Modeling of Breakdown-Limited Endurance in Spin-Transfer Torque Magnetic Memory Under Pulsed Cycling Regime

Roberto Carboni, *Student Member, IEEE*, Stefano Ambrogio, *Member, IEEE*, Wei Chen, Manzar Siddik, Jon Harms, Andy Lyle, Witold Kula, Gurtej Sandhu, and Daniele Ielmini, *Senior Member, IEEE*

Abstract—**Perpendicular spin-transfer torque (p-STT) magnetic memory is gaining increasing interest as a candidate for storage-class memory, embedded memory, and possible replacement of static/dynamic memory. All these applications require extended cycling endurance, which should be based on a solid understanding and accurate modeling of the endurance failure mechanisms in the p-STT device. This paper addresses cycling endurance of p-STT memory under pulsed electrical switching. We show that endurance is limited by the dielectric breakdown of the magnetic tunnel junction stack, and we model endurance lifetime by the physical mechanisms leading to dielectric breakdown. The model predicts STT endurance as a function of applied voltage, pulsewidth, pulse polarity, and delay time between applied pulses. The dependence of the endurance on sample area is finally discussed.**

Index Terms—**Cycling endurance, magnetic tunnel junction (MTJ), reliability analysis, reliability modeling, spin-transfer torque magnetoresistive RAM (STT-MRAM).**

I. INTRODUCTION

MAGNETORESISTIVE random access memory (MRAM) is one of the most promising memory technology due to its fast switching, nonvolatile states, high endurance, CMOS compatibility, and low current operation [1]. Thanks to these characteristics, MRAM is under intense consideration for applications as storage-class memory (SCM) [2]–[5] and embedded nonvolatile

cache, which takes advantage of the nonvolatile behavior to reduce the OFF-state power consumption [6], [7]. Also, the MRAM technology and spintronics, in general, are gaining considerable interest for non-von Neumann computing architectures, such as low-power hybrid MTJ/CMOS logic circuit [8] and beyond-CMOS brain-inspired neuromorphic circuit [9].

The state-of-the-art conceptual implementation of MRAM relies on the magnetic tunnel junction (MTJ), namely a metal–insulator–metal stack consisting of a MgO dielectric barrier ($t_{MgO} \approx 1$ nm) between two CoFeB ferromagnetic electrodes. Of these two ferromagnets (FMs), the pinned layer (PL) has fixed magnetic polarization, whereas the free layer (FL) can change its polarization between parallel (P) and antiparallel (AP) with respect to the PL. The MTJ resistance is dependent on the relative orientation of the magnetic polarization in the two FMs due to the tunnel magnetoresistance (TMR) effect [10], where the P state has a relatively low resistance R_P, while the AP state has a relatively high resistance R_{AP}. Switching from P to AP and *vice versa* takes place by spin-transfer torque (STT), where the spin polarization of the electron flow across the MTJ is transferred to the FL ferromagnetic polarization by momentum conservation [11], [12]. The perpendicular STT (p-STT) concept, where the ferromagnetic polarization lies out of the MTJ plane, allows a smaller switching current at a given retention time, thus enabling low power operation and improved scalability [13].

To drive the switching current across the MTJ, bipolar voltage pulses are applied, which might induce degradation and time-dependent dielectric breakdown (TDDB) in the long term. Although the cycling endurance of STT-MRAM is generally referred to as virtually infinite [14], the repeated electrical stress during cycling induces a breakdown-limited endurance lifetime, which poses a limitation on the applicability of STT-MRAM as working memory or in-memory computing element. Despite the relevant need for high endurance, the characterization methodology, the physical understanding, and the simulation models for breakdown-limited endurance are not yet well established.

In this paper, we address the endurance of p-STT-based memory. We study endurance failure for various pulse amplitude, polarity, and pulsewidth. Then, we present a model for breakdown-limited endurance based on defect generation,
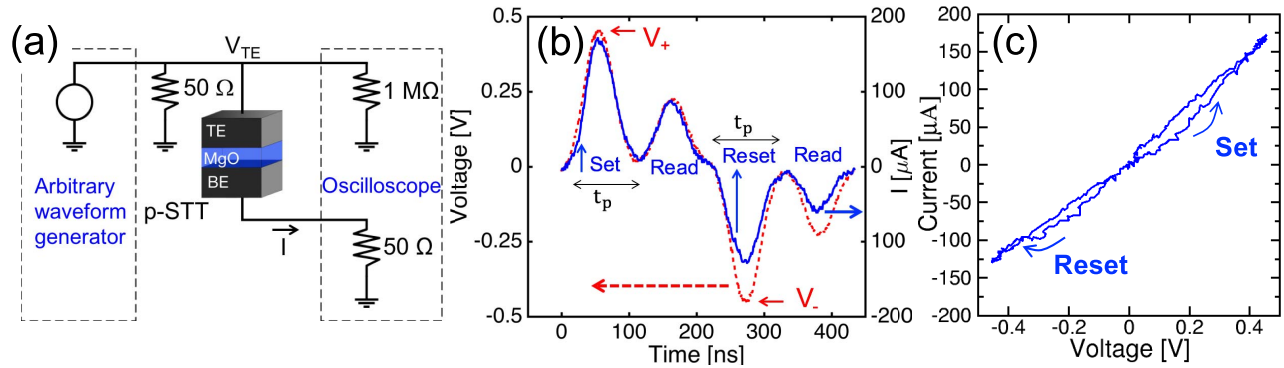
Fig. 1. (a) Experimental setup for real-time monitoring of the $I-V$ curves during ac cycling of the p-STT devices. (b) Measured waveforms of voltage and current. (c) Measured $I-V$ characteristic.

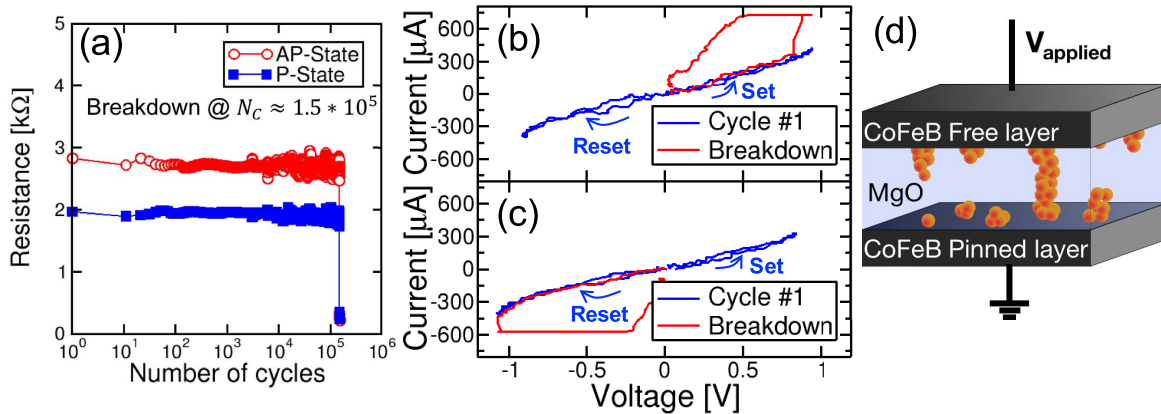Fig. 2. (a) P and AP measured resistances during cycling, showing TMR $\approx 50\%$ and endurance failure after $1.5 * 10^5$ cycles; median values over 10 reads are shown. After the MgO breakdown, the device showed a resistance of 300 Ω, corresponding to the contact resistance. Breakdown happened during (b) positive sweep if $V_+ > |V_-|$ or (c) negative sweep if $|V_-| > V_+$. The apparent current clamping is due to the oscilloscope limiting the visible range. (d) Endurance failure is attributed to the increased defect concentration in the MgO structure after the application of a stress voltage $V_{applied}$.

activation, and diffusion, capable of predicting STT-MRAM lifetime under different cycling conditions. Finally, we discuss the endurance dependence on device area.

A preliminary study of the modeling of STT-MRAM endurance was reported previously in [15]. Here, we provide a fully detailed report, with a deeper investigation of the fundamental mechanisms of defect generation/activation, a direct evidence for polarity-dependent activation, and a study of area-dependent endurance.

II. SAMPLES AND METHODOLOGY

We used p-STT memory devices sketched in Fig. 1(a), consisting of CoFeB PL [bottom electrode (BE)] and FL [top electrode (TE)] with a crystalline MgO dielectric layer. The device cross-sectional area was 47 nm × 47 nm. Fig. 1(a) also shows the experimental setup for the pulsed characterization of STT devices, including a TGA 12102 waveform generator (TTi) to apply triangular pulses for set (transition from AP to P under positive voltage) and reset (transition from P to AP under negative voltage) processes, while the applied V_{TE} voltage and current I across the MTJ were monitored by a 600-MHz LeCroy Waverunner oscilloscope. Fig. 1(b) shows a typical sequence of set, read, reset, and read operations. Each

triangular pulse had a width of $t_P = 100$ ns and a pulse delay of $t_D = 20$ ns, except where noted. The maximum positive voltage during set was V_+, while the maximum negative voltage for reset was V_-. The read current in Fig. 1(b) confirms the different states of the device, namely P state after set and AP state after reset. Fig. 1(c) shows the $I-V$ curve obtained from the collected V and I data [16]. By monitoring the $I-V$ curves at each cycle, we could observe possible degradation phenomena and the exact event of endurance failure. This technique is thus most accurate in reproducing the exact device conditions in real time, instead of unrealistic description by constant/ramped stress [17], [18]. Also, the pulsed signal of Fig. 1(b) enables a comprehensive analysis with respect to various parameters, such as voltage (V_+ and V_-) and time (t_P and t_D) parameters. All measurements were carried out at room temperature.

III. CYCLING ENDURANCE

Fig. 2(a) shows the measured resistance during a typical pulsed experiment under symmetric switching ($V_+ = |V_-|$) as in Fig. 1, as a function of the number of cycles. Data show clearly separate P and AP states with a TMR = $\Delta R / R_P \approx$ 50%, where $\Delta R = R_{AP} - R_P$. Endurance failure is marked
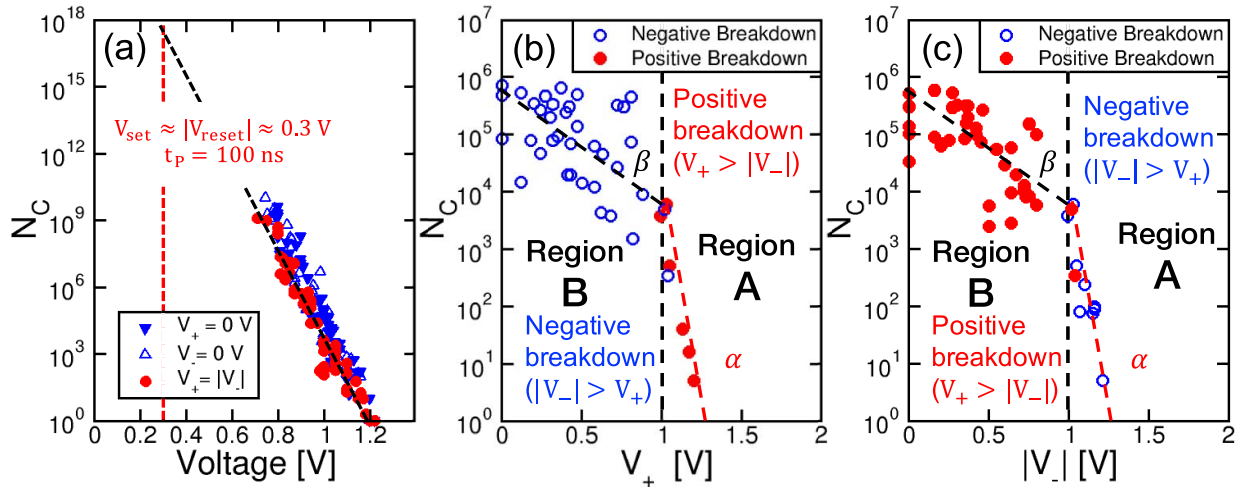
Fig. 3. (a) Number of cycles at endurance failure N_C as a function of the applied stress voltage for symmetric bipolar and for positive/negative unipolar cycling. Measured N_C for asymmetric bipolar cycling at (b) variable V_+ and constant $V_- = -1$ V and (c) variable V_- and constant $V_+ = 1$ V.
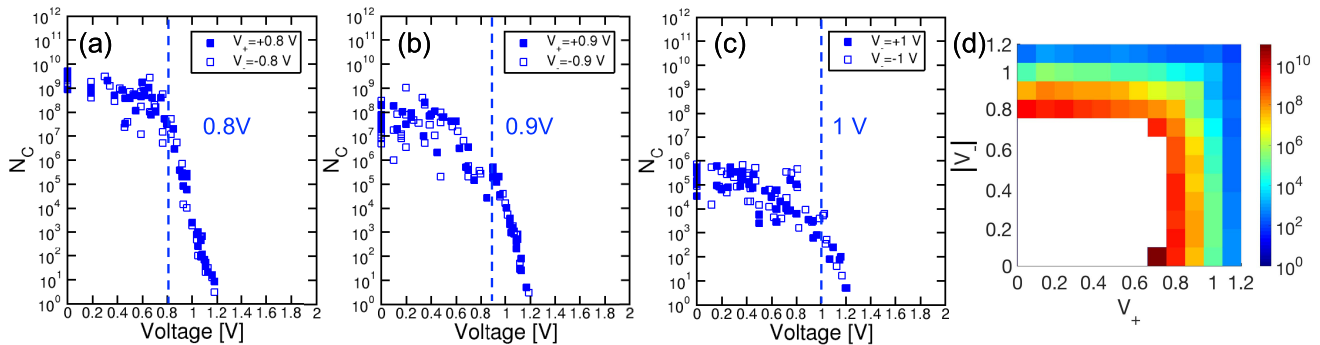

Fig. 4. Measured N_C for asymmetric bipolar cycling for an increasing constant positive/negative voltage of (a) 0.8, (b) 0.9, and (c) 1 V. A map color plot summarizes the measured N_C value as a function of V_+ and $|V_-|$ for $t_P = 100$ ns.

by an abrupt drop of resistance, corresponding to a hard breakdown of the MgO dielectric layer, after a number N_C of cycles. The resistance values R_P and R_{AP} are constant throughout the lifetime, thus indicating no obvious cycling-induced resistance degradation [19]. Also with a cycle-by-cycle observation of the $I-V$ characteristics, allowed by triangular stress pulses, no clear evidence for degradation could be found. Even though some preliminary studies [20] suggested a possible gradual dielectric breakdown for relatively thick MgO layer, in a nanometer thick tunnel barrier, an abrupt breakdown event is typically observed [19]. Breakdown could take place on either voltage polarities, e.g., breakdown during the positive sweep for $V_+ > |V_-|$ [see Fig. 2(b)] or during the negative sweep for $|V_-| > V_+$ [see Fig. 2(c)]. Breakdown can be explained by defect generation in MgO, inducing a percolative path and thermal runaway, as sketched in Fig. 2(d) [21], [22]. After breakdown, the device shows a TMR of 0% and a constant resistance R $\approx$ 300 Ω, which we attribute to the metal contacts and interfaces. No other kinds of cycling-induced failure, such as a degradation of the magnetoresistance ratio due to the cycling-induced degradation of the ferromagnetic layers, were observed, although this might be possible as a result of the thermal runaway right after dielectric break-down. The latter was always responsible for device failure,

consistently with the high electric field causing stress within the MgO barrier.

Fig. 3(a) shows the measured cycling endurance N_C as a function of the applied voltage with a pulsewidth $t_P = 100$ ns and a pulse delay $t_D = 20$ ns. Three cycling conditions are compared in Fig. 3(a), i.e., symmetric bipolar stress with $V_+ = |V_-|$, positive unipolar stress with $V_- = 0$ V, and negative unipolar stress with $V_+ = 0$ V. N_C data for positive and negative unipolar stress show similar behaviors, suggesting a high polarity symmetry of the MTJ structure with respect to degradation and breakdown processes. N_C shows a steep exponential voltage dependence with a slope $\approx$ 50 mV/decade for the three regimes in Fig. 3(a). A simple extrapolation to the switching voltage $V_{set} \approx |V_{reset}| \approx 0.3$ V [15] indicates an estimated $N_C \approx 10^{18}$ at $V = 0.3$ V and $t_P = 100$ ns, which is high enough to comply with most SCM and dynamic random access memory applications. Data indicate a higher N_C value hence reduced degradation, for unipolar stress, compared with the bipolar stress condition. This can be interpreted by considering the MgO–CoFeB interfaces to be the regions of maximum generation of stress-induced defects, and thus, unipolar stress predominantly creates damage at a single interface, whereas both interfaces are affected by bipolar stress-induced degradation.
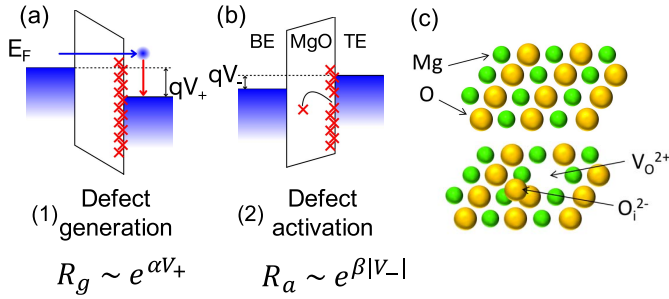
Fig. 5. Schematic of the semiempirical model of MgO breakdown, comprising (a) defect generation phase and (b) their activation. (c) Defects could be considered to be $O_i^{2-} - V_O^{2+}$ Frenkel pairs.
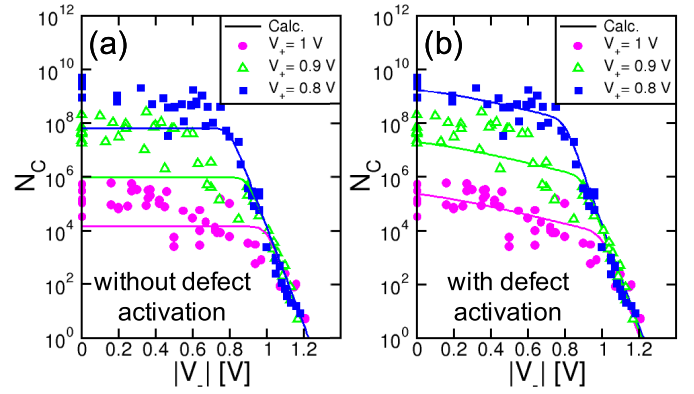

Fig. 6. (a) Measured and calculated N_C taking into account only defect generation process. (b) Calculated cycling endurance considers also defect activation process, demonstrating good agreement with experimental data.
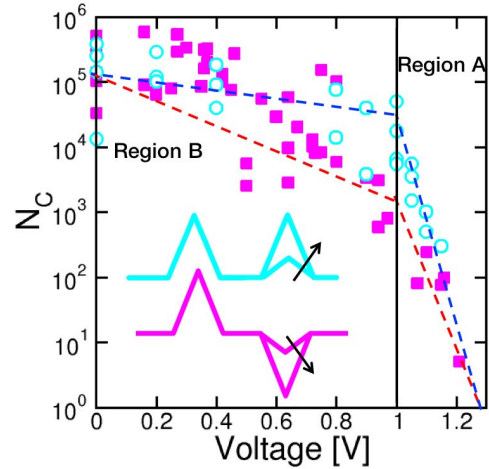

Fig. 7. Measured N_C as a function of the applied voltage for asymmetric bipolar and asymmetric unipolar stress. The different voltage dependence supports the vision where the defect activation consists in a defect displacement rather than a thermal effect.

Fig. 3(b) shows cycling endurance for asymmetric bipolar stress, with variable V_+ and constant $V_- = -1$ V. The voltage dependence of N_C shows two distinct regions, namely: 1) region A for $V_+ > |V_-|$ where data show a steep slope ≈ 50 mV/decade and positive-voltage breakdown, i.e., failure occurs during the set pulse and 2) region B for $|V_-| > V_+$ with a relatively low slope ≈ 600 mV/decade and negative-voltage breakdown, i.e., failure occurs during the reset pulse. Even though breakdown polarity is dictated by the largest applied voltage, surprisingly V_+ influences breakdown in region B, where $V_+ < |V_-|$. This remarkable evidence is further confirmed by Fig. 3(c), showing N_C for asymmetric bipolar stress with variable V_- and constant $V_+ = 1$ V and indicating the same qualitative behavior as in Fig. 3(b).

The same behavior is evidenced by Fig. 4(a)–(c), showing N_C for asymmetric bipolar stress with fixed V_+ and variable V_- or fixed V_- and variable V_+, with constant voltage equal to 0.8, 0.9, and 1 V. Note that, in each figure, N_C for fixed positive voltage and fixed negative voltage overlaps almost exactly, again supporting the high symmetry of the MTJ stack with respect to voltage stress. The presence of two distinct regions A and B is confirmed in all of the three cases. Fig. 4(d) summarizes measured N_C as a function of V_+ and $|V_-|$ in a color map plot, again confirming that the smaller voltage, e.g., $|V_-|$ for $|V_-| < V_+$ or V_+ for $V_+ < |V_-|$, and also contributes in dictating endurance lifetime.

No other input patterns were explored, e.g., a mixed unipolar/bipolar regime, although we expect that the failure mechanism would not change, and the endurance would be intermediate between the unipolar and bipolar cases.

IV. ENDURANCE MODEL

We developed a semiempirical model of endurance, which describes the dependence of N_C on the voltage amplitude and pulsewidth of the applied signal. In the model, N_C is inversely proportional to the defect concentration within the MgO layer, namely $N_C = N_{C0}(n_D/n_{D0})^{-1}$, where N_{C0} and n_{D0} are constant and n_D was calculated as $n_D = n_{D,TE} + n_{D,BE}$, where $n_{D,TE}$ and $n_{D,BE}$ are the defect concentrations originating from the TE interface and the BE interface, respectively. In this physical picture, defects are mostly generated near the interfaces where electrons have the highest kinetic energy and where the structure might display possible degradation

precursors, e.g., dangling bonds or oxygen vacancies. For example, an incomplete Mg oxidation could make unoxidized atoms to move more easily toward anode due to electromigration, thus increasing Mg/O vacancy concentration close to the interface [23]. In addition, boron (B) diffusing from the electrodes toward the tunnel barrier might initiate the creation of pinholes that might short circuit the tunnel conduction [24]. A relatively high density of initial degradation precursors plays also a key role in lowering the electron transport barrier height in MTJ [25]. Defect concentrations are given by $n_{D,TE} = n_{D0} * R_{TE}/R_0$ and $n_{D,BE} = n_{D0} * R_{BE}/R_0$, where R_{TE} and R_{BE} are the generation rates describing the cycling-induced degradation at the TE and BE interfaces, respectively, while R_0 is a constant. For our crystalline MgO layer, defects might be attributed, e.g., to Frenkel pairs of O vacancies V_O^{2+} and O interstitials O_i^{2-} [26].

As shown in Fig. 5, tunneling electrons are considered to have a primary role in MgO degradation according to a two-stage mechanism, including: 1) defect generation [see Fig. 5(a)] and 2) defect activation [see Fig. 5(b)], as detailed
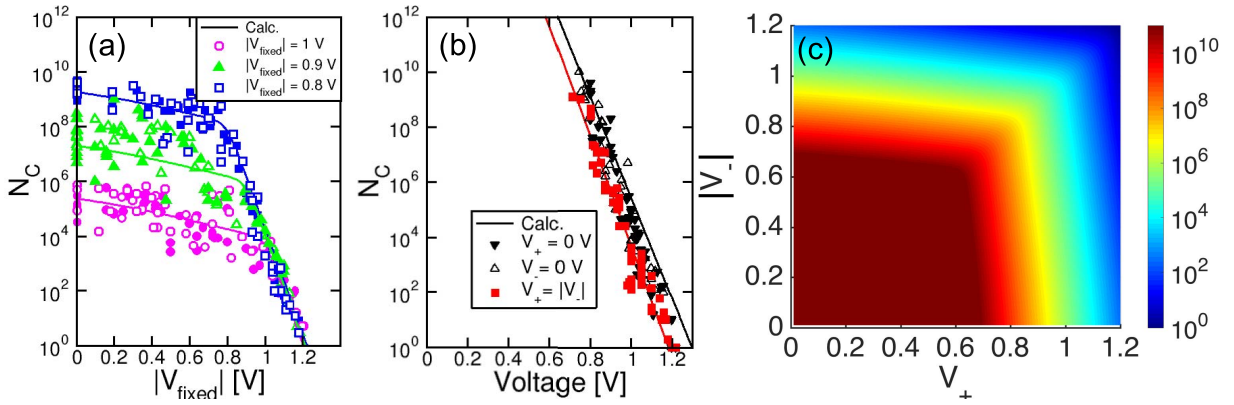
Fig. 8. Measured and calculated N_C for (a) asymmetric cycling at constant V_- and constant V_+ and (b) symmetric bipolar and unipolar cycling. (c) Color plot of N_C as a function of V_+ and $|V_-|$ for $t_P = 100$ ns, obtained from model calculations.

in the following. Even though process variability is of great importance for the STT-MRAM design [27], we did not take into account such effects given the relatively low device-to-device variation of conduction and switching among our samples. The observed variation in cycling endurance for a given voltage might result from intrinsic variability of both the position in the oxide layer and the number of generated defects.

A. Defect Generation

Fig. 5(a) shows the defect generation mechanism in our model. Electrons injected from one interface reach the other with a kinetic energy E given by the difference of the Fermi levels in the two electrodes, i.e., $E = E_{F,\mathrm{TE}} - E_{F,\mathrm{BE}} = qV_+$ for positive voltage applied to the TE and hence electrons injected from the BE. The release of the energy E induces lattice vibrations and defect generation at the TE interface by bond breaking. Even though the strong ionic bond between Mg and O is very energetic, bond breaking is possible due to the extremely high local field and polarization that it will experience, leading to significant bond distortion [28]. This condition can be explained considering its high dielectric susceptibility and dipole moment [29].

In our model, defect generation probability is assumed to increase exponentially with the energy E, and thus, the generation rate is given by $R_{\mathrm{TE}} = R_0 e^{\alpha V_+}$, where α is a constant, in agreement with the E-model of dielectric breakdown [30], [31]. Similarly, the generation rate at the BE interface can be written as $R_{\mathrm{BE}} = R_0 e^{\alpha |V_-|}$.

To test the defect generation model in Fig. 5(a), Fig. 6(a) shows the calculated N_C value for asymmetric bipolar cycling, compared with data from Fig. 4. We assumed $\alpha = 42$ V^{-1} in the calculations. The model correctly describes the steep decay of N_C in region A; however, the model fails to predict the weak voltage dependence in region B. In fact, due to the exponential voltage dependence of R_{BE} and R_{TE}, the defect generation model only attributes degradation to the largest voltage, in contrast to the experimental evidences in Figs. 3 and 4.

B. Defect Activation

To account for the impact of the smaller voltage in the MgO degradation, we considered the defect activation mechanism

displayed in Fig. 5(b). After a positive pulse of voltage V_+, the application of a negative pulse with amplitude $|V_-| < V_+$ can activate the defects generated by the positive semicycle, e.g., by displacing an interstitial O_i^{2-} away from the corresponding O vacancy in the newly created Frenkel pair, as shown in Fig. 5(c), with a rate $R_a/R_0 = k e^{\beta |V_-|}$, where k and β are constants with $\beta < \alpha$. The activation causes an additional damage to the dielectric layer during the low-voltage semicycle, since the separation of the two constituents of the Frenkel pair leads to: 1) a reduced probability of recombination and 2) an increased defect concentration in the bulk of the MgO, supporting the formation of a percolative path [22]. Calculation results from the generation/activation model with $k = 1$ and $\beta = 4$ V^{-1} are shown in Fig. 6(b), indicating better agreement with data in both regions A and B.

To further confirm that the activation process consists of a displacement rather than a thermal effect, e.g., a temperature-induced stabilization of the generated defect, we compared the asymmetric bipolar stress (fixed $V_+ = 1$ V and variable negative V_-) and the asymmetric unipolar stress, where both the fixed and variable voltages were positive. Data shown in Fig. 7 indicate a larger N_C value and a rather flat behavior in region B for the asymmetric unipolar stress, thus suggesting that a positive voltage is not effective in displacing O_i^{2-} from V_O^{2+}. These data confirm that the activation process requires bipolar stress.

To complete our model, we included defect generation and activation at the BE side with the same parameters used for the TE side in view of the high symmetry of our MTJ stack. We also included an explicit dependence on the pulsewidths t_+ and t_- of the positive and negative pulses, respectively. The total defect density due to generation and activation is thus written as

$$n_D = n_{D0} \left[\frac{t_+}{t_0} e^{\alpha V_+} e^{\beta |V_-|} + \frac{t_-}{t_0} e^{\alpha |V_-|} e^{\beta V_+} \right] \quad (1)$$

where $t_0 = 10^{-30}$ s is a constant. The model parameters are summarized in Table I. Fig. 8(a) shows the measured and calculated N_C value for both constant V_+ with variable V_- and constant V_- with variable V_+. Our model is able to predict the different slopes in regions A and B, where n_D can be
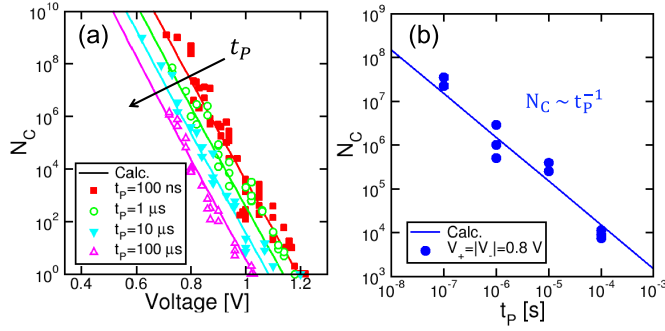
Fig. 9. (a) Measured and calculated N_C for symmetric bipolar cycling for different applied pulsewidths t_P. (b) Corresponding data and calculations for N_C as a function of t_P for $V_+ = |V_-| = 0.8$ V.
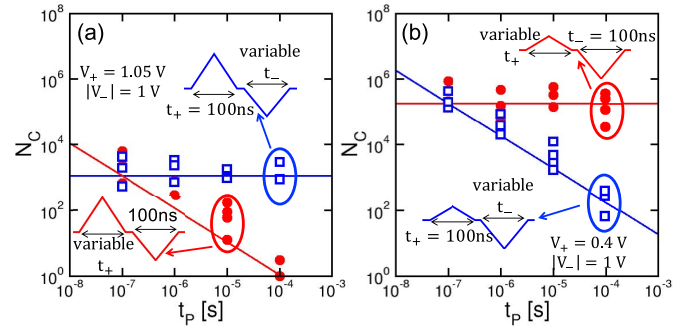

Fig. 10. Measured and calculated N_C as a function of t_+ and t_- for asymmetric bipolar cycling in (a) region A and (b) region B. The maximum number of cycles depends only on the pulsewidth of the highest voltage pulse, which is responsible for the generation step in Fig. 5(a).

approximated as

$$n_D \approx n_{D0}\frac{t_+}{t_0}e^{\alpha V_+}e^{\beta|V_-|} \sim e^{\alpha V_+} \quad (V_+ > |V_-|) \quad (2a)$$

$$n_D \approx n_{D0}\frac{t_-}{t_0}e^{\alpha|V_-|}e^{\beta V_+} \sim e^{\beta V_+} \quad (V_+ < |V_-|) \quad (2b)$$

respectively. Slopes in regions A and B can be directly related to α and β. The model is able to account for N_C for unipolar (positive and negative) and symmetric bipolar stress (i.e., $V_+ = |V_-|$), as shown in Fig. 8(b). Fig. 8(c) shows the simulated voltage-dependent endurance for $t_P = 100$ ns and $t_D = 20$ ns.

V. PULSE-TIME DEPENDENCE OF ENDURANCE

A. Impact of Pulsewidth t_P

To test the impact of the pulsewidth t_P on endurance, Fig. 9(a) shows the measured and calculated N_C value for symmetric bipolar stress ($V_+ = |V_-|$) for increasing t_P from 100 ns to 100 μs. Data indicate that N_C decreases at increasing t_P as $N_C \sim t_P^{-1}$, as also summarized in Fig. 9(b) for stress at $V_+ = |V_-| = 0.8$ V. Calculations accurately account for the t_P dependence, as a result of the dependence on t_+ and t_- in (1), to describe the increase of the defect density with increasing stress time.

To study the distinct impacts of t_+ and t_- in (1), Fig. 10(a) and (b) shows N_C for asymmetric bipolar cycling for fixed $V_- = -1$ V and two distinct values of V_+, namely $V_+ = 1.05$ V corresponding to region A and $V_+ = 0.4$ V corresponding to region B. In these two regions, we measured N_C as a function of t_+ for constant $t_- = 100$ ns or as a function of t_- for constant $t_+ = 100$ ns. Data in region A [see Fig. 10(a)] indicate that N_C decreases as $N_C \sim t_+^{-1}$ while t_- has no impact on N_C. On the other hand, N_C decreases as $N_C \sim t_-^{-1}$ in region B [see Fig. 10(b)] with no role of t_+. Calculations by (2a) and (2b) are also shown, thus demonstrating that our model can predict the distinct dependence on t_+ and t_-.

From our data, N_C shows a dependence only on the width of the pulse of the largest voltage, namely the one that generates defects in the MgO [see Fig. 5(a)]. The duration of the activation pulse is instead not affecting degradation. This is in agreement with a physical picture where activation behaves like a binary event, i.e., resulting in either failure or success.
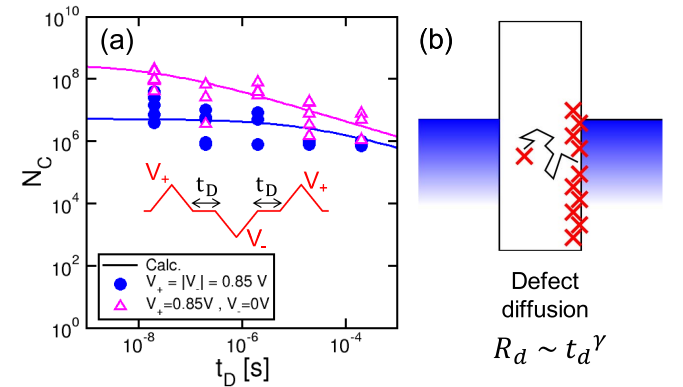

Fig. 11. (a) Measured and calculated N_C as a function of t_D for unipolar and bipolar stress. (b) Schematic of the defect diffusion while no voltage is applied to the device.

B. Impact of Pulse Delay t_D

Fig. 11(a) shows the measured and calculated N_C value as a function of the delay time t_D for both unipolar stress ($V_+ = 0.85$ V and $V_- = 0$ V) and bipolar stress ($V_+ = |V_-| = 0.85$ V). In both cases, N_C slightly decreases for increasing t_D, which can be explained by defect diffusion from the interface region where the defects are generated toward the bulk region of the MgO layer, as shown in Fig. 11(b). Similar to defect activation, delay enhances the degradation of generated defects by preventing recombination of oxygen interstitials and vacancies, and by enhancing the defect concentration within the bulk of the MgO layer, thus supporting the creation of percolation paths [22].

The dependence on t_D was taken into account in the model by adding a diffusive rate $R_d \sim t_D^\gamma$ in (1), leading to

$$n_D = n_{D0}\left[\frac{t_+}{t_0}e^{\alpha V_+}\left(ke^{\beta|V_-|} + \frac{t_D}{t_{D0}}^\gamma\right) \right.$$
$$\left. + \frac{t_-}{t_0}e^{\alpha|V_-|}\left(ke^{\beta V_+} + \frac{t_D}{t_{D0}}^\gamma\right)\right] \quad (3)$$

where t_{D0} and γ are the constant parameters shown in Table I. Calculations by (3) are shown in Fig. 11(a), in close agreement with the experimental results. The results also suggest that the gap between unipolar and bipolar endurance decreases for increasing t_D, which is fully taken into account by our

TABLE I
SUMMARY OF ENDURANCE MODEL PARAMETERS IN (1)–(3)

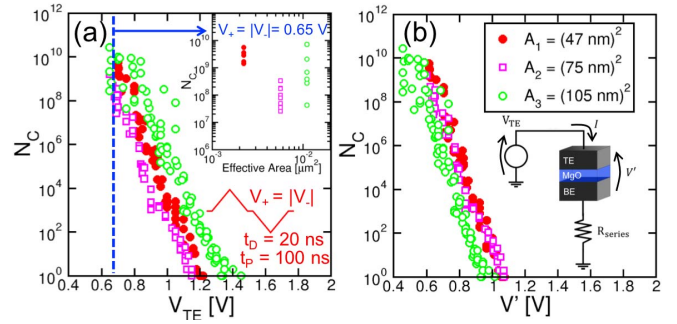| Parameter	Value
N_{C0}	1
α [V^{-1}]	42
β [V^{-1}]	4
k	1
t_0 [s]	10^{-30}
t_{D0} [s$^{-0.5}$]	7.07×10^3
γ	0.5

Fig. 12. (a) Measured N_C as a function of the applied voltage for different sample areas for symmetric bipolar stress condition, the inset shows N_C for three different areas at $V_+ = |V_-| = 0.65$ V. (b) Similar cycling failure data are presented as a function of the actual voltage drop on the MTJ (V'), as shown in the inset.
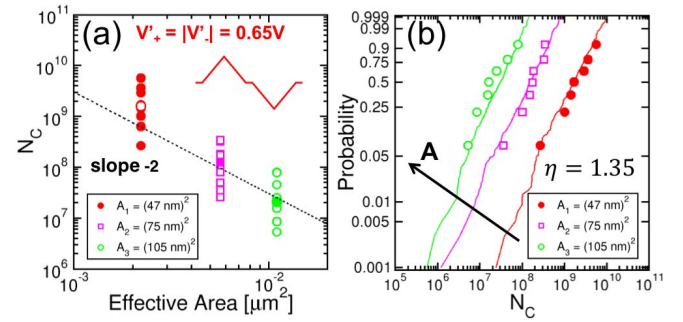

Fig. 13. (a) Measured and calculated N_C as a function of device area for three different device areas. The applied waveform was symmetric bipolar with $t_P = 100$ ns and $V'_+ = |V'_-| = 0.65$ V. (b) Corresponding Weibull plot for measured and calculated N_C. Endurance data can be well reproduced by Weibull statistics even though the area dependence is stronger than the one predicted with Poisson scaling approach: TDDB $\sim A^{-1/\eta}$.

diffusive model. In fact, as t_D increases, defects efficiently diffuse toward the opposite interface, thus making the difference between unipolar and bipolar stress increasingly negligible. Note that the weak dependence of bipolar endurance on t_D is consistent with previous results in [32]. On the other hand, our data for unipolar stress show no dramatic dependence on t_D, in contrast to [32], which might be explained by a different structure or etch damage profile in our MgO layer.

VI. AREA DEPENDENCE

The reduction of device area A in p-STT-MRAM devices allows to decrease the switching current, which is required to minimize the cell area limited by the driving transistor in 1T-1MTJ structures [14], [19]. In addition to reducing the footprint and power consumption, area scaling also allows to improve cycling endurance due to the Poisson area scaling of TDDB [14], [18]. To study the area dependence, Fig. 12 shows the measured N_C value for bipolar cycling as a function of $V_+ = |V_-|$ for increasing area, namely $A = (47\text{ nm})^2$, $(75\text{ nm})^2$, and $(105\text{ nm})^2$.

Data in Fig. 12(a) show an unexpected nonmonotonic behavior, which is summarized in Fig. 12(a) (inset) for the case $V_+ = |V_-| = 0.65$ V. Here, N_C decreases with area but shows an anomalous large value for the largest area. This result was attributed to a series resistance effect, where the actual voltage drop V' across the MTJ decreases with the device area. In fact, V' is given by $V' = V - R_s I$, where R_s is the series resistance associated with the contacts and interfaces and I is the current. As the device area increases, I also increases, thus causing a decrease of V'. We estimated V' by assuming $R_s = 300\ \Omega$, corresponding to the resistance after breakdown in Fig. 2(a). Fig. 12(b) shows N_C as a function of V', evidencing a correct monotonic decrease of N_C with area.

Fig. 13(a) shows N_C as a function of device area for $V_+ = |V_-| = 0.65$ V, evidencing a decrease according to the power law $N_C \sim A^{-2}$. Based on Poisson area scaling, the exponent in the power law should be equal to the inverse of the Weibull shape factor of N_C, namely the slope of the cumulative distribution in the Weibull plot [33]. The latter is shown in Fig. 13(b) for various A, indicating an area-independent Weibull shape factor $\eta = 1.35$ in the formula log(- log(1- F) = ηlog (N_C/N_{C0}). Such a value of the shape parameter η can be explained by intrinsic TDDB processes, such as defect generation controlled by the electrical stress, in contrast to extrinsic breakdown processes for $\eta < 1$ [14], [34]. From Poisson area scaling, we calculate a theoretical

slope in Fig. 13(a) of $-1/\eta \approx -0.75$, in contrast with the experimental slope ≈ -2. This disagreement might be explained by the etching process having beneficial effects on the device lateral surface, resulting in a low probability of breakdown initiation. This effect results in a reduced effective area for breakdown process appearing as a stronger area dependence for relatively small devices [18]. Also, Joule heating effects might contribute to TDDB, thus causing deviation from the field-driven Poisson area scaling for relatively small device area.

VII. CONCLUSION

We show a comprehensive study of breakdown-limited cycling endurance in p-STT-MRAM devices. Cycling endurance is experimentally monitored as a function of the pulse amplitude, polarity, and timing. We developed a semi-empirical model based on generation, activation, and diffusion of defects in the MgO tunnel barrier. The model accounts for the dependence of endurance lifetime on applied voltage, pulsewidth, and pulse delay. Finally, the area scaling of endurance is experimentally analyzed and discussed.

REFERENCES

[1] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive random access memory," *Proc. IEEE*, vol. 104, no. 10, pp. 1796–1830, Oct. 2016, doi: 10.1109/JPROC.2016.2590142.

[2] R. F. Freitas and W. W. Wilcke, "Storage-class memory: The next storage system technology," *IBM J. Res. Develop.*, vol. 52, nos. 4–5, pp. 439–447, 2008, doi: 10.1147/rd.524.0439.

[3] M. Natsui *et al.*, "Nonvolatile logic-in-memory array processor in 90 nm MTJ/MOS achieving 75% leakage reduction using cycle-based power gating," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, Feb. 2013, pp. 194–195, doi: 10.1109/ISSCC.2013.6487696.

[4] H.-S. P. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature Nanotechnol.*, vol. 10, no. 3, pp. 191–194, 2015, doi: 10.1038/nnano.2015.29.

[5] T. Endoh, H. Koike, S. Ikeda, T. Hanyu, and H. Ohno, "An overview of nonvolatile emerging memories—Spintronics for working memories," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 2, pp. 109–119, Jun. 2016, doi: 10.1109/JETCAS.2016.2547704.

[6] K. Ikegami *et al.*, "Low power and high density STT-MRAM for embedded cache memory using advanced perpendicular MTJ integrations and asymmetric compensation techniques," in *IEDM Tech. Dig.*, Dec. 2014, pp. 1–28, doi: 10.1109/IEDM.2014.7047123.

[7] S. Ghosh *et al.*, "Overview of circuits, systems, and applications of spintronics," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 3, pp. 265–278, Sep. 2016, doi: 10.1109/JETCAS.2016.2601310.

[8] E. Deng, Y. Zhang, J.-O. Klein, D. Ravelsona, C. Chappert, and W. Zhao, "Low power magnetic full-adder based on spin transfer torque MRAM," *IEEE Trans. Magn.*, vol. 49, no. 9, pp. 4982–4987, Sep. 2013, doi: 10.1109/TMAG.2013.2245911.

[9] S. Lequeux *et al.*, "A magnetic synapse: Multilevel spin-torque memristor with perpendicular anisotropy," *Sci. Rep.*, vol. 6, p. 31510, Aug. 2016, doi: 10.1038/srep31510.

[10] C. Chappert, A. Fert, and F. N. Van Dau, "The emergence of spin electronics in data storage," *Nature Mater.*, vol. 6, no. 11, pp. 813–823, Nov. 2007, doi: 10.1038/nmat2024.

[11] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *J. Magn. Magn. Mater.*, vol. 159, nos. 1–2, pp. L1–L7, Jun. 1996, doi: 10.1016/0304-8853(96)00062-5.

[12] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Phys. Rev. B, Condens. Matter*, vol. 54, pp. 9353–9358, Oct. 1996, doi: 10.1103/PhysRevB.54.9353.

[13] S. Ikeda *et al.*, "A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction," *Nature Mater.*, vol. 9, pp. 721–724, Jul. 2010, doi: 10.1038/nmat2804.

[14] J. J. Kan *et al.*, "A study on practically unlimited endurance of STT-MRAM," *IEEE Trans. Electron Devices*, vol. 64, no. 9, pp. 3639–3646, Sep. 2017, doi: 10.1109/TED.2017.2731959.

[15] R. Carboni *et al.*, "Understanding cycling endurance in perpendicular spin-transfer torque (p-STT) magnetic memory," in *IEDM Tech. Dig.*, Dec. 2016, pp. 21.6.1–21.6.4, doi: 10.1109/IEDM.2016.7838468.

[16] S. Balatti *et al.*, "Voltage-controlled cycling endurance of HfO$_x$-based resistive-switching memory," *IEEE Trans. Electron Devices*, vol. 62, no. 10, pp. 3365–3372, Oct. 2015, doi: 10.1109/TED.2015.2463104.

[17] K. Kim, C. Choi, Y. Oh, H. Sukegawa, S. Mitani, and Y. Song, "Time-dependent dielectric breakdown of MgO magnetic tunnel junctions and novel test method," *Jpn. J. Appl. Phys.*, vol. 56, no. 4S, p. 04CN02, Feb. 2017, doi: 10.7567/JJAP.56.04CN02.

[18] S. Van Beek *et al.*, "Four point probe ramped voltage stress as an efficient method to understand breakdown of STT-MRAM MgO tunnel junctions," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Apr. 2015, pp. MY.4.1–MY.4.6, doi: 10.1109/IRPS.2015.7112818.

[19] S. Amara-Dababi *et al.*, "Charge trapping-detrapping mechanism of barrier breakdown in MgO magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 99, no. 8, p. 083501, Aug. 2011, doi: 10.1063/1.3615654.

[20] O'C. Robert, G. Hughes, P. Casey, and S. B. Newcomb, "Degradation and breakdown characteristics of thin MgO dielectric layers," *J. Appl. Phys.*, vol. 107, no. 2, pp. 024501–024504, Oct. 2010, doi: 10.1063/1.3265434.

[21] C. Yoshida, M. Kurasawa, Y. M. Lee, K. Tsunoda, M. Aoki, and Y. Sugiyama, "A study of dielectric breakdown mechanism in CoFeB/MgO/CoFeB magnetic tunnel junction," in *Proc. Int. Rel. Phys. Symp.*, Apr. 2009, pp. 139–142, doi: 10.1109/IRPS.2009.5173239.

[22] J. H. Stathis, "Percolation models for gate oxide breakdown," *J. Appl. Phys.*, vol. 86, no. 10, pp. 5757–5766, Nov. 1999, doi: 10.1063/1.371590.

[23] C. Yoshida and T. Sugii, "Reliability study of magnetic tunnel junction with naturally oxidized MgO barrier," in *Proc. IEEE Int. Rel. Phys. Symp.*, Apr. 2012, pp. 2A.3.1–2A.3.5, doi: 10.1109/IRPS.2012.6241773.

[24] K. Komagaki *et al.*, "Influence of diffused boron into MgO barrier on pinhole creation in CoFeB/MgO/CoFeB magnetic tunnel junctions," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3453–3456, Oct. 2009, doi: 10.1109/TMAG.2009.2022189.

[25] P. G. Mather, J. C. Read, and R. A. Buhrman, "Disorder, defects, and band gaps in ultrathin (001) MgO tunnel barrier layers," *Phys. Rev. B, Condens. Matter*, vol. 73, no. 20, p. 205412, May 2006, doi: 10.1103/PhysRevB.73.205412.

[26] B. P. Uberuaga *et al.*, "Structure and mobility of defects formed from collision cascades in MgO," *Phys. Rev. Lett.*, vol. 92, no. 11, p. 115505, Mar. 2004, doi: 10.1103/PhysRevLett.92.115505.

[27] W. Kang, L. Zhang, J.-O. Klein, Y. Zhang, D. Ravelosona, and W. Zhao, "Reconfigurable codesign of STT-MRAM under process variations in deeply scaled technology," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1769–1777, Jun. 2015, doi: 10.1109/TED.2015.2412960.

[28] W. Oepts, H. J. Verhagen, W. J. M. de Jonge, and R. Coehoorn, "Analysis of breakdown in ferromagnetic tunnel junctions," *J. Appl. Phys.*, vol. 86, no. 7, pp. 3863–3872, Oct. 1999, doi: 10.1063/1.371300.

[29] O. Madelung, U. Rössler, and M. Schulz, Eds., *Magnesium Oxide (MgO) Dielectric Constants: Optical and Photoelectric Properties*. Berlin, Germany: Springer, 1999, pp. 1–12.

[30] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO$_2$ thin films," *J. Appl. Phys.*, vol. 84, no. 3, pp. 1513–1523, Jul. 1998, doi: 10.1063/1.368217.

[31] Y. Wang *et al.*, "Compact model of dielectric breakdown in spin-transfer torque magnetic tunnel junction," *IEEE Trans. Electron Devices*, vol. 63, no. 4, pp. 1762–1767, Apr. 2016, doi: 10.1109/TED.2016.2533438.

[32] S. Amara-Dababi, H. Bea, R. Sousa, K. Mackay, and B. Dieny, "Modelling of time-dependent dielectric barrier breakdown mechanisms in MgO-based magnetic tunnel junctions," *J. Phys. D, Appl. Phys.*, vol. 45, no. 29, p. 295002, 2012, doi: 10.1088/0022-3727/45/29/295002.

[33] E. Y. Wu and J. Sune, "On voltage acceleration models of time to breakdown—Part I: Experimental and analysis methodologies," *IEEE Trans. Electron Devices*, vol. 56, no. 7, pp. 1433–1441, Jun. 2009, doi: 10.1109/TED.2009.2021725.

[34] R. Degraeve, J. L. Ogier, R. Bellens, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," *IEEE Trans. Electron Devices*, vol. 45, no. 2, pp. 472–481, Feb. 1998, doi: 10.1109/16.658683.

Roberto Carboni (S'16) received the B.S. and M.S. degrees in electrical engineering from the Politecnico di Milano, Milan, Italy, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering.

His current research interests include the characterization and modeling of resistive-switching memory (RRAM) and spin-transfer torque magnetic memory (STT-MRAM).

Stefano Ambrogio (M'16) received the B.S. degree, the M.S. degree *(cum laude)*, and the Ph.D. degree in electrical engineering from the Politecnico di Milano, Milan, Italy, in 2010, 2012, and 2016, respectively.

He is currently a Post-Doctoral Researcher with IBM Research-Almaden, San Jose, CA, USA. His current research interests include nonvolatile memory and cognitive computing.

Dr. Ambrogio received the IEEE EDS Rappaport Award in 2015.

Wei Chen was with Micron Technology, Inc., Boise, ID, USA, as a Senior Engineer, for five and a half years, where he was involved in thin-film process and device technology in the field of STT-MRAM and dynamic random access memory. He is currently a Senior Technical Staff Member with Spin Transfer Technologies, Fremont, CA, USA, an STT-MRAM technology company.

Witold Kula received the Ph.D. degree in physics from the Institute of Physics, Polish Academy of Sciences, Warsaw, Poland.

He was a fellow with Micron Technology, Inc., Boise, ID, USA. He has held several management and engineering positions at TDK/Headway, Cypress Semiconductor, and IBM Corporation, all in San Jose, CA, USA. He is currently a Chief Technology Officer at Antaios, SAS, Montbonnot-Saint-Martin, France, a semiconductor memory technology company.

Manzar Siddik, photograph and biography not available at the time of publication.

Gurtej Sandhu, photograph and biography not available at the time of publication.

Jon Harms, photograph and biography not available at the time of publication.

Andy Lyle, photograph and biography not available at the time of publication.

Daniele Ielmini (SM'09) is currently a Full Professor with the Dipartimento di Elettronica, Informazione, e Bioingegneria, Politecnico di Milano, Milan, Italy. He current research interests include emerging nanoelectronic devices, such as phase change memory and resistive-switching memory (RRAM), and their application in computing.

Dr. Ielmini received the Intel Outstanding Researcher Award in 2013, the ERC Consolidator Grant in 2014, and the IEEE EDS Rappaport Award in 2015.