), 1450–1453 (1968)
145. Y. Koo, H. Hwang, $Zn_{1-x}Te_x$ ovonic threshold switching device performance and its correlation to material parameters. Sci. Rep. **8**, 11822 (2018)
146. T. Gao, J. Feng, H. Ma, The ovonic threshold switching characteristics in Si_xTe_{1-x} based selector devices. Appl. Phys. A **124**, 734 (2018)
147. M. Lee, D. Lee, S.-H. Cho, J.-H. Hur, S.-M. Lee, D.H. Seo, D.-S. Kim, M.-S. Yang, S. Lee, E. Hwang, M.R. Uddin, H. Kim, U.-I. Chung, Y. Park, I.-K. Yoo, A plasma-treated chalcogenide switch device for stackable scalable 3D nanoscale memory. Nat. Commun. **4**, 2629 (2013)
148. H.Y. Cheng, W.C. Chien, I.T. Kuo, E.K. Lai, Y. Zhu, J.L. Jordan-Sweet, A. Ray, F. Carta, F.M. Lee, P. Tseng, M. Lee, Y. Lin, W. Kim, R. Bruce, C.W. Yeh, C. Yang, M. BrightSky, H. Lung, An ultra high endurance and thermally stable selector based on TeAsGeSiSe chalcogenides compatible with BEOL IC Integration for cross-point PCM, in *IEEE International Electron Devices Meeting (IEDM),* pp. 2.2.1–2.2.4 (2017)
149. M.-J. Lee, D. Lee, H. Kim, H.-S. Choi, J.-B. Park, H.G. Kim, Y.-K. Cha, U.-I. Chung, I.-K. Yoo, K. Kim, Highly-scalable threshold switching select device based on chaclogenide glasses for 3D nanoscaled memory arrays, in *IEEE International Electron Devices Meeting (IEDM),* pp. 2.6.1–2.6.3 (2012)
150. M. Alayan, E. Vianello, G. Navarro, C. Carabasse, S. La Barbera, A. Verdy, N. Castellani, A. Levisse, G. Molas, L. Grenouillet, T. Magis, F. Aussenac, M. Bernard, B. DeSalvo, J.M. Portal, E. Nowak, In-depth investigation of programming and reading operations in RRAM cells integrated with Ovonic Threshold Switching (OTS) selectors, in *IEEE International Electron Devices Meeting (IEDM),* pp. 2.3.1–2.3.4 (2017)
151. S.H. Jo, T. Kumar, S. Narayanan, W.D. Lu, H. Nazarian, 3D-stackable crossbar resistive memory based on Field Assisted Superlinear Threshold (FAST) selector, in *IEEE International Electron Devices Meeting (IEDM),* pp. 6.7.1–6.7.4 (2014)

RRAM Device Characterizations and Modelling

Peng Huang, Bin Gao, and Jinfeng Kang

Abstract Resistive random access memory (RRAM) is one of the most promising candidate for future nanoscale nonvolatile memory. Extensive research efforts have been carried out to facilitate practical use of RRAM as data storage system. However, further improvements, such as reducing the operation voltage and current, suppressing the device variability, etc., are still needed for the commercialization of RRAM. To further optimize the device performance, physical mechanism of resistive switching behavior must be understood and physical model should be developed. This chapter summarizes the current physical mechanisms, which provides an atom view of the resistive switching behavior. Then we will discuss the materials characterization used to identify the origins of switching behaviors, including the high-resolution X-ray photoelectron spectroscopy (XPS), electron energy loss spectrum (EELS), in situ transmission electron microscopy (TEM) and so on. After that, Monte Carlo simulation of the dynamic resistive switching processes is presented, allowing for correlating the observed switching characteristics with the microcosmic physical processes. Besides, compact model for spice simulation of RRAM based circuit is discussed. Finally, we will introduce the electrical characterization of RRAM, such as retention, endurance, RTN and so on.

1 Introduction

Resistance switching phenomenon that the insulator can be switched between different states has been found for over 40 years. The first reports of this phenomenon can be traced back to 1960s [1, 2]. In the late 1990s and the early 2000s, the discovery of hysteresis I-V characteristic in perovskite is reported [3, 4]. Since Samsung

P. Huang (✉) · J. Kang
Institute of Microelectronics, Peking University, No. 5 Yiheyuan Road, Beijing 100871, P. R. China
e-mail: phwang@pku.edu.cn

B. Gao
Institute of Microelectronics, Beijing Innovation Center for Future Chips (ICFC), Tsinghua University, Beijing 100084, P. R. China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_11

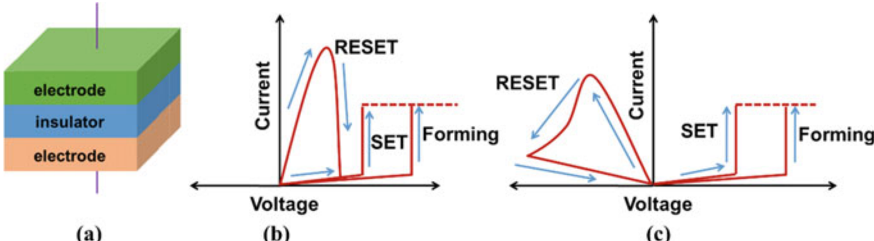


Fig. 1 **a** Schematic of MIM structure for Ox-RRAM, I-V characteristics of **b** unipolar and **c** bipolar I-V characteristics

demonstrated NiO resistance random access memory (RRAM) array in the 2004 [5], research activities on the binary oxides have been blooming. Up to now, more than 30 types of materials, such as TiO_x [6], ZrO_x [7], ZnO_x [8], Ce O_x [9], HfO_x [10], TaO_x [11], AlO_x [12], etc., can exhibit resistive switching phenomenon. There are two types of RRAM devices according to the resistive switching material and mechanism. The first type is the oxide based RRAM (Ox-RRAM), which is based on the redistribution of oxygen vacancies (V_O) in the resistive switching layer [13, 14]. Another type is the conducive-bridge RAM (CBRAM), which relies on the fast-diffusing Ag or Cu ions into the oxide or chalcogenide [15, 16]. In this chapter, we focus on the Ox-RRAM.

First, some basic concepts and terminologies about Ox-RRAM are introduced. Ox-RRAM is a sandwich structure (MIM) as shown in Fig. 1a. The switching behavior not only depends on the resistive switching layer, which is composed of oxide based insulator, but also on the electrodes and their interfaces. Generally, the oxide used as the resistive layer is nonstoichiometric. Ox-RRAM can be switched between two states, low resistance state (LRS) and high resistance state (HRS), or multi-level states [10, 17]. The process of device switching from LRS to HRS is called SET and the converse switching process is called RESET. The voltages triggering the SET and RESET process are called SET voltage and RESET voltage, respectively. For a fresh device, an "electroform" or Forming process is needed to trigger the resistive switching behavior for the subsequent cycles. The voltage used for the Forming process is usually lager than the SET and RESET voltage. There are two switching modes of RRAM, unipolar and bipolar. Figure 1b, c schematically shows the I-V characteristics for the two switching modes. For unipolar switching, the transition direction depends on the amplitude of the applied voltage and current state but not on the polarity of the applied voltage. If the device can symmetrically switch for positive and negative voltages, it is also referred as a nonpolar switching mode. For bipolar device, SET only occurs at one polarity and only the voltage with reverse polarity can trigger the RESET process. A current compliance is often used to avoid the permanent breakdown during the SET and Forming process. To detect the state of device, a small read voltage is applied without disturbance on its state.

2 Physical Mechanism of RRAM

2.1 Resistive Switching Mechanism

Bipolar

The resistive switching of bipolar Ox-RRAM is based on the formation and rupture of conductive filament (CF) consisting of V_O as shown in Fig. 2 [14, 18–20]. The switching characteristic has strong correlation with the distribution of V_O, which will generate and recombine with oxygen ion (O^{2-}) during the resistive switching process. During SET process, the O^{2-} at lattice sites will be randomly thermal activated into the dissociated O^{2-} under a high electrical field. Meanwhile, a V_O leaves at the lattice sites. The dissociated O^{2-} will hop to the top electrode (anode) and be absorbed by the top electrode. The generated V_O can assist the electron hopping from the cathode to anode. So the current will increase and the local electrical field will be enhanced. Meanwhile, the temperature close to the generated V_O sites will increase. It means the generated V_O will accelerate the new generation of V_O at the neighboring sites. Therefore, there is a positive feedback between the V_O generation

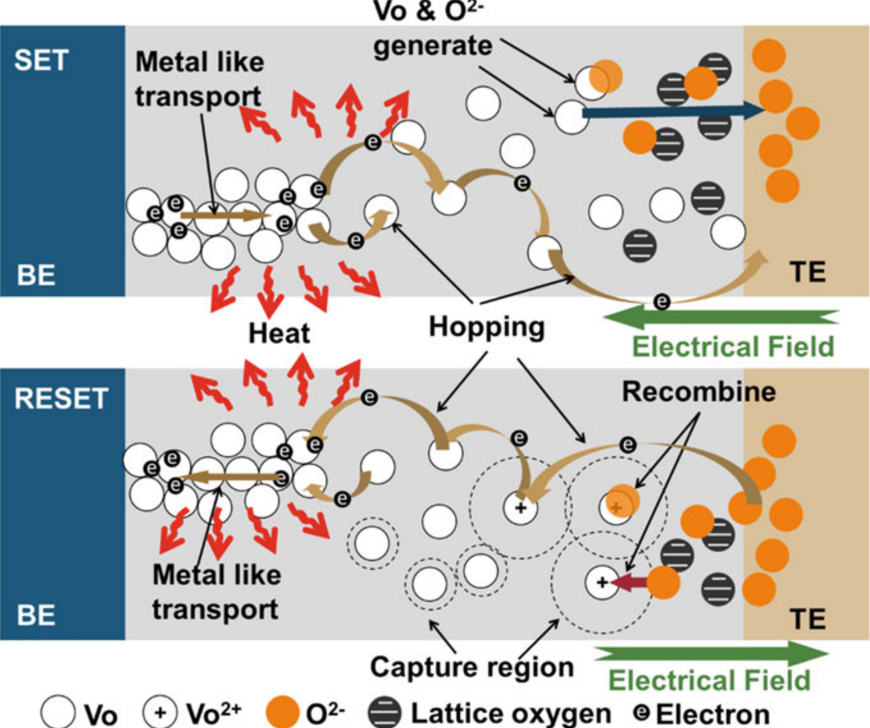


Fig. 2 Schematic physical process of resistive switching in Ox-RRAM

and electrical filed/temperature. Hence the SET process is usually abrupt [21]. As more V_O generated in the insulator layer, a conductive filament consisting of V_O will connect the two electrodes, which results in the cell switching from HRS to LRS. The generation probability (P_g) of V_O under electric field during dt is governed by the following equation [22]:

$$P_g(E, T, \mathrm{d}t) = f\ \mathrm{d}t \exp\left(-\frac{E_a - \alpha_a ZeE}{k_B T}\right) \tag{1}$$

where E is the electric field, T is the local temperature, f is the vibration frequency of oxygen atom, E_a is the average active energy of V_O, α_a is the enhancement factor of the electric field for the lowering of E_a, $Z = 2$ is the charge number of oxygen ion, e is the unit charge and k_B is the Boltzmann constant. The hopping probability (P_h) of the dissociated O^{2-} under the electric field during dt can be described as [23]:

$$P_h(E, T, \mathrm{d}t) = f\ \mathrm{d}t \exp\left(-\frac{E_h - \alpha_h ZeE}{k_B T}\right) \tag{2}$$

E_h is the hopping barrier of O^{2-} and α_h is the enhancement factor of the electric field for the lowering of E_h.

During RESET process, the O^{2-} stored in top electrode will be released and then hop in the oxide layer. The release probability (P_m) of an O^{2-} is governed by the equation [24]:

$$P_m(V, T, \mathrm{d}t) = f\ \mathrm{d}t \exp\left(-\frac{E_i - \gamma ZeV}{k_B T}\right) \tag{3}$$

V is the external voltage, E_i is the energy barrier between the electrode and oxide, γ is the enhancement factor of the external voltage during the O^{2-} release process. The dissociated O^{2-} can recombine with V_O, which can be modeled as an energy relaxation process and governed by the following equation [22]:

$$P_r(T, \mathrm{d}t) = f\ \mathrm{d}t \exp\left(-\frac{\Delta E_r}{k_B T}\right) \tag{4}$$

Pr (T, dt) is the probability of the recombination between V_O and O^{2-} during dt, ΔE_r is the relaxation energy during the recombination process. The recombination between the O^{2-} and V_O will lead to the rupture of conductive filament and a gap region forms between the CF and top electrode. So the current decreases and the device switches from LRS to HRS. As V_O recombined with O^{2-}, the electrical filed in the gap region and temperature will decrease and the hopping of O^{2-} and recombination will become slow. It means that there is a negative feedback between the rupture of conductive filament and the decrease of electrical field and temperature. Therefore, the RESET process is usually gradual and more than hundreds of states can be achieved under proper voltage pulses during the RESET process [21, 25].

The physical mechanism of Forming process is similar to SET process. The intrinsic defect in the fresh samples is usually few. While in the beginning of SET process, there are still some V_O in the resistive switching layer and a remaining defect-rich region near the bottom electrode after RESET process. Hence, the Forming voltage is larger than SET voltage and the conductance after RESET is generally larger than the conductance of fresh state. Often, the remaining defect-rich region is referred to as the virtual electrode [26].

Unipolar

The resistive switching mechanism of unipolar Ox-RRAM is similar to the bipolar type as shown in Fig. 3. The difference is the material storing the dissociated O^{2-}. In the unipolar Ox-RRAM, the dissociated O^{2-} will be absorbed and released from the easily-reduced oxide clusters which are closely neighbored with the conductive filament consisting of V_O. In more details, there are more than one phase of oxide co-exists in the resistive layer [27]. One type of oxide phase is oxygen-rich phase. The oxygen-rich clusters can release O^{2-} due to the thermal decomposition when the local temperature reaches a threshold temperature [27]. The released O^{2-} will recombine with the V_O and rupture the CF. So the device can switches from the LRS to HRS. During SET process, the generation of V_O is the same with bipolar Ox-RRAM. But the dissociated O^{2-} will be absorbed by the around oxygen-rich phase rather than drifting to the top electrode.

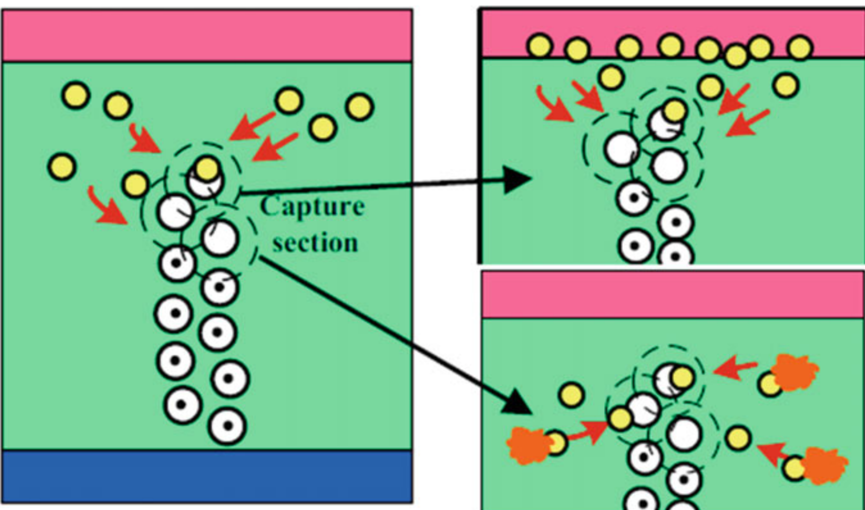


Fig. 3 Comparison between resistive switching mechanism of the unipolar and bipolar Ox-RRAM. For bipolar RESET, the O^{2-} is released by the O^{2-} reservoir assisting by electrical field together with heat. For unipolar RESET, the O^{2-} is released by Joule heating induced thermal decomposition of O^{2-} from the easily-reduced oxide clusters or grains in the capture section region. Reprinted from [14]

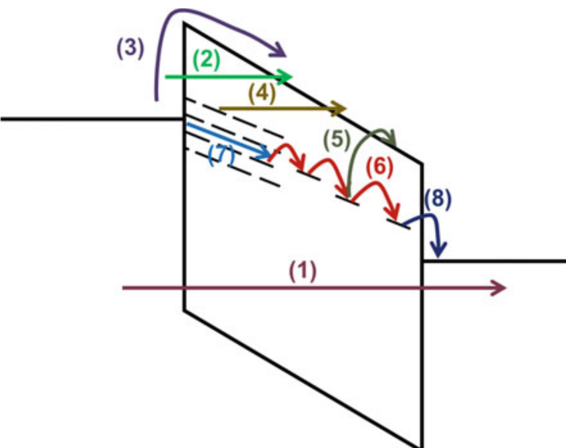


Fig. 4 Possible conduction mechanism in Ox-RRAM. (1) Direct tunnel, (2) FN tunnel, (3) Schottky emission, (4) F-N-like tunneling from trap to conduction band, (5) emission from trap to conduction band, (6) hopping between dispersive traps, (7) semiconductive or metallic-like transport and (8) hopping between the electrode and trap

2.2 Conduction Mechanism

Generally, the resistive switching material is amorphous or polycrystalline. The conduction is correlated with the feature of defects in the resistive switching material. Many different conduction mechanisms, such as semiconductor [28], metallic conduction [22], Poole–Frenkel emission [29], Schottky emission [7], and the space charge limited current (SCLC) [30], have been used to fit the I-V characteristics of current conduction of Ox-RRAM. Figure 4 shows the possible paths for the electron transport from cathode to anode. The resistive switching layer is usually very thin (~10 nm). The electron can transport from cathode to anode by (1) direct tunneling and (2) FN tunneling. The electron in the cathode can be thermally activated over the barrier and injected into the conduction band of resistive switching layer, namely (3) Schottky emission. The electron can also emit from the trap to conduction band (4), which likes the FN tunneling. Besides, the electron in the trap can be thermally activated to the conduction band (5), which is the essential step of Poole–Frenkel emission. The electron transport in resistive switching layer is mainly through the V_O. If the distance between two oxygen vacancies is small, they will be extended states and the conduction mechanism is metallic-like or semiconductive (7). The metallic-like or semiconductive conductivity decreases or increase with temperature (T), according to the Arrhenius law [31]:

$$\sigma = \sigma_0 \exp(E_{AC}/k_B T) \tag{5}$$

where E_{AC} is the activation energy for conduction and σ_0 is the Arrhenius pre-exponential factor for conductivity. The electron transport among the dispersive V_O is hopping (6) and the hopping rate between two vacancies can be calculated by the Mott model [32]:

$$W_{m\to n} = f_{ph}\exp(-2\alpha R_{mn} - E_{mn}/k_B T) \quad (6)$$

where the f_{ph} is the vibration frequency of electron, R_{mn} is the distance between two V_O, α is the attenuation length of the electron wave function, E_{mn} is the change of barrier height induced by the applied external field. Moreover, the electron can hop between the electrode and traps (8).

Although there are many current paths from the cathode to anode, only one or two current mechanism is dominant in a specific Ox-RRAM device. The dominant mechanism is different depending on the properties of resistive switching layer and the interface. Therefore, we can tune the conductance of Ox-RRAM via modulating the properties of resistive switching layer, such as the amount and distribution of V_O. If the concentration of the V_O is high, the contact of electrode and resistive switching layer is Ohmic and there is no barrier. Therefore, symmetrical I-V cure can be observed in LRS. If the concentration of the V_O is low, there is a barrier between the resistive switching layer and the electrode. The height of the barrier for electrode transport between the electrode and traps depends on the energy difference between them. If the energy difference is low, such as TiN/HfO_x/Pt-RRAM, the I-V cure is still symmetrical [33]. The energy difference between the electrode and traps in Ta/TaO_x/TiN RRAM is relatively high, so the electrical characteristics will be asymmetrical [34]. The asymmetrical is beneficial to mitigate the leakage current and increase the read margin of RRAM array.

3 Materials Characterization

Due to the complexity of working principle of Ox-RRAM device, some physical characterization methods are needed to visually display the resistive process. Compared with electrical characterization technology, physical characterization technology is more conducive to analyze material composition and guide the optimization of device structure. Common physical property characterization methods are summarized in Fig. 5.

A. *Spectral Analysis*

Since each atom has its own characteristic line, it could be used to identify the substance and chemical composition based on the spectrum. The common method includes X-ray photoelectron spectroscopy (XPS) and Raman spectroscopy.

Raman spectroscopy is based on Raman effect, and it involves phonon absorption (Stokes process) and phonon emission (anti-Stokes process) [35]. Raman spectroscopy is widely used in structure determination and in situ reaction monitoring in many studies (Fig. 6).

He Tian et al. employed Raman spectroscopy to probe the changes inside the RRAM during the cycling and monitor the oxygen movement at the electrode/oxide interface by inserting single-layer graphene (SLG) [36].

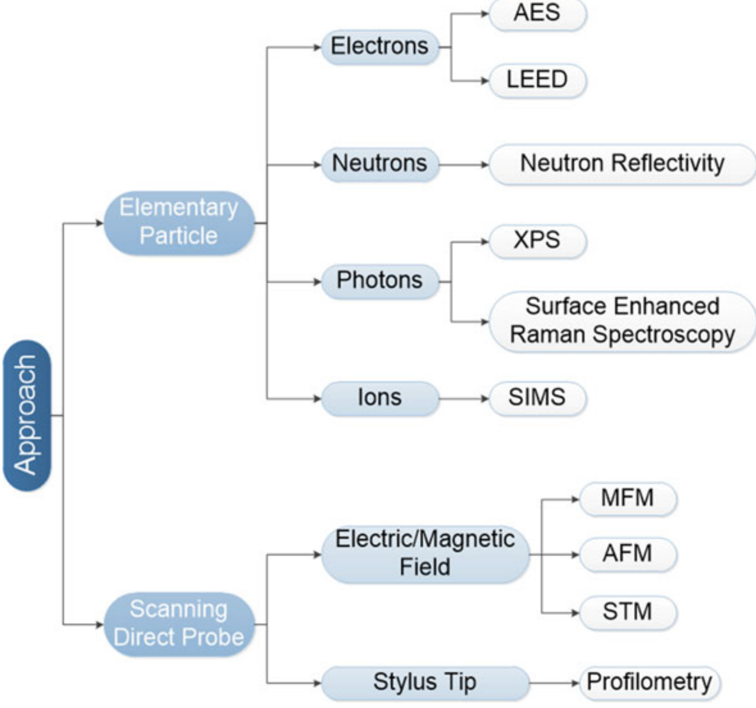


Fig. 5 Summary of the common physical characterization methods

The XPS spectrum could be obtained by measuring the kinetic energy and the amount of electrons escaping from the surface of the material irradiated with the X-ray beam [35]. And the XPS spectrum is used in elemental composition characterization and the electronic state identification of the elements on the surface. Joshua Yang et al. revealed that the Ag was metallic by XPS, which was further confirmed by HRTEM micrographs showing Ag nanocrystals in SiO_xN_y:Ag [37]. P. Calka et al. found that the HfO_2/TiN interface is free from TiO_xN_y oxide-nitride states by XPS [38].

B. ***Electron Microscopy Analysis***

Electron microscopy analysis is the most commonly used method for studying the particle size, chemical composition, relative content and surface morphology of nanomaterials, and the electron microscopy generally includes scanning electron microscopy (SEM), transmission electron microscopy (TEM), and high-resolution transmission electron microscopy (HRTEM) etc.

TEM is primarily employed to visualize the particle size, shape, and position to provide the information to address the performance. For example, Joshua Yang et al. observed that Ag atoms could separate under electrical bias and

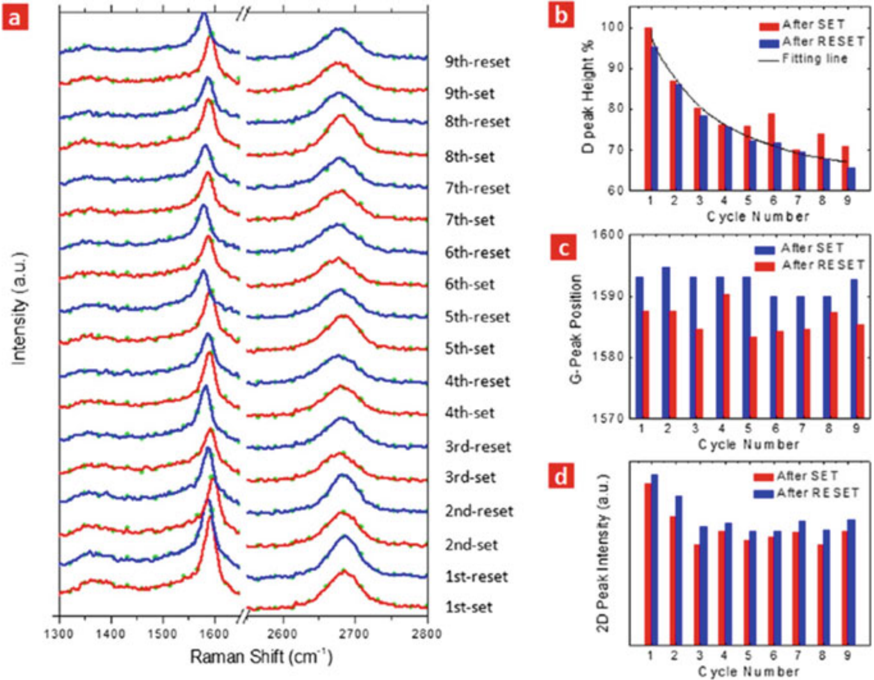


Fig. 6 Single-point Raman measurements of SLG during consecutive electrical programming repeated for nine cycles. **a** Single-point Raman data measured after SET and RESET cycles at the same location, which is 1 μm from the cross-point. The dash lines mark the *D*, *G*, and 2D peak, which will be analyzed in (**b**−**d**). **b** Statistical analysis of *D* peak drop as a function of switching cycle. It is observed that the *D* peak gradually decreases. **c** The position of *G* peak as a function of switching cycle. The *G*-peak position of the SET is always higher than the RESET in each cycle, and reversible shifts of *G* peak position with RRAM programming actions are observed. **d** The intensity of 2D peak as a function of switching cycle. The 2D-peak intensity of the SET is always lower than the RESET in each cycle, and reversible changes of 2D peak intensity with RRAM programming events are observed. Reprinted from [36]

regroup spontaneously under zero bias because of interfacial energy minimization in Au/SiO_xN_y:Ag/Au device [37]. The dynamic process of Ag dispersing and agglomerating is shown in Fig. 7 by in situ TEM.

C. ***Scanning Probe Analysis***

Comparing with the electron microscopy, the novelty of the probe technology is a mechanical change of the piezoelectric device to precisely control the positioning of the probe on individual atoms and scan the atoms one by one, so that it provides the atomic resolved surface analysis based on the principle of van der Waals force [38]. Due to its convenience and reliability, atomic force microscopy (AFM) is regarded as the leading approach in scanning probe analysis method for thickness measurement and the layer number counting. Figure 8a shows the typical CAFM experimental equipment at standard air ambient condition, and the CAFM tip contacts with the

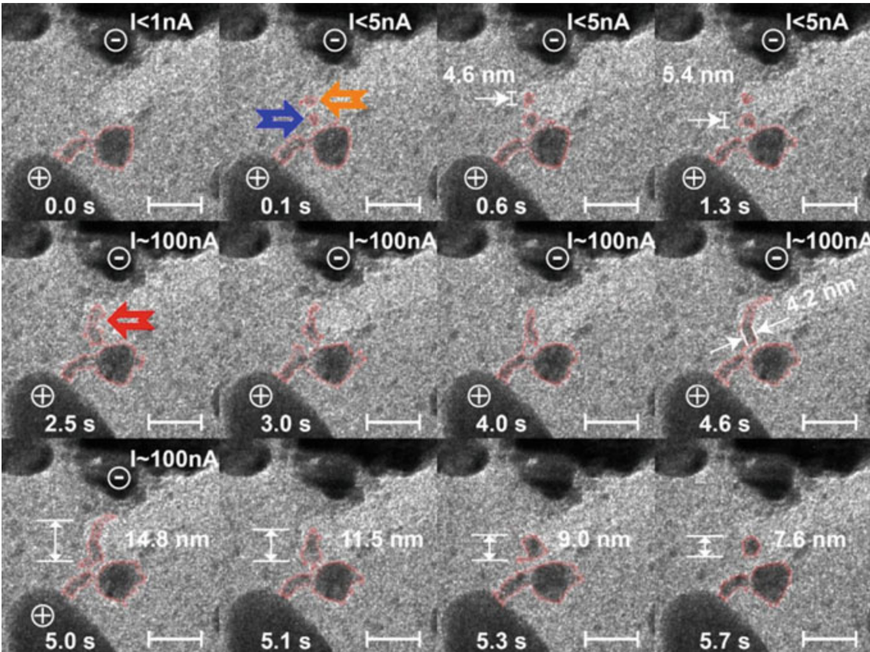


Fig. 7 In situ TEM observation of the threshold switching process of Ag atoms with and without electrical bias. When the external electric field is exerted at the point of time zero, Ag migration is observed at time 0.1 s when two nanocrystals started to form. A clear arc-shaped filament is visible at 2.5 s. When the external biasing is removed at 5.0 s, the filament starts to deform, shrinking to a round spherical nanocluster implying an interfacial energy driven diffusion mechanism. Reprinted from [37]

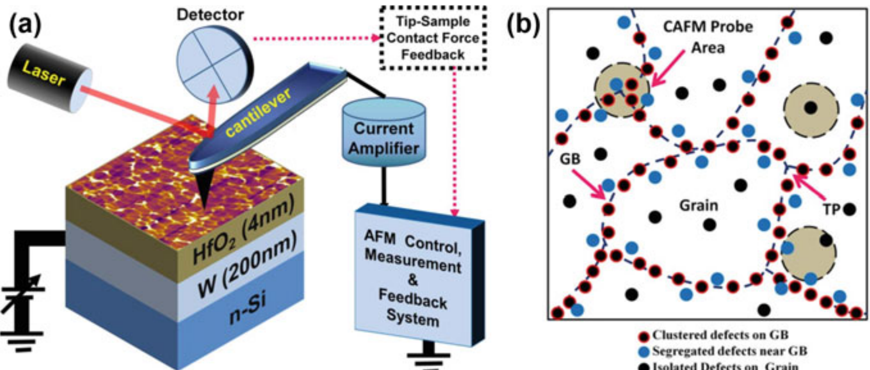


Fig. 8 **a** Schematic of the standard CAFM experimental setup for HfO_2 (~4 nm)/W (200 nm) /n-Si stack and the electrical connections for the measurement. **b** Schematic showing the three different microstructural regions that the CAFM tip might be probing with very different defect density and spatial vacancy distribution patterns. Reprinted from [39]

sample surface in physical contact. A. Ranjan et al. investigate the RTN performance on ultra-thin HfO_2 dielectric films using conductive AFM (CAFM) at a spatial resolution of 10–20 nm. Figure 8b illustrates the spatial inhomogeneity of the defect density distribution, and it may need different measurement patterns of CAFM. Based on the precision detection of CAFM, the metastable V_O defect state and the clustering model at nanoscale level are verified by the experiment [39].

4 Monte Carlo Dynamic Simulation of Resistive Switching Behavior

The Monte Carlo technique is an ideal tool to study the atomic-scale evolution of a system. For RRAM, the Monte Carlo simulation method can be used to self-consistently simulate the microscopic process of the generation and recombination of vacancies and the transport of interstitial oxygen ions under the external electric field and influenced by the local temperature [40]. This benefits to deeply understand the underlying physics of resistive switching behaviors and processes. The simulation algorithm flow is shown in Fig. 9. Take 2D simulation as an example, the simulation is performed on an atomic matrix with a size of m*n, where m and n depend on the fabricated device structure and size. Usually as the inputs of the simulation, the intrinsic defects are introduced and randomly generated when initializing simulation. After that, the potential and current distributions in the resistive switching layer are solved by the Poisson's equations or resistor network, and then the local temperature distribution is calculated. Following these, the probabilities of the physical effects during resistive switching are calculated respectively. By precisely assigning different activation energies (E_a) to particles according to their local structure, the probabilities of the physical processes can be calculated by Eqs. 1–4. The values of these activation energies can be obtained through ab-initio calculations [41] and can also be determined as experimental parameters [42]. After obtaining the probabilities, which physical process occurs can be decided by using Monte Carlo method by comparing the probabilities with the random numbers uniformly distributed between [0 1]. The time and particle distributions have to be updated in every time step. The field and temperature distributions are updated regularly, and are in turn used to update the vacancy/ion configurations. For each time step, all the main physical processes are reevaluated to decide the next move. The above calculations will be repeated as the sweeping voltage changes.

5 Simulation Method

There are two popular methods to calculate the electric field, solving the Poisson's equation and using the resistor network. By applying the boundary condition, the

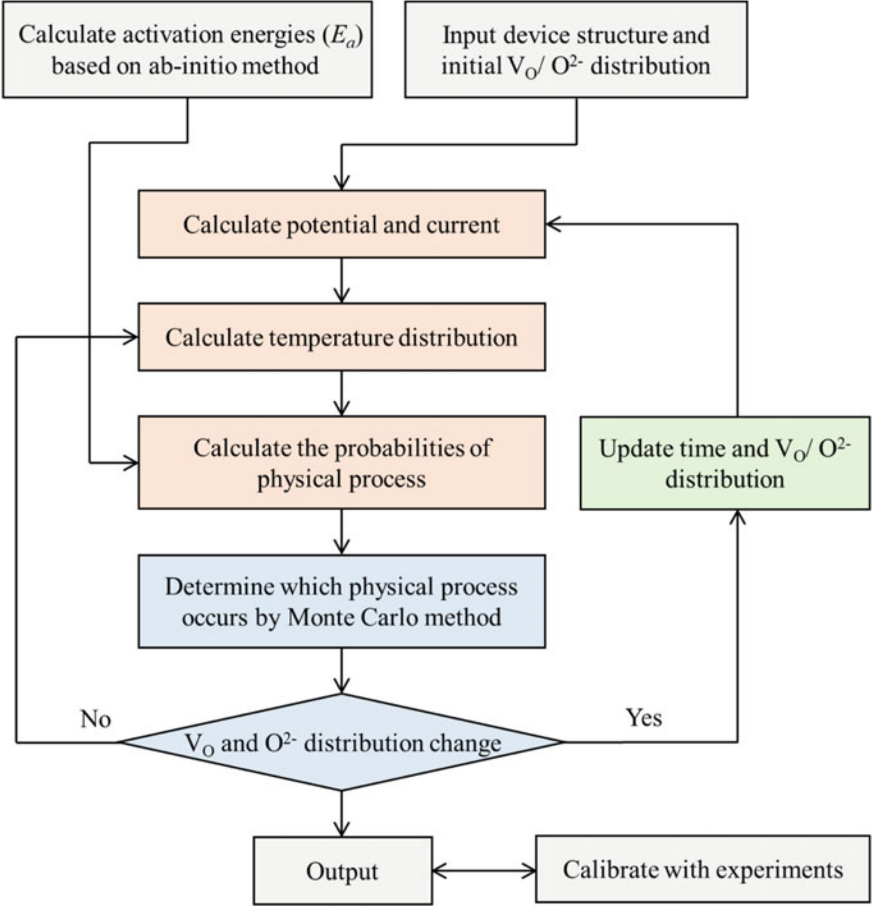


Fig. 9 A typical Monte Carlo simulation algorithm flow for Ox-RRAM

potential can be calculated by solving Poisson's equation as:

$$\nabla^2\phi = -\frac{\rho}{\varepsilon} \tag{7}$$

where ϕ is the potential, ρ is charge density within the oxide material.

The resistor network is based on the percolation theory [32, 43]. An example of a resistor network to calculate the conductance of RRAM is shown in Fig. 10 [44]. The resistive switching layer is divided into the resistor network with cross nodes. Different particles with different resistivity are located in the network following the Monte Carlo processes.

The resistance of the node representing the insulating or semiconductive oxide is nonlinear and can be described as [20]:

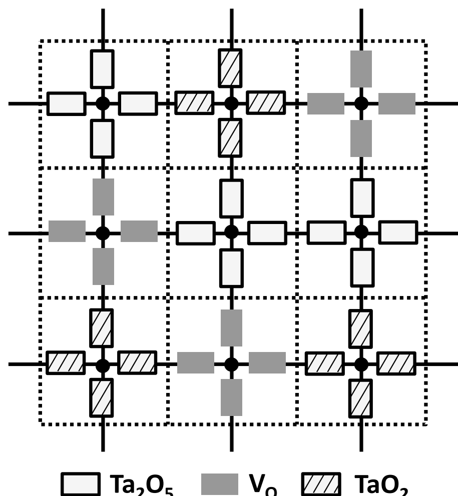


Fig. 10 Example of a resistor network in the TaO_X-RRAM simulator. Reprinted from [44]

$$I = \sinh(\alpha U)/R_O \tag{8}$$

where α is a fitting coefficient and R_O represents the resistance of the oxide in the network. The I-V characteristics of V_O are linear and can be described as:

$$I = U/R_V \tag{9}$$

where R_V represents the resistance value of V_O. Using the resistor network, the potential and current of each node can be calculated by the Kirchhoff law.

The local heat generation is determined by the dot product of the field and current density vectors $\boldsymbol{J}\cdot\boldsymbol{E}$, assuming Joule heating is the dominant mechanism for dissipation. Temperature rise due to self-heating is calculated by the resolution of the time-dependent heat diffusion equation [45]:

$$\nabla[\kappa(\mathbf{r}, T)\nabla T(\mathbf{r}, t)] + g(\mathbf{r}, t) = \rho C \frac{\partial T(\mathbf{r}, t)}{\partial t} \tag{10}$$

where $T(\mathbf{r}, T)$ and $g(\mathbf{r}, t) = \boldsymbol{J} \cdot \boldsymbol{E}$ are the temperature and heat generation, respectively, at a given position $\mathbf{r}$ and time t, $\kappa(\mathbf{r}, T)$ is the temperature-dependent thermal conductivity, ρ is the material density and C is the specific heat capacity.

Simulation of Particle Dynamics and Filament Growth

In thermal equilibrium, vacancies generation is modeled as a random process with the probabilities as Eq. 1, where the lattice oxygen ion barrier E_a is modulated by the local electric field under an external bias. Only when O^{2-} is located at the neighbor of V_O, the recombination process occurs. Ion transportation is governed by the following equation [45]:

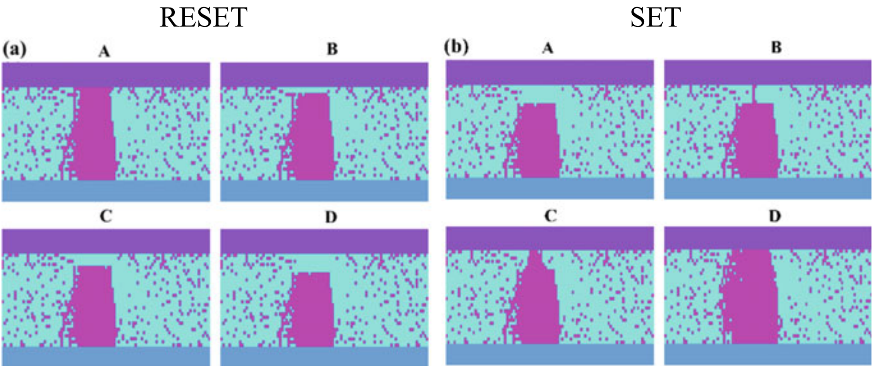


Fig. 11 Simulated CF geometry evolution behaviors during **a** RESET and **b** SET processes in HfO_X-based RRAM. Reprinted from [40]

$$\frac{\partial \mathrm{n}(\mathbf{r},t)}{\partial t} = \nabla[D\nabla n(\mathbf{r},t) - V_{\mathrm{ion}} n(\mathbf{r},t)] + G \tag{11}$$

where n is the O^{2-} concentration, D is the diffusivity, V_{ion} is the ion velocity, and G is the net ion generation, which is equal to generation of V_O. and

$$V_{\mathrm{ion}} = af\exp\left(-\frac{E_m}{k_B T}\right)\sinh\left(\frac{eaE}{2k_B T}\right) \tag{12}$$

Accurate field and temperature distributions are also crucial for evaluating the hopping rates. For the sake of simplicity, only the movement to the nearest neighbor positions is considered.

Simulated Results

The simulated CF geometry evolution during RESET and SET processes in HfO_X-based RRAM is shown in Fig. 11. Corresponding to the same resistance level, the evolutions of filament geometry during SET are quite different from RESET. For RESET, the whole filament firstly disconnects at the top electrode and then extends to the interior as the sweeping voltage increases. For SET, a fine filament forms firstly in the rupture region and then gradually extends along the radius direction as current increases [36]. The corresponding simulated I-V characteristic is shown in Fig. 12. It can be found that the gradual RESET and abrupt SET are reproduced by the proposed model.

3D simulation can also be performed to reproduce the resistive switching characteristics of RRAM. As an example, a 3D volume of SiO_X-RRAM has been simulated, by investigating an oxide thickness of 15 nm and a simulated contact area of 30 nm long and 30 nm wide [45]. Figure 13 shows the vacancy generation process. The initial structure includes defect/Si-rich areas in the form of a pillar as shown in Fig. 13a. At low biases, few vacancies are created. At round 5 V, filament seeds appear and

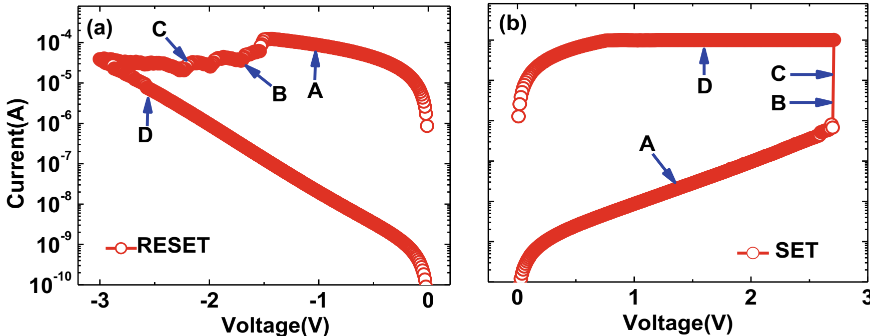


Fig. 12 Simulated I-V characteristics in the RESET (**a**) and SET (**b**) process under the DC sweep mode, where A-D correspond to Fig. 11. Reprinted from [40]

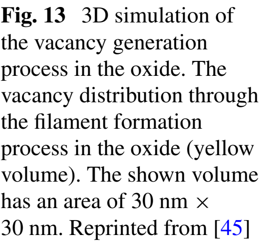
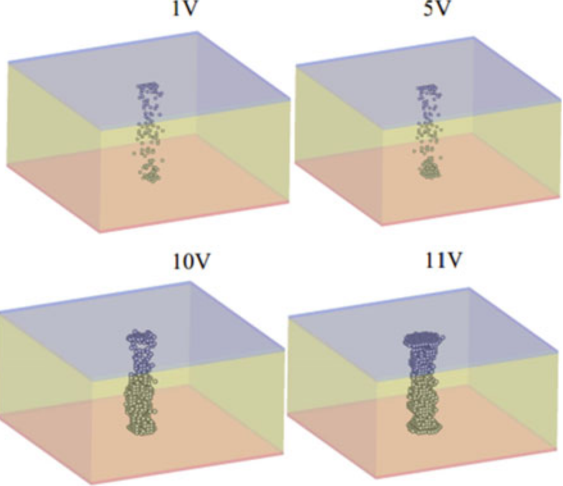


Fig. 13 3D simulation of the vacancy generation process in the oxide. The vacancy distribution through the filament formation process in the oxide (yellow volume). The shown volume has an area of 30 nm × 30 nm. Reprinted from [45]



grow significantly as bias is increased to 10 V. At around 11 V, a self-accelerated generation of oxygen vacancies occurs, which lead to the forming a CF and bridging the electrodes. The CF formation will results in the abrupt current increase, which indicates the transition from the HRS to the LRS.

Figures 14 and 15 show the electric field and lattice temperature distributions in 3D simulation, respectively. Temperature rise due to self-heating is significant when the CF is formed as a significant current starts to flow. The increased temperature due to the self-heating will increase the oxygen ion/vacancy generation probability and diffusion. The temperature maps are much less strongly localized than the electric field maps. Indeed, while percolation paths are characterized by very high power densities, temperature can still be high outside these paths due to heat diffusion.

The Monte Carlo simulation method can also be used to investigate the retention by recording the distribution of V_O and O^{2-} until it exceeds a pre-determined criterion.

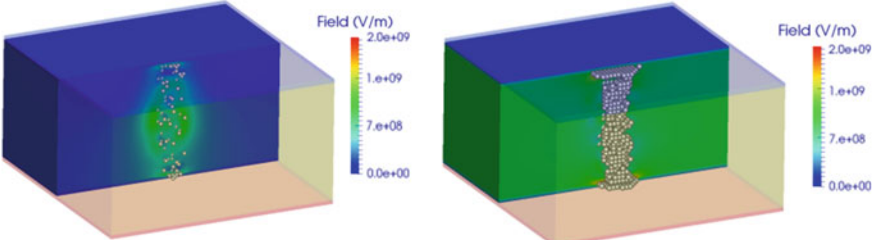


Fig. 14 Typical electric field distributions for (left) HRS and (right) LRS. Reprinted from [45]

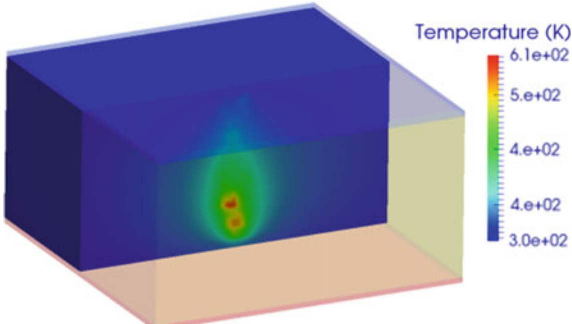


Fig. 15 Typical temperature distribution during the CF formation. Reprinted from [45]

Figure 16 plots the LRS retention behavior of TiN/Hf/HfO_X/TiN RRAM [46] and the corresponding CF geometry. The results are compared with the experimental data with well agreement [47]. The transportation of V_O outside the CF results in

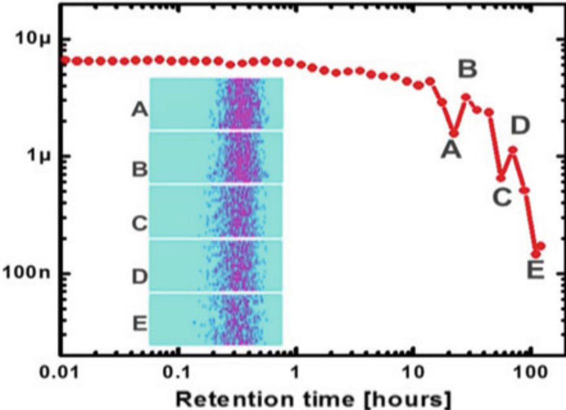


Fig. 16 Simulated LRS retention behavior of TiN/Hf/HfO_X/TiN. Reprinted from [46]

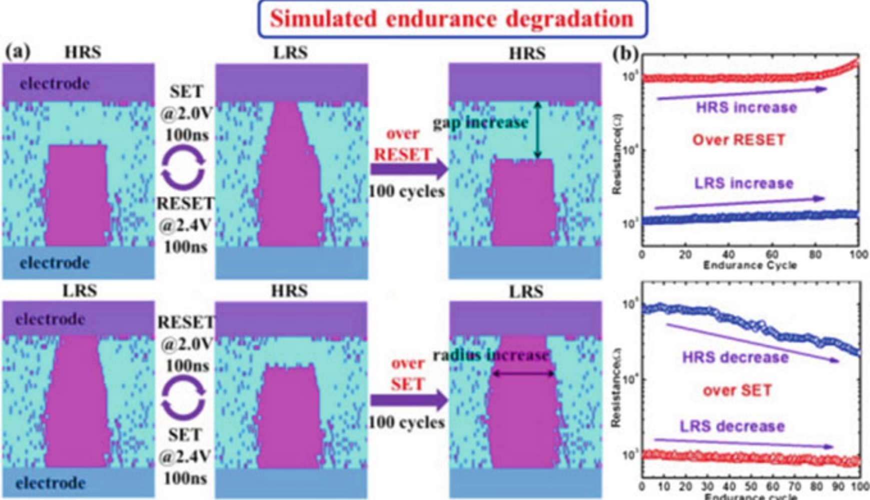


Fig. 17 **a** Simulated CF and **b** the corresponding resistance varies with switching cycle under different operation scheme. It demonstrates that the operation scheme will impact on the degradation type. Reprinted from [48]

the dissolution of CF. Hence, the V_O concentration of CF gradually decrease and the conduction transfer from the metallic to the hopping. Moreover, the O^{2-} release by the electrode and sidewall oxide will recombine the V_O, which will also lead to the decrease of V_O amount as well as the rupture of CF. Therefore, the current decrease as time went on.

The endurance degradation can investigated by the developed Monte Carlo simulator by repeatedly simulate the resistive switching like the endurance test. The CF geometry changing with the switching cycle is simulated and shown in Fig. 17 [48]. The experimental over SET and over RESET endurance degradation behavior are reproduced by the simulator. It can be found that the gap distance increases with cycle for the "over RESET" degradation and radius of CF increase with cycle for the "over SET" degradation. Generally, the simulation of endurance degradation is time consuming due to the fact that so many switching cycles are calculated.

6 Compact Model

Compact model is very important for the development of new electronic devices based on new semiconductor technologies. An accurate and comprehensive compact model acts as an interface between the device physics and system/sub-system design. Beside phenomenological considerations, compatibility with standardized software suites employed in the IC industry is also an essential feature for compact model. For

instance, even if a universal compact model covered all the features of Ox-RRAM is still not available, many models have been proposed with various degrees of accuracy, different features, and mixed results.

The first model of RRAM is the memristor model proposed by Chua [49]. Strukov et al. presented a physical model of a two-terminal electrical device that behaves like a perfect memristor for a certain restricted range of the state variable w [50]. The model provides a simplified explanation for the reports of current–voltage anomalies, including switching and hysteretic conductance, multiple conductance states and apparent negative differential resistance. Guan et al. presented a compact model by considering the generation/migration of V_O/oxygen ion [51]. The proposed model is implemented in Ngspice as a macrocircuit. P. Huang et al. proposed a physical electro-thermal model by modelling the length and width of filament [25, 40]. The model can be implemented into HSPICE for the simulation of large-scale RRAM based circuit by Verilog-A [52]. By modelling the local electrochemical reduction and redox, a self-consistent physical model for SET/RESET operations is built for unipolar RRAM by Bocquet et al. [53]. By gathering local electrochemical reactions and heat equation in a single master equation, M. Bocquet et al. proposed a robust compact model for simultaneously describing Forming, SET, and RESET operations of bipolar RRAM [54]. A self-accelerated thermal dissolution model for RESET operation of unipolar RRAM is proposed by Russo et al. via accurately modelling the joule heat and temperature [55, 56]. Larentis et al. presented a numerical RESET model based on temperature/field-driven ion migration in bipolar RRAM devices [28]. A brief summary of the presented models is listed in Table 1. Three typical compact models of Ox-RRAM will be discussed in detail in this section.

7 Linear Ion Drift

In 2008, RRAM devices were linked with memristor and a couple variable-resistor model is proposed to understand the hysteretic I-V relation by HP Labs [50]. RRAM device is modelled as two films sandwiched between two electrodes as shown in Fig. 18a. The total resistance of the device is determined by two variable resistor connected in series. The doped film with thickness D has a low resistance R_{on}. And the resistance of the other undoped film is relatively high, called as R_{off}. The boundary of this two films will move caused by the drift of dopants when an external bias $v(t)$ is applied. The relation between the external applied voltage $v(t)$ and the current through the device $i(t)$ owing to Ohmic electronic conductance and linear ionic drift in a uniform field with average ion mobility is given by:

$$v(t) = \left(R_{\text{on}}\frac{w(t)}{D} + R_{\text{off}}\left(1 - \frac{w(t)}{D}\right)\right)i(t) \tag{13}$$

The state variable of above equation is the $w(t)$ and can be expressed as:

Table 1 Comparative analysis of the compact models

	Linear ion drift (Hp)	Stanford model	Physical electro-thermal model (PKU)	Marc-Bocquet model	Filament dissolution model
Device type	Bipolar	Bipolar	Bipolar	Uni/bipolar	Uni/bipolar
Mechanism	Diffusion/drift of V_o	Vacancy generation, oxygen migration	Generation and recombination of V_o	Redox and thermally assisted diffusion/electrochemical redox reactions	Joule-heating assisted diffusion/ion migration
Methodology	Gap length	Gap length	Gap length/widths	Gap length and width	Gap width/length
Conduction	Ohmic	Hopping, ohmic	Hopping, ohmic	Ohmic	Semi-conductive, ohmic
Transient characteristic	Yes	Yes	Yes	Not show	Yes
Compatible	Possible with Spice	Verilog-A Spice	Verilog- A Spice	Console/Spice	Console

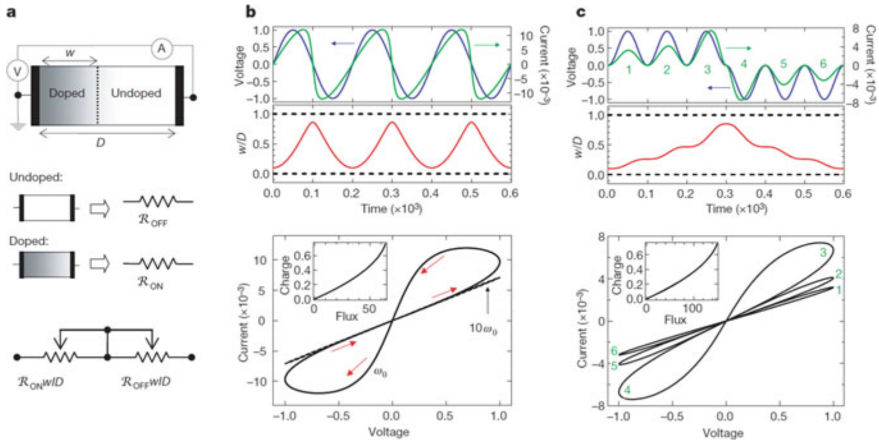


Fig. 18 The coupled variable-resistor model for a memristor. **a** Diagram with a simplified equivalent circuit. V, voltmeter; A, ammeter. **b**, **c**, The applied voltage (blue) and resulting current (green) as a function of time t for a typical memristor. Reprinted from [50]

$$\frac{dw(t)}{dt} = \mu_v \frac{R_{on}}{D} i(t) \tag{14}$$

where μ_v is the average ion mobility. w(t) yields the following formula:

$$w(t) = \mu_v \frac{R_{on}}{D} q(t) \tag{15}$$

where $q(t)$ is the charge. By inserting Eqs. 15 into 13, the so-called memristance of the device can be obtained, which for $R_{\text{off}} \gg R_{\text{on}}$ simplifies to:

$$M(q) = R_{\text{off}} \left(1 - \mu_v \frac{R_{\text{on}}}{D^2} q(t) \right) \tag{16}$$

$M(q)$ is the total resistance of the film. Hence the resistance of the device is proportional to the charge q that passes through the device. Based on above four equations, the double-loop i-v hysteresis can be reproduced by the model as shown in Fig. 18b. In Fig. 18b, the applied voltage is $v_0\sin(w_0 t)$ and the resistance ratio is $R_{\text{off}}/R_{\text{on}} = 160$. And multiple continuous states will also be obtained if there is any sort of asymmetry in the applied bias as shown in Fig. 18c. In Fig. 18c, the applied voltage is $\pm v_0\sin^2(w_0 t)$ and $R_{\text{off}}/R_{\text{on}} = 380$. Equations 13 and 14 can describe the analog bipolar switching and voltages of opposite polarity can switch the device between the LRS and HRS as shown Fig. 18b and c. This model could be attributed to laying the foundations for future RRAM models. Taking this model as the reference, numerous models have been proposed for RRAM devices, such as the non-linear ion drift model [57] and the exponential ion drift model [58].

8 Physical Electro-Thermal Model

In 2012, Huang et al. proposed a comprehensive physical model for bipolar RRAM. Its major feature is linking the resistive switching behavior with the evolution of CF geometry, which depends on multitude of factors. Based on the Monte Carlo simulation of the CF evolution (Fig. 11), 3-D CF evolution process can be modelled as shown in Fig. 19. Both the lateral and vertical growth process are considered in the model. The rupture process is that the whole filament disconnects firstly at the TE then extends towards the interior step by step with increasing voltage. So the gap distance x and the dx/dt are key factors to describe the RESET operation. In CF growing process, the firstly formation of a fine filament in the rupture region connects the tip of the CF and TE then gradually enlarges along the radius direction as the current increase in SET operation. Therefore, x, dx/dt, the width of the newly growth CF w, and dw/dt can be used to characterize the SET process.

The reduction of x in RESET process is determined by the three physical processes: (1) electrode release O^{2-}; (2) O^{2-} hopping in the oxide layer; and (3) recombination between O^{2-} and V_O. The reduction of x is determined by the slowest process among those three processes. If the hopping of O^{2-} is the slowest, dx/dt can be deduced as:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = af \exp\left(-\frac{E_h}{k_B T}\right) \sinh\left(\frac{\alpha_h Z e E}{k_B T}\right) \quad (17)$$

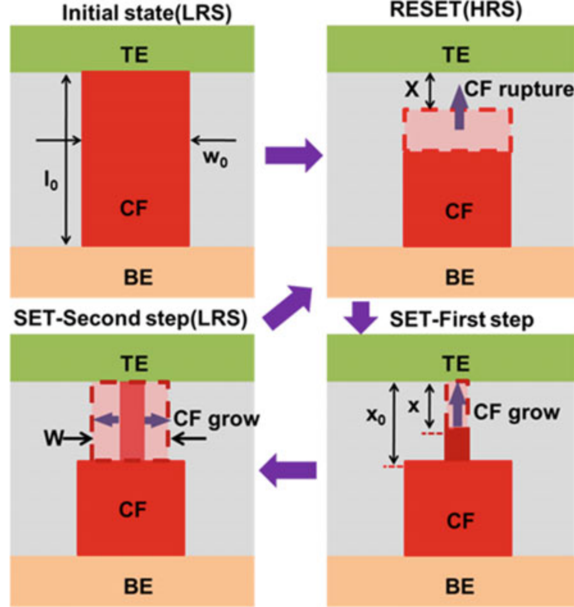


Fig. 19 The schematic of CF evolution. The SET process is divided into two steps. 1st step: CF growth from the rupture CF tip to the electrode. 2nd step: CF extending along the radius direction of formed CF. Reprinted from [22]

Likewise, if the process of O^{2-} released by the electrode is dominant, dx/dt can be written as:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = af\exp\left(-\frac{E_i - \gamma ZeV}{k_B T}\right) \tag{18}$$

and when the process of recombination between O^{2-} and V_O is the slowest, dx/dt is written as:

$$\mathrm{d}x/\mathrm{d}t = af\exp(-\Delta E_r/k_B T) \tag{19}$$

The growth of the CF results from the generation of V_O under external bias. Considering the V_O generation in the gap region, dx can be expressed as the following equation:

$$\mathrm{d}x = -af\exp\left(-\frac{E_a - \alpha_a ZeE}{k_B T}\right)\mathrm{d}t \tag{20}$$

The CF evolution during the switching process can be described based on Eqs. 17, 18, 19 and 20. The temperature plays a key role during the switching process. For the purpose of simplification, uniform temperature is adopted and the description of temperature is:

$$T = T_0 + IVR_{th} \tag{21}$$

where T_0 is the ambient temperature, R_{th} is the effective thermal resistance. The transient effect of thermal capacitance in Eq. 21 is neglected to simply the model. It means the temperature variation is assumed to be much faster than the voltage variation imposed by the setup; in other words, the model is based on a quasi-static approach, where the thermal equilibrium is assumed to be reached once the operation voltage is applied to the cell.

The conduction of CF region is modeled as metallic-like or semiconductive, which depends on the V_O concentration in the CF. The conduction of gap region is hopping conduction. Generally, the hopping current can be calculated with distribution of V_O in real space and energy space based on Eq. 16. Here the hopping current is correlated with the voltage and gap distance by:

$$I = I_0\exp(-x/x_T)\sinh(V/V_T) \tag{22}$$

where I_0 is 10 μA/nm^2, x_T and V_T are the characteristic length and voltage, respectively. Therefore the conduction of HRS and LRS can be modelled as the series – parallel of metallic-like or the semiconductive and hopping as shown in Fig. 20. When the width of the newly grown filament is small, the conduction of LRS is small and the nonlinear I-V characteristic can be reproduced as shown in Fig. 20b. The

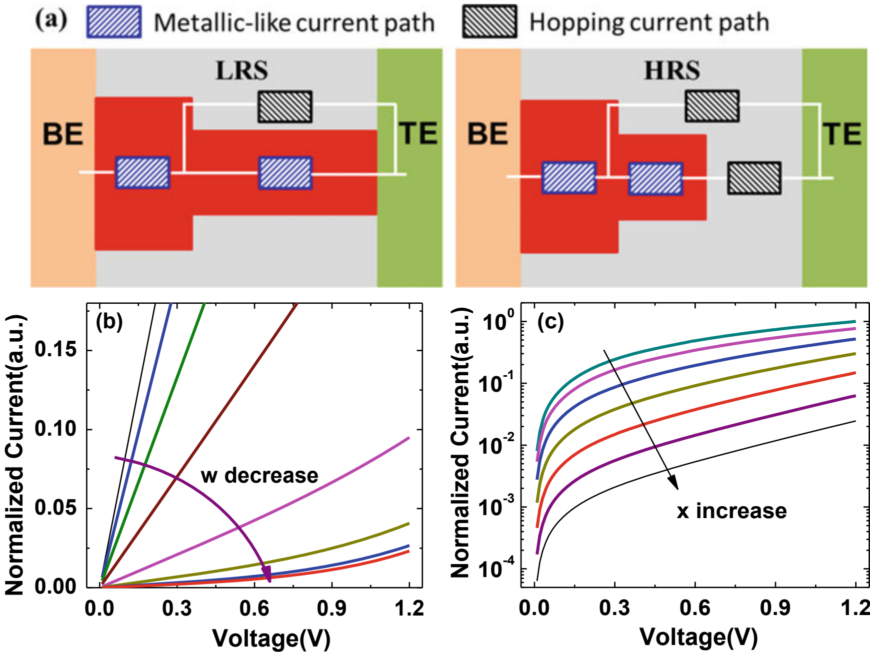
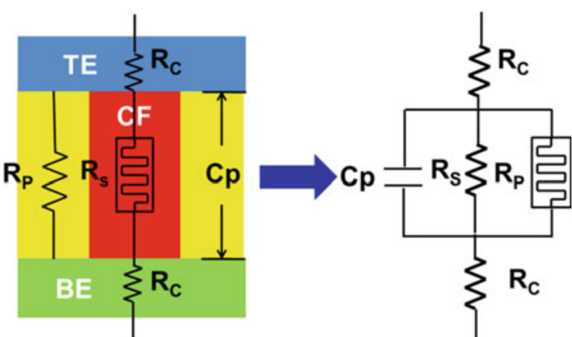


Fig. 20 **a** Equivalent circuit of conduction in LRS and HRS, **b**, **c** LRS/HRS I-V curves reproduced by the conduction model. The non-linear characteristic in LRS is reproduced by the developed conduction model with small LRS (w). Reprinted from [22]

current at HRS exponentially increases with the voltage and exponentially decreases with the gap distance x as shown in Fig. 20c.

The equivalent circuit of a RRAM cell is shown in Fig. 21, which includes a large parallel resistance (R_P), two contact resistances (R_C), a larger parallel capacitance (C_P) and the resistive switching element. *Rp* represents the leakage current between the two electrodes, such as direct tunnel, FN tunnel, trap assisted hopping and so on. *Rc* is the resistance of the electrode and the contact resistance between the CF and

Fig. 21 Equivalent circuit of a RRAM cell with parasitic elements. The parasitic resistance and capacitance are considered in this model. Reprinted from [22]



electrodes. C_P represents the intrinsic capacitance of the MIM structure of RRAM cell.

Figure 22 shows the comparison between the calculated and measured electrical characteristics of RRAM cell. It can be found that the compact model can reproduce the gradual RESET and the abrupt SET processes under DC sweep, which are consistent with the measured data. During the RESET process, there is a negative feedback between the speed of CF rupture and the decrease of electrical field and temperature caused by the current drop and increase of CF gap. Hence, the RESET process is gradual and multiple intermediate states can be achieved by the RESET operation [13] as shown in Fig. 22a. Since different RESET voltages will lead to different gap distance x of CF, multi-level HRS can be achieved with different stop voltages during the RESET process as shown in Fig. 22a. Because of the positive feedback between the CF growth and electrical field/temperature [25], the SET process is abrupt. So, only binary states can be achieved during the SET process as shown in Fig. 22a. The evolution rate of CF depends on the electric field in the gap region during the SET

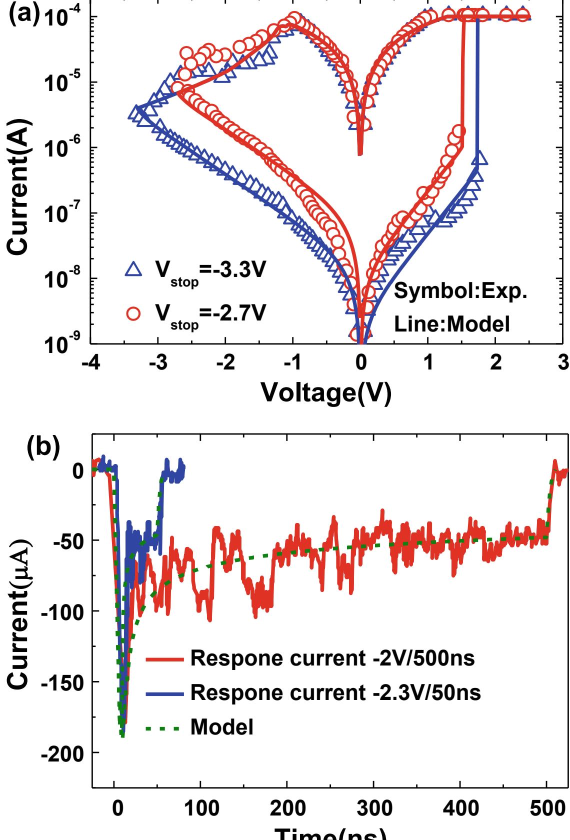


Fig. 22 **a** The modeled DC I-V curves during the RESET and SET operation, together with the measured data. **b** Calculated and measured transient response current waveform for two different RESET programming schemes. Reprinted from [22]

process, so the SET voltage will increase with the initial gap distance × 0 and HRS as shown in Fig. 22a. Two different operation schemes are used to switch the device from ~10 to ~50 kΩ as shown in Fig. 22b. It implies that the switching speed will exponentially increase with operation voltage during the RESET process. By tuning parts of the parameters used in the model, the electrical characteristics of different RRAM device with different material and structures can be reproduced [22]. Excellent agreement between the modelled and measured data in different RRAM devices indicates the validity and universality of the developed compact model to describe the main features of the bipolar RRAM cell operations.

The model can be implemented into the HSPICE for the co-design and optimization of the RRAM based circuit. A memory array with 1T1R configuration is simulated using the presented model for RRAM cell and the BSIM3v3 [59] model for the NMOS as shown in Fig. 23. The simulation is based on the worst case scenario: the selected cell is located at the farthest corner and the other cells are in LRS [60]. The simulated results indicate that the maximum current (I_{max}) of the selected cell decreases with array size due to the degradation of gate-source voltage (V_{GS}) of the selected transistor, as shown in Fig. 23b. Accordingly, the parasitic RC delay increases dramatically with array size during the RESET process resulting from the increase of the interconnect capacitance, as shown in Fig. 23c. The RC delay in

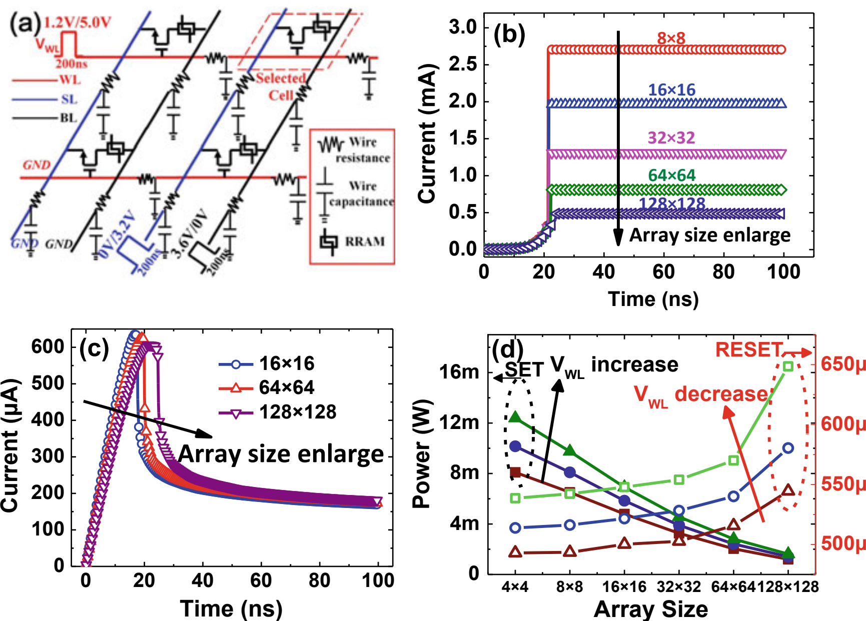


Fig. 23 Simulated the 1T1R RRAM array with the presented compact model. **a** Diagram of 1 T–1R array configuration. Transient response during **b** SET operation and **c** RESET operation of 1 T–1R array, with the array enlarging from 8 × 8 to 128 × 128. **d** Simulated average dynamic power dissipation with V_{WL} changing during write operation of varisized arrays. Reprinted from [52]

the array is more significant during RESET than during SET, since the initial larger current flow in RC network leads to a larger latency feedback. The power dissipation during the operation can also be investigated by the compact model as shown in Fig. 23d. The average power dissipation decreases with the array size due to the decrease of the V_{GS} and I_{max} of the selected cell, which is caused by the increased voltage drop along the signal lines. However, the power dissipation increases with the array size resulting from the increase of sneak current path.

9 Marc-Bocquet Model

Based on the creation and destruction of CF assisted by the joule heat and the electrical field, Marc-Bocquet presented the models for the bipolar and unipolar RRAM [53, 54]. The models discussed above regard to the bipolar RRAM, so here we only focus on the compact for the unipolar RRAM. As shown in Fig. 24, the CF connecting two electrodes will grow and rupture, which depends on the applied voltage. The temperature of the cell will increase due to the joule heat. In one dimension geometry, the heat equation can be simply written as:

$$\sigma_{CF}(x) \cdot \xi(x)^2 = -K_{th\,CF} \frac{\partial^2 T_{CF}}{\partial x^2} + h \frac{T_{CF}(x) - T_{ox}}{t_{ox}} \tag{23}$$

where σ_{CF} is the CF conductivity, $T_{CF}(x)$ is the CF temperature, T_{ox} is the oxide temperature, K_{thCF} is the CF thermal conductivity, t_{ox} is the oxide thickness, and h is the heat transfer coefficient between CF and surrounding oxide. The CF is assumed to be metallic with a high conductivity σ_{CF}. In contrast, the resistance of the oxide gap is higher. The analytical expression of the electric field ξ can be derived:

$$\xi(x) = I_0 \cdot R(x) \tag{24}$$

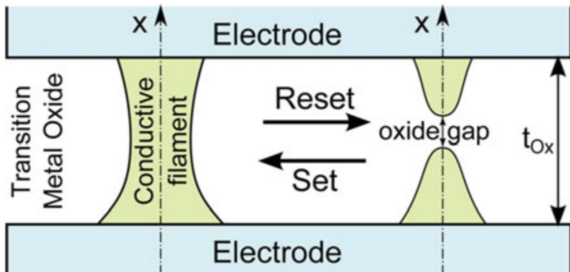


Fig. 24 Schematic representation of a formed and dissolved conductive filament resulting from SET and RESET operations, respectively, in MIM structure. During RESET, the inner temperature is raised by Joule effect and leads to the destruction of the CF. In contrast, during SET, an electrochemical local reduction process enables restoring CF. Reprinted from [53]

where the resistance per unit length $R(x)$ is:

$$R(x) = \frac{1}{r_{\mathrm{CF}}^2(x)\pi(\sigma_{\mathrm{CF}}(x) - \sigma_{ox}) + r_{\mathrm{CF\,max}}^2 \pi \sigma_{ox}} \tag{25}$$

$$I = \frac{V_{cell}}{\int_0^{t_{ox}} R(x)\mathrm{d}x} \tag{26}$$

I_0 is the Ohmic current in the MIM structure, V_{cell} is the applied voltage, σ_{ox} is the conductivity of oxide gap, $\mathrm{r_{CF}(x)}$ and $\mathrm{r_{CFmax}}$ denote the radius and the maximal radius of the CF in cylindrical geometry, respectively. The conductivity of metallic CF is temperature dependence and can be described as [61]:

$$\sigma_{\mathrm{CF}}(x) = \frac{\sigma_{\mathrm{CF0}}}{1 + \alpha_T \cdot (T_{\mathrm{CF}}(x) - T_{\mathrm{amb}})} \tag{27}$$

Two distinct mechanisms are considered to model the creation and destruction of CF: the redox and diffusion process. The classical redox process can be expressed as:

$$Ni^{2+} + 2 \cdot e^- \underset{red}{\overset{ox}{\rightleftharpoons}} Ni \tag{28}$$

The reaction velocities for both the reduction and oxidation processes can adopt the classical Butler Volmer equation [62]:

$$v_{\mathrm{red}} = k_0 \mathrm{e}^{-\frac{\Delta r G_0 + 2(1-\alpha)F(E - E_{\mathrm{eq}})}{RT_{ox}}}(1 - C_{Ni}) \tag{29}$$

$$v_{ox} = k_0 \mathrm{e}^{-\frac{\Delta r G_0 - 2\alpha F(E - E_{\mathrm{eq}})}{\mathrm{RT_{CF}}(x)}} C_{\mathrm{Ni}} \tag{30}$$

α is the asymmetry factor, k_0 is the kinetics constant of electrochemical reaction, R is the ideal gas constant, F is the Faraday constant, ΔrG_0 is the free energy of the reaction at equilibrium, and finally E_{eq} is the equilibrium potential. Here it is assumed that the redox of Ni is isotropic and the potential E is $-|V_{\mathrm{Cell}}|$. C_{Ni} is the dimensionless concentration of metallic species, where can be obtained based on the self-consistent master equation:

$$\frac{\mathrm{d}C_{\mathrm{Ni}}}{\mathrm{d}t} = v_{\mathrm{red}} - v_{ox} - v_{\mathrm{diff}} \tag{31}$$

v_{diff} is the diffusion velocity and can be written as:

$$v_{\mathrm{diff}} = k_{\mathrm{diff}} \cdot \mathrm{e}^{-\frac{E_a}{k_b \cdot T_{\mathrm{CF}}(x)}} \cdot C_{\mathrm{Ni}} \tag{31}$$

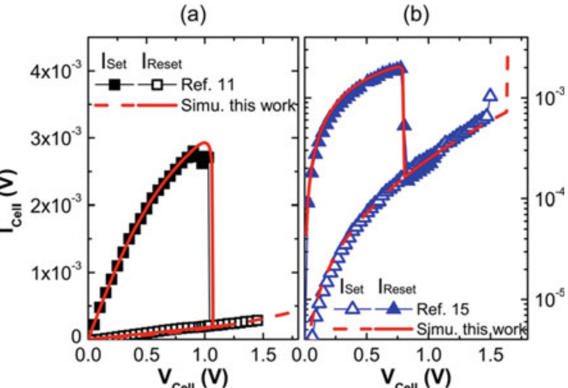


Fig. 25 The measured and calculated I-V characteristic of NiO based unipolar RRAM. Reprinted from [53]

where E_a is the activation energy, governing the thermally assisted diffusion of metallic species.

Figure 25 shows the calculated quasi-static SET and RESET *I*-*V* characteristics, together with the measured data. Since both the CF creation and destruction are described by the self-consistent kinetics equation, the model can continuously calculate the *I*-*V* characteristics of the switching process. Hence, the CF profile obtained after SET operation is used as the initial state of the subsequent RESET operation. The presented model excellently agrees with the experimental data in Fig. 25, which demonstrates the validity of the model.

10 Electrical Characterization

RRAM utilizes the reversible resistive transition effect of materials to achieve information storage. SET and RESET programming are considered as the basic and necessary processes for resistive switching. In this section, the electrical characterization techniques of Ox-RRAM are discussed. Typical measurement methods will be discussed to illustrate the electrical characteristics of Ox-RRAM. Meanwhile, the reliability issues such as retention, endurance and read disturb behaviors obtained by these measurement methods would be introduced in detail.

A. Measurement Method

DC measurement refers to the process of applying gradually increasing voltage sweeps to both ends of the device and obtaining current. It could be used to achieve I-V curve and data retention of RRAM. The typical I-V curves of the memory device in 1R and 1 T-1R configuration after forming process are shown in Fig. 26a, and the typical retention test at high temperature shown in Fig. 26b. When applying increasing voltage to a RRAM cell, the resistance could be calculated based on Ohm's law. In general, the read voltage is approximately one tenth of the switching voltage, and

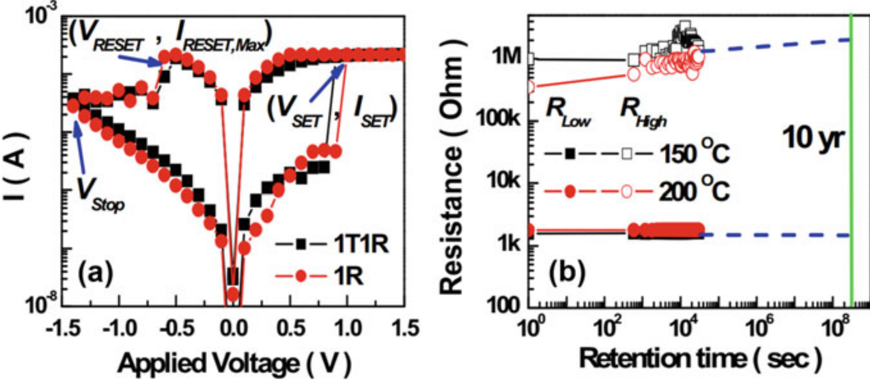


Fig. 26 **a** Typical I-V curve of bipolar resistive switching characteristic in 1R and 1T1R configuration of the TiN/TiO_X/HfO_X/TiN devices, **b** retention properties of resistance of LRS (R_{Low}) and HRS (R_{High}) at 150 and 200 °C. The result predicts 10 years lifetime of stored bit. Reprinted from [63]

the resistance corresponding to the read voltage is recorded as the current resistance of RRAM cell. Besides, the window in I-V curve could reflect the ability of the resistive switching and the health level of RRAM cell. In addition, it is also called on–off ratio and it is proved that cycling endurance could result in the degradation of the on–off ratio. DC test could be measured with a comprehensive semiconductor device parameter analyzer, such as Agilent 4156, Agilent B1500, and Keithley 4200, etc.

Pulse measurement is widely used in the research of RRAM characteristics. In practical applications, the excitation signal of RRAM is pulse signal. Figure 27 shows the typical pulse cycling test waveform: after performing a SET pulse operation, a read operation is followed; and after performing a RESET pulse operation, a read

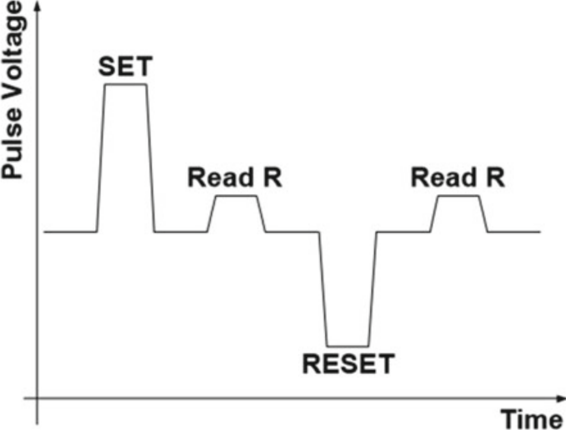


Fig. 27 The typical pulse test waveform for RRAM device

operation is followed. The fast pulse test consists of several these test waveforms, and it could be applied to study the transition speed and cycle endurance characteristics. Figure 28a shows the high operation speed of $TiN/Ti/HfO_2/TiN$ device with 10-ns pulse mode. The 10^6 cycles endurance with binary switching is shown in Fig. 28b.

B. *Application-Cycling Endurance*

For the reliability of RRAM device, endurance and retention are considered as the most important performance to characterize the device. In general, endurance refers to how many switching cycles the RRAM cell could perform. In other words, endurance means the programmable number before the failure of the device. As mentioned before, one SET and one RESET consist a cycle unit in the endurance test. Typical endurance degradation modes could be divided into two scenes: the final failure state of RRAM is fixed in either LRS or HRS, shown in Fig. 29a. The mechanism of endurance degradation could be explained that extra V_O generating or recombining lead to the change of the structure of CF so that it deviates from the original stable

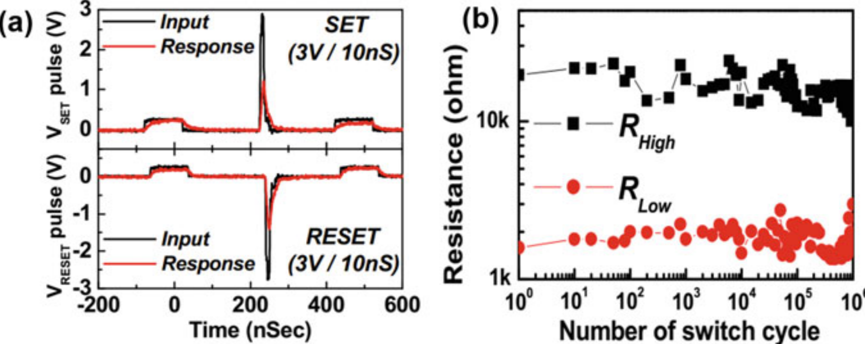


Fig. 28 **a** Test of operation speed with 10-ns pulse width of $TiN/Ti/HfO_2/TiN$ device, **b** endurance switch endurance of 10^6 cycles in $TiN/Ti/HfO_2/TiN$ forming-free memory. Reprinted from [64]

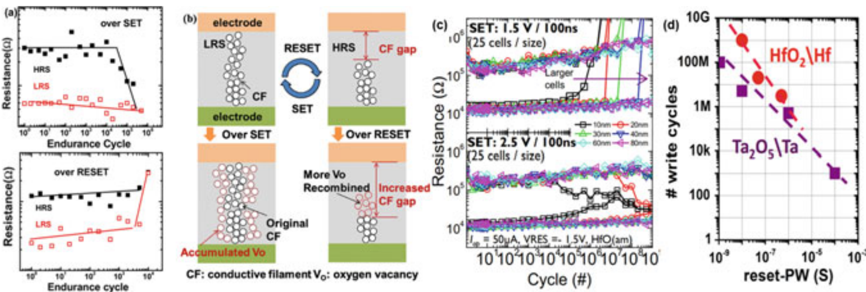


Fig. 29 **a** Typical endurance degradation modes, **b** schematic of endurance degradation mechanisms [48], **c** endurance test of amorphous HfOx with different SET voltage and cell size [65], **d** summary plot of the PW RESET dependence of endurance lifetime obtained from two different dielectric stack [66]

structure, shown in Fig. 29. In detail, under the action of electric field, extra V_O generation results in enlarging the size of CF during SET process. LRS and HRS gradual decrease firstly, then sharp decrease and stuck in LRS causing the loss of ratio, which is called as "over SET". While during RESET process, extra recombination between O^{2-} and V_O results in an enlarged gap. And finally the resistance of both LRS and HRS increases and the device cannot be switched to LRS with the primary SET voltage, which is called as "over RESET" [48].

Cycling endurance degradation is affected by multiple factors. Fantini et al. demonstrate that it depends on the cell size and layer thickness, shown in Fig. 29c, and it could be explained that the degradation of endurance after scaling down could be attributed to the reduced number of ions in the switching layer [65]. Chen et al. study the impact of RESET amplitude and duration on the endurance degradation, and it is found that using shorter write pulse, low O-affinity electrode, and high O-affinity dielectric could obtain significant improvement in endurance performance [66]. What's more, the programming condition also has a significant influence on the endurance degradation [67].

C. *Application-Data Retention*

Another important performance for evaluating the reliability is data retention, and it has been reported by several researchers to obtain optimization method of retention characteristic. Retention refers to the length of time the data can be maintained in the memory. As a universal metric, the commercial applications of the memory should ensure that data could be stored for more than 10 years at 85 °C. Based on the Arrhenius equation, there is an inverse correlation between data retention and temperature, which the retention time decreases with the temperature increasing. In order to efficiently evaluate the retention of RRAM cell or RRAM array, baking the RRAM at high temperature and extrapolating down to the desired temperature is the common accepted method. Because the failure of retention is a completely stochastic process, it makes sense to study the retention behavior of a large number of RRAM device.

Ming Liu et al. demonstrate a temperature-dependent LRS retention failure behavior of RRAM based on TiO_x/Al_2O_3 bilayer structure. The retention behavior at 300μA under different temperatures is shown in Fig. 30a. The retention failure mechanism is attributed to the relationship between the LRS conductivity σ and the temperature T. Besides, the carriers conduction in CF is similar to the nonmetallic path formation [66]. More remarkably, σ is related to T with a simple power law equation, as shown in Fig. 30b. Finally, the resistance deviates from the initial value, so that the RRAM has a retention failure (Fig. 30c).

Meiran Zhao et al. have investigated the statistical behavior of retention in the 1 Kb filamentary analog RRAM array, and evaluate the impact of retention in the neural network. The analog switching ability refers to the conductance of the RRAM varying continuously. The analog data storage could increase the level numbers to make the analog RRAM a promising emerging device for neuromorphic computing. In his work, it is found that the conductance distribution of all conductive levels follows normal distribution with the standard deviation increasing linearly with the

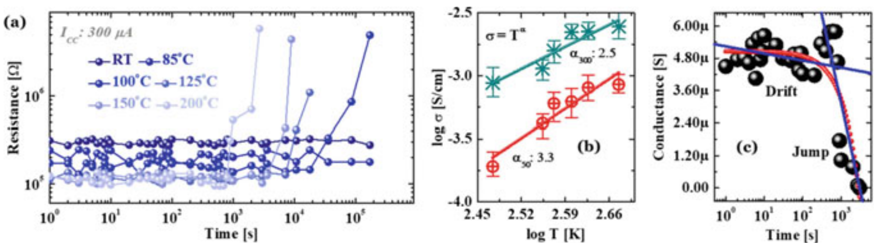


Fig. 30 Retention characterization of TiO_x/Al_2O_3 RRAM. **a** Temperature-dependent LRS retention failure at 300 μA, **b** formation of nonmetallic filament, **c** typical retention failure process starts with a conduction drift followed by a certain jump. Reprinted from [68]

square root of baking time, and the distribution changes with baking time [69]. In detail, Fig. 31 introduces the important results in this work. The current distribution is tight at the beginning and becomes wider as Fig. 31a shown. Figure 31b–d shows that the conductance distribution of each level at a specific time can be fitted as normal distribution. A compact model based on the standard deviation of different levels is proposed from Fig. 31e, f. When applying the compact model to the neural network, it is found that the recognition rate decreases due to the retention (Fig. 31g).

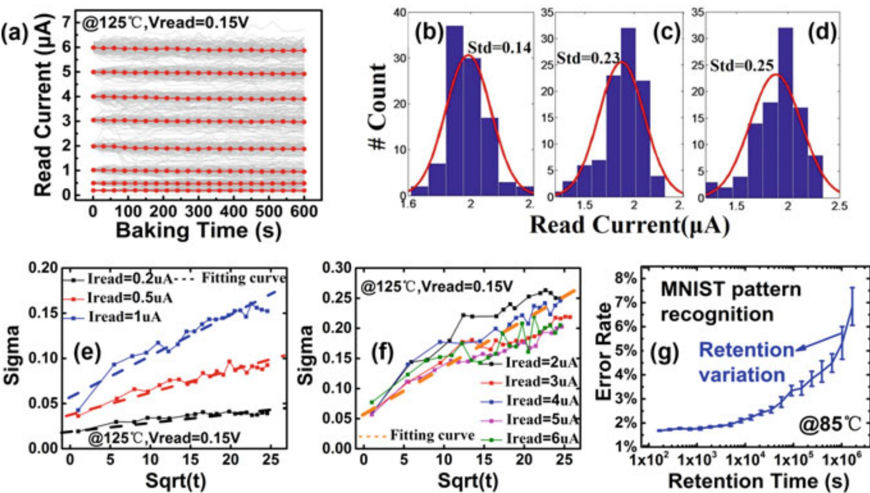


Fig. 31 **a** Read current distribution of 8 levels as a function of baking time at 125 °C, **b**–**d** the normal distribution fitting for read current in (**a**) ~2 μA at 30, 300, and 600 s; standard deviation (sigma) of normal distribution fitted from (**a**) as a function of square root of baking time in **e** 3 h can be fitted independently, **f** 5 LRS can be only fitted with one line, **g** recognition rate as a function of retention time at 85 °C. Reprinted from [69]

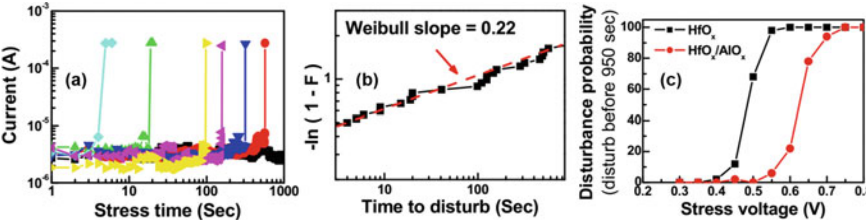


Fig. 32 **a** Dispersive read disturb time of R_{HIGH} under CVS, **b** the Weibull plot of time-to-disturb in R_{high} under CVS. The Weibull slope is 0.22, **c** dependence of disturbance probability on stress voltage. The stress voltage is improved to 0.5 V in HfO_x/AlO_x devices. Reprinted from [70]

D. *Read Disturb*

In addition to endurance and retention, read disturb is another stability concern of RRAM [70]. Read disturb refers to the RRAM cell changing its resistance value after numerous read operations. Y. S. Chen et al. proposed that a thin AlO_x buffer layer under HfO_x layer can enhance the read disturb immunity. It is because that the AlO_x buffer layer could share the voltage with HfO_x layer to elevate the time-to-disturb voltage to 0.5 V, as Fig. 32 shown [70]. In addition to the stress voltage, another important parameter to measure the suppression of read disturb is the read disturb time. Figure 32a shows the read disturb test on time in HRS under constant voltage stress (CVS). And the Weibull distribution of read-disturb-time in HRS under CVS is shown in Fig. 32b. By means of engineering barrier with the AlO_x buffer layer, the read disturb is improved.

References

1. T.W. Hickmott, Low-frequency negative resistance in thin anodic oxide films. J. Appl. Phys. **33**(9), 2669 (1962)
2. J.F. Gibbons, W.E. Beadle, Switching properties of thin NiO films. Solid-State Electron. **7**(11), 785 (1964)
3. Y. Watanabe, J.G. Bednorz, A. Bietsch, Ch. Gerber, D. Widmer, A. Beck, S.J. Wind, B, Current-driven insulator-conductor transition and nonvolatile memory in chromium-doped $SrTiO_3$ single crystals. Appl. Phys. Lett. **78**(23), 3738 (2001)
4. A. Beck, J.G. Bednorz, C. Gerber, C. Rossel, D. Widmer, Reproducible switching effect in thin oxide films for memory applications. Appl. Phys. Lett. **77**(1), 139 (2000)
5. G. Baek, M.S. Lee, S. Seo, M.J. Lee, D.H. Seo, D.-S. Suh, J.C. Park, S.O. Park, H. S. Kim, I.K. Yoo, U.-In. Chung, J.T. Moon, Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 587 (2004)
6. A. Rohde, B.J. Choi, D.S. Jeong, S. Choi, J.S. Zhao, C.S. Hwang, Identification of a determining parameter for resistive switching of TiO_2 thin films. Appl. Phys. Lett. **86**(26), 262907 (2005)
7. Y.A. Lin, S.Y. Wang, D.Y. Lee, T.Y. Tseng, Electrical properties and fatigue behaviors of ZrO_2 resistive switching thin films. J. Electrochem. Soc. **155**(8), H615 (2008)

8. W.Y. Chang, Y.C. Lai, T.B. Wu, S.F. Wang, F. Chen, M.J. Tsai, Unipolar resistive switching characteristics of ZnO thin films for nonvolatile memory applications. Appl. Phys. Lett. **92**(2), 022110 (2008)
9. X. Sun, B. Sun, L. Liu, N. Xu, X. Liu, R. Han, J. Kang, G. Xiong, T.P. Ma, Resistive switching in CeOx films for nonvolatile memory application. IEEE Electr. Dev. Lett. **30**(4), 334 (2009)
10. H.Y. Lee, P.S. Chen, T.Y. Wu, Y.S. Chen, C.C. Wang, P.J. Tzeng, C.H. Lin, F. Chen, C.H. Lien, M.-J. Tsai, Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 297 (2008)
11. Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, K. Kawai, S. Muraoka, S. Mitani, S. Fujii, K. Katayama, M. Iijima, T. Mikawa, T. Ninomiya, R. Miyanaga, Y. Kawashima, K. Tsuji, A. Himeno, T. Okada, R. Azuma, K. Shimakawa, H. Sugaya, T. Takagi, R. Yasuhara, K. Horiba, H. Kumigashira, M. Oshima, Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 293 (2008)
12. Y. Wu, B. Lee, H.P. Wong, Al_2O_3-based RRAM using atomic layer deposition (ALD) with 1-μA RESET current. IEEE Electr. Dev.Lett. **31**(12), 1449 (2010)
13. R. Waser, R. Dittmann, G. Staikov, K. Szot, Redox-based resistive switching memories: Nanoionic mechanisms, prospects, and challenges. Adv. Mater. **21**(25), 2632 (2009)
14. B. Gao, J.F. Kang, Y.S. Chen, F.F. Zhang, B. Chen, P. Huang, L.F. Liu, X.Y. Liu, Y.Y. Wang, X.A Tran, Z.R. Wang, H.Y. Yu, A. Chin, Oxide-based RRAM: unified microscopic principle for both unipolar and bipolar switching. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p.417 (2011)
15. R. Valov, J.R. Waser, M.N. Jameson, Kozicki, Electrochemical metallization memories: Fundamentals, applications, prospects. Nanotechnology **22**(25), 254003 (2011)
16. Q. Liu, J. Sun, H. Lv, S. Long, K. Yin, N. Wan, Y. Li, L. Sun, M. Liu, Real-time observation on dynamic growth/dissolution of conductive filaments in oxide-electrolyte-based ReRAM. Adv. Mater. **24**(14), 1844 (2012)
17. M. Terai, Y. Sakotsubo, S. Kotsuji, H. Hada, Resistance controllability of Ta_2O_5/TiO_2 stack ReRAM for low-voltage and multilevel operation. IEEE Electr. Dev. Lett. **31**(3), 204 (2010)
18. X.M. Guan, S.M. Yu, H.-S. Philip Wong, On the switching parameter variation of metal-oxide RRAM—Part I: Physical modeling and simulation methodology. IEEE Trans. Electr. Dev. **59**(4), 1172 (2012)
19. L. Vandelli, A. Padovani, L. Larcher, G. Broglia, G. Ori, M. Montorsi, G. Bersuker P. Pavan, Comprehensive physical modeling of forming and switching operations in HfO_2 RRAM devices. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 421 (2011)
20. P. Huang, B. Gao, B. Chen, F.F. Zhang, L. F. Liu, G. Du, J.F. Kang, X.Y. Liu, Stochastic simulation of forming, SET and RESET process for transition metal oxide-based resistive switching memory. In *International Conference on Simulation of Semiconductor Processes and Devices*, p. 312 (2012)
21. S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, H.-S.P. Wong, A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. Adv. Mater. **25**(12), 1774 (2013)
22. P. Huang, X.Y. Liu, B. Chen, H.T. Li, Y.J. Wang, Y.X. Deng, K.L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang, J.F. Kang, A physics based compact model of metal oxide based RRAM DC and AC operation. IEEE Trans. Electr. Dev. **60**(12), 1114 (2013)
23. B. Gao, S. Yu, N. Xu, L.F. Liu, B. Sun, X.Y. Liu, R.Q. Han, J.F. Kang, B. Yu, Y.Y. Wang, Oxide-based RRAM switching mechanism: a new ion-transport-recombination model, in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 563 (2008)
24. B. Gao, J. Kang, L. Liu, X. Liu, B. Yu, A physical model for bipolar oxide-based resistive switching memory based on ion-transport-recombination effect. Appl. Phys. Lett. **98**(23), 232108 (2011)
25. P. Huang, D. Zhu, S. Chen, Z. Zhou, Z. Chen, B. Gao, L. Liu, X. Liu, J. Kang, Compact model of HfO_X-Based electronic synaptic devices for neuromorphic computing. IEEE Trans. Electr. Dev. **64**(2), 614 (2017)

26. H.S.P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F.T. Chen, M.-J. Tsai, Metal-oxide RRAM. Proc. IEEE **100**(6), 1951 (2012)
27. Y.S. Chen, J.F. Kang, B. Chen, B. Gao, L.F. Liu, X.Y. Liu, Y.Y. Wang, L. Wu, H.Y. Yu, J.Y. Wang, Q. Chen, E.G. Wang, Microscopic mechanism for unipolar resistive switching behaviour of nickel oxides. J. Phys. D: Appl. Phys. **45**(6), 065303 (2012)
28. S. Larentis, F. Nardi, S. Balatti, D.C. Gilmer, D. Ielmini, Resistive switching by voltage-driven ion migration in bipolar RRAM—Part II: modeling. IEEE Trans. Electr. Dev. **59**(9), 2468 (2012)
29. Y.M. Kim, J.S. Lee, Reproducible resistance switching characteristics of hafnium oxide-based nonvolatile memory devices. J. Appl. Phys. **104**(11), 114115 (2008)
30. Q. Liu, W.H. Guan, S.B. Long, R. Jia, M. Liu, J.N. Chen, Resistive switching memory effect of ZrO2 films with Zr+ implanted. Appl. Phys. Lett. **92**(1), 012117 (2008)
31. E. Ielmini, F. Nardi, C. Cagli, A.L. Lacaita, Size-dependent retention time in NiO-based resistive-switching memories. IEEE Electr. Dev. Lett. **31**(4), 353 (2010)
32. N.F. Mott, E.A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, U.K., 1979)
33. S. M. Yu, Y. Wu, H.-S. Philip Wong, Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory. Appl. Phys. Lett. **98**(10), 103514(2011)
34. M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U-I. Chung, In-K. Yoo, K. Kim, A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-X}/TaO_{2-X} bilayer structures. Nat. Mater. **10**(8), 625 (2011)
35. M. Hu, Z. Yao, X. Wang, Characterization techniques for graphene-based materials in catalysis. AIMS Mater. Sci. **4**(3), 755 (2017)
36. H. Tian, H.-Y. Chen, B. Gao, S. Yu, J. Liang, Y. Yang, D. Xie, J. Kang, T.-L. Ren, Y. Zhang, H.-S.P. Wong, Monitoring oxygen movement by raman spectroscopy of resistive random access memory with a graphene-inserted electrode. Nano Lett. **13**(2), 651 (2013)
37. Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J.P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H.L. Xin, R.S. Williams, Q. Xia, J.J. Yang, Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. Nature Mater. **16**(1), 101 (2016)
38. P. Calka, E. Martinez, D. Lafond, H. Dansas, S. Tirano, V. Jousseaume, F. Bertin, C. Guedj, Resistance switching in HfO_2-based OxRRAM devices. Microelectr. Eng. **88**(7), 1140 (2011)
39. A. Ranjan, N. Raghavan, K. Shubhakar, R. Thamankar, J. Molina, S.J. O'Shea, M. Bosman, K.L. Pey, CAFM based spectroscopy of stress-induced defects in HfO_2 with experimental evidence of the clustering model and metastable vacancy defect state. in *Reliability Physics Symposium*, 7A-4–1 (2016)
40. P. Huang, X.Y. Liu, W.H. Li, Y.X. Deng, B. Chen, Y. Lu, B. Gao, L. Zeng, K.L. Wei, G. Du, X. Zhang, J.F. Kang, A physical based analytic model of RRAM operation for circuit simulation. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 605 (2012)
41. R. Pornprasertsuk, T. Holme, F.B. Prinz, Kinetic Monte Carlosimulations of solid oxide fuel cell. J. Electrochem. Soc. **156**(12), B1406 (2009)
42. S. Yu, X. Guan, H.-S. P. Wong, On the stochastic nature of resistive switching in metal oxide RRAM: physical modeling, Monte Carlo simulation, and experimental characterization. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 413 (2011)
43. S.C. Chae, J.S. Lee, S. Kim, S.B. Lee, S.H. Chang, C. Liu, B. Kahng, H. Shin, D.-W. Kim, C.U. Jung, S. Seo, M.-J. Lee, T.W. Noh, Random circuit breaker network model for unipolar resistance switching. Adv. Mater. **20**(6), 1154 (2008)
44. Y.D. Zhao, P. Huang, Z. Chen, H. Li, B. Chen, W. Ma, F. Zhang, B. Gao, X. Liu, J. Kang, Modeling and optimization of bilayered TaOx RRAM based on defect evolution and phase transition effects. IEEE Trans. Electr. Dev. **63**(4), 1524 (2016)
45. T. Sadi, L. Wang, L. Gerrer, V. Georgieva, A. Asenov, Self-consistent physical modeling of SiOx-based RRAM structures. In *International Workshop on Computational Electronics* (IWCE) (2015)

46. P. Huang, B. Chen, H. Li, Z. Chen, B. Gao, X. Liu, Parameters extraction on HfO_X based RRAM. In *European Solid-State Device Research Conference (ESSDERC)*, p. 250 (2014)
47. Y. Y. Chen, M. Komura, R. Degraeve, B. Govoreanu, L. Goux, A. Fantini, N. Raghavan, S. Clima, L. Zhang, A. Belmonte, A. Redolfi, G. S. Kar, G. Groeseneken, D. J. Wouters, M. Jurczak, Improvement of data retention in HfO_2/Hf 1T1R RRAM cell under low operating current. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 252 (2013)
48. P. Huang, B. Chen, Y. Wang, F. Zhang, L. Shen, R. Liu, L. Zeng, G. Du, X. Zhang, B. Gao, J. Kang, X. Liu, X. Wang, B.B. Weng, Y.Z. Tang, G.-Q. Lo, D.-L. Kwong, Analytic model of endurance degradation and its practical applications for operation scheme optimization in metal oxide based RRAM. in *Tech. Dig. IEEE Int. Electron Devices Meeting*, p. 597 (2013)
49. L. Chua, Memristor—the missing circuit element. IEEE Trans Circuit Theory **18**(5), 507 (1971)
50. D.B. Strukov, G.S. Snider, D.R. Stewart, R.S. Williams, The missing memristor found. Nature **453**(7191), 80 (2008)
51. X. Guan, S. Yu, H.-S.P. Wong, A SPICE compact model of metal oxide resistive switching memory with variations. IEEE Electr. Dev. Lett. **33**(10), 1405 (2012)
52. H. Li, P. Huang, B. Gao, B. Chen, X. Liu, J. Kang, A SPICE model of resistive random access memory for large-scale memory array simulation. IEEE Electr. Dev. Lett. **35**(2), 211 (2014)
53. M. Bocquet, D. Deleruyelle, C. Muller, J.-M. Portal, Self-consistent physical modeling of set/reset operations in unipolar resistive-switching memories. Appl. Phys. Lett. **98**(26), 263507 (2011)
54. M. Bocquet, D. Deleruyelle, H. Aziza, C. Muller, J.-M. Portal, T. Cabout, E. Jalaguier, Robust compact model for bipolar oxide-based resistive switching memories. IEEE Trans. Electr. Dev. **61**(3), 674 (2014)
55. U. Russo, D. Ielmini, C. Cagli, A.L. Lacaita, Filament conduction and reset mechanism in NiO-based resistive-switching memory (RRAM) devices. IEEE Trans. Electr. Dev. **56**(2), 186 (2009)
56. U. Russo, D. Ielmini, C. Cagli, A.L. Lacaita, Self-accelerated thermal-dissolution model for reset programming in unipolar resistive switching memory (RRAM) devices. IEEE Trans. Electr. Dev. **56**(2), 193 (2009)
57. J.J. Yang, M.D. Pickett, X. Li, D.A.A. Ohlberg, D.R. Stewart, S.S. Williams Memristive switching mechanism for metal/oxide/metal nanodevices. Nat. Nanotech. **3**(7) 429 (2008)
58. D.B. Strukov, R.S. Williams, Exponential ionic drift: fast switching and low volatility of thin-film memristors. Appl Phys Mater Sci Process **94**(3), 515 (2009)
59. V.-A. Manual, *Accellera Org* (Inc., Napa, CA, USA, 2009)
60. Y.X. Deng, P. Huang, B. Chen, X. Yang, B. Gao, J. Wang, L. Zeng, G. Du, J. Kang, X. Liu, RRAM crossbar array with cell selection device: a device and circuit interaction study. IEEE Trans. Electr. Dev. **60**(2), 719(2013)
61. U. Russo, D. Ielmini, C. Cagli, A. L. Lacaita, S. Spiga, C. Wiemer, M. Perego, M. Fanciulli, in *IEEE International Electron Devices Meeting*, p. 775 (2007)
62. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications* (Wiley, New York, 2001)
63. H.Y. Lee, P.S. Chen, T.Y. Wu, Y.S. Chen, C.C. Wang, P.J. Tzeng, C.H. Lin, F. Chen, C.H. Lien, M.-J. Tsai, Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM, in *IEEE International Electron Devices Meeting*, p. 1 (2008)
64. Y.S. Chen, T.-Y. Wu, P.-J. Tzeng, P.-S. Chen, H.-Y. Lee, C.-H. Lin, F. Chen, M.-J. Tsai, Forming-free HfO2 bipolar RRAM device with improved endurance and high speed operation. in *International Symposium on Vlsi Technology, Systems, and Applications*, p. 37 (2009)
65. L.G.A. Fantini, A. Redolfi, R. Degraeve, G. Kar, Y.Y Chen, M. Jurczak, Lateral and vertical scaling impact on statistical performances and reliability of 10nm TiN/Hf(Al)O/Hf/TiN RRAM devices. in *2014 Symposium on VLSI Technology Digest of Technical Papers*, p. 1 (2014)
66. C.Y. Chen, L. Goux, A. Fantini, A. Redolfi, Understanding the impact of programming pulses and electrode materials on the endurance properties of scaled Ta_2O_5 RRAM cells. in *IEEE International Electron Devices Meeting*, p. 355 (2014)

67. Y.Y. Chen, B. Govoreanu, L. Goux, R. Degraeve, Balancing SET/RESET Pulse for Endurance in 1T1R Bipolar RRAM. IEEE Trans. Electr. Dev. **59**(12), 3243 (2012)
68. W. Banerjee, N. Lu, Y. Yang, L. Li, H. Lv, Q. Liu, S. Long, M. Liu, Investigation of retention behavior of TiOx/Al2O3 resistive memory and its failure mechanism based on meyer-neldel rule. IEEE Trans. Electr. Dev. **65**(3), 957 (2018)
69. M. Zhao, H. Wu, B. Gao, Q. Zhang, W. Wu, S. Wang, Y. Xi, D. Wu, N. Deng, S. Yu, H.-Y. Chen, H. Qian, Investigation of statistical retention of filamentary analog RRAM for neuromophic computing. in *IEEE International Electron Devices Meeting*, p. 872 (2017)
70. 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, M.-J. Tsai, Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity. in *IEEE International Electron Devices Meeting*, p. 105 (2009)

RRAM-Based Neuromorphic Computing Systems

Putu Andhita Dananjaya, Roshan Gopalakrishnan, and Wen Siang Lew

Abstract New computational paradigms have been widely investigated in order to further improve the approach in handling the exponentially increasing amount of data generated across the globe as well as various emerging hardware requirements to execute complex tasks, e.g., pattern recognition, speech classification, etc. Neuromorphic computing has emerged as one of the most extensively investigated among these approaches. RRAM devices with their desired characteristics have been rigorously engineered to meet the synaptic element requirements to realize such robust yet power efficient platform. Among the performance parameters necessary to achieve an ideal synaptic device in context of RRAM device, there are certain dependency and potential trade-offs. In this chapter, different type of RRAM, i.e., anion and cation, devices based on their underlying physical mechanism with various advantages and disadvantages are discussed. Different techniques that have been implemented to improve the device synaptic characteristics from material viewpoint and programming approach followed by several system level simulations demonstrating the projected performance of these devices are provided in detail. Different algorithms available for the RRAM synapse implementation are also discussed.

1 Introduction

Human brain performs massively parallel and low power operations. It can outperform present age microprocessors on many tasks involving pattern recognition and input classification. The underlying neurons are heavily inter-connected; on average each neuron is connected to 10,000 (or up to 100,000) other neurons [1–3]. Despite

P. A. Dananjaya · W. S. Lew (✉)
Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Jurong West 637371, Singapore
e-mail: wensiang@ntu.edu.sg

R. Gopalakrishnan
Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Fusionopolis Way 138632, Singapore

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_12

the complexity of the human brain and the on-going research to gain the full understanding of it, significant amount of efforts and progress has been made in closely emulating it in terms of its functionalities to execute various complex tasks.

In recent years, RRAM devices have emerged as one of the most promising candidates as a major component in neuromorphic engineering to mimic the functionality of synapses in the human brain [5]. With its relatively simple building blocks of two-terminal device structure, it has demonstrated various synaptic behaviors, e.g. potentiation, depression, and spike-timing dependent plasticity (STDP). Furthermore, its two-terminal nature enables the implementation of high-density crossbar array architecture. The RRAM-based crossbar array can be used both as a high-density memory component as well as a computation unit.

An ideal synaptic device is one of the major elements required to realize a robust neuromorphic computing platform. It plays major role in determining the interconnectivity strength among neurons in the system by storing the weight values. These values are updated according to the learning rules implemented during training. In the case of spiking neural network (SNN), the tuneable resistive state of RRAM-based synapses is analogous to the synaptic plasticity of the brain. The electrical connection between a presynaptic neuron and a postsynaptic neuron (as shown in Fig. 1) changes, strengthening or weakening the synaptic impulses, thus mimicking brain-like functionalities. With RRAM excellent device and array scalability, highly connected crossbar architecture, shown in Fig. 2, can potentially be implemented in the large neural network. Input axons are the input connections from the output neurons in the previous convolution layer which was mapped onto another neuromorphic core. Output neurons are spiking neurons. Spiking neurons receive input current from many other spiking neurons and fire a spike when the integrated current input reaches the neuron threshold. These building blocks like axons, neurons and synapses together can perform mathematical operations. Matrix dot vector multiplications can be performed efficiently with these crossbar structure [6]. Each column in a crossbar produces the sum of product of input from axons and the weights stored at each RRAM synapses. With the available models and algorithms in the field, there are several core requirements a device must have in order to realize an ideal synaptic device.

From neural network (NN) accuracy and robustness viewpoint, the most important requirements for the synaptic device are analog deterministic as well as symmetric linear weight update. Ideally, each synaptic device should exhibit non-overlapping

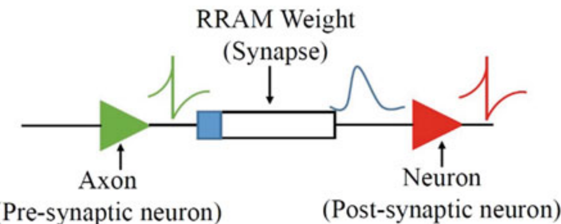


Fig. 1 RRAM-based synapse between an axon and a neuron. This figure is adapted from [4]

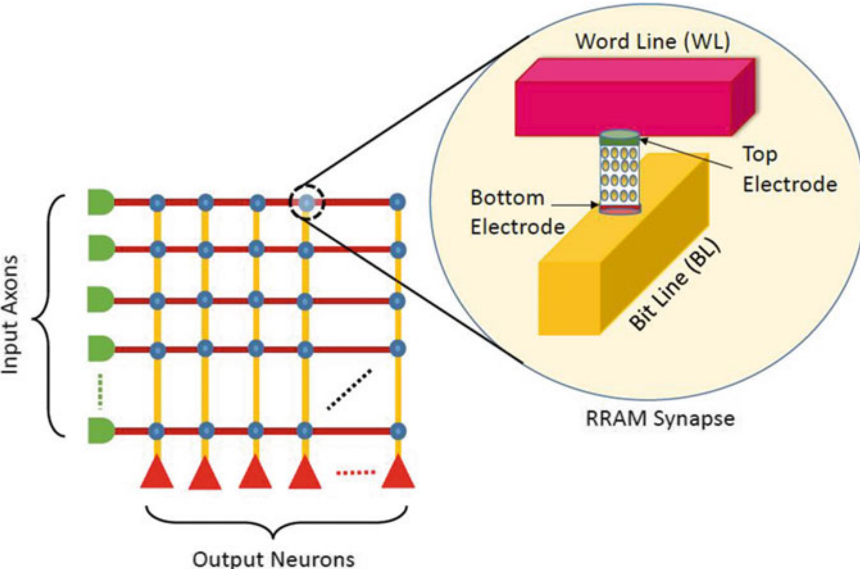


Fig. 2 Crossbar array of two-terminal synaptic devices in a neuromorphic core, allowing a direct mapping of the weight values to the algorithm. This figure is adapted from [4]

multilevel conductance characteristics of at least 32 levels (5-bit). However, due to inherent cycle-to-cycle and device-to-device variation of RRAM devices, trade-off in the number of bits per cell might be required to accommodate the state variation allowing sufficient read margin in between the states. Tighter distribution of the states can be achieved by implementation of write-verify scheme at the expense of programming energy and overall speed. The more conductance levels obtained within a single synapse will enhance the network immunity towards input noise, thus realizing higher learning and test accuracy. Despite that, the number of bits required per cell is still subjected to the network architecture and algorithms implemented. Linearity of the weight update is associated with the relationship between the change of weight value for every programming cycle, while the symmetry is referring to the change of weight value during potentiation and depression cycle. Symmetric linear weight update feature will allow convenient direct mapping of the device conductance and the algorithm weight values. Furthermore, it will enable more efficient training process through state-independent weight update. However, due to the two-terminal nature of RRAM devices, asymmetric nonlinear change of conductance is a huge challenge. This undesired feature has been shown to significantly reduce the network learning accuracy. Thus, different techniques from materials and circuits perspective as well as hardware-algorithm co-optimizations have been investigated.

Other requirements from key device performance parameters consist of endurance characteristics of $\geq 10^9$, long data retention of ≥ 10 years, low programming energy of ≤ 10 fJ, high scalability of ≤ 10 nm, and maximum dynamic ratio of ≥ 100. High

endurance capability is required to allow more training cycles for the network. This is especially important for on-chip learning implementation. On the other hand, long retention is important to accommodate more inference processes, in which the weight values are read with minimum read disturbance. If the data retention of the device is poor, the number of inferences can be done without refreshing the weight value will be relatively lower. Trade-off between endurance and retention in RRAM devices has been reported, thus optimization from materials and programming point of view must be thoroughly considered. In order to not only mimic human brain functionalities but also its efficiency, the device must be able to operate in the order of ~10 fJ per synaptic event. This is one of the most challenging aspects in synaptic device engineering, especially in highly scalable two-terminal devices since programming and reading of the states are done through the same terminals. This leads to another trade-off with device retention. A long data retention requires high state energy barrier to reduce the effect of external disturbance such as heat and electric field, however at the same time this energy barrier must be sufficiently low to achieve low programming energy requirement. Like high density storage devices, highly scalable device footprint is also desired for synaptic device applications to enable large scale neural network within compact chip dimension. In order to fully utilize the high scalability of RRAM devices, a two-terminal select device is required to facilitate the real crossbar array implementation. This can potentially add on to the challenging task of achieving linear and symmetrical weight update. Dynamic ratio is defined as the ratio of the highest conductance value to the lowest one. Higher dynamic ratio can be translated into more superior mapping capability of the network. It also enables larger network connectivity in which maintaining sufficient read margin is crucial. These correlations among the device parameters post an enormous challenge in finding a reliable device that can provide excellent scalability while maintaining high synaptic performances. Thus, device, circuit, and algorithm-level co-optimization is needed (Fig. 3).

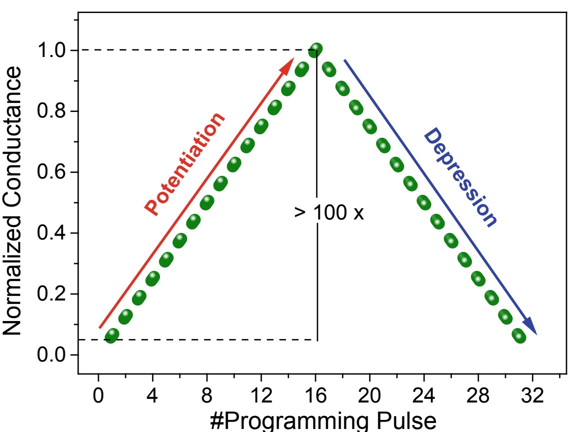


Fig. 3 Ideal analog synapse properties with gradual, linear, and symmetrical weight modulation under identical programming pulse condition with sufficient margin between the states and large dynamic ratio

RRAM-based Synaptic Devices and System Level Simulations

Different types of RRAM devices have been introduced in Chap. 10, i.e., cation and anion devices. Their promising performances as synaptic devices have been widely investigated and demonstrated on different neuromorphic computing platforms. The underlying mechanism of different RRAM structures might lead to a huge difference in macroscopic behavior of the device. Through structural engineering and rigorous optimization of device programming schemes, RRAM devices have demonstrated highly stochastic memory behavior to significantly more deterministic features. To accommodate the different synaptic behaviors of these devices, various learning rules have also been implemented. In this section, synaptic properties of different RRAM devices under various learning rules with several system-level simulations are discussed.

2 Anion-Based Synaptic Devices

The underlying mechanism of anion devices is based on the oxygen vacancy defects modulation within the oxide layer under external electric field. The fundamental structure of an anion device consists of an oxide switching layer coupled with an inert electrode on one side and oxygen reservoir system on the other side, which can be in the form of reactive electrode (Ti, Hf, Ta, etc.) or oxygen-deficient oxide layer. Anion devices have been reported to have high scalability of sub-10 nm [7–9], excellent reliability (endurance as high as 10^{12} and retention of more than 10 years) [10–12], multibit per cell capability, and low energy consumption. Anion devices initially emerged as one of the most promising candidates in non-volatile memory technology as both embedded memory and standalone memory for high density storage applications. In recent years, these devices have also attracted interest from neuromorphic computing and engineering community due to their desired characteristics. They have then been extensively studied and implemented as synaptic device for various NN applications, mainly taking advantage of their high scalability and analog memory characteristic. Anion-based devices can be categorized into two major classes based on the switching nature of the device, i.e., localized (filamentary) and non-localized (non-filamentary) switching class. The difference between these two device classes is mainly on the active area involved during switching operation, with the former involves significantly smaller area than the latter.

a. ***Filamentary Devices***

In general, the filamentary anion devices have an abrupt SET process, i.e., transition from low to high conductance state, while having a gradual RESET process, i.e., transition from high to low conductance state. The gradual RESET process holds the main advantage of anion device over its cation counterpart as a synaptic device to achieve gradual depression. This is because achieving a gradual SET process during potentiation can be performed by controlling the compliance current level

in 1-transistor-1-RRAM (1T1R) structure during the weight update. Thus, both gradual potentiation and depression can be achieved in anion devices. However, this approach still requires non-identical programming pulses and resistance state verification before the programming step, thus overhead on circuitry is needed during the learning process. Furthermore, the inherent oxygen vacancy defects variation within the device structure makes achieving symmetrical and linear weight update with sufficient read margin remain a huge challenge in these devices. Several structures, i.e., AlO_x, HfO_x and TaO_x—based structures, have been comprehensively investigated, improved, and implemented as synaptic devices to meet the requirements of an ideal synaptic device.

Aluminum Oxide (AlO_x)-based Devices

AlOx-based RRAM devices have been investigated as both digital and analog memory devices under different systems, i.e., Ti/Al_2O_3/Pt [13], TiN/Al_2O_3/Pt [14], Ni/Al_2O_3/Pt [15], CNT/AlO_x/CNT [16], Ti/AlO_x/TiN [17], and Al/AlO_x/Pt [18]. In general, the reported AlO_x-based devices have high dynamic ratio (ranging from 10 to 1000), high scalability (down to 36 nm^2 device active area) [16], and low switching energy (below 2 pJ) [14, 17]. The potentiation and depression characteristic of the AlO_x-based structure was experimentally tested in Ti/AlO_x/TiN [17]. The linear gradual conductance change in both directions was achieved under non-identical pulses scheme. Different compliance currents (CCs) from 50 to 900 μA under 1.5 V, 500 μs voltage pulse were imposed during potentiation while different pulse amplitudes from -1 V to -1.6 V with 500 μs duration were used in depression mode. The device was able to achieve an average of 1.2 and 1.7% conductance change per programming pulse with 85 potentiation and 60 depression steps while maintaining ~10 dynamic ratio. However, due to non-identical pulses scheme required during the operation for both SET and RESET, significant overhead must be implemented on the peripheral circuit, which is not ideal for on-chip learning application.

AlO_x has relatively high oxygen scavenging immunity, which is reported to result in significantly smaller filament dimension [19]. While it is a desired property to achieve high ON/OFF ratio, high speed, and excellent uniformity, it also raises a challenge in achieving linear and symmetrical weight update. Thus, rather than being implemented as the main switching layer, AlO_x has been more widely used as an insertion layer interfacing the main switching layer to improve the synaptic performance of the RRAM devices in terms of uniformity and linearity of the conductance update [20–27].

Hafnium Oxide (HfO_x)-based Devices

One of the first structures explored for synaptic device applications is HfO_x-based device with Ti oxygen reservoir electrode. Different weight update schemes have been demonstrated for this system, i.e., identical and non-identical pulses. Under identical pulses scheme, the device can only achieve gradual depression while still having abrupt potentiation. The gradual potentiation can be achieved under non-identical pulse condition in which the current flowing through the 1T1R device is closely controlled by pulsing the transistor's gate, resulting in a well-control filament

formation. Several approaches have been implemented to enable gradual weight update in both directions under identical pulse condition, i.e., insertion of an oxide layer with less defect's mobility and thermal enhancement layer (TEL) [25, 28].

The first approach interfaced HfO_2 layer with AlO_x at the inert electrode side of the structure [25]. AlO_x layer has higher oxygen vacancy diffusion barrier compared to HfO_2 layer, which induced filament constriction at the AlO_x/HfO_2 interface. This promoted lateral filament modulation during potentiation and depression process. The devices are able to achieve gradual conductance change in both directions and further improved their linearity. In comparison with HfO_2/Ti structure, the improvement on the linearity of the potentiation in AlO_x/HfO_2/Ti system, i.e., from α value of 16.53 to −0.01, carried a noticeable trade-off in dynamic range of the device conductance (reduced from 10 to 3). The AlO_x/HfO_2/Ti synaptic device properties were input into simulated NN implementing multilayer perceptron algorithm. It was evaluated under the Mixed National Institute of Standards and Technology (MNIST) dataset to test the NN accuracy in performing pattern recognition. It was shown that the improvement in linearity of the conductance change of the synapse was translated into significant increase in pattern recognition accuracy, i.e., from ~10% for HfO_2/Ti to ~90% for AlO_x/HfO_2/Ti structure.

The insertion of TEL in HfO_x-based device was designed based on gradual SET process observed during high temperature programming of the device [28]. HfO_x/Ti anion device was observed to exhibit abrupt SET under room temperature programming condition and gradual SET during cell programming at 150 °C. In order to obtain gradual SET process at room temperature operation, oxygen deficient TaO_x layer was introduced as TEL and oxygen reservoir in the structure replacing Ti electrode. This layer has significantly lower thermal conductivity compared to Ti electrode, which induced localized Joule heating effect across the active filament region during the switching process. This shifted the device switching property from predominantly electric field to thermally induced switching. This resulted in the formation of multiple weak filaments instead of single filament switching, converting the abrupt into analog SET process, while maintaining dynamic ratio of 10. While it provides a promising solution to mitigate abrupt potentiation issue, the multiple weak filaments system has a trade-off in read disturb and retention of the conductance state. This will have negative impact on the amount of inferences the NN can perform while maintaining the weight values within acceptable deviation. This synaptic device has been experimentally demonstrated on a 1 k-bit 1T1R array to carry out human face classification.

Tantalum Oxide (TaO_x)-based Devices

Another oxide system that has been widely investigated for synaptic device applications is TaO_x-based devices. One of the first reports on TaO_x-based devices was Ta_2O_{5-x}/TaO_{2-x} system that demonstrated an excellent digital memory endurance capability of 10^{12} cycles under 10 ns operating speed [11]. Ta_2O_{5-x} was implemented as the oxide switching layer with an oxygen deficient TaO_{2-x} acted as the oxygen reservoir in the structure. Multilevel cell capability of this structure was demonstrated with an improved ON/OFF ratio of ~1000 with well separated 4 conductance levels (2

bits/cell) and 10 years extrapolated retention [29]. The application of this structure as synaptic device was demonstrated through rigorous optimization of pulse amplitude, pulse width, and the interval between subsequent pulses during operation to achieve gradual potentiation and depression [30]. This system was demonstrated as 2nd order memristor to realize synaptic plasticity in which different parameters involved during switching operation were considered. The modulation of conductive filament dimension, w, that directly results in the device conductance change was referred as 1st order parameter, utilizing memristor as a simple programmable memory device. On the other hand, the 2nd order memristor used the local temperature, T, within the active switching region that governed the evolution of the 1st order parameter, w, capturing the dynamics aspect of the device. T provides time-dependent variable that abruptly increases with the applied pulse and spontaneously decays after its removal. T enables the system to bio-realistically demonstrate activity-dependent plasticity, which is analogous to Ca^+ concentration that regulates the weight-state variable.

Other TaO_x-based devices used Ti or TiO_x layer as oxygen reservoir in the system. Compared to the HfO_x-based devices, the proposed switching mechanism is based on predominant lateral modulation of filament width instead of vertical modulation of filament gap connecting the electrodes [31]. This has been proven to be crucial in achieving gradual and linear weight update. Ta_2O_5/TiO_x synaptic device was implemented on simulated multilayer perceptron neural network under on-chip training condition by back-propagation algorithm. Even with dynamic ratio of ~5, the system was able to achieve almost 90% recognition accuracy using MNIST training data set [31] (Fig. 4).

b. ***Non-filamentary Devices***

Non-filamentary anion device utilizes interfacial defects movement between two layers of material, i.e., oxides and/or metals, which uniformly occurs across the entire device area. The change in the structural defects configuration under external electric field modulates the Schottky barrier at the interface causing significant change

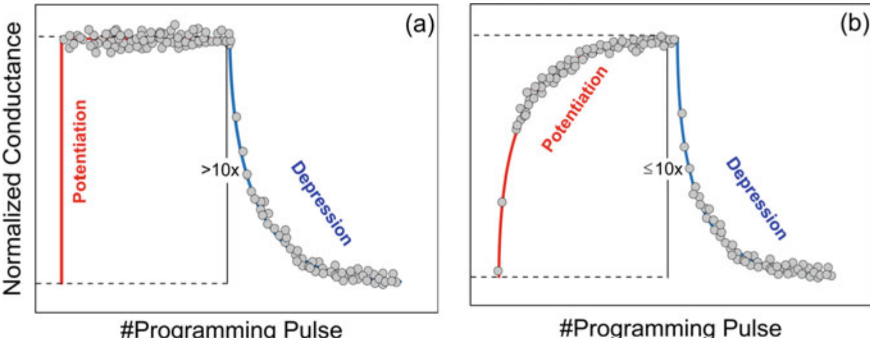


Fig. 4 **a** General synaptic behavior of conventional filamentary anion devices under identical pulse condition, abrupt potentiation, and gradual depression. **b** Gradual potentiation can be achieved through structural engineering in the expense of the device dynamic ratio

in device conductance, i.e., interfacial type switching. During the SET or RESET operation, it also alters the thickness of tunneling gap in the system, allowing higher or lower number of electrons flowing through the device. These enable gradual conductance change during the switching operation, mitigating the issue encountered in most of the filamentary synaptic devices. Non-filamentary switching has been widely reported in many oxide structures, e.g., TiO_x, TaO_x, WO_x, and PCMO.

Titanium Oxide (TiO_x)-based Devices

Analog characteristic in TiO_x systems have been demonstrated by the tuning the oxide layer stoichiometry and the use of oxygen gathering electrode. One of the early structures investigated for synaptic device was TiO_x/TiO_y bilayer oxides system [32]. It was composed of ~50 nm sol-gel TiO_x layer grown on top of 6 nm TiO_y layer with defect ratio of ~0.23 and ~0.17, respectively. This created active interface between the two oxide layers in which the exchange of oxygen content occurred under external electric field. Gradual potentiation (4 MV cm^{-1}, 10 ms) and depression (-2 MVcm^{-1},10 ms) were obtained with dynamic ratio of ~10. The excellent device characteristics enabled its implementation on weight change, STDP, and STDP triple model.

Engineering at the interface of the between the electrode and the active TiO_x switching layer have been reported to successfully improve the dynamic ratio as well as reduce the switching current of the TiO_x-based devices. Insertion of thin Al_2O_3 layer (~2 nm) at the interface between TiO_2 and TiN electrode could achieve memory window of >100 with <10 μA switching current [33]. Further improvement of switching current, i.e., down to ~1 μA was demonstrated by replacing Al_2O_3 with a-Si layer [34, 35]. a-Si played role as an oxygen gathering layer facilitating the movement of oxygen ions at the interface. The semi-insulating property of a-Si layer enabled nonlinear IV cell characteristics, which caused amplification of the energy barrier modulation leading to the large dynamic ratio of the device. The synaptic characteristics of a-Si/TiO_2-based devices were input into a simulated 3-layer ANN and the pattern recognition accuracy of the NN was tested using MNIST database. The focus on the demonstration was on investigating the effect of read noise, i.e., random telegraph noise (RTN), on the pattern recognition accuracy. a-Si/TiO_2-based device achieved much better accuracy compared to filamentary TaO_x-based devices due to lower RTN amplitude value and distribution with much less noise occurrence rate [36].

TiO_x has also been implemented on various bilayer oxide systems [37], i.e., AlO_x, TaO_x, WO_x, HfO_x, ZnO_x, and SiO_x. Different dynamic ratio and multibit capability performances were achieved with different pairing oxide with TiO_2. The most promising multilevel states property was found in AlO_y/TiO_x bilayer oxide structure, which was able to achieve non-overlapping conductance states of 6.5 bit per cell despite of less than 10 dynamic ratios. This was attributed to AlO_x property being the oxide with lowest oxygen ions mobility among the pairing oxide layers tested. Although no synaptic characteristics were especially discussed, well control conductance update can be achieved with different pulse schemes. Specific conductance level can be achieved from the same starting value using train of identical pulses or single

pulse with optimized amplitudes and durations. Together with the non-overlapping conductance states achieved, this showed a promising feature towards an ideal analog synaptic device characteristic.

The electrode engineering in TiO_x system was associated with the symmetry characteristics of the system. $TiN/TiO_x/Mo$ system was found to improve the symmetry of the system as compared to $TiN/TiO_x/Pt$ [38]. It was due to the work function difference between the corresponding two electrodes. 64 conductance levels with excellent distribution was achieved. The device potentiation exhibited more linear conductance change as compared to its depression. In order to improve the linearity of the device depression and thus achieving more symmetrical weight update, current pulse scheme was adopted. The hybrid scheme of voltage (potentiation) and current (depression) pulse was able to improve the pattern recognition accuracy by around 10%.

Tantalum Oxide (TaO_x)-based Devices

$Ta/TaO_x/TiO_2/Ti$ structure was initially proposed as 3D-integrated storage class memory [39, 40]. It has high endurance of 10^{12}, forming free, self-compliant, and self-rectifying characteristics that significantly simplify the peripheral circuit required during the operation. It works based on homogenous Schottky barrier modulation due to oxygen vacancy defects migration at Ta/TaO_x interface under external electric field. In this structure, TiO_2 layer provided diode-like effect that resulted in self-rectifying characteristics in the structure with rectification ratio of ~10^5. The switching mechanism was confirmed by simulation to accurately reproduce experimentally obtained DC and AC characteristics of the device [41]. Moreover, its synaptic characteristics, i.e., long term potentiation (LTP), long term depression (LTD), STDP and paired-pulse facilitation (PPF), have also been experimentally investigated [41, 42]. The structure exhibited non-linear gradual potentiation and depression with dynamic ratio of >2 under identical pulse scheme (LTP: +3 V/5 ms, LTD: −3 V/5 ms, and read: −1.5 V/1 ms). The training pulse duration was found to linearly scale with pulse amplitude required to maintain similar synaptic plasticity [42]. Extremely low <10 fJ per synaptic event was experimentally recorded [43]. The nonlinearity of the weight update could be improved under two different pulse schemes, i.e., state-independent unipolar pulse scheme (UPS) and bipolar pulse scheme (BPS) [42]. UPS used single pulse (positive or negative) to move the weight value up or down, while BPS utilized a pair of pulses of different polarities (positive-high, negative-low or negative-high, positive low) to run one cycle of weight update. The linearity of the weight update was improved from 0.6–0.81 (UPS) to 0.42–0.54 (BPS) with ~50% trade-off in the dynamic ratio of the weights. This device characteristics were implemented in the simulation of the training evolution of 8 × 8 binary pattern. BPS achieved ~90% accuracy, which was significantly higher than ~75% accuracy attained under UPS. This showed the importance of weight update linearity in the long run to provide more immunity to input noise.

Other than insertion of TiO_2 layer, non-filamentary TaO_x devices have also been paired with Al_2O_3 barrier layer. Different deposition techniques were used. i.e., electron beam evaporation, post-rapid thermal annealing in O_2 ambient, and ALD.

Different deposition techniques resulted in significantly different initial resistance values. When both films are deposited via e-beam evaporation (AlO_x 3 nm and TaO_x 5 nm), the initial resistance of the structure was found to be around 200 Ω, which is extremely low. This was due to the loss of oxygen content during the deposition. On the other hand, the films deposited via post-rapid thermal annealing in O_2 ambient and ALD consistently started from highly insulating state. This agrees with the two techniques tend to result in stoichiometric films. Other than RESET process required to initiate the switching operation in the non-stoichiometric structures, the switching polarity and mechanisms involved during the operation remained the same. From current-voltage characteristics, area-dependent LRS, and elemental analysis, the change in conductance during operation was attributed to the tunneling barrier modulation induced by oxygen ions migration across the whole area of TaO_x/electrode interface. With the tunable gradual SET/RESET feature of the device, LTP/LTD, PPF, and STDP were demonstrated. LTP and LTD were characterized under identical pulse scheme (50 pulses) with different pulse amplitude (4.5–5.5 V) and duration (1–100 μs). Estimated energy of 50 pJ per spike of programming pulse was recorded. Increase in linearity of conductance change was observed with the decrease in both pulse amplitude and duration in the expense of the weight dynamic ratio. Improvement in linearity and dynamic ratio could be achieved under non-identical training pulse scheme with increasing pulse amplitude (2–6 V of 100 μs pulse). Under this scheme, rough and fine tuning to achieve certain weight value from any randomly chosen weight with excellent <1% variation was also shown.

Tungsten Oxide (WO_x)-based Devices

Another extensively studied oxide structure with underlying mechanism of homogenous anions migration across device active area is WO_x. The migration of the oxygen ions enables the system to tune the interchanging role of Schottky barrier emission and tunneling as predominant conduction mechanism during the operation. In Pt/WO_x/Ti structure [44], the Schottky barrier formed at the interface of Pt/WO_x due to the higher work function of Pt compared to Ti. During the SET process in which the Pt electrode was positively biased, the oxygen ions migrated towards the Pt electrode and got accumulated at the Pt/WO_x interface. This reduced the Fermi level near the WO_x surface and at the same time decreased the Schottky barrier height between Pt and WO_x, resulting in the increase of device conductance. This specific structure was demonstrated on a flexible substrate. The synaptic properties of the device, i.e., excitatory postsynaptic current (EPSC), PPF, STP/LTP, and STDP were characterized and no performance degradation occurred under large angle bending or 100 times bending tests. EPSC property of the device was experimentally obtained through device dynamic response upon receiving 2 V, 50 ms programming pulse. Immediately after the removal of the electric field, the conductance of the device started to drop and eventually relaxed back to the initial conductance value after ~400 ms. PPF was determined through the ratio of EPSC peaks obtained by sending two identical pulses (2 V, 50 ms). The correlation of the PPF and the interval between the subsequent pulses was recorded up to 1 s. It was well fitted with double exponential function containing the initial facilitation magnitudes and characteristic

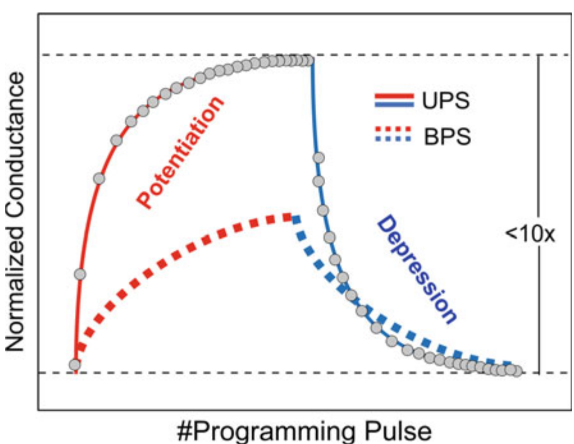


Fig. 5 Gradual potentiation and depression with more deterministic nature can be achieved in most of the reported non-filamentary anion devices. However, the excellent synaptic properties in terms of weight update linearity and symmetry are usually accompanied by low dynamic ratio, extremely low device conductance, and high switching voltage

relaxation times of the PPF. The same function can also be used to correlate the retention characteristics of the device. The extracted time constants described the transition between STP to LTP under different number of subsequent programming pulses. STDP function of the structure was characterized by sending a pair of pulses with opposite polarity (+2 V and −2 V, 50 ms) to top and bottom electrode as pre- and post- synaptic spike. The relative change in weight value was recorded under different interval of the pulse pairs.

Another structure that has been investigated as synaptic device even earlier than Pt/WO_x/Ti was Pd/WO_x/W [45, 46]. Despite of the difference in the electrodes implemented, similar homogenous switching and conduction mechanisms were obtained. However, under lower programming voltage of 1.3 V with shorter 1 ms duration, this structure was able to achieve better retention characteristics. This could be attributed to the smaller difference in electrode work functions of Pd-W as compared to Pt–Ti pair [47]. Thus, the choice of electrodes used in the structure plays critical role in determining the operating voltage and the temporal dynamics of the device.

Despite of the promising performance in terms of gradual weight update symmetry and linearity, non-filamentary anion devices tend to have high programming voltages, which might not be suitable for 1T1R integration. This currently limits the implementation of the various non-filamentary devices only on small scale neural network. Furthermore, the devices have significant trade-off between device latency and retention capability. Thus, more optimization is still required to get closer to ideal synaptic device characteristics (Fig. 5).

3 Cation-Based Synaptic Devices

Cation-based devices work based on the formation and dissolution of metallic filaments within switching layer under external electric field. These devices are

also known as conductive-bridge RAM (CBRAM) or electrochemical-metallization memory (ECM). The most commonly used active metal electrodes are Ag and Cu with electrochemically inert electrodes such as Au, Pt, and Ir. Wide variety of compounds have been investigated as switching layer, which can be classified into three major groups, i.e., solid electrolytes, oxides, and nitrides. They have been known to have promising characteristics in terms of scalability, switching speed, and programming power. In general, they also have lower operating voltage compared to their anion devices counterpart. These desirable properties are due to the high mobility of Cu and Ag ions within the switching layer. While having high ions mobility is beneficial in terms of programming speed and power, it also raises challenges in device reliability, i.e., achieving high endurance and long retention. The device failure has been reported to mainly due to excessive amount of metal species residing inside the switching host. Furthermore, it also leads to generally abrupt and stochastic switching operation. These challenges have especially been hindering the cation-based devices application as artificial synapses in NN.

Based on the amount of metal cations involved during the switching operation, the cation-based devices can be divided into two categories, i.e., infinite and finite cations source devices.

a. ***Infinite Cations Source***

Infinite cations source devices refer to devices that rely on active metal electrodes as the source of the cations to facilitate the switching operation, as depicted on Fig. 6. This configuration virtually enables infinite amounts of cations responsible for the conductance change during the device operation. In agreement with the aforementioned challenges, the amount of metal species migrating within the switching layer in this type of devices plays a critical role in the uniformity and reliability of the device, especially in obtaining multilevel conductance characteristics for analog synaptic device applications.

From the device programming viewpoint, multilevel conductance switching has been demonstrated in cation-based devices by implementing different compliance current values during the device operation [48–50]. Different compliance currents lead to different amount of active metal ions injected and different conductive filament dimensions, allowing the device to have different values of conductance. This operating scheme requires the use of a transistor to work in tandem with the RRAM device to provide a precise current control through the device. Thus, it limits the array level implementation to active array (1T1R) in which the footprint of a single synapse will be limited by the transistor size. To achieve multibit per cell capability in the device, constant drain to source voltage is required, while different voltage pulse amplitudes are implemented to allow different current level flowing through the RRAM device. This weight update scheme will require prior reading of the conductance state before moving upward or downward on the weight level. This will significantly slow down the training process and increase the amount of programming energy due to additional overhead on the network circuitry. While this architecture provides solution to achieve gradual long-term potentiation behavior during SET

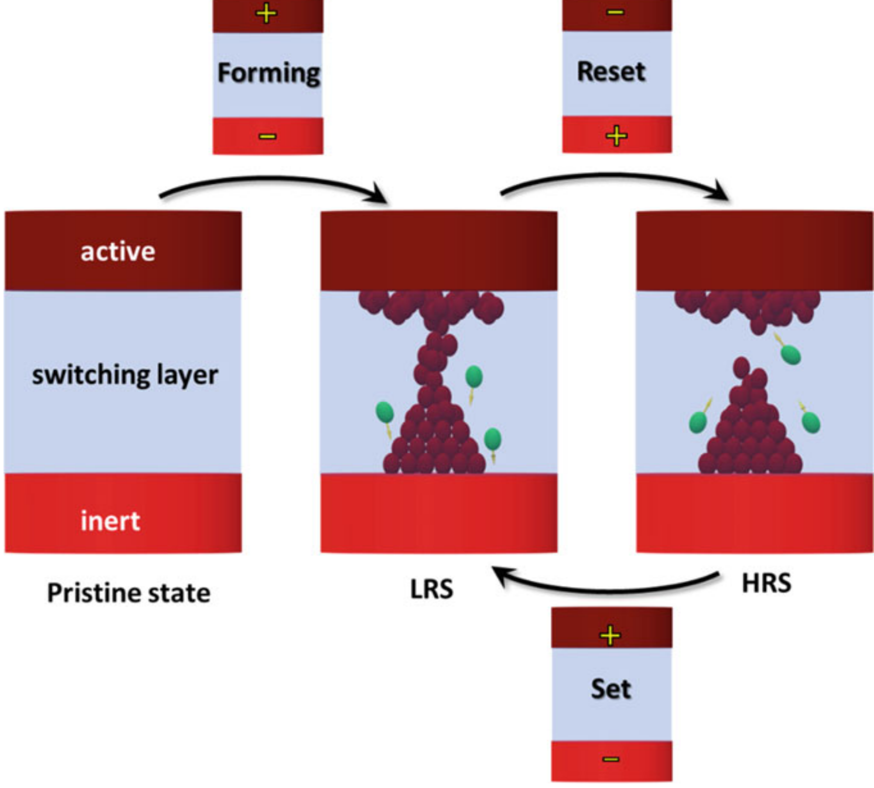


Fig. 6 The switching operation of filamentary cation-based devices with infinite cation supply. The active electrode (e.g., Ag and Cu) acts as the infinite cation source during device operation

process, emulating the same characteristics for long-term depression during RESET process remains a challenge.

Different approaches from materials design and engineering perspective have also been investigated to achieve a better control over the amount of the metal species driven under external electric field to mitigate the stochastic switching nature of the device as well as abrupt RESET process. This is extremely important towards realizing the ideal analog deterministic synapse characteristic. The first approach is done by scaling down the active device area involved during the switching operation. This can significantly reduce the amount of active metal species injected into the switching layer under external electric field. The use of plug structure to scale down the electrode to sub-20 nm area have been evidently improved the switching uniformity and reliability [50–53]. The scaling was further extended to switching layer area of the device to sub-30 nm dimension [54]. With smaller switching area, the electrochemical reaction and the movement of the active metal species becomes more restricted, which resulted in improved uniformity and data retention [54].

The second approach utilizes a thin insertion layer to either prevent unwanted oxide formation at the active electrode/oxide interface or to obtain a better control of cations injection and filament formation during device operation. Insertion of Ti at the interface of Cu/TaO_x-based devices reduced the cycle-to-cycle and device-to-device variation with significant improvement on device dynamic ratio (from ~10 to ~100) [55]. This was attributed to the formation of TiO_x instead of CuO_x at the interface of Cu/Ti/TaO_x structure. Insertion of thin TiW layer at the interface of Cu/AlO_x has also been shown to improve the overall performance of the device [56]. This barrier layer helped to maintain the cell structural integrity up to BEOL processing temperature of 400 °C. It also prevented gradual drifting of conductance states due to parasitic diffusion effects, resulted in excellent cycling control. The W\Al_2O_3\TiW\Cu RRAM cell fabricated on 90 nm W plug exhibited high voltage-disturb immunity with high dynamic ratio of >100 and fast switching operation of ~10 ns with <3 V pulse amplitude. The dynamic ratio was further enhanced to ~1000 by the insertion of WO_x by thermal oxidation of the W plug at 500 °C. It was ascribed to a filament constriction at WO_x/Al_2O_3 interface obtaining an hourglass conductive filament shape that enabled deeper RESET process. In Al/Cu/$GeSe_x$/TaO_x/W, TaO_x [57] insertion layer at the inert electrode side of the $GeSe_x$ switching layer provided an additional layer with lower Cu mobility to alter the filament shape and dimension during the switching. Improvement in switching stability was attributed to nanofilament confinement within TaO_x layer.

The third approach uses a mixture in the form of metal alloy as the source of cations. The first example is copper tellurium (Cu_xTe_{1-x}) as active ions source. It was first demonstrated on 180 nm CMOS technology in CuTe/GdO_x/W structure [49]. It was able to achieve excellent 2-bit memory property with excellent retention under different compliance current levels. Dynamic ratio of ~1000 (10 MΩ/10 kΩ) was achieved under programming parameters of 3 V, 110 μA, and 5 ns for SET and −1.7 V, 125 μA, and 1 ns for RESET. The device also showed potential of gradual RESET, but further optimization of pulse amplitude and width was required. The effect of Cu and Te composition on the active electrode was investigated in Cu_xTe_{1-x}/Al_2O_3/Si cells. It was found that the RRAM cells exhibit volatile switching (SET), non-volatile switching with gradual RESET, and non-volatile switching with abrupt RESET as the Te content decreases. This was associated with higher energy barrier to inject Cu into Al_2O_3 for Cu-Te phase compared to pure Cu. This provides a very useful insight on how the cation source characteristic is able to tune the overall RRAM cell property for specific applications.

An alternative implementation of the cation devices was proposed through stochastic STDP learning rules. Instead of trying to precisely control the switching operation to produce analog deterministic behavior, the binary probabilistic switching nature of the device is being exploited under these learning rules. This approach provides the equivalent system level functionalities to that of network utilizing the analog synaptic devices under deterministic learning rules [58]. Supervised and unsupervised NNs have been demonstrated using the binary probabilistic synapses [59, 60]. The unsupervised NN was demonstrated using 1T1R and 1R system with Ag/GST RRAM structure as synaptic device. Strong (pulse duration of ≥10 μs)

and weak (pulse duration of ≤1 μs) programming conditions were used to toggle between relatively more deterministic and probabilistic switching operation within the same device. In the strong programming condition, it was observed that high resistance state (HRS) of the device had larger distribution compared to the low resistance state (LRS), which could be attributed to uncontrollable metal filaments dissolution during reset process. With each device carries certain degree of stochasticity, in array level, this characteristic is amplified with the presence of device-to-device variation. The 1T1R and 1R synapse crossbar array was implemented in the core circuit, together with input/output CMOS neuron and pseudo-random number generator (PRNG) circuit under leaky integrate and fire (LIF) neuron model. Specific programming schemes for 1T1R and 1R were implemented to handle asynchronous analog streams of data for unsupervised pattern extraction and recognition. Excellent performance parameters were achieved for auditory pattern sensitivity (>2.5) and video detection rate (95%), while maintaining extremely low power dissipation of 0.55 and 74.2 μW for audio and video demonstrator respectively [60].

b. ***Finite Cations Source***

Another type of device has also been engineered to improve analog properties of the cation-based devices, in which a fixed amount of metal species within the switching layer is used rather than an active electrode as a source of metal ions. This approach prevents the formation of a localized conductive path during the switching operation, unlike the conventional cation-based devices. This technique was first implemented by sandwiching Ag-doped amorphous Si in between two inert electrodes [61]. The structure was fabricated using co-sputtering technique of Ag and Si to form a gradient mixture of Ag:Si across the switching layer. This results in the presence of rich and poor Ag region that can be modulated under external electric field. The structure successfully achieved gradual conductance change in both potentiation and depression process under identical programming pulse scheme. The device was also integrated with CMOS-based neuron circuits to demonstrate spike timing dependent plasticity (STDP) learning rules. The same approach was successfully adopted in Ag:TiO_x [62]. The device was able to demonstrate the learning and memory functionalities including STDP, PPF, and STP to LTP transition, with an improved timescale of hundreds of nanoseconds as compared to microseconds pulse used in Ag:Si devices.

A slightly different approach was implemented in Ag-doped WO_x [63] and TaO_x [64, 65] with uniform Ag content across the switching layer. The Ag-doped WO_x demonstrated the tunability of the device characteristics with different Ag content. Low Ag concentration within WO_x leads to volatile switching behavior that was able to mimic the forgetting effect of human memory. While, the devices with relatively higher Ag contents enable analog non-volatile switching properties. The Ag:TaO_x device was fabricated via self-doping during the sputtering process. The TaO_x layer was deposited on top of Ag electrode, resulting in the intermixing layer at the interface. The presence of the Ag:TaO_x layer at the interface of Ag and TaO_x layer caused a double switching behavior under different external electric field. The

device successfully demonstrated the key synaptic behaviors such as STP, LTP, and spike-rate-dependent plasticity (SRDP).

NN Algorithms and Architectures

In order to reach the goal of having compact system that can facilitate robust large scale NN, co-optimization from all aspects of the NN, i.e., device, circuit, and algorithm, must be thoroughly considered. With the device desired characteristics and drawbacks discussed in the previous sections, suitable algorithms and architectures must be implemented to efficiently utilize the specific device-circuit system.

As mentioned in [66], there are two ways of looking at RRAM based neuromorphic algorithms. From the deep learning perspective, one is to design algorithms for inference only, i.e., to map the pre-trained deep learning models which fulfil certain hardware constraints onto the RRAM based neuromorphic hardware without any further training. While another way is to perform on-chip training on the RRAM based neuromorphic hardware, which will require additional interface circuitry, that is usually unique to the algorithm implemented. Inference alone requires the conversion of existing pre-trained deep learning algorithms in high precision digital domain to the binary event-based (or spiking) domain to allow the mapping onto RRAM based neuromorphic hardware. Whereas, on-chip training can be implemented at the RRAM synapse in the neuromorphic hardware by emulating local spike timing-based algorithms such as spike timing dependent plasticity or its variants. These two methods belong to a computational paradigm known as spiking deep neural network (SDNN).

Other than the aforementioned learning algorithms that can be implemented on RRAM based neuromorphic hardware, low precision convolutional neural networks (CNN), such as the binarized neural network [67], binaryNet [68], XNOR-NET[69], and DoReFa-NET [70], can be mapped onto a chip containing RRAM based synaptic crossbar array [71]. In such approach, the computations performed in the CNN can be converted to bitwise operations, such as bitwise convolution, batch normalization and pooling etc. [71]. Contrary to other paradigms, mapping process is relatively simpler with such approach as it does not involve spiking neurons. Irrespective of the mapping algorithms implemented on the RRAM based neuromorphic hardware, one should expect a drop-in accuracy due to hardware noise, especially the noise inherent in RRAM synapses (Set or reset variability [72], Random Telegraph Noise (RTN) [73], etc.). One plausible approach to mitigate the drop in accuracy is to account for the noise itself during training, which may help to alleviate the accuracy loss to some extent.

4 Computation in a Crossbar Array of RRAM Synapses and Experimental Demonstration

The convolution operation in a CNN can be performed by using the crossbar array of synapses in a neuromorphic core. For an example, a 3×3 convolution kernel is

obtained through the ex situ training process of a CNN. The predetermined kernel weights are labelled as, W, as shown in Fig. 7. If the 3 × 3 kernel array to be converted into the conductance values of the synapses in a single column in RRAM crossbar array, consistent mapping rules are required. If the kernel weight can either take positive or negative values, separating these values into two different columns might be easier for the network to process, i.e., positive weights (marked in blue) and negative weights (marked in pink). Similarly, for the input, *X*, it is divided into two matrices one is positive (marked in green) and another is negated input, −*X* (marked in orange). Moreover, each of the kernel weight must be connected with the correct input node.

Once the weight matrix and input matrix values and arrangement are ready, the weight values can be programmed into the RRAM synapse conductance, i.e., positive weights occupy the top of the crossbar column while negative weights occupy the bottom part. The corresponding conductance values (σ) can be mapped from the

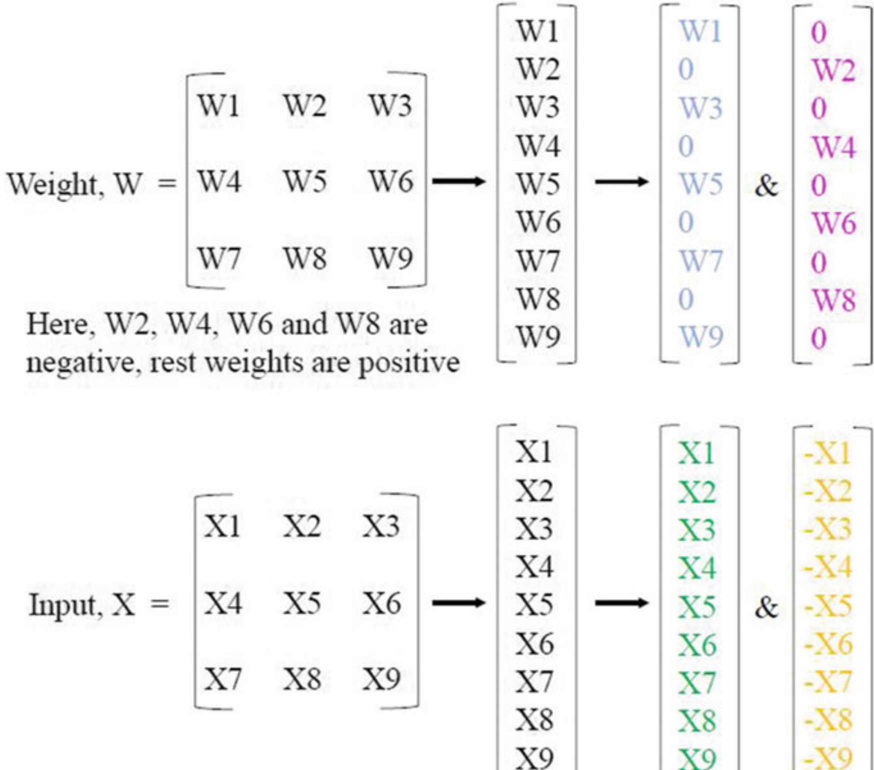


Fig. 7 The weights and input activations used in the crossbar architecture of a neuromorphic core. Different coloring corresponds to positive and negative weights and inputs. This figure is adapted from [4]

determined weights by the following equation,

$$\sigma^{\pm} = \frac{(\sigma_{max} - \sigma_{min})}{|W_{max}|} \times W^{\pm} + \sigma_{min}$$

On the other hand, the input values are also respectively fed into the axons, i.e., positive inputs are given to the top part of the axons and negative inputs to the bottom part of the axons. The convolution operation in a crossbar array between the inputs and the weights kernel is explicitly illustrated in Fig. 8 as mentioned in [74]. Each crossbar column is connected to an operational amplifier (op-amp) that can be used to implement sigmoid activation function as well as scaling up the output voltage after multiply-and-accumulate (MAC) operations on the total conductivity within one column.

One of the disadvantages of such implementation is the utilization of double the amount of input axons (2 × filter size) needed as well as double the number of RRAM

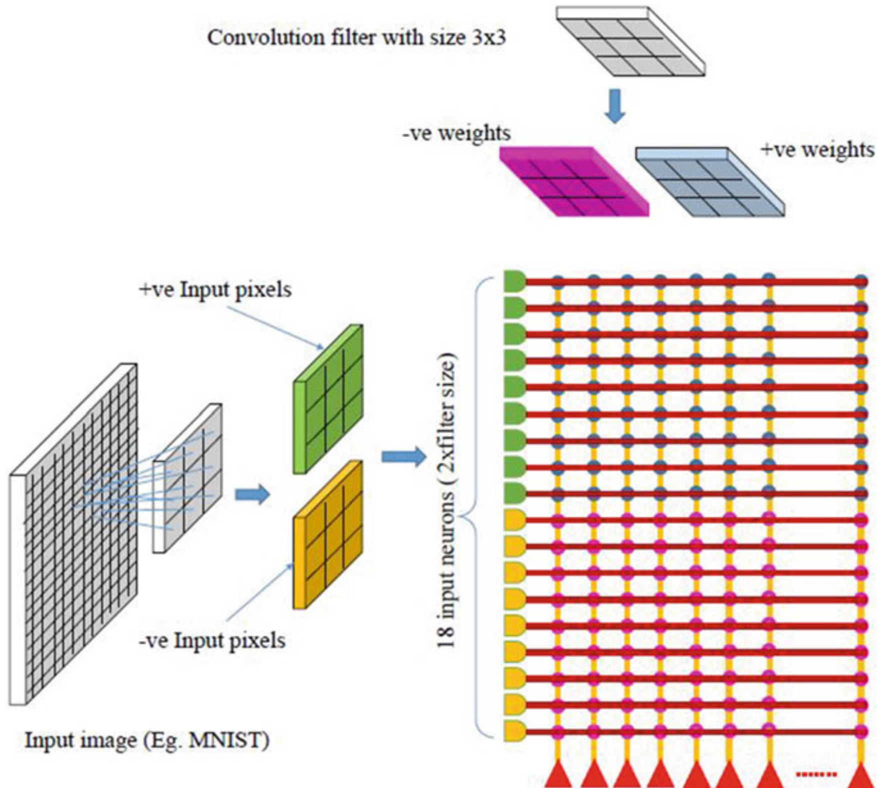


Fig. 8 Illustration of computation in a crossbar array of synapses in a neuromorphic core. This figure is adapted from [4]

synapses. Half of the RRAM synapses has to be written with low conductance state. This work is also extended to make the architecture extremely parallel by stretching the separated weight matrices as in the toeplitz matrix [75]. But, the same disadvantages of poor utilization of axons and synapses as mentioned above will remain. A slightly different approach of implementation is utilized in IBM's Truenorth chip [76]. They have only ternary weights (-1, 0, $+1$) and uses two crossbar synapses in a column as a single synapse to implement ternary weights. This will also end up using double the number of physical synapses on neuromorphic chip compared to actual number of synapses in a weight kernel. Hence, Truenorth also has a disadvantage of poor utilization of axons and synapses. Truenorth's actual physical core size, meaning number of axons $\times$ number of neurons, is 256×256, but literally their core size is only 128×256 to implement ternary weights.

MAC or weighted-sum operation is considered as one of the most tedious and yet important processes that involves heavy calculation tasks during the learning step of a neuromorphic chip. While, the RRAM crossbar array provides a promising platform to facilitate this process in high density architecture, its implementation on a large scale NN is still a rather challenging matter. During the MAC operation of each crossbar column discussed earlier, the problem raised by the inherent sneak-path current issue of the real crossbar array is negligible. However, this issue must be carefully considered during the programming of the device conductance. The neighboring cells of the selected RRAM synapse in the array are partially selected during the operation. In order to mitigate this issue while maintaining the synaptic device dimension, two-terminal select device is required. The development of the RRAM synapse alone to meet the requirements of an ideal synaptic device has been of a great challenge. This leads to an even bigger challenge to develop another select device that can compatibly work in tandem with the RRAM-based synapse (1S1R).

An alternative of the 1S1R architecture for the real crossbar implementation is the 1T1R integration. 1T1R synaptic device enables pseudo crossbar array design in which the RRAM synapse selection in the array is fully controlled by the transistor gate activation, in the expense of the device dimension. However, 1T1R also has an advantage over 1S1R synapse in terms of controlling the current flows across the RRAM synapse. This approach has been largely implemented due to the maturity of the CMOS technology. Several experimental demonstrations of the 1T1R-based neuromorphic chip have been reported [77, 75]. The first demonstration was performed by using anion-based TiN/TaO_x/$HfAl_yO_x$/TiN analog RRAM structure with 1.2 μm CMOS technology node as cell selector and word line (WL) decoder. The programming of the synaptic weights into the 1T1R cells were performed with and without write-verify scheme on 128×8 pseudo crossbar array synapses. The hardware implementation adopted single layer perceptron network to classify the image data sets into 3 different categories. It consists of 320 input neurons and 3 output neurons with the hyperbolic "tanh" as the activation function. The device conductance states were mapped to accommodate 256 different weight values (0–255). The network successfully achieved 91.67 and 87.50% classification accuracy employing programming with and without write verify scheme. The write-verify scheme required longer total programming time of 422.4 μs as compared to the

34.8 μs needed without write-verify scheme. However, the network without write-verify scheme requires 48 more iterations to converge compared to the one with the write-verify scheme, resulting in more energy consumption during the weight update phase (197.98–61.16 nJ). This shows that not only the write-verify scheme helped to mitigate the RRAM variability issue, it also reduced the overall power consumption during the weight update phase.

The second demonstration was done on cation-based Ta/HfO_2/Pt RRAM device grown on 2 μm technology node transistor. Two-layer perceptron was constructed with 128 × 64 1T1R synapse array with 64 input, 54 hidden, and 10 output neurons. The network was trained on the handwritten digits from MNIST dataset. The synaptic weight programming was done by non-identical synchronized pulse scheme to bias the transistor gate and the top or bottom electrode (TE or BE) of the RRAM cell. The potentiation was obtained by constant pulse amplitude of 500 μs on the TE with increasing gate voltage to allow different current flowing through the device. On the other hand, the depression was performed by resetting the device to the lowest conductance, by applying 5 μs synchronized pulse to the BE and the gate voltage, and then implement the conductance increase scheme used during the potentiation with decreasing transistor gate voltage. This improved the linearity and symmetry of the device weight update significantly with excellent cycle-to-cycle and device-to-device uniformity [78, 79]. Thus, the network has a promising potential to accommodate in situ training with various learning algorithm. The demonstrated network was trained by stochastic gradient descent (SGD) to execute the classification task. The network is initialized through inference by softmax function to obtain the log-probability of each output label before updating the weight within each layer for every new training dataset. The network underwent 1600 training cycles on 80,000 images from the database. It successfully achieved 91.71% classification accuracy from 10,000 images.

5 Inference on RRAM Based Neuromorphic Hardware: From Deep Neural Network (DNN) to Spike-Based Domain Architecture

a. *DNN and SNN*

Deep learning has made significant progress in recent years so much so that it has even outperformed humans in certain tasks, for instance, AlphaGo computer program managed to defeat the human GO world champion. DNN or deep CNN has achieved state-of-the-art accuracy in many image classifications or pattern recognition tasks such as handwritten digit recognition [80], several other datasets such as CIFAR [81], and ImageNet [82]. However, these networks typically need large amount of labelled training data; ImageNet has over 1 million labelled images for training.

A conventional CNN is shown in Fig. 9. It comprises of mainly three blocks: the first block is made up of convolution layers, the second of fully connected

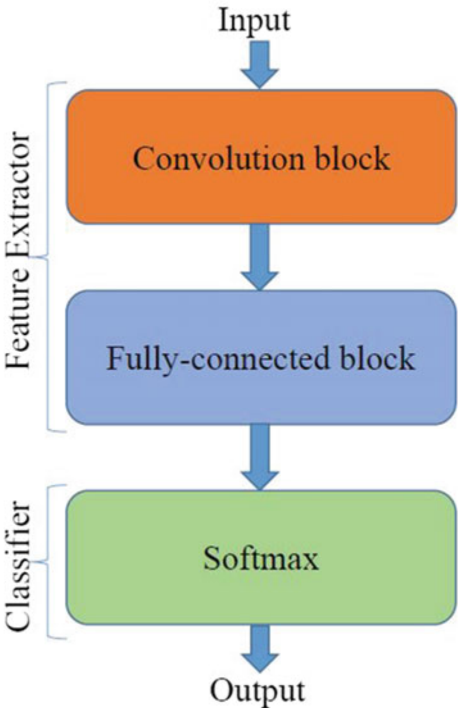


Fig. 9 Block diagram of a conventional deep neural network architecture (the convolutional neural network). It comprises of mainly three blocks: the convolutional block, the fully connected block and the softmax layer. This figure is adapted from [4]

layers and the third is the softmax layer. The convolution block contains convolution layers that perform the convolution operation on intermediate output activations. The convolution block also contains other layers that perform batch normalization or pooling. The fully connected block contains several layers of fully connected neural network. These two blocks are mainly for feature extraction. The final layer is a fully connected classifier which gives an output based on the softmax function. A typical learning algorithm used in a CNN is backpropagation of errors with stochastic gradient descent. The network parameters such as weights and biases are adjusted during training to predict the object label of an input image during testing.

Spiking neural network (SNN) is considered as the third generation of neural networks [83]. SNN is inspired by biological neural networks while the DNN less so; hence the DNN is also commonly referred to as artificial neural networks (ANN). DNN does not have any biological roots apart from the hierarchical structure it possesses [84]. SNN is event based: neural activations are communicated through spikes. Spiking neurons integrate incoming input spikes and emit a spike which is a threshold crossing event, as and when new information needs to be processed or communicated. These spikes are communicated through synapses which are associated with a weight quantity.

A neuron in an SNN and its hardware implementation is shown in Fig. 10. Figure 10a shows a single neuron (as part of an SNN) with its input and output

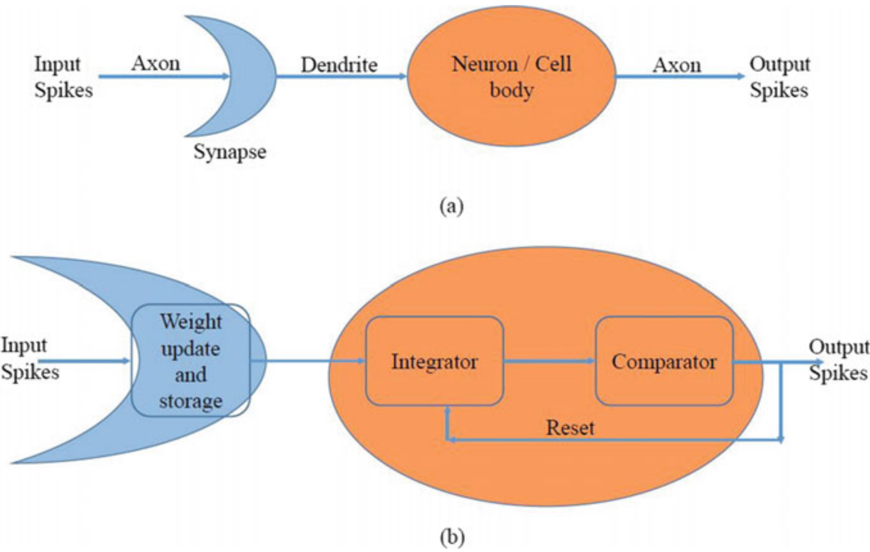


Fig. 10 Block diagram of a neuron in a SNN: implemented in blocks as shown, namely, the synapse and the neuron. This figure is adapted from [4]

mechanisms. The synapse is the connection between the axon of a pre-synaptic neuron and the dendrite of a post-synaptic neuron. A neuron integrates the incoming spikes received through its dendrites and may then emit a spike in the event of threshold crossing through its axon to its post-synaptic neurons. Figure 10b shows a block diagram representation of the biological model as in Fig. 10a. The synapse is a storage element with input spikes and output current. Neuron computation is done using an integrator and a comparator. The integrator accumulates the input currents in terms of potential difference, which emulates the membrane potential in biological neurons. The comparator then checks if the membrane potential crosses the voltage threshold; a spike is emitted if crossed and the membrane potential is then reset to its baseline value.

b. ***Conversion of DNN to Spiking Deep Neural Network (SDNN)***

In a conventional CPU or GPU, it requires more time and energy to run a SDNN, whereas the power consumption and computational latency in neuromorphic analog or digital dedicated hardwares [85–87] are orders of magnitude less. The substantial computational cost incurred during training and inference in a deep network for real world practical applications has created a need for specialized hardware acceleration and a new computational paradigm [88]. One emerging approach is to convert the pre-trained DNN into SNN (while retaining its parameters) so that it can be mapped directly onto a neuromorphic hardware with little performance loss.

The spike-based computation in the SNN consumes much less power compared to the high precision digital computation in the DNN. DNN has better classification accuracy compared to SNN. Hence, mapping a deep CNN to a SDNN potentially

allows us to achieve better accuracy with high energy efficiency. If we say, SNN corresponds to low energy networks and DNN corresponds to network with better accuracy then, spiking deep neural network (SDNN) will be Better Accuracy Low Energy (BALE) neural network. While it is difficult to achieve in a mapped SDNN the same level of accuracy as the DNN, research is ongoing to develop better mapping techniques.

SDNN background

DNN to SNN conversion techniques were developed in the ongoing research to map a trained neural network in conventional frame-based vision system representation to an event-based one [89]. Neurons in the frame-based CNN were converted to event-based neurons with leak, membrane potential reset and refractory periods.

One of the first research paper on CNN to SNN conversion is [66]. The conventional CNN is first converted into a tailored CNN which fulfils the requirements of the SNN. This tailored CNN is then trained. Finally, this tailored CNN is converted into a spiking CNN, while retaining the trained weights. The requirements imposed by the SNN on the tailored CNN are as follows:

1. Using RELU [29] as activation functions
2. Removing biases from convolution and fully connected layers
3. Using spatial linear subsampling in place of maxpooling.

The work in [66] was extended in [88] by adding weight normalization techniques to improve the conversion accuracy. The approximation errors in SNNs due to either excessive or too little spikes are avoided by rescaling of weights. Model and data-based weight normalization techniques were proposed; data-based normalization gives no loss in conversion accuracy for classification of MNIST dataset.

The integrate and fire (IF) neuron model was extensively used in SDNN until [30] demonstrated that a CNN can also be mapped onto a SDNN made up of leaky integrate and fire (LIF) neurons which are more biological plausible. This is achieved by using a modified LIF neuron known as the softened LIF neuron and by training the network with noise so as to improve network robustness against the variability inherent in spikes.

The hardware constrained neuromorphic algorithm is implemented in [76] on the IBM Truenorth neuromorphic chip. The hardware constraints are namely, low precision weights and restricted connectivity among spiking neurons.

Adapting SNN is introduced in [90], which is based on adaptive spiking neurons. Asynchronous pulsed sigma-delta coding scheme is used by these spiking neurons to efficiently encode information in spike trains, while homeostatically optimizing the firing rate. This method uses an order of magnitude less spikes compared to other SDNN approaches; the RELU neurons in an ANN could be directly mapped to adaptive spiking neurons during conversion.

General steps for conversion

The conversion of a pre-trained DNN to the event-based domain is for inference purposes. The principle of the conversion technique as mentioned in [66] is that the

time averaged firing rate of a spiking neuron must be correlated with the activation value of the corresponding neuron in the ANN. The generic steps involved for network conversion is in the following:

1. Choose a CNN to train.
2. Use ReLU for activation functions in the CNN.
3. Fix the bias to zero throughout training using stochastic gradient descent.
4. Save all the weights after training.
5. Replace neurons in the CNN with integrate and fire neurons without refractory period.
6. Map the saved weights to the SNN.
7. Convert the input image to Poisson spike trains with firing rates proportional to each pixel intensity value.

Challenging Factors and Solutions in Achieving High Conversion Accuracy

The issues affecting conversion accuracy as mentioned in [66] are: in the CNN the weights and biases can be negative. Since input integration is a weighted sum of inputs and the bias, the output can be negative. If the sigmoid function is used for activation it may also be negative. It is difficult to represent negative activations in the CNN on a SNN. It is also difficult to represent biases in the SNN. Two-layer neural network is needed to implement spatial maxpooling in the SNN.

CNN to SNN mapping requires the input image to be converted to Poisson spike trains with firing rates proportional to the pixel intensity value. As a result, the loss of accuracy during conversion can happen due to the factors [88]: Input spikes are not enough to result in threshold crossing, hence no output spike is emitted when activation values in the CNN are below threshold. If the spiking neuron receives too many input spikes in a single timestep or if some of its synaptic weights are higher than threshold, then the spiking neuron should emit more than one spike per timestep, which it cannot, and hence introducing error in the process. Due to the non-uniformity of the spike trains or the stochastic nature of the spiking input, a specific feature set could be over- or under- activated by incoming spikes.

An analysis of conversion and its theory is proposed in [91]. One on one mapping of the spiking neuron and the activation function of the CNN reveals that during threshold crossing, the membrane potential reached maybe of any value above threshold. This error would accumulate over time. The solution to the above-mentioned issues are the following:

1. As mentioned in [66], are to remove biases from convolution layers, use ReLU as activation function and use spatial linear subsampling instead of maxpooling.
2. As mentioned in [88] use weight normalization.
3. As mentioned in [91] use reset by subtraction instead of reset to zero for spiking neurons. Instead of removing biases from convolutional layers, a constant input current can be applied to emulate the biases. Also apply normalization techniques.
4. As mentioned in [90], to reduce the variability of input spikes, the multi-bit values of the input maybe fed directly into the first hidden layer and spikes are then output henceforth.

5. As mentioned in [92], pooling layers can be avoided in deep neural network. Hence, even though there are techniques to convert pooling layers in SDNN, we can remove these layers from the DNN for simplicity sake.

6 On-Chip Learning on RRAM Based Neuromorphic Hardware: Spike Based Algorithms

In the past decade, spike timing dependent plasticity (STDP) has been a popular unsupervised learning method due to its biological plausibility [93–95]. STDP mechanism depends on the timing difference between the pre-synaptic and post-synaptic spikes to adjust the synaptic weight. In the simple, doublet STDP [96], when a post-synaptic spike happens after a pre-synaptic spike has arrived (pre-post event), then the weight of the synapse increases i.e. synaptic potentiation takes place; whereas, if a post-synaptic spike happens before a pre-synaptic spike (post-pre event), then the weight of the synapse decreases, i.e. depotentiation takes place. Like the doublet STDP, there is another variant of STDP called the triplet STDP [97], whereby, three spike events are considered (pre-post-pre, post-pre-post etc.). Apart from the time based STDP there are other popular onchip learning rules based on spike rates like spike driven synaptic plasticity (SDSP) [98], Bienenstock, Cooper, and Munro (BCM) rule [99].

Over the past decade, researchers have tried to implement the above-mentioned plasticity rules onto integrated circuits [100–106]. There are CMOS devices such as the floating gate MOSFET or nano-technology devices such as the memristors, Resistive Random-Access Memories (RRAM), Phase Change Memories (PCM) and Spin-Transfer Torque Magnetic Random-Access Memories (STT-MRAMs) used for the implementation of artificial synapses. One of the challenges is to integrate these nano-technology devices with CMOS. The characteristics of high synaptic density on neuromorphic hardware has to be compromised. Interfacing circuitry or voltage generators play a major role in mimicking the plasticity behaviours mentioned above using a single two or three terminal devices. Understanding the device physics becomes the key for the implementation of artificial synapses, especially while using any of the technologies such as floating gate MOSFET, memristors or the more recent spin devices to implement plasticity rules.

7 Conclusion and Outlook

The current state-of-the-art devices have demonstrated promising characteristics to fulfil certain aspects of an ideal synaptic device. However, the presence of the trade-offs among the device properties has raised a significant challenge in the array and system level implementation, especially towards applications that require a relatively large scale NN. The importance of the gradual weight update with linear and symmetrical weight modulations has been emphasized and thoroughly discussed for

the analog synapse systems, which in general consists of different type of anion-based devices. Several approaches have been implemented to mitigate the conventional anion device limitations, i.e., conductive filament constriction engineering, an additional thermal enhanced layer, and rigorous programming pulse tuning. On the other hand, the stochastic nature of the cation-based devices has also been utilized under stochastic STDP learning rules, which enable a platform with equivalent functionalities of analog synapse systems.

Various neuromorphic algorithms on RRAM based devices have been comprehensively studied. While research is on-going to develop SNN to facilitate on-chip learning, the current reliable approach for a real-world application is to do on-chip inference based on a converted DNN that is pre-trained off-chip. During the mapping of a DNN to neuromorphic hardware, the hardware constraints, e.g., number of neurons and synapses, core size, fan in-fan out degrees, routing, and spike traffic congestion, must be taken into consideration. Should any of the above constraints are not met, the DNN architecture will have to be modified accordingly to fit into a specific neuromorphic hardware. Given its small form factor and energy efficiency, neuromorphic hardware is well suited for edge computing applications such as in the fields of robotics, surveillance, and unmanned aerial vehicles. It may also be worth investigating conversion of DNN using different encoding schemes such as temporal coding or latency coding instead of just rate coding. This would reduce the number of spikes required to represent an input and result in more efficient computing. In the long run however, hardware compatible SNN algorithms should be developed that enable on-chip learning and inference for various applications. This will eliminate the need for conversion of DNN to SNN; the challenge would be how one may improve the accuracy of such SNN algorithms.

With the aforementioned challenges have been encountered from hardware and algorithms viewpoint, more research works should adopt the approach that involves co-design and co-development of hardware, software, and middleware aspects in the system. This will allow a much more efficient utilization of the devices with certain advantages and disadvantages for specific target applications.

References

1. C. Koch, *Biophysics of computation: information processing in single neurons (computational neuroscience series)*. Oxford University Press, Inc. (2004)
2. G.W. Burr et al., Neuromorphic computing using non-volatile memory. Adv Phys **2**(1), 89–124 (2017)
3. S.B. Laughlin, T.J. Sejnowski, Communication in neuronal networks. Science **301**(5641), 1870 (2003)
4. R. Gopalakrishnan, *RRAM based neuromorphic algorithms*, *arXiv preprint* arXiv:1903.02519 (2019)
5. G. Indiveri, E. Linn, S. Ambrogio, ReRAM-based neuromorphic computing. *Resistive Switching*, pp. 715–736, 2016/06/22 (2016)

6. M. Hu et al., Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication, in *2016 53nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1–6 (2016)
7. B. Govoreanu et al., 10 × 10 nm^2 Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation, in *2011 International Electron Devices Meeting*, pp. 31.6.1–31.6.4 (2011)
8. L. Kai-Shin et al., Utilizing sub-5 nm sidewall electrode technology for atomic-scale resistive memory fabrication, in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, pp. 1–2 (2014)
9. M.J. Kim et al., Low power operating bipolar TMO ReRAM for sub 10 nm era, in *2010 International Electron Devices Meeting*, pp. 19.3.1–19.3.4 (2010)
10. H.Y. Lee et al., Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM, in *2008 IEEE International Electron Devices Meeting*, pp. 1–4 (2008)
11. M.-J. Lee et al., A fast, high-endurance and scalable non-volatile memory device made from asymmetric $Ta_2O_5 - x/TaO_2 - x$ bilayer structures. Nature Mater. **10**(8), 625–630, 2011/08/01 (2011)
12. D.J.J. Loy, P.A. Dananjaya, X.L. Hong, D.P. Shum, W.S. Lew, Conduction mechanisms on high retention annealed MgO-based resistive switching memory devices. Sci. Rep. **8**(1), p. 14774, 2018/10/03 (2018)
13. C.-Y. Lin, C.-Y. Wu, C.-Y. Wu, C. Hu, T.-Y. Tseng, Bistable resistive switching in Al_2O_3 memory thin films. J. Electrochem. Soc. **154**(9), G189–G192 (2007)
14. Y. Wu, S. Yu, B. Lee, P. Wong, Low-power TiN/Al_2O_3/Pt resistive switching device with sub-20 μA switching current and gradual resistance modulation. J. Appl. Phys. **110**(9), 094104 (2011)
15. B. Sarkar, B. Lee, V. Misra, Understanding the gradual reset in Pt/Al2O3/Ni RRAM for synaptic applications, Semicond. Sci. Technol. 30(10), 105014, 2015/08/24 (2015)
16. Y. Wu, Y. Chai, H.-Y. Chen, S. Yu, H.-S.P. Wong, Resistive switching AlO_x-based memory with CNT electrode for ultra-low switching current and high density memory application, in *2011 Symposium on VLSI Technology-Digest of Technical Papers*, pp. 26–27: IEEE (2011)
17. Y. Wu et al., AlO_x-based resistive switching device with gradual resistance modulation for neuromorphic device application, in *2012 4th IEEE International Memory Workshop*, pp. 1–4: IEEE (2012)
18. K. Park, J.-S. Lee, Reliable resistive switching memory based on oxygen-vacancy-controlled bilayer structures, *RSC Advances*. https://doi.org/10.1039/c6ra00798h 6(26), 21736–21741 (2016)
19. L. Goux et al., Understanding of the intrinsic characteristics and memory trade-offs of sub-μA filamentary RRAM operation, in *2013 Symposium on VLSI Technology*, pp. T162–T163: IEEE (2013)
20. G.C. Adam, B.D. Hoskins, M. Prezioso, F. Merrikh-Bayat, B. Chakrabarti, D.B. Strukov, 3-D memristor crossbars for analog and neuromorphic computing applications. IEEE Trans. Electron Devices **64**(1), 312–318 (2016)
21. W. Banerjee, X. Xu, H. Lv, Q. Liu, S. Long, M. Liu, Variability improvement of tio x/Al_2O_3 bilayer nonvolatile resistive switching devices by interfacial band engineering with an ultrathin Al_2O_3 dielectric material. ACS Omega **2**(10), 6888–6895 (2017)
22. K.-C. Chuang et al., Impact of the stacking order of HfOx and AlOx dielectric films on RRAM switching mechanisms to behave digital resistive switching and synaptic characteristics. IEEE J. Electr. Dev. Soc. **7**, 589–595 (2019)
23. Goux et al., Asymmetry and switching phenomenology in TiN\(Al_2O_3)\HfO_2\Hf systems. ECS Solid State Lett. **1**(4), P63–P65 (2012)
24. W. Song et al., Analog switching characteristics in TiW/Al_2O_3/Ta_2O_5/Ta RRAM devices. Appl. Phys. Lett. **115**(13), 133501 (2019)
25. J. Woo et al., Improved synaptic behavior under identical pulses using AlO x/HfO_2 bilayer RRAM array for neuromorphic systems. IEEE Electr. Dev. Lett. **37**(8), 994–997 (2016)

26. S. Yu, Y. Wu, Y. Chai, J. Provine, H.-S. P. Wong, Characterization of switching parameters and multilevel capability in HfO x/AlO x bi-layer RRAM devices, in *Proceedings of 2011 International Symposium on VLSI Technology, Systems and Applications*, pp. 1–2: IEEE (2011)
27. Y. Sun et al., A Ti/AlO x/TaO x/Pt analog synapse for memristive neural network. IEEE Electr. Dev. Lett. **39**(9), 1298–1301 (2018)
28. W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, H. Qian, Improving analog switching in HfO x-based resistive memory with a thermal enhanced layer. IEEE Electr. Dev Lett. **38**(8), 1019–1022 (2017)
29. S. R. Lee et al., Multi-level switching of triple-layered TaOx RRAM with excellent reliability for storage class memory, in *2012 Symposium on VLSI Technology (VLSIT)*, pp. 71–72: IEEE (2012)
30. S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W.D. Lu, Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. Nano Lett. **15**(3), 2203–2211 (2015)
31. J. Woo, A. Padovani, K. Moon, M. Kwak, L. Larcher, H. Hwang, Linking conductive filament properties and evolution to synaptic behavior of RRAM devices for neuromorphic applications. IEEE Electron Dev. Lett. **38**(9), 1220–1223 (2017)
32. K. Seo et al., Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device, *Nanotechnology* 22(25), p. 254023, 2011/05/16 (2011)
33. B. Govoreanu et al., Vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM): an area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell, in *2013 IEEE International Electron Devices Meeting*, pp. 10.2.1–10.2.4 (2013)
34. B. Govoreanu et al., Advanced a-VMCO resistive switching memory through inner interface engineering with wide (>102) on/off window, tunable μA-range switching current and excellent variability, in *2016 IEEE Symposium on VLSI Technology*, pp. 1–2 (2016)
35. B. Govoreanu et al., A-VMCO: A novel forming-free, self-rectifying, analog memory cell with low-current operation, nonfilamentary switching and excellent variability, in *2015 Symposium on VLSI Technology (VLSI Technology)*, pp. T132–T133 (2015)
36. Z. Chai et al., Impact of RTN on pattern recognition accuracy of RRAM-based synaptic neural network. IEEE Electr. Dev. Lett. **39**(11), 1652–1655 (2018)
37. S. Stathopoulos et al., Multibit memory operation of metal-oxide bi-layer memristors, Sci. Rep. **7**(1), 17532, 2017/12/13 (2017)
38. J. Park, M. Kwak, K. Moon, J. Woo, D. Lee, H. Hwang, TiO_x-Based RRAM synapse With 64-levels of conductance and symmetric conductance change by adopting a hybrid pulse scheme for neuromorphic computing. IEEE Electr. Dev. Lett. **37**(12), 1559–1562 (2016)
39. C. Hsu et al., Self-rectifying bipolar TaOx/TiO2 RRAM with superior endurance over 1012 cycles for 3D high-density storage-class memory, in *2013 Symposium on VLSI Technology*, pp. T166–T167 (2013)
40. C.-W. Hsu et al., Homogeneous barrier modulation of TaOx/TiO2 bilayers for ultra-high endurance three-dimensional storage-class memory, Nanotechnology **25**(16), 165202, 2014/03/25 (2014)
41. Y.-F. Wang, Y.-C. Lin, I. T. Wang, T.-P. Lin, and T.-H. Hou, “Characterization and Modeling of Nonfilamentary Ta/TaOx/TiO2/Ti Analog Synaptic Device,” *Scientific Reports,* Article vol. 5, p. 10150, 05/08/online 2015
42. I.T. Wang, C.-C. Chang, L.-W. Chiu, T. Chou, T.-H. Hou, 3D Ta/TaOx/TiO_2/Ti synaptic array and linearity tuning of weight update for hardware neural network applications. Nanotechnology **27**(36), 365204, 2016/08/02 (2016)
43. I. Wang, Y. Lin, Y. Wang, C. Hsu, and T. Hou, “3D synaptic architecture with ultralow sub-10 fJ energy per spike for neuromorphic computation,” in *2014 IEEE International Electron Devices Meeting*, 2014, pp. 28.5.1–28.5.4

44. Y. Lin et al., Transferable and flexible artificial memristive synapse based on WO_x schottky junction on arbitrary substrates. Adv. Electron. Mater. **4**(12), 1800373 (2018)
45. T. Chang, S.-H. Jo, W. Lu, Short-term memory to long-term memory transition in a nanoscale memristor. ACS Nano **5**(9), 7669–7676, 2011/09/27 (2011)
46. T. Chang, S.-H. Jo, K.-H. Kim, P. Sheridan, S. Gaba, W. Lu, Synaptic behaviors and modeling of a metal oxide memristive device. Appl. Phys. A **102**(4), 857–863, 2011/03/01 (2011)
47. S. Jabeen, M. Ismail, A. M. Rana, E. Ahmed, Impact of work function on the resistive switching characteristics of M/ZnO/CeO2/Pt devices. Mater. Res. Expr. **4**(5), 056401, 2017/05/16 (2017)
48. U. Russo, D. Kamalanathan, D. Ielmini, A.L. Lacaita, M.N. Kozicki, Study of multilevel programming in programmable metallization cell (PMC) memory. IEEE Trans. Electr. Dev. **56**(5), 1040–1047 (2009)
49. K. Aratani et al., A novel resistance memory with high scalability and nanosecond switching, in *2007 IEEE International Electron Devices Meeting* pp. 783–786 (2007)
50. M. Kund et al., Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20 nm, in *IEEE International Electron Devices Meeting, 2005. IEDM Technical Digest.* pp. 754–757 (2005)
51. S. Sills et al., A copper ReRAM cell for storage class memory applications, in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, pp. 1–2 (2014)
52. S. Yasuda et al., A cross point Cu-ReRAM with a novel OTS selector for storage class memory applications, in *2017 Symposium on VLSI Technology*, pp. T30–T31 (2017)
53. J. Guy et al., Investigation of the physical mechanisms governing data-retention in down to 10 nm nano-trench Al2O3/CuTeGe conductive bridge RAM (CBRAM), in *2013 IEEE International Electron Devices Meeting*, pp. 30.2.1–30.2.4 (2013)
54. S. Fujii et al., Scaling the CBRAM switching layer diameter to 30 nm improves cycling endurance. IEEE Electr. Dev. Lett. **39**(1), 23–26 (2018)
55. S. Z. Rahaman et al., Excellent resistive memory characteristics and switching mechanism using a Ti nanolayer at the Cu/TaO_x interface, (in eng). Nanoscale Res. Lett. **7**(1), 345–345 (2012)
56. A. Belmonte et al., 90nm $W\backslash Al_2O_3\backslash TiW\backslash Cu$ 1T1R CBRAM cell showing low-power, fast and disturb-free operation, in *2013 5th IEEE International Memory Workshop*, pp. 26–29 (2013)
57. S. Z. Rahaman et al., Impact of TaO_x nanolayer at the $GeSe_x/W$ interface on resistive switching memory performance and investigation of Cu nanofilament. J. Appl.Phys. **111**(6), 063710 (2012)
58. E.O. Neftci, B.U. Pedroni, S. Joshi, M. Al-Shedivat, G. Cauwenberghs, Stochastic synapses enable efficient brain-inspired learning machines. Front. Neurosci. **10**, 241 (2016)
59. J.H. Lee, K.K. Likharev, Defect-tolerant nanoelectronic pattern classifiers. Int. J. Circ. Theory Appl. **35**(3), 239–264 (2007)
60. M. Suri et al., Bio-inspired stochastic computing using binary CBRAM synapses. IEEE Trans. Electr. Dev. **60**(7), 2402–2409 (2013)
61. S.H. Jo, T. Chang, I. Ebong, B.B. Bhadviya, P. Mazumder, W. Lu, Nanoscale memristor device as synapse in neuromorphic systems. Nano Lett. **10**(4), 1297–1301, 2010/04/14 (2010)
62. X. Yan et al., Memristor with Ag-cluster-doped TiO_2 films as artificial synapse for neuroinspired computing. Adv. Func. Mater. **28**(1), 1705320 (2018)
63. T.D. Dongale, S.V. Mohite, A.A. Bagade, R.K. Kamat, K.Y. Rajpure, Bio-mimicking the synaptic weights, analog memory, and forgetting effect using spray deposited WO3 memristor device. Microelectr. Eng**. 183–184**, 12–18, 2017/11/05 (2017)
64. J.H. Yoon et al., Truly electroforming-free and low-energy memristors with preconditioned conductive tunneling paths. Adv. Func. Mater. **27**(35), 1702010 (2017)
65. Y. Wang et al., Self-doping memristors with equivalently synaptic ion dynamics for neuromorphic computing. ACS Appl. Mater. Interf. **11**(27), 24230–24240, 2019/07/10 (2019)
66. Y. Cao, Y. Chen, D. Khosla, Spiking deep convolutional neural networks for energy-efficient object recognition. Int. J. Comput. Vis. **113**(1), 54–66, 2015/05/01 (2015)
67. I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, Y. Bengio, "Binarized neural networks, Presented at the Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain (2016)

68. M. Courbariaux, Y. Bengio, BinaryNet: training deep neural networks with weights and activations constrained to $+1$ or -1, *CoRR,* vol. abs/1602.02830 (2016)
69. M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi, XNOR-Net: ImageNet classification using binary convolutional neural networks, in *Computer Vision—ECCV 2016*, Cham, pp. 525–542: Springer International Publishing (2016)
70. S. Zhou, Y. Wu, Z. Ni, X. Zhou, H. Wen, Y. Zou, DoReFa-Net: training low bitwidth convolutional neural networks with low bitwidth gradients, *arXiv preprint* arXiv:1606.06160 (2016)
71. L. Ni, Z. Liu, H. Yu, R.V. Joshi, An Energy-Efficient Digital ReRAM-Crossbar-Based CNN With Bitwise Parallelism. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits **3**, 37–46 (2017)
72. S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, D. Ielmini, Statistical fluctuations in HfO_x resistive-switching memory: Part I—set/reset variability. IEEE Trans. Electr. Dev. **61**(8), 2912–2919 (2014)
73. S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, D. Ielmini, Statistical fluctuations in HfOx resistive-switching memory: part II—random telegraph noise. IEEE Trans. Electr. Dev. **61**(8), 2920–2927 (2014)
74. C. Yakopcic, M.Z. Alom, T.M. Taha, Memristor crossbar deep network implementation based on a convolutional neural network, in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 963–970 (2016)
75. C. Yakopcic, M. Z. Alom, T.M. Taha, Extremely parallel memristor crossbar architecture for convolutional neural network implementation, in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1696–1703 (2017)
76. S.K. Esser et al., Convolutional networks for fast, energy-efficient neuromorphic computing. Proc. Natl. Acad. Sci. **113**(41), 11441 (2016)
77. P. Yao et al., Face classification using electronic synapses. Nat. Commun. **8**(1), 15199, 2017/05/12 (2017)
78. C. Li et al., Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. Nat. Commun. **9**(1), 2385, 2018/06/19 (2018)
79. H. Jiang et al., Sub-10 nm Ta channel responsible for superior performance of a HfO2 memristor. Sci. Rep. **6**(1), 28525, 2016/06/23 (2016)
80. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
81. A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images. Citeseer (2009)
82. J. Deng, W. Dong, R. Socher, L. Li, L. Kai, F.-F. Li, ImageNet: a large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (2009)
83. W. Maass, Networks of spiking neurons: the third generation of neural network models, Neural Netw. **10**(9), 1659–1671, 1997/12/01 (1997)
84. K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,. Biol. Cybern. **36**(4), 193–202, 1980/04/01 (1980)
85. B.V. Benjamin et al., Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. Proc. IEEE **102**(5), 699–716 (2014)
86. L.A. Plana et al., SpiNNaker: design and implementation of a GALS multicore system-on-chip. J. Emerg. Technol. Comput. Syst. **7**(4), 1–18 (2011)
87. F. Akopyan et al., TrueNorth: design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **34**(10), 1537–1557 (2015)
88. P.U. Diehl, D. Neil, J. Binas, M. Cook, S. Liu, M. Pfeiffer, Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing, in *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8 (2015)

89. J.A. Pérez-Carrasco et al., Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing-application to feedforward ConvNets. IEEE Trans. Patt. Anal. Mach. Intell. **35**(11), 2706–2719 (2013)
90. D. Zambrano, S.M. Bohte, Fast and efficient asynchronous neural computation with adapting spiking neural networks. *arXiv preprint* arXiv:1609.02053 (2016)
91. B. Rueckauer, I.-A. Lungu, Y. Hu, M. Pfeiffer, Theory and tools for the conversion of analog to spiking convolutional neural networks. *arXiv preprint* arXiv:1612.04052 (2016)
92. J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for simplicity: the all convolutional net. *arXiv preprint* arXiv:1412.6806 (2014)
93. G.-q. Bi, M.-m. Poo, Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell Type. J. Neurosci. **18**(24), 10464 (1998)
94. L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, M.-m. Poo, A critical window for cooperation and competition among developing retinotectal synapses. Nature **395**(6697), 37–44, 1998/09/01 (1998)
95. L. F. Abbott, S. B. Nelson, Synaptic plasticity: taming the beast. Nat. Neurosci. **3**(11), 1178–1183, 2000/11/01 (2000)
96. http://www.scholarpedia.org/article/Spike-timing_dependent_plasticity
97. J.-P. Pfister, W. Gerstner, Triplets of spikes in a model of spike timing-dependent plasticity. J. Neurosci. **26**(38), 9673 (2006)
98. J. M. Brader, W. Senn, S. Fusi, Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. Neural Comput. **19**(11 November 2007), 2881–2912. http://dx.doi.org/10.1162/neco.2007.19.11.2881 (2007)
99. http://www.scholarpedia.org/article/BCM_theory
100. S. Fusi, M. Annunziato, D. Badoni, A. Salamon, D. J. Amit, Spike-driven synaptic plasticity: theory, simulation, vlsi implementation. Neural Comput. **12**(10), 2227–2258, 1 Oct. 2000 (2000)
101. S. Ramakrishnan, P. Hasler, C. Gordon, Floating gate synapses with spike time dependent plasticity, in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (2010), pp. 369–372
102. R. Gopalakrishnan, A. Basu, Robust doublet STDP in a floating-gate synapse, in *2014 International Joint Conference on Neural Networks (IJCNN)* (2014), pp. 4296–4301
103. R. Gopalakrishnan, A. Basu, On the non-STDP behavior and its remedy in a floating-gate synapse. IEEE Trans. Neural Netw. Learn. Syst. **26**(10), 2596–2601 (2015)
104. M. R. Azghadi, S. Al-Sarawi, D. Abbott, N. Iannella, A neuromorphic VLSI design for spike timing and rate based synaptic plasticity. Neural Netw. **45**, 70–82 (2013)
105. R. Gopalakrishnan, A. Basu, Triplet spike time-dependent plasticity in a floating-gate synapse. IEEE Trans. Neural Netw. Learn. Syst. **28**(4), 778–790 (2015)
106. R. Gopalakrishnan, A. Basu, Triplet spike time-dependent plasticity in a floating-gate synapse. IEEE Trans. Neural Netw. Learn. Syst. **28**(4), 778–790 (2017)

An Automatic Sound Classification Framework with Non-volatile Memory

Jibin Wu, Yansong Chua, Malu Zhang, Haizhou Li, and Kay Chen Tan

Abstract Environmental sounds form part of our daily life. With the advancement of deep learning models and the abundance of training data, the performance of automatic sound classification (ASC) systems has improved significantly in recent years. However, the high computational cost, hence high power consumption, remains a major hurdle for large-scale implementation of ASC systems on mobile and wearable devices. Motivated by the observations that humans are highly effective and consume little power whilst analyzing complex audio scenes, a biologically plausible ASC framework is introduced, namely SOM-SNN. The emerging dense crossbar array of non-volatile memory (NVM) devices have been recognized as a promising approach to emulate such distributed, massively-parallel and densely connected neuromorphic computing systems. This chapter presents the general structure of this framework for sound event and speech recognition, demonstrating attractive computational benefits and suitableness with an NVM implementation.

1 Introduction

Automatic sound classification (ASC) generally refers to the automatic identification of the ambient sounds in the environment. Environmental sounds, complementary to visual cues, provide a great amount of information about our surrounding environment and is an essential part of our daily life. ASC technologies enable a wide range of applications such as content-based sound classification and retrieval [1], audio surveillance [2], sound event classification [3] and disease diagnosis [4].

J. Wu (✉) · M. Zhang · H. Li
Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore
e-mail: jibin.wu@u.nus.edu

Y. Chua
Institute for Infocomm Research, A*STAR, Singapore, Singapore

K. C. Tan
Department of Computer Science, City University of Hong Kong, Hong Kong, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_13

Driven by large-scale deep learning models, abundant training data and powerful parallel graphics processing units (GPUs), deep learning has made remarkable progress in automatic sound classification. Despite compelling performance has been demonstrated with these deep learning models [5], the required high-performance computing, which usually comes along with high power consumption, prevent the deployment of these models on the pervasive mobile and wearable devices. Furthermore, the performance of these deep learning based system degrades significantly with the increasing amount of environmental noise.

Humans perform efficiently and robustly in various auditory perception tasks, whereby spectral contents of the acoustic signal are encoded asynchronously using sparse and highly-parallel spiking impulses. Remarkably, despite the fact that spiking impulses in biological neural systems transmit at several orders of magnitude slower than the signals in modern transistors, humans are able to analyze complex audio scenes effortlessly with much lower energy consumption [6]. Moreover, humans learn to distinguish sounds with only sparse supervision, with occasional labeled data as in zero-shot or one-shot learning [7, 8].

The event-based computation, as has been observed in the human brain and sensory systems, relies on asynchronous and highly parallel spiking events to efficiently represent information. In contrast to traditional frame-based machine vision and auditory systems, event-based biological neural systems represent and process information in a much more energy efficient manner with energy consumed only during generation and transmission of spikes. Notably, neuromorphic computing has emerged with the vision to mimic such event-based biological neural systems. Neuromorphic computing leverages on low-power, densely-connected parallel computing units to perform complex perceptual and cognitive tasks; the inherent collocating memory and processing effectively address the problem of low bandwidth between the CPU and memory (i.e., von Neumann bottleneck) [9].

The emerging dense crossbar array of non-volatile memory (NVM) devices have been recognized as a promising approach to emulate such distributed, massively-parallel and densely connected neuromorphic computing systems in hardware[10–12]. The compact, low power NVM devices with continuous, near-linear conductance dynamic range are attractive for implementing synapses and neurons, instances include Phase Change Memory (PCM) and Resistive RAM (RRAM) etc. When integrated with event-based sensors, such as the DVS [13], DAVIS [14] and DAS [15], such event-based neuromorphic computing systems are attractive for real-time, adaptive and energy efficient applications [16, 17].

In this chapter, we introduce a novel neuromorphic automatic sound classification framework based on the spiking neural network (SNN). Moreover, we explain how to implement this framework with emerging NVM devices, organically integrating the compelling algorithmic power with efficient hardware for real-world applications. The rest of this chapter is organized as follows: we first describe the system architecture of the proposed ASC framework, which uses unsupervised self-organizing map (SOM) for feature representation and SNN for temporal classification, namely SOM-SNN. Subsequently, we present the experiments designed to evaluate the classification performance and robustness to noise of the proposed framework; compare

it against other state-of-the-art deep learning and SNN-based models. Finally, we conclude and discuss the computational benefits of the proposed framework as well as the considerations when using NVM devices to implement synapses and neurons in this framework.

2 SOM-SNN ASC Framework

In this section, the feedforward SOM-SNN ASC framework is described. As shown in Fig. 1, we adopt a biologically plausible auditory front-end (using logarithmic mel-scaled filter bank that resembles the functionality of the human cochlea) to first extract low-level spectral features. After which, the unsupervised self-organizing map (SOM) [18] is used to generate an effective and sparse mid-level feature representation. The best-matching units (BMUs) of the SOM are activated over time and the corresponding spatiotemporal spike patterns are generated, which represent the characteristics of each sound event. Finally, the Maximum-Margin Tempotron temporal learning rule [19] is used to train SNN so as to classify the spike patterns into different sound categories.

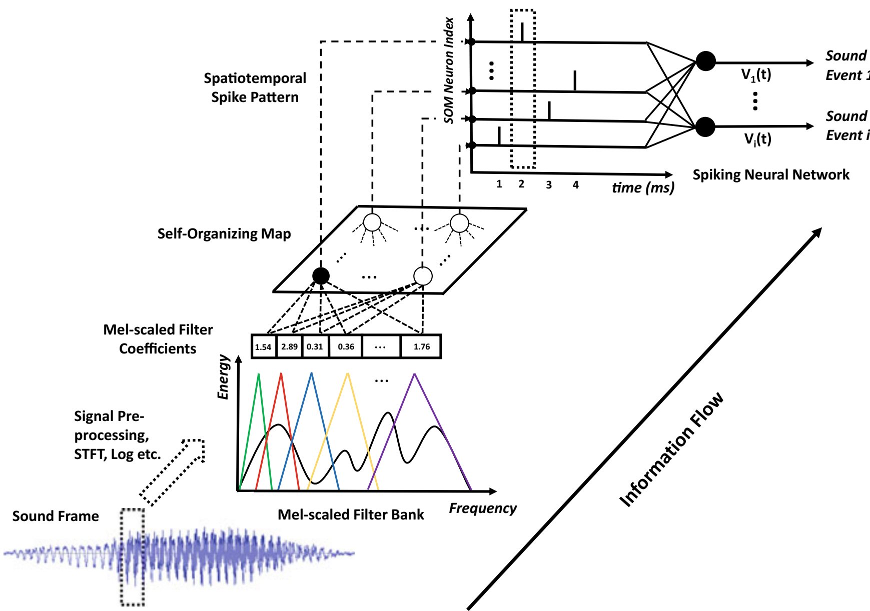


Fig. 1 The details of the proposed SOM-SNN ASC framework. The sound frames are pre-processed and analyzed using mel-scaled filter banks. Then, the SOM generates discrete BMU activation sequences which are further converted into spike trains. All such spike trains form a spatiotemporal spike pattern to be classified by the SNN

2.1 Auditory Front-End

Human auditory front-end consists of the outer, middle and inner ear. In the outer ear, sound waves travel through air and arrive at the pinna, which also embeds the location information of the sound source. From the pinna, the sound signals are then transmitted via the ear canal, which functions as a resonator, to the middle ear. In the middle ear, vibrations (induced by the sound signals) are converted into mechanical movements of the ossicles (i.e., malleus, incus, and stapes) through the tympanic membrane. The tensor tympani and stapedius muscles, which are connected to the ossicles, act as an automatic gain controller to moderate mechanical movements under the high-intensity scenario. At the end of the middle ear, the ossicles join with the cochlea via the oval window, where mechanical movements of the ossicles are transformed into fluid pressure oscillations which move along the basilar membrane in the cochlea [20].

The cochlea is a wonderful anatomical work of art. It functions as a spectrum analyzer which displaces the basilar membrane at specific locations that correspond to different frequency components in the sound wave. Finally, displacements of the basilar membrane activate inner hair cells via nearby mechanically gated ion channels, converting mechanical displacements into electrical impulse trains. The spike trains generated at the hair cells are transmitted to the cochlear nuclei through dedicated auditory nerves. Functionally, the cochlear nuclei act as filter banks, which also normalize activities of saturated auditory nerve fibers over different frequency bands. Most of the auditory nerves terminate at the cochlear nuclei where sound information is still identifiable. Beyond the cochlear nuclei, in the auditory cortex, it remains unclear how information is being represented and processed [21].

The understanding of the human auditory front-end has a significant impact on machine hearing research and inspires many biologically plausible feature representations of acoustic signals, such as the MFCC and GTCC. Here, we adopt the MFCC representation. As shown in Fig. 1, we pre-processed the sound signals by first applying pre-emphasis to amplify high-frequency contents, then segmenting the continuous sound signals into overlapped frames of suitable length so as to better capture the temporal variations of the sound signal, and finally applying the Hamming window on these frames to reduce the effect of spectral leakage. To extract the spectral contents in the acoustic stimuli, we perform Short-Time Fourier-Transform (STFT) on the sound frames and compute the power spectrum. After that, we apply 20 logarithmic mel-scaled filters on the resulting power spectrum, generating a compressed feature representation for each sound frame. The mel-scaled filter banks emulate the human perception of sound that is more discriminative towards the low frequency as compared to the high frequency components.

2.2 Feature Representation Using SOM

Feature representation is critical in all ASC systems; state-of-the-art ASC systems input low-level MFCC or GTCC features into the GMM-HMM or deep learning models so as to extract higher-level representations. In our initial experiments, we observe that existing SNN temporal learning rules cannot discriminate latency [22] or population [23] encoded mel-scaled filter bank outputs effectively. Therefore, we propose to use the biologically inspired SOM to form a mid-level feature representation of the sound frames. The neurons in the SOM form distinctive synaptic filters that organize themselves tonotopically and compete to represent the filter bank output vectors. Such tonotopically organized feature maps have been found in the human auditory cortex in many physiological experiments [24].

As shown in Fig. 1, all neurons in the SOM are fully connected to the filter bank and receive mel-scaled filter outputs (real-valued vectors). The SOM learns acoustic features in an unsupervised manner, whereby two mechanisms: competition and cooperation, guide the formation of a tonotopically organized neural map. During training, the neurons in the SOM compete with each other to best represent the input frame. The best-matching unit (BMU), with its synaptic weight vector closest to the input vector in the feature space, will update its weight vector to become closer to the input vector. Additionally, the neurons surrounding the BMU will cooperate with it by updating their weight vectors to move closer to the input vector. The magnitude of the weight update of neighboring neurons is inversely proportional to its distance to the BMU, effectively facilitating the formation of neural clusters. Eventually, the synaptic weight vectors of neurons in the SOM follow the distribution of input feature vectors and organize tonotopically, such that adjacent neurons in the SOM will have similar weight vectors.

During the evaluation, as shown in Fig. 1, the SOM (through the BMU neuron) emits a single spike at each sound frame sampling interval. The sparsely activated BMUs encourage pattern separation and enhance power efficiency. The spikes triggered over the duration of a sound event form a spatiotemporal spike pattern, which is then classified by the SNN into one of the sound classes. The mechanisms of SOM training and testing are provided in Algorithm 1 (for more details see [18]). This seminal work [18] trained the SOM for a phoneme recognition task, which then used a set of hand-crafted rules to link sound clusters of the SOM to actual phoneme classes. Here, we use an SNN-based classifier to automatically categorize the spatiotemporal spike patterns into different sound events.

2.3 *Supervised Temporal Classification*

2.3.1 Neuron Model

For the SNN-based temporal classifier, we adopt the current-based leaky integrate-and-fire neuron model [25], which utilizes the kernel function to describe the effect of pre-synaptic spikes on the membrane potential of post-synaptic neurons. When there is no incoming spike, the post-synaptic neuron i remains at its resting potential V_{rest}. Each incoming spike from the pre-synaptic neuron j at t_j will induce a post-synaptic potential (PSP) on the post-synaptic neuron as described by the following kernel function:

Algorithm 1: The Self-Organizing Map Algorithm

Input:
The randomly initialized weight vector $w_i(0)$ for neuron i = 1, ..., $M \cdot N$, where M and N are the length and width of the SOM
The training set that is formed by framewise filter bank output vectors
The initial width of the neighborhood function $\sigma(0) = \sqrt{M^2 + N^2}/2$
The number of training epochs E, initial learning rate η_0 and time constant of the time-varying width $\tau_1 = E/\log[\sigma(0)]$

Output:
The final weight vectors $w_i(E)$ for neuron $i = 1, ..., M \cdot N$

Train:
for $e \in [0, 1, 2, ..., E - 1]$ **do**

1. Randomly choose an input vector $x_{train} = [x_1, x_2, x_3, ..., x_n]$ from the training set, where n is the total number of mel-scaled filters
2. Determine the winner neuron k that has a weight vector closest to the current input vector x_{train}:

$$k = \arg\min_i ||w_i(e) - x_{train}|| \tag{1}$$

3. Update the learning rate $\eta(e)$, the time-varying width $\sigma(e)$ and the Gaussian neighborhood function $h_{i,k}(e)$ for all neurons $i = 1, ..., m$:

$$\eta(e) = \eta_0 \cdot \exp(-e/E) \tag{2}$$

$$\sigma(e) = \sigma(0) \cdot \exp(-e/\tau_1) \tag{3}$$

$$h_{i,k}(e) = \exp\{-||w_i(e) - w_k(e)||^2/[2 \cdot \sigma(e)^2]\} \tag{4}$$

4. Update $w_i(e + 1)$ for all neurons $i = 1, ..., M \cdot N$:

$$w_i(e + 1) = w_i(e) + \eta(e) \cdot h_{i,k}(e) \cdot [x_{train} - w_i(e)] \tag{5}$$

Test:
Given any input vector x_{test} from the testing set, label it with the winner neuron k that has weight vector closest to x_{test}:

$$k = \arg\min_i ||w_i(E) - x_{test}|| \tag{6}$$

$$K(t - t_j) = K_0 \left[\exp(-\frac{t - t_j}{\tau_m}) - \exp(-\frac{t - t_j}{\tau_s}) \right] \theta(t - t_j) \quad (7)$$

where K_0 is a normalization factor that ensures the maximum value of the kernel $K(t - t_j)$ is 1. τ_m and τ_s correspond to the membrane and synaptic time constants, which jointly determine the shape of the kernel function. In addition, $\theta(t - t_j)$ represents the Heaviside function to ensure that only pre-synaptic spikes emitted before time t are considered.

$$\theta(x) = \begin{cases} 1, & if\ x \geq 0 \\ 0, & otherwise \end{cases} \quad (8)$$

At time t, the membrane potential of the post-synaptic neuron i is determined by the weighted sum of all PSPs triggered by incoming spikes before time t:

$$V_i(t) = \sum_j w_{ji} \sum_{t_j < t} K(t - t_j) + V_{rest} \quad \forall t \in [0, T] \quad (9)$$

where w_{ji} is the synaptic weight between the pre-synaptic neuron j and post-synaptic neuron i, and T is the duration of the simulation. Whenever the membrane potential $V_i(t)$ of the post-synaptic neuron i reaches the firing threshold, it emits a spike. For the single-spike based classifier used in this work, the membrane potential of the post-synaptic neuron then smoothly relaxes back to V_{rest} after spiking by shunting all subsequent input spikes (i.e., input spikes arriving after the post-synaptic spike have no effect on the membrane potential of the post-synaptic neuron). Since these input spikes would not contribute to any learning in the single-spike based classifier, the unnecessary post-spike computations can be safely ignored.

2.3.2 Maximum-Margin Tempotron Learning Rule

For the classification of spatiotemporal patterns as illustrated by the SNN in Fig. 1, we use a modified version of the biologically plausible Tempotron [25] learning rule to train the classifier, which has been successfully used in several ASC tasks [19, 26–28]. The original Tempotron rule is designed for a binary classification task, such that a neuron emits a spike when it observes a spike pattern from its desired class, and remains quiescent otherwise. For a multi-class classification task, we adopt the one-against-all strategy to train one output neuron to respond to each class.

During training, for neuron i that represents the ith class, we treat all training samples with class label i as positive samples, and all others as negative. During testing, we monitor the membrane potential of all output neurons and classify the test sample as follows: (1) If no output neuron fires over the sample duration, we select the output neuron with the highest membrane potential as the correct class. (2) If only a single output neuron fires, the class label corresponding to this neuron is selected. (3) Otherwise, if two or more neurons fire, we label the test sample with the

earliest firing neuron, which signals the detection of the earliest local discriminative feature (a property of the Tempotron).

The Tempotron learning rule follows a stochastic gradient descent method for synaptic weight updates: the desired output neuron triggers a weight update whenever it fails to fire on samples with matching class label or when the wrong output neurons fire erroneously on samples from other classes. When the desired output neuron i fails to fire, long-term potentiation (LTP) update with cost function V_{thr} - $V_{t_i^{\max}}$ is triggered. Similarly, long-term depression (LTD) update with cost function $V_{t_i^{\max}}$ - V_{thr} is triggered when the wrong output neuron fires erroneously. The Tempotron update rule is defined as follows:

$$\Delta w_{ij} = \begin{cases} \lambda \sum\limits_{t_j^{(f)} < t_i^{\max}} K(t_i^{\max} - t_j^{(f)}), & if\ LTP \\ -\lambda \sum\limits_{t_j^{(f)} < t_i^{\max}} K(t_i^{\max} - t_j^{(f)}), & if\ LTD \\ 0, & otherwise \end{cases} \tag{10}$$

where λ denotes a constant learning rate and $t_i^{\max}$ refers to the time instant when the postsynaptic neuron i reaches its maximum membrane potential over the pattern duration. The $t_j^{(f)}$ are spike times of spike emitted by the pre-synaptic neuron j. The synaptic weights are only updated at the time instant of $t_i^{\max}$. For LTD weight update, the $t_i^{\max}$ is also the spike time since the post-spike computations are ignored.

Inspired by the maximum-margin classifier [19, 29], we introduce a hard margin Δ to the V_{thr} and denote this learning rule as the Maximum-Margin Tempotron. During the training phase, the Δ term is either added to or deducted from the V_{thr} of the desired or wrong output neurons, respectively. Consequently, for the desired neuron i, a spike is generated at t if

$$V_i(t) = V_{thr} + \Delta \quad and \quad \frac{d}{dt} V_i(t) > 0 \tag{11}$$

For the other (wrong) neurons, a spike is generated if

$$V_i(t) = V_{thr} - \Delta \quad and \quad \frac{d}{dt} V_i(t) > 0 \tag{12}$$

The desired output neuron will fire only when it has observed strong evidence that causes its $V_{t_{max}}$ to rise above V_{thr} by a margin of Δ. Similarly, the other neurons will be discouraged to fire and maintain its membrane potential by a margin Δ below V_{thr}. This additional margin Δ imposes a harder constraint during training and encourages the SNN classifier to find more discriminative features in the input spike patterns. Therefore, during testing, when the hard margin Δ is removed from

$V_i(t)$ as described in Equation 9 , the neurons are encouraged to respond with the desired spiking activities. This strategy helps to prevent overfitting and improves classification accuracy.

2.4 Multi-condition Training

Although state-of-the-art deep learning based ASC models perform reasonably well under the noise-free condition, it remains a challenging task for these models to recognize sound robustly in noisy real-world environments. To address this challenge, we investigated training the proposed SOM-SNN model with both clean and noisy sound data, as per the multi-condition training strategy.

The motivation for such an approach is that with training samples collected from different noisy backgrounds, the trained model will be encouraged to identify the most discriminative features and become more robust to noise. This methodology has been proven to be effective for Deep Neural Network (DNN) and SVM models under the high noise condition, with some trade-off in performance for clean sound data [30]. Here, we investigate its generalizability to SNN-based temporal classifiers under noisy environments.

3 Experimental Results

Here, we first introduce two standard benchmark datasets used to evaluate the classification accuracies of the proposed SOM-SNN framework, which are made up of environmental sounds and human speech. After which, we describe the experiments conducted on the RWCP dataset to evaluate model performance pertaining to the effectiveness of feature representation using the SOM, early decision making capability and noise robustness of the classifier.

3.1 Training and Evaluation Setup

3.1.1 Evaluation Datasets

The Real World Computing Partnership (RWCP) [31] sound scene dataset was recorded in a real acoustic environment at a sampling rate of 16 kHz. For a fair comparison with other SNN-based systems [26, 27], we used the same 10 sound event classes from the dataset: 'cymbals', 'horn', 'phone4', 'bells5', 'kara', 'bottle1', 'buzzer', 'metal15', 'whistle1', 'ring'. The sound clips were recorded as isolated samples with duration of 0.5s to 3s at high SNR. There are also short lead-in and lead-out silent intervals in the sound clips. We randomly selected 40 sound clips

from each class, of which 20 are used for training and the remaining 20 for testing, giving a total of 200 training and 200 testing samples.

The TIDIGITS [32] dataset consists of reading digit strings of varying lengths, and speech signals are sampled at 20 kHz. The TIDIGITS dataset is a publicly available dataset from the Linguistic Data Consortium, which is one of the most commonly available speech datasets used for benchmarking speech recognition algorithms. This dataset consists of spoken digit utterances from 111 male and 114 female speakers. We used all of the 12,373 continuous spoken digit utterances for the SOM training and the rest of the 4950 isolated spoken digit utterances for the SNN training and testing. Each speaker contributes two isolated spoken digit utterances for all 11 classes (i.e., ‘zeros’ to ‘nine’ and ‘oh’). We split the isolated spoken digit utterances randomly with 3950 utterances for training and the remaining 1000 utterances for testing.

3.1.2 SOM-SNN Framework

The SOM-SNN framework, as shown in Fig. 1, consists of three processing stages organized in a pipeline. These stages are trained separately and then evaluated in a single, continuous process. For the auditory front-end, we segment the continuous sound samples into frames of 100 ms length with 50 ms overlap between neighboring frames for the RWCP dataset. In contrast, we use a frame length of 25 ms with 10 ms overlap for the TIDIGITS dataset. These values are determined empirically to sufficiently discriminate the signals without excessive computational load. We utilize 20 mel-scaled filters for the spectral analysis, ranging from 200 to 8000 Hz and 200 to 10,000 Hz respectively for the RWCP and TIDIGITS datasets. The number of filters is again empirically determined, such that more filters do not improve classification accuracy.

For feature representation learning in the SOM, we utilize the SOM available in the MATLAB Neural Network Toolbox. The Euclidean distance is used to determine the BMUs, which are subsequently converted into spatiotemporal spike patterns. The output spikes from the SOM are generated per sound frame, with an interval as determined by the frame shift (i.e., 50 ms for RWCP dataset and 15 ms for TIDIGITS dataset).

We initialize the SNN by setting the threshold V_{thr}, the hard margin Δ and learning rate λ to 1.0, 0.5 and 0.005, respectively. The time constants of the SNN have determined empirically such that the PSP duration is optimal for the particular dataset, and we set τ_m to 750, 225 ms and τ_s to 187.5, 56.25 ms for the RWCP and TIDIGITS datasets, respectively. We train all the SNNs for 10 epochs by when convergence is observed. The initial weights for the neurons in the SNN classifier are drawn randomly from the Gaussian distribution with a mean of 0 and standard deviation of 10^{-3}. Parameters used in all our experiments are as above unless otherwise stated.

3.1.3 Traditional Artificial Neural Networks

To facilitate comparison with other traditional ANN models trained on the RWCP dataset, we implement four common neural network architectures, namely the Multi-Layer Perceptron (MLP) [33], the Convolutional Neural Network (CNN) [34], the Recurrent Neural Network (RNN) [35] and the Long Short-Term Memory (LSTM) [36] using the Pytorch library. For a fair comparison, we implement the MLP with 1 hidden layer of 500 ReLU units, and the CNN with two convolution layers of 128 feature maps each followed by 2 fully-connected layers of 500 and 10 ReLU units. The input frames to the MLP and CNN are concatenated over time into a spectrogram image. Since the number of frames for each sound clip varies from 20 to 100 and cannot be processed directly by the MLP or CNN, we bilinearly rescale these spectrogram images into a consistent dimension of 20×64.

We implement both the RNN and LSTM with two hidden layers containing 100 hidden units each, and a dropout layer with a probability of 0.5 is applied after the first hidden layer to prevent overfitting. The input to the RNN and LSTM are the 20-dimensional filter bank output vectors. The weights for all networks are initialized with orthogonal conditions as suggested in [37]. The deep learning networks are trained with the cross-entropy criterion and optimized using the Adam [38] optimizer. The learning rate is decayed to 99% of the original value after every epoch, and all networks are trained for 100 epochs, except for the CNN (50 epochs), by when convergence is observed. Simulations are repeated 10 times for each model, with random weight initialization.

3.1.4 Noise Robustness Evaluation

Environmental Noise We generate noise-corrupted sound samples by adding "Speech Babble" background noise from the NOISEX-92 dataset [39] to the clean RWCP sound samples. This selected background noise represents a non-stationary noisy environment with predominantly low-frequency contents, hence making a fair comparison with the noise robustness tests performed in LSF-SNN [26] and LTF-SNN models [27]. For each training or testing sound sample, a random noise segment of the same duration is selected from the noise file and added at 4 different SNR levels of 20, 10, 0 and -5 dB separately, giving a total of 1000 training and 1000 testing samples. The SNR ratio is calculated based on the energy level of each sound sample and the corresponding noise segment in our experiments. Training is performed over the whole training set, while the testing set is evaluated separately at different SNR levels.

We perform multi-condition training on all the MLP, CNN, RNN, LSTM and SOM-SNN models. Additionally, we also conduct experiments whereby the models are trained with clean sound samples but tested with noise-corrupted samples (the mismatched condition).

Neuronal Noise We also consider the effect of neuronal noise which is known to exist in the human brain, emulated by spike jittering and deletion. Given that the human auditory system is highly robust to these noises, it motivates us to investigate the performance of the proposed framework under such noisy conditions.

For spike jittering, we add Gaussian noise with zero mean and standard deviation σ to the spike timing t of all input spikes entering the SNN classifier. The amount of jitter is determined by σ which we sweep from 0.1 T to 0.8 T, where T is the spike generation period. In addition, we also consider spike deletion, where a certain fraction of spikes are corrupted by noise and not delivered to the SNN. For both types of neuronal noise, we trained the model without any noise and then tested it with jittered (of varying standard deviation σ) or deleted (of varying ratio) input spike trains.

3.2 Classification Results

3.2.1 RWCP Dataset

As shown in Table 1, the SOM-SNN model achieved a test accuracy of 99.60%, which is competitive compared with other deep learning and SNN-based models. As described in the experimental set-up, the MLP and CNN models are trained using spectrogram images of fixed dimensions, instead of explicitly modeling the temporal transition of frames. Despite their high accuracy on this dataset, it may be challenging to use them for classifying sound samples of long duration; the temporal structures will be affected inconsistently due to the necessary rescaling of the spectrogram images [40]. On the other hand, the RNN and LSTM models capture the temporal transition explicitly. These models are however hard to train for long sound samples due to the vanishing and exploding gradient problem [41].

Table 1 Comparison of the classification accuracy of the proposed SOM-SNN framework against other ANNs and SNN-based frameworks on the RWCP dataset. The average results over 10 experimental runs with random weight initialization are reported

Model	Accuracy (%)
MLP	99.45
CNN	99.85
RNN	95.35
LSTM	98.40
LSF-SNN [26]	98.50
LTF-SNN [27]	97.50
SOM-SNN (ReSuMe)	97.00
SOM-SNN (Maximum-Margin Tempotron)	99.60

LSF-SNN [26] and LTF-SNN [27] classify the sound samples by first detecting the spectral features in the power spectrogram, and then encoding these features into a spatiotemporal spike pattern for classification by a SNN classifier. In our framework, the SOM is used to learn the key features embedded in the acoustic signals in an unsupervised manner, which is more biologically plausible. Neurons in the SOM become selective to specific spectral features after training, and these features learned by the SOM are more discriminative as shown by the superior SOM-SNN classification accuracy compared with the LSF-SNN and LTF-SNN models.

3.2.2 TIDIGITS Dataset

As shown in Table 2, it is encouraging to note that the SOM-SNN framework achieves an accuracy of 97.40%, outperforming all other bio-inspired systems on the TIDIGITS dataset. In [42–44], novel systems are designed to work with spike streams generated directly from the AER silicon cochlea sensor. This event-driven auditory front-end generates spike streams asynchronously from 64 bandpass filters spanning over the audible range of the human cochlea. Anumula et al. [43] provide a comprehensive overview of the asynchronous and synchronous features generated from these raw spike streams, once again highlighting the significant role of discriminative feature representation in speech recognition tasks.

Tavanaei et al. [45, 46] proposes two biologically plausible feature extractors constructed from SNNs trained using the unsupervised spike-timing-dependent plasticity (STDP) learning rule. The neuronal activations in the feature extraction layer are then transformed into a real-valued feature vector and used to train a traditional classifier, such as the HMM or SVM models. In our work, the features are extracted using the SOM and then used to train a biologically plausible SNN classifier. These differ-

Table 2 Comparison of the classification accuracy of the proposed SOM-SNN framework against other baseline frameworks on the TIDIGITS dataset

Model	Accuracy (%)
Single-layer SNN and SVM [45][a]	91.00
Spiking CNN and HMM [46][a]	96.00
AER Silicon Cochlea and SVM [43][b]	95.58
AER Silicon Cochlea and Deep RNN [44][b]	96.10
AER Silicon Cochlea and Phased LSTM [42][b]	91.25
Liquid State Machine [47][c]	92.30
MFCC and GRU RNN [42][c]	97.90
SOM and SNN (this work)[c]	97.40

[a]Evaluate on the Aurora dataset which was developed from the TIDIGITS dataset.
[b]The data was collected by playing the audio files from the TIDIGITs dataset to the AER Silicon Cochlea Sensor.
[c]Evaluate on the TIDIGITS dataset

ent biologically inspired systems represent an important step towards an end-to-end SNN-based automatic speech recognition system.

We note that the traditional RNN based system offers a competitive accuracy of 97.90% [42]; our proposed framework, however, is fundamentally different from traditional deep learning approaches. It is worth noting that the network capacity and classification accuracy of our framework can be further improved using multi-layer SNNs.

3.3 Early Decision Making Capability

We note that the SNN-based classifier can identify temporal features within the spatiotemporal spike pattern and generate an output spike as soon as enough discriminative evidence is accumulated. This cumulative decision-making process is more biologically plausible, as it mimics how human makes decisions. A key benefit of such a decision-making process is low latency. As shown in Fig. 2a, the SNN classifier makes a decision before the whole pattern has been presented. On average, the decision is made when only 50% of the input is presented.

Additionally, we conduct experiments on the SOM-SNN, RNN and LSTM models, whereby they are trained on the full input patterns but tested with only a partial presentation of the input. The training label is provided to the RNN and LSTM models at the end of each training sequence by default as it is not clear beforehand when enough discriminative features have been accumulated. Likewise, the training labels are provided at the end of input patterns for the SNN classifier. For testing, we increase the duration of the test input pattern presented from 10 to 100% of the actual duration, starting from the beginning of each pattern. As shown in Fig. 2b, the classification accuracy as a function of the input pattern percentage increases more rapidly for the SNN model. It achieves a satisfactory accuracy of 95.1% when only 50% of the input pattern is presented, much higher than the 25.7% and 69.2% accuracy achieved by the RNN and LSTM models respectively. For the RNN and LSTM models to achieve early decision-making capability, one may require that the models be trained with partial inputs or output labels provided at every time-step. Therefore, SNN-based classifiers demonstrate great potential for real-time temporal pattern classification, compared with state-of-the-art deep learning models such as the RNN and LSTM.

3.4 Feature Representation of the SOM

To visualize the features extracted by the SOM, we plot the BMU activation sequences and their corresponding trajectories on the SOM for a set of randomly selected samples from class 'bell5', 'bottle1' and 'buzzer' in Fig. 3. We observe low intra-class

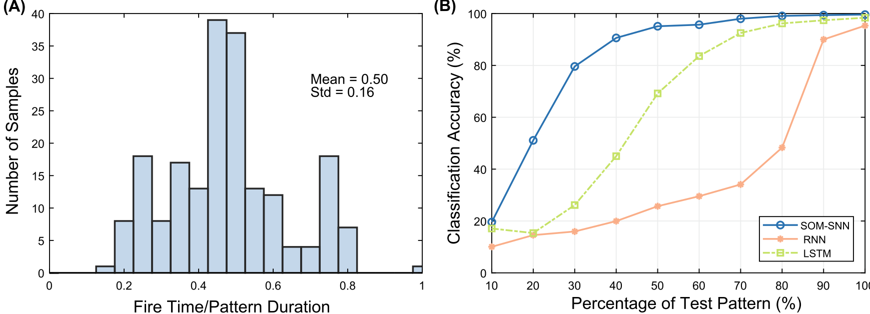


Fig. 2 The demonstration of the early decision making capability of the SNN-based classifier. **a** The distribution of the number of samples as a function of the ratio of decision time (spike timing) to sample duration on the RWCP test dataset. On average, the SNN-based classifier makes the classification decision when only 50% of the pattern is presented. **b** Test accuracy as a function of the percentage of test pattern input to different classifiers (classifiers are trained with full training patterns)

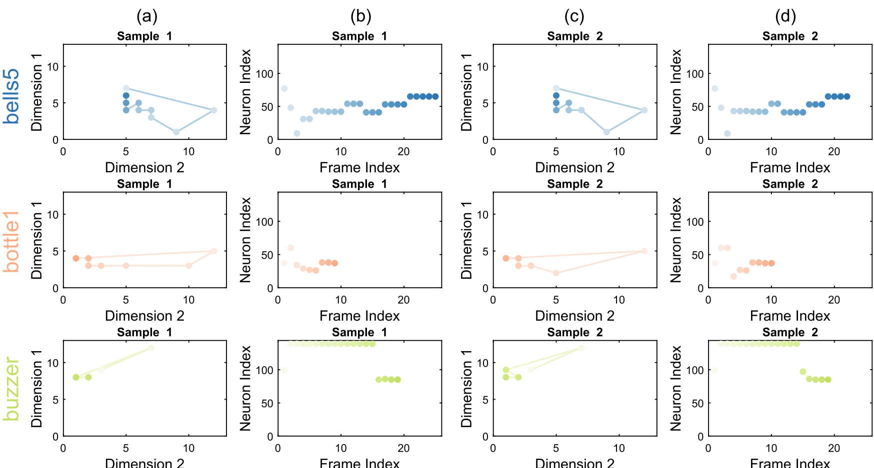


Fig. 3 BMU activation trajectories of the SOM (**a**, **c**) and BMU activation sequences (**b**, **d**) for randomly selected sound samples from classes ‘bell5’ (Row 1), ‘bottle1’ (Row 2) and ‘buzzer’ (Row 3) of a trained 12 × 12 SOM on the RWCP dataset. For BMU activation trajectories, the lines connect activated BMUs from frame to frame. The activated BMUs are highlighted from light to dark over time. For BMU activation sequences, the neurons of the SOM are enumerated along the y-axis and color matched with neurons in the BMU activation trajectories. The low intra-class variability and high inter-class variability for the BMU activation trajectories and sequences are observed

variability and high inter-class variability in both the BMU activation trajectories and sequences, which are highly desirable for pattern classification.

We note that the time-warping problem exists in the BMU activation sequences, whereby the duration of sensory stimuli fluctuates from sample to sample within the same class. However, the SNN-based classifier is robust to such fluctuations as shown in the classification results. The decision to fire for a classifying neuron is made based on a time snippet of the spiking pattern; such is the nature of the single spike-based temporal classifier. As long as the BMU activation sequence stays similar, duration fluctuations of input sample will not affect the general trajectory of the membrane potential in each output neuron; the right classification decision, therefore, can be guaranteed.

To investigate whether the feature dimension reduction of the SOM is necessary for the SNN classifier to learn different sound categories, we performed experiments that directly input the spike trains of the latency-encoded (20 neurons) [22] or population-encoded (144 neurons) [23] mel-scaled filter bank outputs into the SNN for classification. We find that the SNN classifier is unable to classify such low-level spatiotemporal spike patterns, and only achieve 10.2% and 46.5% classification accuracy for latency- and population-encoded spike patterns, respectively. For both latency- and population-encoded spike patterns, as all encoding neurons spike in every sound frame, albeit with different timing, the synaptic weights therefore either all strengthen or all weaken in the event of misclassification as defined in the Tempotron learning rule. Such synchronized weight updates make it challenging for the SNN classifier to find discriminative features embedded within the spike pattern.

The learning rules for the SNN can be categorized into either membrane-potential based [25, 48, 49] or spike-time based [50, 51]; the Maximum-Margin Tempotron learning rule belongs to the former. To study the synergy between the SOM-based feature representation and spike-time based learning rule, we conducted an experiment using the ReSuMe [50] learning rule to train the SNN classifier. For a fair comparison with the Maximum-Margin Tempotron learning rule, we use one output neuron to represent each sound class and each neuron has a single desired output spike. To determine the desired spike timing for each output neuron, we first present all training spiking patterns from the corresponding sound class to the randomly initialized SNN; and monitor the membrane potential trace of the desired output neuron during the simulation. We note the time instant when the membrane potential trace reaches its maximum (denoted as T_{max}) for each sound sample, revealing the most discriminative local temporal feature. We then use the mean of T_{max} across all 20 training samples as the desired output spike time. As shown in Table 1, the SNN trained with ReSuMe rule achieves a classification accuracy of 97.0%, which is competitive with other models. This, therefore, demonstrates the compatibility of features extracted by the SOM and spike-time based learning rules, whereby the intra-class variability of sound samples is circumvented by SOM feature extraction such that a single desired spike time for each class suffices.

3.5 *Robustness to Noise*

Environmental Noise We report the classification accuracies over 10 runs with random weight initialization in Tables 3 and 4 for mismatched and multi-condition training respectively.

We note that under the mismatched condition, the classification accuracy for all models degrades dramatically with an increasing amount of noise and falls below 50% with SNR at 10 dB. The LSF-SNN and LTF-SNN models use local key points on the spectrogram as features to represent the sound sample, and are therefore robust to noise under such conditions. However, the biological evidence for such spectrogram features is currently lacking.

As shown in Table 4, multi-condition training effectively addresses the problem of performance degradation under noisy conditions, whereby MLP, CNN, LSTM, and SOM-SNN models have achieved classification accuracies above 95% even at the challenging 0 dB SNR. Similar to observations made in [30], we note that the improved robustness to noise comes with a trade-off in terms of accuracy for clean sounds, as demonstrated in the results for the ANN models. However, the

Table 3 Average classification accuracy of different models under the mismatched-condition. Experiments are conducted over 10 runs with random weight initialization. The bold values indicate the best classification accuracies under different SNRs

SNR	MLP (%)	CNN (%)	RNN (%)	LSTM (%)	SOM-SNN (%)
Clean	99.45 ± 0.35	**99.85 ± 0.23**	95.35 ± 1.06	98.40 ± 0.86	99.60 ± 0.15
20 dB	55.05 ± 4.30	61.5 ± 4.71	25.15 ± 8.86	47.20 ± 5.36	**79.15 ± 3.70**
10 dB	32.10 ± 8.38	**42.70 ± 5.84**	11.85 ± 2.06	34.50 ± 10.61	36.25 ± 1.25
0 dB	24.60 ± 4.94	**28.40 ± 6.60**	10.10 ± 1.64	22.35 ± 6.63	26.50 ± 1.29
−5 dB	18.40 ± 4.58	**22.65 ± 5.08**	9.20 ± 1.98	16.60 ± 7.00	19.55 ± 0.16
Average	45.92	51.02	30.33	43.81	**52.21**

Table 4 Average classification accuracy of different models with multi-condition training. Experiments are conducted over 10 runs with random weight initialization. The bold values indicate the best classification accuracies under different SNRs

SNR	MLP (%)	CNN (%)	RNN (%)	LSTM (%)	SOM-SNN (%)
Clean	96.10 ± 1.18	97.60 ± 0.89	94.30 ± 3.04	98.15 ± 0.71	**99.80 ± 0.22**
20 dB	98.45 ± 0.61	99.50 ± 0.22	94.30 ± 2.70	99.10 ± 0.89	**100.00 ± 0.00**
10 dB	99.35 ± 0.45	99.70 ± 0.33	95.25 ± 2.49	99.05 ± 1.25	**100.00 ± 0.00**
0 dB	98.20 ± 1.45	99.45 ± 0.75	93.65 ± 2.82	95.80 ± 3.93	**99.45 ± 0.55**
−5 dB	92.50 ± 1.53	98.35 ± 0.78	86.85 ± 5.20	91.35 ± 4.82	**98.70 ± 0.48**
Average	96.92	98.92	92.87	96.69	**99.59**

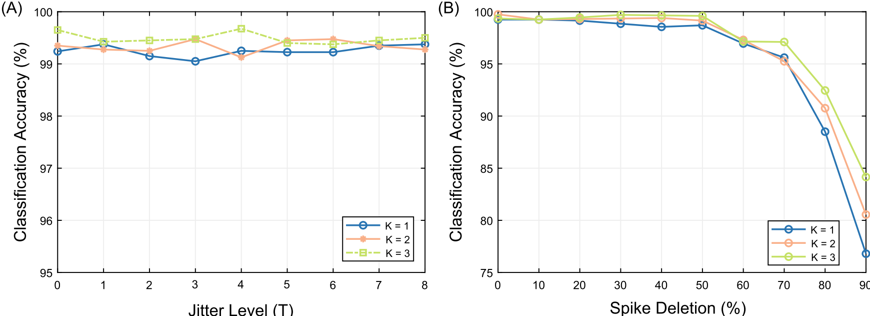


Fig. 4 The effect of spike jittering and spike deletion on the classification accuracy. **a** Classification accuracy as a result of spike jitter added at the input to the SNN classifier. The amount of jitter is added as a fraction of the spike generation period T (i.e., 50 ms used for the RWCP dataset). The classifier is robust to spike jitter, maintaining a high accuracy with different amount of jitter. **b** Classification accuracy as a result of spike deletion at the input to the SNN classifier. The accuracy of the classifier remains stable for spike deletion ratio below 60% and decays with increased spike deletion

classification accuracies improve across the board for the SOM-SNN model under all acoustic conditions using the multi-condition training, achieving an accuracy of 98.7% even for the challenging case of -5 dB SNR. The SOM-SNN model hence offers an attractive alternative to other models especially when a single trained model has to operate under varying noise levels.

Spike Jittering As shown in Fig. 4a, the SOM-SNN model is shown to be highly robust to spike jittering and maintains a high accuracy independent of the number of neurons activated per sound frame in the SOM. We suspect that given only a small subset of neurons in the SOM are involved for each sound class, the requirement of the SNN for precise spike timing is relaxed.

Spike Deletion As shown in Fig. 4b, the SOM-SNN model maintains a high classification accuracy when spike deletion is performed on the input to the SNN. As only a small subset of pre-synaptic neurons in the SOM deliver input spikes to the SNN for each sound class, with high inter-class variability, the SNN classifier is still able to classify correctly even with some input spike deletion. The peak membrane potential value is used in some cases to make the correct classification.

4 Discussion and Conclusion

In this chapter, we propose a biologically plausible SOM-SNN framework for automatic sound classification. This framework integrates the auditory front-end, feature representation learning and temporal classification in a unified framework. Biological plausibility is a key consideration in the design of this framework, which dis-

tinguishes it from many other machine learning frameworks. Here, we discuss the computational advantages of this framework and key considerations when implement this framework on the dense crossbar array of NVM devices.

4.1 Computational Benefits of SOM-SNN Framework

The SOM-SNN framework is organized in a modular manner, whereby acoustic signals are pre-processed using a biologically plausible auditory front-end, the mel-scaled filter bank, for frequency content analysis. This framework emulates the functionality of the human cochlea and the non-linearity of human perception of sound [20]. Although it is still not clear how information is represented and processed in the auditory cortex, it has been shown that certain neural populations in the cochlear nuclei and primary auditory cortex are organized in a tonotopic fashion [24, 52]. Motivated by this, the biologically plausible SOM is used for the feature extraction and representation of mel-scaled filter bank outputs. The selectivity of neurons in the SOM emerges from unsupervised training and organizes in a tonotopic fashion, whereby adjacent neurons share similar weight vectors. The SOM effectively improves pattern separation, whereby each sound frame originally represented by a 20-dimensional vector (mel-scaled filter bank output coefficients) is translated into a single output spike. The resulting BMU activation sequences are shown to have the property of low intra-class variability and high inter-class variability. Consequently, the SOM provides an effective and sparse representation of acoustic signals as observed in the auditory cortex [53].

Acoustic signals exhibit large variations not only in their frequency contents but also in temporal structures. State-of-the-art machine learning based ASC systems model the temporal transition explicitly, using the HMM, RNN or LSTM, while our work focuses on building a biologically plausible temporal classifier based on the SNN. For efficient training, we use supervised temporal learning rules, namely the membrane-potential based Maximum-Margin Tempotron and spike-timing based ReSuMe. As demonstrated in our experiments, the SOM-SNN framework achieves comparable classification results on both the RWCP and TIDIGITS datasets against other deep learning and SNN-based models.

The SNN-based classifier has an early decision making capability: making a classification decision when only part of the input is presented. In our experiments, the SNN-based classifier achieves an accuracy of 95.1%, significantly higher than those of the RNN and LSTM (25.7% and 69.2% respectively) when only 50% of the input pattern is presented. This early decision making capability can be further exploited in noisy environments, as exemplified by the cocktail party problem [54]. The SNN-based classifier can potentially identify discriminative temporal features and classify accordingly from a time snippet of the acoustic signals that are less distorted, which is desirable for an environment with fluctuating noise.

Environmental noise poses a significant challenge to the robustness of any sound classification systems: the accuracy of many such systems degrade rapidly with an

increasing amount of noise as shown in our experiments. Multi-condition training, whereby the model is trained with noise-corrupted sound samples, is shown to overcome this challenge effectively. In contrast to the DNN and SVM classifiers [30], there is no trade-off in performance for clean sounds in the SOM-SNN framework with multi-condition training; probably because the classification decision is made based on local temporal patterns. Additionally, noise is also known to exist in the central nervous system [55, 56] which can be simulated by spike jittering and deletion. Notably, the SOM-SNN framework is shown to be highly robust to such noises introduced to spike inputs arriving at the SNN classifier.

4.2 SOM-SNN Framework with Non-volatile Memory

Recent research of novel nanoscale materials with non-volatile properties opens up a compelling avenue for emulating the synaptic plasticity and state-holding properties that appeared in the real biological synapses and neurons[11, 57]. Comparing with NVM-free CMOS Neuromorphic implementations, the crossbar array of NVM devices offer high device density, low-power consumption and efficient signal integration. These novel NVM devices, therefore, would boost the algorithm development and enable the deployment of data-centric applications onto pervasive low-power wearable and Internet-of-Things devices [10].

The computational cost and memory bandwidth requirements of the proposed framework would be key concerns for neuromorphic system implementation. For the auditory front-end, our implementation is similar to that of the MFCC. As evaluated in [42], the MFCC implementation is computationally more costly compared to the spike trains generated directly from the neuromorphic cochlea sensor. Our recent work [58] proposes a novel time-domain frequency filtering scheme which addresses the cost issue in MFCC implementation. Hence, we expect the SOM to be the main computational bottleneck of the proposed framework. Nevertheless, the parallel matrix multiplication operations demanded by the SOM can be effectively implemented with arrays of analog resistive memory devices, whereby multiplication operations are naturally performed at every crosspoint following the Ohm's law. Hence, such parallel and in-memory analog computation offered by resistive memory devices demonstrate drastic time and energy saving for loading data and computation against high-performance GPUs implementations[10, 59].

The recently introduced spiking-SOM implementations [60, 61] support online training and real-time adaptation, whereby the winner neuron spikes the earliest and inhibits all other neurons from firing (i.e., a winner-takes-all mechanism) and hence by construction, the BMU. Notably, NVM devices adapt their conductance depending on the history of electrical stimuli applied, making them particularly suitable to emulate the biologically plausible local spike-time-dependent-plasticity (STDP) learning rules used in those works. The spiking-SOM also facilitates the implementation of the whole framework on neuromorphic hardware. In tandem with

the SNN classifier, a fully SNN-based framework when implemented with NVM devices would translate to significant power saving and speed up.

Besides emulating synaptic functionalities, several works also explored using NVM to emulate the neuronal dynamics, including membrane potentials maintenance, transient dynamic, spike generation and transmission [62–65]. Neuromorphic systems implemented fully with PCM have been reported in [64, 65], in which the neurons are integrated seamlessly with dense synaptic arrays using PCM. For NVM implemented synapses, each conductance state of an NVM is important and impacts the synaptic integration performed at the post-synaptic neurons. In contrast, the requirements for continuum and linearity of conductance states are less stringent when implementing leaky integrate-and-fire neurons. The biological neuron fires whenever the membrane potential crosses the firing threshold, hence, it is sufficient that accumulated electrical pulses trigger the ‘firing’ after receiving a certain number of pulses, without explicit conductance update for every electrical pulse. Furthermore, the non-volatility property of certain materials is beneficial to implement the leaky dynamic. It also worth noting that large changes on conductance state, appeared in the neuron ‘firing’ process, pose severe challenges to device endurance. While the sparse activation demonstrated with the proposed framework would greatly improve the system lifetime.

Although NVM devices offer attractive prospects for energy efficient, distributed and parallel neuromorphic computing systems, large-scale system implementation with NVM devices are still lacking. Substantial research efforts are demanded to overcome the engineering challenges, including device variability and non-ideality, power dissipation, limited conductance dynamic range and asymmetric response of conductance[10].

References

1. G. Guo, S.Z. Li, Content-based audio classification and retrieval by support vector machines. IEEE Trans. Neural Networks **14**(1), 209–215 (2003). Jan
2. A. Rabaoui, M. Davy, S. Rossignol, N. Ellouze, Using one-class SVMS and wavelets for audio surveillance. IEEE Trans. Inf. Forensics Secur. **3**(4), 763–775 (2008). Dec
3. J. Dennis, H.D. Tran, H. Li, Spectrogram image feature for sound event classification in mismatched conditions. IEEE Signal Process. Lett. **18**(2), 130–133 (2011). Feb
4. C. Kwak, O.W. Kwon, Cardiac disorder classification by heart sound signals using murmur likelihood and hidden Markov model state likelihood. IET Signal Proc. **6**(4), 326–334 (2012). June
5. R.V. Sharan, T.J. Moir, An overview of applications and advancements in automatic sound recognition. Neurocomputing **200**, 22–34 (2016)
6. P.A. Merolla, J.V. Arthur, R. Alvarez-Icaza, A.S. Cassidy, J. Sawada, F. Akopyan, B.L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S.K. Esser, R. Appuswamy, B. Taba, A. Amir, M.D. Flickner, W.P. Risk, R. Manohar, D.S. Modha, A million spiking-neuron integrated circuit with a scalable communication network and interface. Science **345**(6197), 668–673 (2014)
7. F.F. Li, R. Fergus, P. Perona, One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 594–611 (2006). April

8. M. Palatucci, D. Pomerleau, G.E. Hinton, T.M. Mitchell, Zero-shot learning with semantic output codes. Adv. Neural Inf. Process. Syst. pp. 1410–1418 (2009)
9. C.D. Schuman, T.E. Potok, R.M. Patton, J.D. Birdwell, M.E. Dean, G.S. Rose, J.S. Plank, A survey of neuromorphic computing and neural networks in hardware. arXiv preprint arXiv:1705.06963 (2017)
10. G.W. Burr, R.M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M.i Ishii, P. Narayanan, A. Fumarola, et al., Neuromorphic computing using non-volatile memory. Adv. Phys.: X **2**(1), 89–124 (2017)
11. G. Indiveri, S.-C. Liu, Memory and information processing in neuromorphic systems. Proc. IEEE **103**(8), 1379–1397 (2015)
12. M. Ziegler, Ch. Wenger, E. Chicca, H. Kohlstedt, Tutorial: Concepts for closely mimicking biological learning with memristive devices: Principles to emulate cellular forms of learning. J. Appl. Phys. **124**(15), 152003 (2018)
13. T. Delbrück B. Linares-Barranco, E. Culurciello, C. Posch, Activity-driven, event-based vision sensors, in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, May 2010, pp. 2426–2429
14. C. Brandli, R. Berner, M. Yang, S.C. Liu, T. Delbruck, A 240× 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. IEEE J. Solid-State Circuits **49**(10), 2333–2341 (2014). Oct
15. S.C. Liu, A. van Schaik, B.A. Minch, T. Delbruck, Asynchronous binaural spatial audition sensor with 2644 channel output. IEEE Trans. Biomed. Circuits Syst. **8**(4), 453–464 (2014). Aug
16. A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, D. Modha, A low power, fully event-based gesture recognition system, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 7388–7397
17. T. Serrano-Gotarredona, B. Linares-Barranco, F. Galluppi, L. Plana, S. Furber, Convnets experiments on spinnaker, in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, pp. 2405–2408
18. T. Kohonen, The self-organizing map. Neurocomputing **21**(1–3), 1–6 (1998)
19. J. Wu, Y. Chua, M. Zhang, H. Li, K.C. Tan, A spiking neural network framework for robust sound classification. Front. Neurosci. **12**, 836 (2018)
20. M. Bear, B. Connors, M. Paradiso, *Neuroscience: Exploring the brain*, 4th edn. (Wolters Kluwer, Philadelphia, 2016)
21. A. R. Møller, *Hearing: Anatomy, Physiology, and Disorders of the Auditory System*, (Plural Publishing, 2012)
22. Q. Yu, H. Tang, K.C. Tan, H. Li, Rapid feedforward computation by temporal encoding and learning with spiking neurons. IEEE Trans. Neural Networks Learn. Syst. **24**(10), 1539–1552 (2013)
23. S.M. Bohte, J.N. Kok, H. La Poutre, Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing **48**(1–4), 17–37 (2002)
24. C. Pantev, O. Bertrand, C. Eulitz, C. Verkindt, S. Hampson, G. Schuierer, T. Elbert, Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. Electroencephalogr. Clin. Neurophysiol. **94**(1), 26–40 (1995)
25. R. Gütig, H. Sompolinsky, The tempotron: a neuron that learns spike timing-based decisions. Nat. Neurosci. **9**(3), 420 (2006)
26. J. Dennis, Q. Yu, H. Tang, H.D. Tran, H. Li, Temporal coding of local spectrogram features for robust sound recognition, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 803–807
27. R. Xiao, R. Yan, H. Tang, K.C. Tan, *A Spiking Neural Network Model for Sound Recognition* (Springer, Singapore, 2017), pp. 584–594
28. J. Wu, Y. Chua, H. Li, A biologically plausible speech recognition framework based on spiking neural networks, in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–8

29. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
30. I. McLoughlin, H. Zhang, Z. Xie, Y. Song, W. Xiao, Robust sound event classification using deep neural networks. IEEE/ACM Trans. Audio, Speech, Language Process. **23**(3), 540–552 (2015). March
31. T. Nishiura, S. Nakamura, An evaluation of sound source identification with RWCP sound scene database in real acoustic environments, in *Proceedings. IEEE International Conference on Multimedia and Expo*, 2002, vol. 2, pp. 265–268
32. R.G. Leonard, G. Doddington, *Tidigits Speech Corpus* (Linguistic Data Consortium, Philadelphia, 1993)
33. N. Morgan, H. Bourlard, Continuous speech recognition using multilayer perceptrons with hidden markov models, in *1990 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, New York, 1990), pp. 413–416
34. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. pp. 1097–1105 (2012)
35. A. Graves, A. Mohamed, G.E. Hinton, Speech recognition with deep recurrent neural networks, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, New York, 2013), pp. 6645–6649
36. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
37. A.M. Saxe, J.L. McClelland, S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv preprint arXiv:1312.6120 (2013)
38. D. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
39. A. Varga, H.J. Steeneken, Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. Speech Commun. **12**(3), 247–251 (1993)
40. R. Gütig, H. Sompolinsky, Time-warp-invariant neuronal processing. PLoS Biol. **7**(7), e1000141 (2009)
41. K. Greff, R.K. Srivastava, J. Koutnk, B.R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey. IEEE Trans. Neural Networks Learn. Syst. **28**(10), 2222–2232 (2017)
42. J. Anumula, D. Neil, T. Delbruck, S.C. Liu, Feature representations for neuromorphic audio spike streams. Front. Neurosci **12**, 23 (2018)
43. M. Abdollahi, S.-C. Liu, Speaker-independent isolated digit recognition using an AER silicon cochlea, in *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Nov 2011, pp. 269–272
44. D. Neil, S.-C. Liu, Effective sensor fusion with event-based sensors and deep network architectures, in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2016, pp. 2282–2285
45. A. Tavanaei, A. Maida, Bio-inspired multi-layer spiking neural network extracts discriminative features from speech signals, in *International Conference on Neural Information Processing* (Springer, Berlin, 2017), pp. 899–908
46. A. Tavanaei, A. Maida, A spiking network that learns to extract spike signatures from speech signals. Neurocomputing **240**, 191–199 (2017)
47. Y. Zhang, P. Li, Y. Jin, Y. Choe, A digital liquid state machine with biologically inspired learning and its application to speech recognition. IEEE Trans. Neural Networks Learn. Syst. **26**(11), 2635–2649 (2015)
48. M. Zhang, H. Qu, A. Belatreche, X. Xie, EMPD: An efficient membrane potential driven supervised learning algorithm for spiking neurons. IEEE Trans. Cogn. Dev. Syst. **99**, 1–1 (2017)
49. R. Gütig, Spiking neurons can discover predictive features by aggregate-label learning. Science **351**(6277), 4113 (2016)
50. F. Ponulak, A. Kasiński, Supervised learning in spiking neural networks with Re Su Me: sequence learning, classification, and spike shifting. Neural Comput. **22**(2), 467–510 (2010)

51. Q. Yu, H. Tang, K.C. Tan, H. Li, Precise-spike-driven synaptic plasticity: Learning hetero-association of spatiotemporal spike patterns. PLoS ONE **8**(11), 1–16 (2013). Nov
52. D. Bilecen, K. Scheffler, N. Schmid, K. Tschopp, J. Seelig, Tonotopic organization of the human auditory cortex as detected by bold-FMRI. Hear. Res. **126**(1), 19–27 (1998)
53. T. Hromádka, M.R. DeWeese, A.M. Zador, Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol. **6**(1), e16 (2008)
54. S. Haykin, Z. Chen, The cocktail party problem. Neural Comput. **17**(9), 1875–1902 (2005)
55. M.C.W. Van Rossum, B.J. O'Brien, R.G. Smith, Effects of noise on the spike timing precision of retinal ganglion cells. J. Neurophysiol. **89**(5), 2406–2419 (2003)
56. E. Schneidman, *Noise and information in neural codes*, Ph.D. thesis, Hebrew University, (2001)
57. M.A. Zidan, A. Chen, G. Indiveri, W.D. Lu, Memristive computing devices and applications. J. Electroceram. **39**(1–4), 4–20 (2017)
58. Z. Pan, Y. Chua, J. Wu, H. Li, An event-based cochlear filter temporal encoding scheme for speech signals, in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–8
59. G.W. Burr, P. Narayanan, R.M. Shelby, S. Sidler, I. Boybat, C. di Nolfo, Y. Leblebici, Large-scale neural networks implemented with non-volatile memory as the synaptic weight element: Comparative performance analysis (accuracy, speed, and power), in *2015 IEEE International Electron Devices Meeting (IEDM)*, Dec 2015, p. 4.4
60. T. Rumbell, S.L. Denham, T. Wennekers, A spiking self-organizing map combining STDP, oscillations, and continuous learning. IEEE Trans. Neural Networks Learn. Syst. **25**(5), 894–907 (2014)
61. H. Hazan, D. Saunders, D.T. Sanghavi, H. Siegelmann, R. Kozma, Unsupervised learning with self-organizing spiking neural networks, in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018
62. T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, Stochastic phase-change neurons. Nat. Nanotechnol. **11**(8), 693 (2016)
63. W. Gerstner, W.M. Kistler, R. Naud, L. Paninski, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press, Cambridge, 2014)
64. T. Tuma, M. Le Gallo, A. Sebastian, E. Eleftheriou, Detecting correlations using phase-change neurons and synapses. IEEE Electron Device Lett. **37**(9), 1238–1241 (2016)
65. A. Pantazi, S. Woźniak, T. Tuma, E. Eleftheriou, All-memristive neuromorphic computing with level-tuned neurons. Nanotechnology **27**(35), 355205 (2016)

Correction to: Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions

P. K. Muduli, Raghav Sharma, Dhananjay Tiwari, Naveen Sisodia, Afshin Houshang, O. G. Heinonen, and Johan Åkerman

Correction to: Chapter "Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions" in: W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*, https://doi.org/10.1007/978-981-15-6912-8_1

The original version of the book was published without incorporating the chapter "Microwave Oscillators and Detectors Based on Magnetic Tunnel Junctions" corrections provided by the corresponding author. The correction chapter and book have been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-15-6912-8_1

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

W. S. Lew et al. (eds.), *Emerging Non-volatile Memory Technologies*,
https://doi.org/10.1007/978-981-15-6912-8_14