Validity of the thermal activation model for spin-torque switching in magnetic tunnel junctions

R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall, A. B Kos

National Institute of Standards and Technology (NIST), Boulder, Colorado 80305

*Publication of NIST, an agency of the US government, not subject to copyright.
Abstract

We have performed spin transfer switching experiments on nanoscale MgO magnetic tunnel junctions, with a large number of trials (up to $10^7$ switching events) as functions of pulse voltage amplitude and duration, and compared the results with both macrospin simulations and a simple thermal model of switching. Three different methods derived from the thermally activated spin-torque-assisted switching model are used to determine the thermal stability factor and the intrinsic switching voltage from the switching distributions. Specifically, their values are determined from the read disturb rate, switching voltage vs. pulse duration, and switching voltage distribution measurements. Our conclusion is that the data obtained from the first two methods agree well with each other as well as with the values from quasistatic measurements, assuming that for the region of voltages over which the data is taken, the voltage is small compared to the intrinsic switching voltage, and the influence from other factors contributing to the switching is negligible (such as current-induced heating and field-like torque). The third method (switching voltage distribution measurements) is shown to be impractical, because the applied voltage surpasses the intrinsic switching voltage within the time-scale (< 1 μs) at which the experiments are performed.
Current-induced magnetization switching forms the basis of spin torque transfer random access memory (STT-RAM) technology. The thermally activated model that includes the effects of spin torque has been important in characterizing current-induced magnetization switching in nanoscale magnetic tunnel junctions (MTJs) [1-3]. The model provides an analytical estimate of the switching probability as a function of pulse amplitude, pulse width, and device parameters, and can be used to determine the thermal stability factor $\Delta$ and intrinsic switching voltage (current) $V_{C_0}$ ($I_{C_0}$) from switching probability measurements [4, 5]. Quick and accurate determination of these parameters is needed for the development of next generation of STT-RAM devices. In this paper, we present current-induced magnetization switching experiments of in-plane magnetized MgO MTJs with a large number of trials to obtain enough statistical data to test the thermally-activated spin-torque-assisted switching model and determine the voltage (current) region over which it is valid. As discussed below, we find that this simple model works well in the small voltage (current) region where $V/V_{C_0}$ ($I/I_{C_0}$) is significantly less than 1, and fails when the voltage (current) approaches and surpasses $V_{C_0}$ ($I_{C_0}$).

The MgO MTJ wafers were prepared by the Singulus company [6], and processed into devices at NIST using optical and electron-beam lithography and ion milling to define the pillar structures. The MTJ stack structure is 5 nm Ta/15 nm PtMn/2.3 nm Co$_{70}$Fe$_{30}$/0.8 nm Ru/2.5 nm Co$_{40}$Fe$_{20}$B$_{40}$/0.87 nm MgO/1.5 nm Co$_{40}$Fe$_{40}$B$_{20}$/10 nm Ta/7 nm Ru, with both free and reference layers having an easy magnetization axis in the film plane. The free layers in both devices have significant perpendicular anisotropy ($H_k^z \sim 80 \%$ of $M_s$) in order to reduce the switching current [7]. The ion milling was stopped at the MgO barrier, meaning that only the free ferromagnetic layer is patterned. Wafer level current-in-plane-tunneling (CIPT) measurements [8] and measurements on processed devices give resistance-area products of 3 $\Omega \mu$m$^2$ and tunneling magnetoresistance (TMR) ratios of about 60%. The devices have elliptical shapes with an aspect ratio of 3 to 1 and sizes ranging from 50 nm x 150 nm to 100 nm x 300 nm. All data in this paper come from one 50 nm x 150 nm device, while all other devices from the wafer show the similar features and the device-to-device variations were minor. Devices with RA products ranging from 3 $\Omega \mu$m$^2$ to 25 $\Omega \mu$m$^2$, having various sizes, compositions and thermal stability factors, show similar behavior and are in agreement with the conclusions presented here.

The measurement setup to efficiently measure the switching probability (probability that the relative orientation of the free and reference layers will change under an applied voltage pulse) with large statistics ($> 10^7$ trials) is shown in Fig. 1a. It consists of a 10 GS/s arbitrary waveform generator (AWG) used to create a sequence of rectangular voltage pulses, a set of bias tees, a test device
embedded into a coplanar waveguide structure, and a 13 GHz digital storage oscilloscope. The pulse sequence (Fig. 1b) consists of a reset pulse used to initialize the device state (parallel or antiparallel alignment), a long duration (\(\sim 50 \, \mu s\)), low-voltage read pulse to verify the initial state of the device, a short write pulse of duration 3 ns to 1 \(\mu s\) of opposite polarity than the reset pulse, and another read pulse to measure the state of the device after the write. Prior to the pulse sequence, the device is in a quiescent state, meaning that there is no bias applied to the device. Using the bias tees, we separate the low frequency signal into a high-impedance path and send it to the oscilloscope. The rise times of the write pulse and the read pulse are about 125 ps and 40 \(\mu s\), respectively, hence there is negligible cross-talk between the high and the low frequency circuit paths. The data are then analyzed and plotted as a switching probability, as shown in Fig. 1c. The data in Fig. 1c are for antiparallel-to-parallel switching (positive voltage pulse) in zero total magnetic field. We also measured and analyzed parallel-to-antiparallel switching, which agrees with the results and conclusions from antiparallel-to-parallel switching.

In the simplest thermally activated spin-transfer switching model, the switching probability is given by [2, 3]:

\[
P_{sw} = 1 - e^{\left(-\frac{t_p}{\tau}\right)}
\]  

(1), where \(t_p\) is the duration of the voltage pulse,

\[
\tau = \tau_0 \exp\left(\Delta(1 - \frac{V}{V_{C0}})\right)
\]

(2) is the thermal activation lifetime, \(\tau_0\) is the attempt time, usually taken to be near 1 ns [9, 10], \(\Delta = \frac{E_{barrier}}{k_BT}\) (3) is the thermal stability factor [4], \(V\) is the amplitude of the voltage pulse, and \(V_{C0}\) is the intrinsic switching voltage [5]. While this model excludes the effects of device heating and the field-like torque term [11], it is commonly used to fit experimental data in order to obtain values of \(\Delta\) and \(V_{C0}\), which are important figures of merit of an STT-RAM device, and yields correct results when used under certain restrictions as we discuss below. Starting from the data shown in Fig. 1c, we focus on the region where \(t_p \geq 100 \, ns\) \(\gg \tau_0\), and analyze three different methods based on equations (1-3) to determine the values of \(\Delta\) and \(V_{C0}\). We then compare the methods with each other and with the value of the intrinsic thermal stability factor determined from magnetostatic measurements \(\Delta_{static} = 42 \pm 2\) [12].

The first method to determine \(\Delta\) and \(V_{C0}\) is fitting the slope of the read disturb rate (RDR), which is the probability of accidental switching when reading the resistance state of the device, i.e., probability of switching of the magnetization in the limit of small applied voltages. Starting from Eq. 1, and using
\( \frac{t_p}{\tau} << 1 \), we perform a Taylor expansion and arrive at the following expression to describe the RDR:

\[
\ln(P_{sw}) = \ln\left(\frac{t_p}{\tau_0}\right) - \Delta(1 - \frac{V}{V_{c0}}) \quad (4).
\]

To check the validity of our assumption that \( \frac{t_p}{\tau} << 1 \), we take \( V/V_{c0} \leq 0.8 \) and \( \Delta = 42 \) (which are justified below), \( \tau_0 = 1 \text{ ns} \), and using Eq. 2, we find

\[
\frac{t_p}{\tau} = \frac{t_p}{\tau_0 \times 4400} << 1, \text{ for } t_p << 4400 \text{ ns}, \text{ which is valid at our pulse width of 100 ns.}
\]

Fig. 2a shows a plot of the switching probability distribution at 100 ns, with the \( P_{sw} \) shown on a logarithmic scale to emphasize the low voltage region, which is the RDR. Also shown is the fit to Eq. 4, which gives \( \Delta = 42.55 \pm 0.84 \) and \( V_{c0} = 0.246 \pm 0.003 \text{ V} \) if performed in the range \( 10^{-7} < P_{sw} < 10^{-3} \) and \( 0.52 < V/V_{c0} < 0.76 \). The measured value of \( \Delta \) is in good agreement with the determined value of \( \Delta_{\text{static}} \). Thus, in our case, this method produces valid results for \( \Delta \), and so presumably for \( V_{c0} \) as well, in the range \( V/V_{c0} < 0.8 \). If the range of the fit is increased to higher voltages (higher switching probability) the values of \( \Delta \) are significantly reduced, indicating that the fit must be performed below a certain threshold, which is discussed in more detail below. Fits to the RDR at other pulse times (50 ns - 200 ns) gave the same results.

The second method, which we label as ‘time domain’, is widely used [13-21] to determine these parameters as well. Starting with Eq. 2, the switching voltage can be expressed as a function of pulse duration \( V_c = V_{c0}[1 - \frac{1}{\Delta} \ln(\frac{t_p}{\tau_0})] \) \( (5) \), which is strictly valid when \( P_{sw} = 1 - e^{-1} = 0.63 \). From data shown in Fig. 1c, we determined the value of the pulse amplitude \( V_c \) that corresponds to 63% switching probability for pulse durations between 3 ns - 1 \( \mu \)s, and plotted this on a semi-logarithmic scale versus time (Fig. 2b). The pulse amplitude data is linear for \( t_p > 300 \text{ ns} \), curves slightly for \( 70 \text{ ns} < t_p < 300 \text{ ns} \), and deviates strongly from linear for \( t_p < 70 \text{ ns} \). Similar behavior is also seen in [16]. Using Eq. 5, we performed two linear fits of the data in the regions \( 300 \text{ ns} < t_p < 1 \text{ \( \mu \)s} \) (labeled as fit 1) and \( 70 \text{ ns} < t_p < 1 \text{ \( \mu \)s} \) (labeled as fit 2). Fit 1 gives values \( \Delta = 41.5 \pm 1.2, V_{c0} = 0.256 \pm 0.001 \text{ V} \), while fit 2 gives values \( \Delta = 32.6 \pm 1.3, V_{c0} = 0.269 \pm 0.002 \text{ V} \). The value of \( \Delta \) from fit 1 agrees well with the value of \( \Delta_{\text{static}} \), while the fit 2 gives \( \Delta \) that is significantly less than the value of \( \Delta_{\text{static}} \). The slight curvature in the data for \( 70 \text{ ns} < t_p < 300 \text{ ns} \) might not be obvious, thus taking the data from this region and performing a linear fit might look correct, however the results are erroneous. Assuming \( V_{c0} = 0.25 \text{ V} \) (as determined through fitting of the RDR), we find that the region in fit 1 is \( 0.85 < V/V_{c0} < 0.88 \), while the fit 2 is over
the region \(0.85 < V/Vc_0 < 0.94\). We conclude that in our case the ‘time domain’ analysis of the data in the region \(V/Vc_0 < 0.88\) produces valid results, in agreement with the previous results from analysis of RDR, while analysis of the data at pulse amplitudes above this value gives incorrect results. Initially, one might think that the range of validity of Eq. 5 is determined only by the requirement that \(t_p > \tau_0\), i.e., that the switching occurs well above the “precessional” regime \((t_{precessional} \leq 1 \text{ ns})\) [22]. However, as shown in Fig. 2b, the data suggest that restrictions on the pulse amplitude must be met as well in order for Eq. 5 to be applicable.

The third method to determine the values of \(\Delta\) and \(Vc_0\) is based on the calculation of the cumulative switching voltage probability and its derivative [14]. The switching voltage distribution is obtained from the derivative of Eq. 1,

\[
\frac{dP_{sw}}{dV} = \frac{\Delta}{Vc_0} \frac{t_p}{\tau} \exp\left(-\frac{t_p}{\tau}\right) \quad (6)
\]

Here we take the switching probability data at 100 ns (Fig. 1c), differentiate it with respect to the pulse amplitude, and use Eq. 6 to fit the data, which is shown in Fig. 2c on a semi-logarithmic scale so that the tails of the distribution are as equally weighted as the transition from not-switching to switching. A fit to the full distribution gives \(\Delta = 23.39 \pm 0.41\) and \(Vc_0 = 0.297 \pm 0.001\) V, which are not in agreement with the values obtained from previous methods, nor with the value of \(\Delta_{\text{static}}\). Furthermore, it can be seen that Eq. 6 does not fit the data well, especially the tails of the distribution. The reason for the poor fit is that the fitting region of the switching voltage distribution spans the values \(0.52 < V/Vc_0 < 1.08\) (assuming \(Vc_0 = 0.25\) V), thus the high voltage region \(V > Vc_0\), invalidates the use of the model and the fitting results. In Fig. 2c, we also plotted Eq. 6 with values \(\Delta = 42,\ Vc_0 = 0.25\) V, and we see that the data and the analytical expression overlap in the region \(V/Vc_0 < 0.8\), which is basically the RDR region. We conclude that ‘switching voltage distribution’ method is not valid at the time scale of most pulsed-voltage switching measurements (< 1 \(\mu\)s). It is likely only valid at very long pulse times, where both low and high switching probability tails lie in the region \(V/Vc_0 < 0.8\), therefore this method is impractical from experimental point of view since very long pulse times (> 1 s) may be required to obtain the full switching distribution.

The first two methods (‘read disturb’ and ‘time domain’) to obtain the value of \(\Delta\) seem to agree well with each other and with the value of \(\Delta_{\text{static}}\), provided the fits are performed within a certain voltage limit. The disagreement seen for the fits close to \(V = Vc_0\), as well as the disagreement between the fitted values of the full switching voltage distribution and the static data, could be attributed to current-induced heating and field-like torque, since these effects are not included in the simple thermal model. To check
this, we performed numerical simulations of spin-torque-induced switching using a macrospin model, which includes finite temperature effects [23] but not those of the field-like torque or current-induced heating, and used the above mentioned methods to determine the values of \( \Delta \) and \( I_c \) [5]. The simulation parameters are: \( \mu_0 M_s = 1.38 \) T, \( \alpha = 0.01 \), \( \mu_0 H_z^c = 0.73 \) T (perpendicular anisotropy), \( 150 \) nm x \( 50 \) nm x \( 1.5 \) nm, rectangular shape, \( T = 300 \) K, \( \mu_0 H_{\text{ext}} = 0 \) T (no external field), \( P = 0.8 \) (polarization of spin injection current), \( \Lambda = 1 \) (spin-torque asymmetry), which give values of \( \Delta = 63 \) and \( I_c = 0.28 \) mA [24]. Both the spin-current polarization and the free layer easy axis lie in the film plane. We obtain equivalent results from simulations with other values of \( \Delta \) with and without perpendicular anisotropy.

At each current value the switching probability was determined using about 10,000 switching trials, and more for the RDR (at low currents).

Fig. 3a shows the simulated switching probability data at \( t_p = 100 \) ns plotted on a semi-logarithmic scale vs. normalized current units \( I/I_c \). Fitting the data in the RDR region \( (I/I_c < 0.85) \) to Eq. 4 gives values of \( \Delta = 63.34 \pm 0.62 \) and \( I_c = 0.285 \pm 0.0008 \) mA, in a reasonable agreement with the expected values. A ‘time domain’ plot of current values for \( P_{\text{sw}} = 63\% \), and the corresponding fit in the region 700 ns to 10 \( \mu \)s is presented in Fig. 3b. The linear fit performed over the region of currents smaller than about 0.263 mA gives values of \( \Delta = 61.9 \pm 0.1 \) and \( I_c = 0.294 \pm 0.004 \) mA, also in close proximity to the correct values. The slight discrepancy is likely due to the smaller number of simulation trials at long times (200 trials at pulse times from 1 \( \mu \)s to 10 \( \mu \)s) to get the most accurate \( P_{\text{sw}} \). We note that the data start deviating from linear behavior roughly below 600 ns. This also shows, in agreement with the experimental data, that although the switching time scale is well above the time scale of the “precessional” regime, the value of \( I/I_c \) must not approach 1 otherwise the switching process can no longer be described by Eq. 5. In Figs. 3c and 3d, we compare the simulated switching probability data and the derivative of switching current probability at \( t_p = 100 \) ns with calculated values using Eq. 1 and Eq. 6, respectively, with the values of \( \Delta = 63, I_c = 0.28 \) mA, which are the input parameters in the simulations. Fig. 3c shows that the simulated data deviate from the Eq. 1 for \( I/I_c > 0.85 \), thus fitting the whole switching probability distribution to obtain the values of \( \Delta \) and \( I_c \) would be incorrect. Similarly, performing a fit to the data in Fig. 3d over the full current range of \( 0.8 < I/I_c < 1.12 \), gives incorrect values of \( \Delta = 33.53 \pm 0.54 \) and \( I_c = 0.325 \pm 0.001 \) mA. A plot of calculated switching current distribution in Fig. 3d shows that a correct fit can only be performed in the low current RDR region.

The macrospin simulations show that the fitting region must be carefully chosen so that it falls within the region in which the thermally activated model is valid, and that the discrepancy in the fit to
the switching current distribution is not due to the current-induced heating or field-like torque. In the experiments, we do not know whether heating or field-like torque significantly influence the switching process, but since they both depend on voltage/current (field-like torque has a significant contribution only at high bias [25]), the low voltage/current region will likely show minimal influence of heating and field-like torque on the spin-torque switching.

In conclusion, we showed that the thermal activation model can be used to describe STT-RAM switching in the long time, low voltage limit, and to extract useful STT-RAM parameters such as the stability factor and the critical voltage and current. Care must be taken to only apply the model in the appropriate range of validity and in the region where the contribution of other parameters that influence the switching is negligible. Applying the model outside the long time, low voltage limit will result in considerable errors in determining STT-RAM parameters.
Acknowledgements

The authors acknowledge support by OMP, ARRA and DARPA. The authors thank Dr. Jürgen Langer of the Singulus Technologies AG for providing the wafers with the starting thin-film stack.

References


4. Thermal stability factor $\Delta = E_{\text{barrier}}/k_B T$, $E_{\text{barrier}}$ is the switching energy barrier, which depends on the free layer magnetization and anisotropy, $k_B$ is Boltzmann’s constant, and $T$ is the device temperature.

5. $V (I)$ is the applied switching voltage (current) at $T = 300$ K and $V_{C0} (I_{C0})$ is the intrinsic long-time switching voltage (current) at $T = 0$ K. While the actual experimental measurements were performed using voltage pulses of different amplitude, in simulations we use current pulses since, in the simple model of spin transfer, the current is more directly related to the current induced torque.


11. In the MTJ configuration with both the fixed and the free layers lying in the film plane, the field like torque, or perpendicular torque, is the component of the spin-transfer torque that is perpendicular to the free layer magnetic moment, and can play important role in spin-transfer switching under certain bias conditions.

12. Measured value of Co$_{40}$Fe$_{40}$B$_{20}$ free layer magnetization of our MgO MTJ (though SQUID measurements): $\mu_0 M_s \approx 1.27$ T. Measured values of 150 nm x 50 nm device parameters: $\mu_0 H_k \approx 0.037$ T (through thermal FMR measurements), average device resistance $R \approx 510 \ \Omega$, $RA \approx 3 \ \Omega \mu m^2$ (through
CIPT measurements). Estimated device area $A \approx 0.0059 \, \mu m^2$ (by comparing the actual device resistance and the $RA$ product). Using these values we obtain value of $\Delta^{\text{static}} = 42 \pm 2$.


24 Intrinsic switching current $I_{Co}$ in the simulations was determined by extrapolating switching currents to $T = 0 \, K$ and infinite switching time.

Figure 1. a) Schematic of the experimental setup. The low-frequency part of the circuit acts to include the MTJ device as a variable resistor in a high-impedance voltage-divider network. The state of the device is read using applied voltages of less than 35 mV, ensuring that the device is not accidentally written during this process. b) Sequence of pulses. c) Switching probability $P_{sw}$ for antiparallel-to-parallel switching as a function of pulse amplitude and pulse duration. Each data point is an average of $10^5$ to $10^7$ trials.
Figure 2. Evaluation of $\Delta$ and $V_{C0}$ from three methods: a) read disturb rate, b) time domain c) switching voltage distribution. The arrows in 2b indicate the fitting region. The green plot in 2c is the calculated switching voltage distribution (Eq. 6) with $\Delta = 42$ and $V_{C0} = 0.25$ V.
Figure 3. Macrospin simulations of current-induced magnetization switching probability using $\Delta = 63$ and $Ic_0 = 0.28$ mA: a) ‘read disturb rate’, b) ‘time domain’, c) comparison of simulated $P_{sw}$ and calculated $P_{sw}$ using the Eq. 1 with $\Delta = 63$ and $Ic_0 = 0.28$ mA, d) comparison of simulated switching current distribution and the calculated using Eq. 6 with $\Delta = 63$ and $Ic_0 = 0.28$ mA.