Spin-torque switching efficiency in CoFeB-MgO based tunnel junctions

J. Z. Sun,[1,*] S. L. Brown,[1] W. Chen,[2] E. A. Delenia,[3] M. C. Gaidis,[1] J. Harms,[2] G. Hu,[1] Xin Jiang,[3] R. Kilaru,[1] W. Kula,[2] G. Lauer,[1] L. Q. Liu,[1] S. Murthy,[2] J. Nowak,[1] E. J. O'Sullivan,[1] S. S. P. Parkin,[3] R. P. Robertazzi,[1] P. M. Rice,[3] G. Sandhu,[2] T. Topuria,[3] and D. C. Worledge[1]

[1]*IBM-Micron MRAM Alliance, IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA*

[2]*IBM-Micron MRAM Alliance, Micron, Boise, Idaho 83707, USA*

[3]*IBM-Micron MRAM Alliance, IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120, USA*

(Received 7 July 2013; published 26 September 2013)

It is convenient to define the spin-torque switching efficiency in nanostructured magnetic tunnel junctions as the ratio between the free-layers thermal activation barrier height E_b and the threshold switching current I_{c0}. Recent device exploration has led to occasional observations of spin-torque induced magnetic switching efficiency in magnetic tunnel junctions that exceeds the macrospin limit by a factor of 2–10. In this paper we examine the possible origins for such enhancement, and materials properties that may allow the full realization of such enhancements.

DOI: 10.1103/PhysRevB.88.104426 PACS number(s): 75.78.Jp, 85.75.Dd, 75.76.+j, 73.40.−c

I. INTRODUCTION

Spin-transfer torque switched magnetic tunnel junction is a basic building block for a new class of solid-state technologies as represented by the so-called spin-transfer-torque magnetic random access memory, or STT-MRAM.[1] For the approach to be competitive commercially, it is important to scale the device structures below 30 nm while maintaining a strong magnetic anisotropy energy for data retention. At the same time, one needs to keep the switching current minimized for reducing the load on the bit-selection transistor and for power savings. The magnetic anisotropy energy E_b needs to stay above 60 or so k_BT where T is the ambient temperature. The spin-torque switching involves a threshold current I_{c0} (or voltage V_{c0} for a tunnel junction) whose value is, in an ideal macrospin limit, proportional to E_b. A macrospin limit is equivalent to setting the exchange energy $A_{ex} \to +\infty$. For an ideal tunnel junction in a pure uniaxial anisotropy potential, the threshold *voltage* is symmetric in magnitude for parallel-to-antiparallel (P-AP) and antiparallel-to-parallel (AP-P) switching.[2,3] At low tunnel junction bias when the conductances are nearly bias independent, this voltage is approximately[2,3]

$$V_{c0} \approx \left(\frac{2e}{\hbar\eta}\right)\left(\frac{\alpha}{G_P}\right) m H_k, \tag{1}$$

where m is the total magnetic moment of the macrospin, H_k is the uniaxial anisotropy field of the macrospin, whose axis is here assumed to be in collinear alignment with the spin current's polarization, and $\eta = \sqrt{m_r(m_r+2)}/2(m_r+1)$ is the spin-polarization factor related to the tunnel magnetoresistance (TMR) value $m_r = (R_{ap} - R_p)/R_p$.[4] Obviously, $I_{c0} \approx G_p V_{c0}$, and the threshold current density

$$J_{c0} \approx \left(\frac{4e}{\hbar\eta}\right)\alpha\left(\frac{1}{2}M_s d H_k\right) \approx \left(\frac{4e}{\hbar\eta}\right)\alpha\Sigma_{\text{eff}}, \tag{2}$$

where M_s is the free-layer film's saturation magnetization, and d is its thickness. Note the quantity $\frac{1}{2}M_s d H_k \to \Sigma_{\text{eff}}$ can be viewed sometimes as an areal density of anisotropy energy, especially in the ultrathin film limit discussed below.

The spin-torque switching efficiency κ is empirically defined as the ratio $\kappa = E_b/I_{c0}$. In the macrospin limit,

$$\kappa \equiv \frac{E_b}{I_{c0}} = \left(\frac{\hbar}{4e}\right)\left(\frac{\eta}{\alpha}\right) \approx \left(\frac{\hbar}{4e}\right)\frac{\sqrt{m_r\,(m_r+2)}}{2\alpha\,(m_r+1)} \tag{3}$$

for the P-AP switching.[3]

The validity of the macrospin assumption depends on a comparison of length scales. A finite exchange energy A_{ex} of the ferromagnet gives rise to several length scales, originating from the competition between an exchange energy penalty due to spatially inhomogeneous magnetization orientation and the change of various other energy terms. A basic expression of the total energy density at a given point $\mathbf{r}$ inside a ferromagnet with order parameter $\mathbf{M} = M_s\mathbf{n}_m(\mathbf{r})$ is

$$K_v(\mathbf{n}_m,\mathbf{r}) = K_{\text{ani}} - M_s\mathbf{n}_m\cdot\mathbf{H}_{\text{eff}} + A_{ex}|\nabla\mathbf{n}_m|^2, \tag{4}$$

where M_s is the ferromagnet's saturation magnetization and is assumed to be uniform throughout the material, $\mathbf{n}_m(\mathbf{r})$ is the local magnetization direction, assumed to be position ($\mathbf{r}$) dependent. K_{ani} is the total materials-related anisotropy which gives rise to a local $\mathbf{n}_m$ dependent energy, and $\mathbf{H}_{\text{eff}}$ is the total effective magnetic field acting on the volume element at location $\mathbf{r}$, including any applied field $\mathbf{H}_{\text{appl}}$ and dipolar field $\mathbf{H}_{\text{dipole}}$ from the ferromagnetic moment elsewhere. Note that K_{ani} and $\mathbf{H}_{\text{eff}}$ can both be position $\mathbf{r}$ dependent by themselves in addition to the position dependence in $\mathbf{n}_m$. The spatial derivative in the exchange energy density term is[5] $|\nabla\mathbf{n}_m|^2 = (\nabla n_{mx})^2 + (\nabla n_{my})^2 + (\nabla n_{mz})^2$ in a Cartisian coordinates system $\{\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z\}$ where $\nabla = \partial_x\mathbf{e}_x + \partial_y\mathbf{e}_y + \partial_z\mathbf{e}_z$, and $\mathbf{n}_m = n_{mx}\mathbf{e}_x + n_{my}\mathbf{e}_y + n_{mz}\mathbf{e}_z$. Within the effective magnetic field $\mathbf{H}_{\text{eff}}$ in the context of a perpendicularly magnetized thin film nanomagnet, an important contributor is the effective demagnetization field from $\mathbf{n}_m$, which in the thin-film limit where film thickness d is small compared to the rest of the structures is often reduced to a simple $-4\pi M_s$ in the direction normal to the film thickness when $\mathbf{n}_m$ is perpendicular, giving rise to a so-called demagnetization energy $2\pi M_s^2$ to the $-M_s\mathbf{n}_m\cdot\mathbf{H}_{\text{eff}}$ term in Eq. (4). The competition between exchange energy and this demagnetization gives rise to a

dipolar exchange length for such thin films that is

$$\lambda_d = \sqrt{A_{ex}/2\pi M_s^2}. \tag{5}$$

In thin films with perpendicular magnetic anisotropy (PMA), $K_{\rm ani}$ in Eq. (4) contains a driving anisotropy term that is perpendicular to film surface, and is to the lowest order uniaxial in symmetry, meaning the total energy varies as $\sin^2\theta$, where θ is the angle between $\mathbf{n}_m$ and the film normal. This may originate from either the bulk or the interfaces of the thin film. One can in either case write an effective uniaxial volume PMA energy density as $K_\perp$. Then it is convenient to write the total perpendicular anisotropy for such a thin film as K_{pma} in the form of

$$\begin{cases} K_{pma} = K_\perp - 2\pi M_s^2 = \frac{1}{2} M_s H_k, \\ Q = K_\perp / \left(2\pi M_s^2\right) = \frac{H_k}{4\pi M_s} + 1, \end{cases} \tag{6}$$

where one has defined the thin film's net measurable hard-axis anisotropy field as H_k, and the "quality factor" Q that describes the strength of PMA in relation to demagnetization. Obviously, $Q = 1$ describes the strength of $K_\perp = 2\pi M_s^2$ which is necessary to bring the moment out of the plane. The larger Q is, the stronger the total PMA density. The competition of the exchange energy with this total PMA strength leads to a PMA exchange length

$$\lambda_{pma} = \sqrt{\frac{A_{ex}}{K_{pma}}} = \sqrt{\frac{2A_{ex}}{M_s H_k}}. \tag{7}$$

It follows that $\lambda_{pma} = \lambda_d/\sqrt{Q-1}$. An ultrathin-film limit is defined as the film thickness $d \ll \min(\lambda_{pma}, \lambda_d)$. For a typical CoFeB-based thin-film MTJ considered here, one has $d \approx 2$ nm, $M_s \approx 900$ emu/cm^3, $H_k \approx 4.5$ kOe, with A_{ex} within a range 2–6 $\times 10^{-6}$ erg/cm,[6] thus one has $Q \sim 1.4$, $\lambda_d \approx 6$–10 nm, and $\lambda_{PMA} \approx 10$–17 nm. So $d \ll \lambda_d$ is satisfied, and one is within the ultrathin-film limit where the magnetic moment can normally be considered position independent along the direction of the film norm.

In more detailed analysis, the locational nature of $K_\perp$ could also be significant, especially when $K_\perp$ is dominated by *interface*-originated anisotropy energies. In such a situation, an interface anisotropy energy density Σ_s would introduce another nontrivial length scale:

$$\lambda_{spma} = \frac{A_{ex}}{\Sigma_s} \tag{8}$$

whose significance will become more apparent in later discussions.

When device sizes are comparable or larger than these length scales, the free-layer typically exhibits nonmacrospin behavior, which also causes a coherent thermal fluctuation length that is shorter than the device size, resulting in a reduced thermal activation energy. For such relatively large devices, the thermal activation energy is limited by shorter length-scale fluctuations mostly related to λ_{pma},[6–9] that could nucleate magnetic reversal events, and is not simply proportional to the lateral area of the device. As discussed above, λ_{pma} for our CoFeB-based free layer is of the order of 10–17 nm—a small lateral size only became lithographically accessible fairly recently.

For the smaller devices, the more fundamental macrospin-dictated switching efficiency Eq. (3) is being approached.[10,11] Since Eq. (3) contains only two relatively well-known materials and device parameters ($\alpha \sim 0.005$–0.01 for most free layer materials in use today, and $m_r \approx 0.5$–1.0 for a high-MR, low-RA MTJ), the macrospin-limit efficiency typically evaluates to $\kappa \approx 1\ k_B T/\mu A$ where T is the ambient temperature of 300 K.

For devices with sizes below 30 nm (conductances below about 0.1 $1/k\Omega$), a range of efficiencies between 1 and 10 $k_B T/\mu$A has been seen. The lower end of this range, about 1 $k_B T/\mu$A, agrees well with the macrospin-based estimate Eq. (3). The high end of the efficiencies are about three times higher than the macrospin model would comfortably expect. Similarly high efficiency values have also been reported in recent literature on individual junction devices by other researchers.[12,13]

This paper will attempt to analyze the dependence of efficiency κ on MTJ size and other properties, and propose a few plausible mechanisms that may account for the larger than macrospin values for κ.

II. EXPERIMENT

Tunnel junctions with MgO barrier and CoFeB as free layer with perpendicular magnetic anisotropy were produced in ways similar to what has been reported earlier.[11] These include the sputter deposition, at ambient temperature, a CoFeB-based MTJ materials stack. The wafers are then postdeposition annealed in vacuum at 300 °C for 1 h prior to being lithographically patterned down to sizes ranging from about 15 nm to >100 nm in diameter for circular-shaped devices in this series of samples. A reactive ion etch is used for the main junction etching step, followed by a low-energy (<200 eV), grazing incidence Ar ion-beam etch for trimming the junction sides to the desired dimensions. The finished structures are characterized for their spin-torque switching properties using methods described in Ref. 8.

These junction devices are used to deduce a set of junction-size dependent quantities including thermal activation energy E_b, spin-torque switching threshold voltage V_{c0}, spin-torque switching threshold current I_{c0}, as evaluated from V_{c0}/R_p, and the switching efficiency E_b/I_{c0}. A detailed description of these measurements in our experiments can be found in Refs. 7 and 8.

Figure 1 shows a set of switching efficiency vs parallel-state junction conductance G_p and vs junction TMR plots. Data include devices from multiple wafers of the general materials parameters as described above, with the dominant variation of G_p originating from junction size differences, although a junction RA variation of about 5–20 $\Omega\mu$m^2 is also part of the variables convoluted in this summary. For larger devices where E_b is relatively constant against size variations (consistent with earlier findings[7,8]), their efficiency κ is inversely proportional to the junction area, reflecting a relatively size-independent E_b, and a junction I_{c0} scaling with junction area. In the small G_p region below 0.1 (kΩ)$^{-1}$, corresponding roughly to junctions of about 30 nm or so in size and below, the data in Fig. 1(a) suggest a change of slope, caused by a decrease of E_b for smaller junction size. The exact mapping of junction size at the small-size end from G_p however may be unreliable, as there are variations of device-level junction resistance-area product

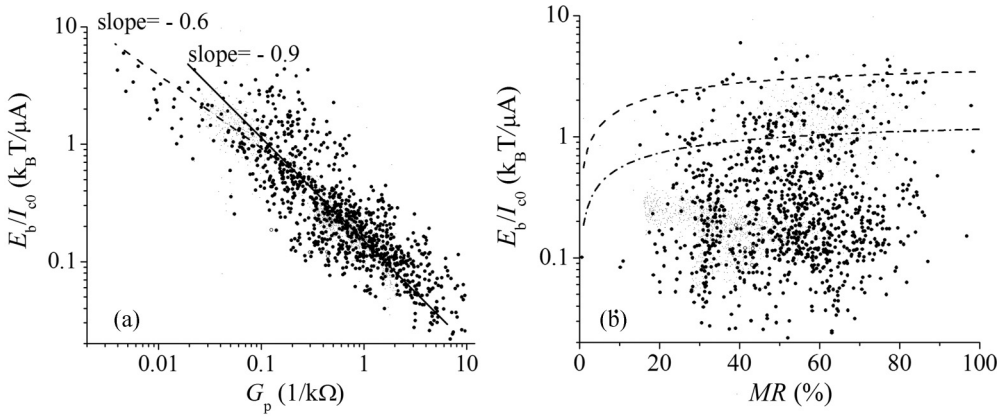

FIG. 1. An overview plot of the measured efficiencies of spin-torque switchable MTJs of different sizes. (a) The dependence of observed efficiency κ vs junction size. Here the parallel-state conductance G_p is used to represent the junction area, with an averaged junction resistance-area product ranging between 5 and 20 $\Omega\mu m^2$. Junctions are all of circular shape. (b) The same set of devices showing the efficiency vs junction's MR. The two dashed lines are comparisons to Eq. (3) with $\alpha = 0.005$ (upper) and 0.015 (lower).

that are process and junction size dependent. Regardless, the switching efficiency is seen to increase with decreasing device size. This is expected as the device size is decreased towards the magnetic exchange lengths. Data in Fig. 1 suggest that for the majority of devices the macrospin-level efficiency of $\kappa > 1\ k_BT/\mu A$ is not reached. Most of that is due to the device size being larger than the exchange length. However, for those very small devices at the lowest conductance end in Fig. 1(a), and on the top boundary of Fig. 1(b), they have κ values exceeding $3\ k_BT/\mu A$, some even with apparent values exceeding $5\ k_BT/\mu A$. For such small devices it is important to unambiguously establish the physical size of the same junction the transport measurements were performed on, so as to ascertain the relationship between E_b, κ, and device size.

Figure 2 gives a subset of devices shown in Fig. 1. Here the transport-measured properties are directly plotted against their individual physical device sizes as determined from cross-sectional transmission-electron microscopy (TEM) images obtained post-transport measurement. Figures 2(a) and 2(b) indicate the thermal barrier E_b decreases with decreasing device size when devices are smaller than about 40–50 nm in diameter. For such small devices, Fig. 2(b) shows that E_b scales almost linearly with device diameter, rather than with device area—as a comparison with Fig. 2(a) would demonstrate. On the other hand, for switching current threshold I_{c0}, Fig. 2(c) shows a scaling over this entire junction size range that is essentially linear with device area, with a near-zero intercept. This particular set of experimental observations, namely with $E_b \propto a$ and $I_{c0} \propto a^2$ (where a is the device diameter), leads to an efficiency $\kappa \propto 1/a$, and is the apparent reason the observed efficiency keeps increasing with decreasing junction size a, as shown in Fig. 2(d). At the small device sizes around 17 nm, one observes an efficiency around $3.5\ k_BT/\mu A$ which stretches the best estimate one could expect from a macrospin-based model according to Eq. (3).

In sections following, one examines these two size-dependent scaling behaviors in more detail—that of E_b vs a, and I_{c0} vs a^2. This examination leads one to conclude that the $E_b \propto a$ scaling results mostly from an edge-demagnetization field correction, whereas the $I_{c0} \propto a^2$ scaling may be indicative of a spin-wave instability threshold that is governed by the combined interface and bulk anisotropy area density.

III. NEARLY LINEAR DEPENDENCE OF E_b ON DIAMETER

In the thin-film limit it is often assumed that the total PMA energy density described by Eq. (6) is sufficient to estimate the total thermal activation barrier height in the macrospin limit. This would give rise to an $E_b = K_{pma} d(\frac{\pi}{4})a^2$ for a circularly shaped film disk where d is the film thickness and a is the film diameter. Such E_b should scale with disk area. The experimentally observed E_b however shows predominantly a *linear* dependence on a for the size ranges between around 15 and 50 nm with nearly zero intercept, as shown in Fig. 2(b).

There could be several plausible sources for an E_b that would depart from simple area dependence on disk size. Below one evaluates a few of the most likely causes quantitatively to identify the leading cause.

A. Reversal through a domain-wall sweep across the disk

Here one is concerned mostly with disk-shaped ferromagnetic thin films with diameter a approaching that of the exchange length λ_d. In this size range one candidate for magnetic reversal is a domain-wall-like object sweeping across the disk.[14] This would produce a total thermal activation energy roughly proportional to the length of such a wall-like structure, and could potentially explain the linear size dependence of the observed E_b. A closer examination of this mechanism however shows that it is unlikely to be the dominant reason for the linear size dependence of E_b in this set of experiments.

The maximum energy point of such a wall sweep process is the symmetric situation illustrated in the inset of Fig. 3, where the center of the wall-like structure crosses the disk, lying along its diameter in, say, the $\mathbf{e}_y$ direction. In fact, if one ignores long-range dipolar interaction, but includes the related thin-film demagnetization into the local PMA

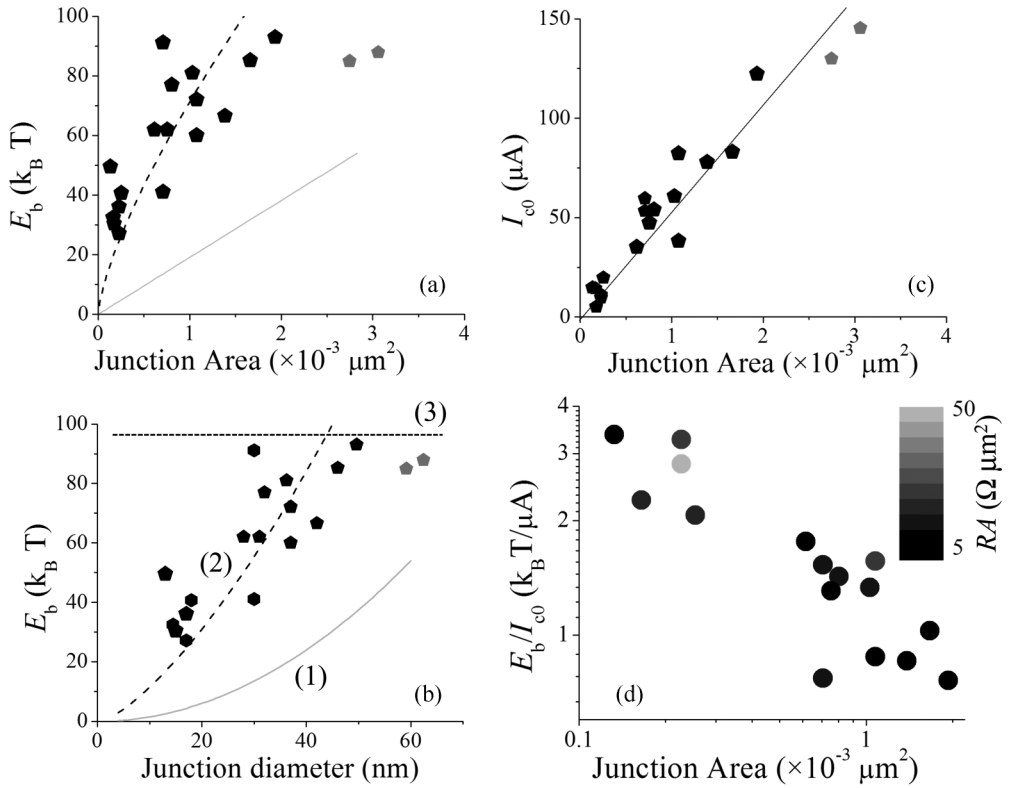

FIG. 2. Size dependence of junction properties. Here for most devices [except the relatively large ones marked in gray pentagons in (a)–(c)], the size of the actual junction is confirmed post-transport measurement with cross-section transmission electron microscopy (TEM) images. (a) The dependence of thermal activation energy on junction area, showing the initial rise followed by saturation at larger junction area. Note that the initial rise does not follow a simple linear area dependence with zero intercept but rather there is a significant intercept at zero device area. (b) The dependence of E_b on the diameter of junction, showing a large region of nearly linear dependence with zero intercept before reaching saturation E_b (Refs. 6–9) as indicated by line (3). (c) The dependence of switching current on junction area showing the expected linear dependence with zero intercept. (d) The dependence of spin-torque efficiency $\kappa = E_b/I_{c0}$ on junction area. The gray scale represents the actual measured resistance-area product RA of the junction. It shows that the effect of RA on κ is minor compared to κ's very strong size dependence.

energy, the one-dimensional domain-wall solution of $\theta(x,y) = 2\tan^{-1}[e^{(x/\lambda)}]$[16] would be a good approximation for a thin-film disk with uniform total perpendicular anisotropy density K_{pma}. Here θ is the polar angle of the local magnetization with respect to the easy-axis (film-perpendicular) direction of $\mathbf{e}_z = \mathbf{e}_x \times \mathbf{e}_y$; $-a/2 < x < a/2$ with x being the horizontal coordinate, and $\lambda = \lambda_{pma} = \sqrt{A_{ex}/K_{pma}}$. Integrating throughout the volume of the disk in the thin-film limit with thickness $d \ll \lambda$, and one gets a wall-sweeping activation energy (measured in reduced unit of $K_{pma}d\lambda^2$) as a function of the disk diameter a as the straight line 0-A in Fig. 3. The cross point at 4.63λ is the diameter above which domain-wall sweep yields a lower barrier height energy than macrospin rotation, and thus becomes the preferred mode of reversal.

This reveals a significant problem with the estimate. The crossover point thus estimated (point A in Fig. 3) lies well above the subvolume saturation value of E_b. This is illustrated in Fig. 3. The subvolume saturation E_b, roughly estimated to be of the order of $4\pi A_{ex}d$ (which was already an overestimate when compared with experiment), is seen to be crossed by the macro-spin branch of the $E_b(a)$ at point B, with $a \sim 4\lambda < 4.63\lambda$, the latter being what is for crossing over into domain-wall sweeplike reversal. If this simple picture were correct, one would not see the presence of a domain-wall sweep-mediated reversal mode, and the macrospin mode would directly cross into subvolume reversal with saturated $E_b \sim 4\pi A_{ex}d$. Thus, if the domain-wall sweep mode is what gave rise to the linear E_b vs a scaling, to be consistent with experimental observations, the $E_b(a)$ relationship has to have a shallower slope than predicted by this simple model described above, and intersect the macrospin reversal E_b at a much shorter length than 4.63λ as illustrated in Fig. 3.

Although a more careful inclusion of long-range dipolar interactions into such a picture could moderate the DW-like structure's maximum energy, it is still unlikely that it could be the leading-order effect to explain our observation. This is especially the case when an alternative factor appears to be more significant in controlling the leading order E_b vs size behavior, as will be discussed below.

B. Edge-enhancement of K_{pma}

As described by Eq. (6), the net PMA energy volume density K_{pma} is a balance between the perpendicular anisotropy energy $K_\perp$ (bulk or interface, here represented in volume density; if interface, then $K_\perp$ would contain a $1/d$ dependence

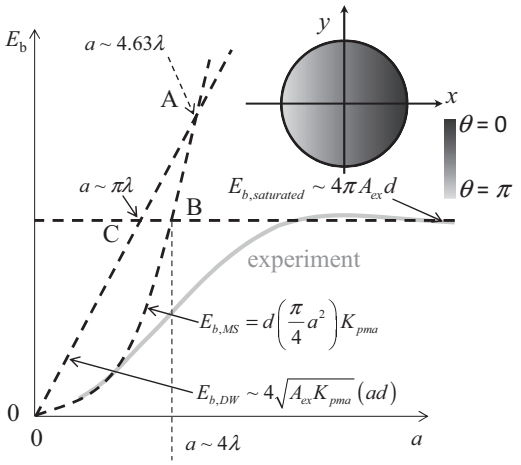

FIG. 3. A sketch of the various model predictions of thermal activation energy vs disk diameter a compared with experimental observations. Here d is the disk thickness. $E_{b,MS}$ is the macrospin energy barrier, $E_{b,DW}$ is the energy barrier for a domain-wall-like object sweeping across the disk, $E_{b,\text{saturated}} \approx 4\pi A_{ex} d$ is the subvolume related saturation activation energy (Refs. 6–9). The gray curve illustrates our experimentally observed behavior. The inset shows the energy maximum state of a domain-wall-like object sweeping across the disk. The gray scale represents the polar angle of $\mathbf{n}_m$ with disk-film normal $\mathbf{e}_z = \mathbf{e}_x \times \mathbf{e}_y$. The choice of K_{pma} here, and the related relationship between the macrospin branch and the experimental curve is only for illustrating the difference in trend. Scaling relations with "$\sim$" symbols are only crude estimates, while those with "$=$" are exact predictions. A more quantitative comparison is given by the model calculation presented in Fig. 2.

inside) and the shape-dictated dipolar demagnetization energy density $-2\pi M_s^2$. Such definitions are based on *extended* films over a length scale far greater than the film thickness. In any patterned structures with finite size, the edge demagnetization field is reduced. Based simply on symmetry, a half-infinite plane of film would have the edge demagnetization field only of 1/2 the value of that deep inside the film. Thus an edge demagnetization energy density at most would be only about $-\pi M_s^2$. Therefore, even if the intrinsic PMA energy density $K_\perp$ is uniform and position independent, one would have a total PMA energy density K_{pma} showing an enhancement near sample edges due to the *reduction* of demagnetization.

One may estimate the amount of PMA gain from this mechanism crudely by integrating the edge-fringe field related loss to dipolar demagnetization energy, assuming the disk diameter is far greater than the film thickness. The demagnetization field distortion near the film edge extends at least the order of a film thickness d into the lateral direction of the film. Thus the total energy gain from this mechanism is of the order of $\eta_g \pi^2 (M_s d)^2 a$ where η_g is a geometry dependent integration constant of the order $\sim$1. This would give

$$E_b \approx \left(\frac{\pi}{4}\right) a^2 d K_{PMA} + \eta_g a (\pi M_s d)^2 \tag{9}$$

giving a leading order linear dependence in the small a limit of

$$\frac{\delta E_b}{\delta a} = \eta_g \pi^2 (M_s d)^2 . \tag{10}$$

The geometry constant η_g can be roughly estimated by using a long stripe thin film with width $L \to +\infty$, thickness d, with perpendicularly magnetized moment density M_s. In such a case, with $-L < x < 0$ and $x = 0$ defined as one edge, the near edge-dipole related field profile can be calculated to be, for the edge near $x = 0$,

$$H(x) \approx \left[2\pi + 4\cos^{-1}\left(\frac{d}{\sqrt{d^2 + 4x^2}}\right)\right] M_s \tag{11}$$

and thus the per-unit edge-length dipole demagnetization energy to be, at distance s into the plate in the $-x$ direction,

$$\begin{aligned} E_{ed} &= -\left(\frac{d M_s^2}{2}\right) \int_{-s}^{0} H(x) dx \\ &= \left(2\pi M_s^2\right)(sd) - (M_s d)^2 \left[1 + \ln\left(\frac{2s}{d}\right)\right], \end{aligned} \tag{12}$$

where the first term is the uniform demagnetization field $4\pi M_s$'s contribution, and the second term is the edge-dipole related energy reduction per unit length. If one crudely assumes the integration depth s to be around the radius $(a/2)$ of our disk, this gives an estimate to the η_g value in Eq. (9) as

$$\eta_g \approx \frac{1}{\pi}\left[2 + \ln\left(\frac{a}{d}\right)\right]. \tag{13}$$

For MTJs in this study, $M_s \approx 900$ emu/cm^3, $d \approx 2$ nm. Thus $\frac{\delta E_b}{\delta a} \approx 1.3\, k_B T$/nm. While this is an extremely crude estimate considering the logarithmic nature of the edge dipole energy's distance dependence, it is of the same order of magnitude as the experimental value of around 2 $k_B T$/nm, as shown in Fig. 2(b).

A slightly more refined way of estimating the edge-dipole field reduction of a disk geometry is to directly integrate out the total demagnetization energy for the disk shape. This can be done numerically, again assuming a uniform local magnetization orientation leading to an effectively uniform surface magnetic charge density M_s, giving a magnetic field inside the disk in the $\mathbf{e}_z$ direction along the disk axial direction as $H_z(\mathbf{r}_2) = M_s \int 2(\mathbf{r}_{12} \cdot \mathbf{e}_z / |\mathbf{r}_{12}|^3) d^2 r_1$, where $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ is the relative position vector. $\mathbf{r}_1$ resides on one of the disk surfaces while $\mathbf{r}_2$ is inside the disk. With a rigid local magnetization assumed to be perpendicularly aligned, one only concerns the $\mathbf{e}_z$ component of the demagnetization field. Integrating once more over the disk's entire volume for $\mathbf{r}_2$, one has the disk's total contribution to energy barrier height due to demagnetization as

$$U_{\text{demag}}(d,a) = -\left(2\pi M_s^2\right)(da^2\pi/4)\left[\frac{3}{2} N_z\left(\frac{2d}{a}\right) - \frac{1}{2}\right], \text{ with}$$

$$N_z(\xi) = \frac{1}{2\pi^2} \int \frac{2(\mathbf{r}_{12} \cdot \mathbf{e}_z)}{|\,\mathbf{r}_{12}\,|^3} d^2 r_1 d^3 r_2, \tag{14}$$

where $d^2 r_1$ gets integrated through a surface of the disk, and $d^3 r_2$ gets integrated through the volume of the disk. The integral runs through a disk of normalized radius 1 and the normalized thickness $\xi = 2d/a$, and can be numerically evaluated. $\lim_{\xi \to 0} N(\xi) = 1$ is the $\mathbf{e}_z$ direction demagnetization factor of the disk in the limit of zero thickness. For the first line in Eq. (14) one also employed an approximate sum rule[17] of $2N_x + N_z = 1$ for the demagnetization factors as a way to offset the effect of in-plane demagnetization of the disk at

finite thickness when the moments are rigidly aligned in the plane of the disk—a position that corresponds to the energy maximum during reversal of a macrospin state.

This estimate would give, in place of Eq. (9):

$$E_b = (\pi a^2 d/4)K_\perp + U_{\text{demag}}(d,a). \tag{15}$$

For our experimental data, assuming $M_s \approx 900$ emu/cm^3, $d = 2$ nm, and $K_\perp = 5.5 \times 10^6$ erg/cm^2, one has the calculated E_b from Eq. (15) shown in Figs. 2(a) and 2(b) as the short-dashed line labeled (2) in Fig. 2(b). This produces a size dependence far from quadratic in diameter a, and agrees with the experimental data. The gray curve (1) in Fig. 2(b) is the E_b of the same parameter set, but assuming a simple uniform demagnetization energy of $-2\pi M_s^2$. Of this model calculation the only adjustable parameter without prior independent measurement is the value of $K_\perp$.

Based on these estimates, one concludes that the edge dipolar field reduction-related PMA enhancement is significant, and it could largely explain the observed size dependence of E_b in our device size range. Further refinement of the energy calculation would need to include both exchange and the edge enhancement that would produce more realistic size dependencies over wider size ranges. These however are too complex for analytical estimates, and would have to be evaluated with micromagnetic computation.

IV. SCALING OF SWITCHING THRESHOLD I_{c0} ON JUNCTION DIAMETER

The previous sections have established, both in terms of experimental observations and in investigations of mechanisms, a thermal activation energy's scaling relation with device size that is not areal but more linear. It then begs the question of how would the spin-torque switching threshold's current scale with device size. In the experimental data shown in Fig. 2(c), the junction's switching current I_{c0} is seen to scale robustly as a^2. This implies a constant current *density* scaling, without being much moderated by the reduction of the measured E_b in small devices. This is clearly an unexpected behavior from a macrospin point of view, as in macrospin limit, the switching current is expected to scale as the barrier height E_b. The observed nonareal scaling of E_b apparently does not get translated into a nonareal scaling of I_{c0}.

If I_{c0} is not scaling with the barrier height as macrospin would dictate, what determines it? The slope of I_{c0} vs device area of the data shown in Fig. 2(d) gives a threshold current density of $J_{c0} \approx 5.4 \times 10^6$ A/cm^2. The question then becomes, what determines this experimentally observed threshold current density, if not E_b? The rest of this section is going to address this question from a spin-wave excitation threshold point of view.

Before going into such detailed model considerations, one simple observation can be made here with the experimental data. If one uses the J_{c0} expression derived based on macrospin type of model as in Eq. (2), one sees that J_{c0} can be related to an areal anisotropy energy density Σ_{eff}. In macrospin limit, that is the areal density of the net uniaxial anisotropy $K_{pma}d$ where d is the film thickness. In our case here, using the slope from Fig. 2(c) and $m_r \sim 0.5$, $\alpha \sim 0.005$ as assumed before, and using Eq. (2), it appears the experimentally obtained effective areal energy density is about $\Sigma_{\text{eff}} \sim 0.66$ erg/cm^2. Note that our extended-film measured $H_k \sim 4.5$ kOe would give a net PMA energy density of $\Sigma_{pma} = M_s d H_k/2 \sim 0.41$ erg/cm^2, and the demagnetization energy density corresponding to $M_s \sim 900$ emu/cm^3 is $\Sigma_{\text{demag}} \approx 1.02$ erg/cm^2. In an extended film limit, $\Sigma_{\text{eff}} = \Sigma_\perp - \Sigma_{\text{demag}}$, which leads to an intrinsic PMA energy density of the order $\Sigma_\perp \triangleq K_\perp d \approx 1.43$ erg/cm^2. These areal energy densities are within a factor of 2 or so of what is deduced from J_{c0}, possibly implying a correlation.

A. Spin-torque excitation with interface and bulk perpendicular anisotropy

As well known in the past,[19,20] magnetic anisotropies concentrated at interfaces of thin-film ferromagnets can alter their spin-wave excitation behaviors, and under favorable conditions encourage the formation of interface spin waves. More recently, interface-localized perpendicular anisotropy has been shown to induce a particular type of softening in spin-wave modes that could reduce the spin-torque excitation threshold for inducing spin-wave instabilities in thin slabs of YIG crystals.[21] Below the same methodology is explored to relate our observations above to such interface-anisotropy related possible softening of spin-wave modes.

To this end in a classical-dynamics limit one can start with the continuous-medium Landau-Lifshitz-Gilbert equation for a description of the magnetodynamics of the free layer of the MTJ:

$$-\frac{1}{\gamma}\left[\left(\frac{\partial \mathbf{n}_m}{\partial t}\right) - \alpha \mathbf{n}_m \times \left(\frac{\partial \mathbf{n}_m}{\partial t}\right)\right] = \mathbf{n}_m \times \mathbf{H} + \left(\frac{2A_{ex}}{M_s}\right)\mathbf{n}_m \times \nabla^2 \mathbf{n}_m, \tag{16}$$

where $\mathbf{n}_m = \mathbf{n}_m(x,y,z)$ is the unit vector of the local magnetic moment. At an interface of the ferromagnet with a nonmagnet, one writes the boundary conditions for $\mathbf{n}_m$ to be[19–21]

$$\mathbf{n}_m \times \left[\left(\frac{2\gamma A_{ex}}{M_s}\right)\left(\frac{\partial \mathbf{n}_m}{\partial \mathbf{n}}\right) + \left(\frac{2\gamma \Sigma_s}{M_s}\right)(\mathbf{n}_m \cdot \mathbf{n})\,\mathbf{n} - h_s\,(\mathbf{n}_s \times \mathbf{n}_m)\right] = 0, \tag{17}$$

where $\mathbf{n}$ is the interface normal unit vector pointing *into* the ferromagnet. Σ_s is the interface anisotropy energy. $\Sigma_s > 0$ in this sign convention of $\mathbf{n}$ defines a perpendicular interface anisotropy. h_s is the spin current in units of $(\gamma/M_s)\,J_s$ where J_s is the conventionally defined magnetic moment flow density, in magnetic moment per unit area per unit time. One assumes here that the spin-torque term applies only to the bottom interface—a simplifying assumption that is justified by the very short spin-decoherence length in ferromagnets with strong s-d exchange coupling. $\gamma = g\mu_B/\hbar \approx 2\mu_B/\hbar > 0$ is the gyromagnetic ratio, here defined as a positive number. To keep the problem to a manageable size, one ignores dipolar interactions between the magnetic free layer and the reference layers inside an MTJ, and focuses instead on the basic dynamic problem of the free layer itself.

Further for simplicity, here one only examines the result in the thin-film limit, with film thickness $d \ll \min\{\lambda_d, \lambda_{spma}\}$, and attempts to find the characteristics of spin-torque excitation threshold at such an interface as it depends on the

materials parameters. To this end, assume that $\mathbf{n}_m \sim \mathbf{n} \parallel \mathbf{e}_z$, with $\mathbf{e}_z = \mathbf{n}$ defined as the film-normal direction. Assume further that only the bottom interface has a perpendicular anisotropy through Σ_s. The top interface is assumed to be free of either spin-current or interface anisotropy, for simplicity.

Equations (16) and (17) are linearized in the small amplitude limit of $\mathbf{n}_{mxy}$ in the x-y plane, with $|\mathbf{n}_{mxy}| \ll 1$, and $\mathbf{n}_m \approx \mathbf{e}_z + \mathbf{n}_{mxy}$. Fourier transform them into a plane-wave state and write $\mathbf{n}_{mxy} = \mathbf{n}_{m0} \exp[i(\omega t + \mathbf{k}\cdot\mathbf{r})]$, together with its accompanied oscillating magnetic field[22] $\mathbf{h}_{xy} = -4\pi M_s(\mathbf{k}\cdot\mathbf{n}_{mxy})(\mathbf{k}/k^2)$, satisfying the conditions of $\nabla\cdot(\mathbf{H} + 4\pi M_s\mathbf{n}_m) = 0$ and $\nabla\times\mathbf{H} = 0$ as required by the Maxwell equations, with $\mathbf{H} = (H_{\rm eff} - 4\pi M_s)\mathbf{e}_z + \mathbf{h}_{xy}$ the total magnetic field in Eq. (16), and $\mathbf{H}_{\rm eff} = \mathbf{H}_a$ the *volume* uniaxial anisotropy field of the film, also assumed to be along $\mathbf{e}_z$. The eigenvalue problem for this set of equations gives $\omega(k)$ dispersion relations whose imaginary parts' zeros yield the onset of magnetic instability. When expanded to the first order in damping α and spin current h_s and second order in the in-plane spin-wave wave vector k_x, the resulting instability threshold for spin-current density $J_{sc} = (M_s/\gamma)h_s$ is

$$\begin{aligned} J_{sc} &= 2\alpha\left(2\pi M_s^2 d\right)\left[Q + Q_1 - 1 + k_x^2 Q\left(\frac{A_{ex}d}{\Sigma_s}\right)\right. \\ &\quad\left.\times\left(1 + \frac{2Q + Q_1 - 1}{2Q^2}\right)\right] \\ &= 2\alpha d\left[K_{pma} + (2\nu A_{ex})k_x^2\right] \end{aligned} \tag{18}$$

with $Q = \Sigma_s/(2\pi M_s^2 d)$, and $Q_1 = H_a/(4\pi M_s)$ describing the interface and bulk perpendicular anisotropy quality factor for the thin film, respectively. The second line in Eq. (18) writes the total PMA anisotropy $K_{pma} = \Sigma_s/d + K_u - 2\pi M_s^2$ as discussed before, and $\nu = 1/2 + [Q + (Q_1 - 1)/2]/Q^2 \sim 1$ as a dimensionless parameter depending on details of Q and Q_1.

B. Spin-wave modes laterally confined by an edge enhancement of K_{pma}

At $k_x = 0$, Eq. (18) leads to $J_{sc} = 2\alpha d[\Sigma_s + M_s(H_a - 4\pi M_s)/2]$, which is the macrospin threshold spin-current density. For finite k_x excitations the threshold current density increases according to the dispersion relation of k_x as expected. Assuming a circular disk with diameter a, the net charge-threshold current from Eq. (18) for the lowest spin-wave branch of wavelength corresponding to $2a$ is

$$\begin{aligned} I_c &\approx \left(\frac{4e\alpha}{\hbar\eta}\right)\left[K_{pma}\left(\frac{\pi}{4}a^2 d\right) + \frac{\pi^3\nu}{2}A_{ex}d\right] \\ &= I_{c0} + I_{c1}, \end{aligned} \tag{19}$$

where $I_{c0} = (\frac{4e\alpha}{\hbar\eta})K_{pma}(\frac{\pi}{4}a^2 d)$, and $I_{c1} = (\frac{4e\alpha}{\hbar\eta})\frac{\pi^3\nu}{2}A_{ex}d$, which is the first available confined spin-wave mode, and its threshold current is an additional I_{c1} above that of a macrospin. Thus in normal conditions, the macrospin mode with $k_x = 0$ is the lowest-lying mode in terms of its spin-torque threshold.[23]

The situation could be different if K_{pma} is strongly position dependent. As discussed above, in the case where edge-demagnetization reduction plays a significant role in determining the total PMA energy, K_{pma} can become significantly enhanced near the edges of a circular disk. Consider an extreme limit, where most of the PMA energy for E_b is concentrated near the edge of the disk, and for the interior of the disk K_{pma} is essentially zero. In such a limit, the threshold for exciting the first spin-wave mode would be of the order I_{c1}, while the macrospin-threshold current for E_b would be I_{c0}. Assuming an excitation of the spin-wave mode allows sufficient spin transport into the precessing moments that could lead to an eventual reversal of the total magnetization, then, even though the moments are edge-pinned to a total potential of E_b, the threshold current for negative-damping related thermal activation would become I_{c1}, whose ratio to the macrospin threshold is

$$\frac{I_{c0}}{I_{c1}} \approx \frac{E_b}{\frac{\pi^3\nu}{2}A_{ex}d}. \tag{20}$$

This could contribute to a possible efficiency enhancement over a certain size range—due to the excitation of the first finite wavelength mode spin wave, which under the specific conditions of a K_{pma} being weak near disk center could result in $I_{c1} < I_{c0}$.

This mechanism however has the fundamental difficulty of I_{c1} being a size-independent quantity, and could thus not be consistent with the range of data observed in our experiments described above, where the threshold involves a relatively constant *current density*, while E_b scales roughly linearly with sample diameter. This dilemma could not be easily resolved by invoking the low-lying spin-wave modes of nonzero k_x, confined by an edge-enhanced PMA potential.

C. Interface PMA-related mode softening

Another materials-related possibility is a significant reduction of the exchange constant A_{ex} near the interfaces, which under certain conditions could create a softened interface spin-wave mode.[21] Assume an interface layer of thickness d_1 separating the bottom interface which Eq. (17) addresses and the main slab of the ferromagnet with a thickness of $d \gg d_1$. Further assume the bottom d_1 interface layer has an apparent A_{ex} that is reduced from the interior of the slab by a factor of β with $0 < \beta \ll 1$. Solving the linearized LLG equations with the properly defined boundaries gives a threshold current density J_{sc} for the simpler case of $k_x = 0$:

$$J_{sc} \approx 2\alpha\left(2\pi M_s^2 d\right)\left[Q + Q_1 - 1 + \left(\frac{\beta_c}{\beta}\right)\left(\frac{Q}{2} + Q_1 - 1\right)\right], \tag{21}$$

where $Q = \Sigma_s/(2\pi M_s^2 d)$, $Q_1 = K_u/(2\pi M_s^2)$ with K_u being the volume PMA of the interior of the ferromagnetic slab not including demagnetization energy, and $\beta_c \approx 2Qdd_1/\lambda_d^2 \ll 1$ is the critical softening value of β for the interface exchange energy βA_{ex}.

If the anisotropy strengths are such that $Q_1 < 1 - Q/2 < 1$, as $\beta \to \beta_c$, a large amount of reduction in J_{sc} due to this spin-wave mode softening could be expected, hence an increase of the efficiency factor κ. Quantitatively, if one writes J_{scms} as the macrospin threshold current, and J_{scsw} that of the spin-wave mode, then

$$\frac{J_{scms}}{J_{scsw}} \approx \frac{Q + Q_1 - 1}{Q + Q_1 - 1 + \left(\frac{\beta_c}{\beta}\right)\left(\frac{Q}{2} + Q_1 - 1\right)} \tag{22}$$

could become greater than 1 as $\beta \to \beta_c$.

The origin of this mode softening is the coupling strength limit of the weakened bottom layer exchange βA_{ex} in maintaining perpendicular moment of the full slab. The eigenfrequency of the full slab at zero spin current can be written as

$$\omega \approx \gamma\,(4\pi M_s)\left(Q + Q_1 - 1 - \frac{\beta_c}{\beta}Q\right). \qquad (23)$$

It shows the effect of $\beta < 1$ and approaching β_c as a reduction of the resonance frequency.

A more quantitative analysis would however need to proceed cautiously, bearing in mind the uncertainties of the materials model and parameters as well as the myriads of assumptions made for such a simplified model. For one thing, to safely stay within the ultrathin-film limit one would need to confine the discussion to a length scale where the interface anisotropy and exchange energy related length scale is beyond that of the materials thicknesses. That is to require $\frac{A_{ex}\beta}{\Sigma_s} \gg d_1$, or $\beta \gg d_1/\lambda_{spma}$. Since one could write $\beta_c = 2d_1/\lambda_{spma}$, it follows that staying within the thin-film limit requires $\beta \gg \beta_c$, thus the region of $\beta \to \beta_c$ cannot be approached too closely without violating the ultrathin-film assumption used in this toy model.

In reality of course the situation is even more complex. The softening of the interface spin-wave mode would have some consequences on the total thermal activation energy barrier as well which is not investigated here. However this extremely simplified analysis points to the possible importance of a weakened exchange energy at the interface region where both a majority portion of the PMA is induced as well as where the spin torque is absorbed—an optimal match of this exchange energy with the rest of the film would likely provide a threshold current less than that of the simple macrospin threshold.

Such lowering of the threshold spin current for negative damping instability does not necessarily translate into a lowering of threshold current for the *reversal* of the magnetic moment in the layer. What this simple analysis showed is only a reduction of the threshold for inducing an instability in the linear regime. The large-angle dynamics are too complex to be analytically predicted. Numerical simulations would be necessary to quantify the parameter space within which such reduction of threshold instability could correspond to complete reversal.

In addition to relating the linear instability threshold to reversal, for finite-time magnetic switching, there remains the fundamental requirement of angular-momentum conservation, and the slope of switching speed's dependence on spin current. The simple and robust relationship between switching time τ and driving spin current I_s (Ref. 24) remains. In the limit of $I_s \gg I_{sc}$, $\tau I_s \sim m$ where m is the total magnetic moment of the free layer, although the detailed form of the finite-temperature switching probability vs the spin-current intensity might be significantly altered by such interface spin-wave excitations.

Nevertheless, this scenario provides a plausible explanation to the fact of a threshold current that is area scaling, as the spin wave dictates a threshold current *density* regardless of the actual thermal activation barrier height's size-scaling behavior—even if the activation energy is saturated at large device sizes by subvolume excitations.[6–9] This could account for the separate scaling properties of E_b and that of I_{c0} observed experimentally as discussed earlier.

Furthermore, this conceptual model points out some basic trends on how efficiency enhancement may depend on materials parameters. This type of spin-wave-softening induced efficiency increase only occurs when interface PMA dominates and the bulk PMA is insufficient by itself to overcome demagnetization (i.e., $Q_1 < 1$, but $Q + Q_1 > 1$). Within this constraint, an increase of bulk PMA strength (increasing Q_1) would reduce the amount of possible reduction in J_{scsw}, and thus lessen the enhancement of the efficiency κ. These qualitative trends may be of value for materials and device optimization work.

V. CONCLUSIONS

An in-depth examination of spin-torque switched MTJs' threshold current in relation to same-device thermal activation energy suggests that the experimentally evaluated spin-torque switching efficiency borders on the maximum value a macrospin model would yield. In all likelihood additional mechanisms are at play. A set of size-dependence scaling behaviors were experimentally observed. For the thermal activation energy E_b, it is primarily linearly dependent on junction diameter. For threshold current, it scales with the area of the junction, over the size range 15 to over 100 nm. The linear scaling of E_b can be semiquantitatively accounted for by including an edge-fringe-field related reduction of demagnetization energy. The area scaling of threshold current is consistent with a spin-wave dictated threshold that may involve additional softening due to the interplay between an interface perpendicular magnetic anisotropy term and a net in-plane magnetic anisotropy from the interior of the free layer. An interface-dominant PMA with a net easy-plane anisotropy for the interior could provide a mechanism for a softened spin-wave mode that reduces its threshold current density to values below what is required by a macrospin.

ACKNOWLEDGMENTS

Work done with the MRAM group at IBM T. J. Watson Research Center was supported in part by partnerships with TDK/Headway Technologies and with Micron Technologies. Also acknowledged are fruitful scientific discussions with Prof. A. D. Kent's group at New York University, and with Dr. G. Chavez-Flynn, now at New Jersey Institute of Technology.

*jonsun@us.ibm.com

[1] N. D. Rizzo, D. Houssameddine, J. Janesky, R. Whig, F. B. Mancoff, M. DeHerrera, J. J. Sun, K. Nagel, S. Deshpande, M. L. Schneider, H.-J. Chia, S. M. Alam, T. Andre, S. Aggarwal, and J. M. Slaughter, IEEE Trans. Magn. **49**, 4441 (2013).

[2] J. C. Slonczewski, Phys. Rev. B **71**, 024411 (2005).

[3]J. C. Slonczewski and J. Z. Sun, J. Magn. Magn. Mater. **310**, 169 (2007).

[4]Strictly speaking, Eqs. (1) and (2) are approximations taken in the limit of large $m_r \gg 1$. See discussion in, for example, Ref. 3.

[5]C. Kittel, Phys. Rev. **110**, 836 (1958).

[6]M. Yamanouchi, A. Jander, P. Dhagat, S. Ikeda, F. Matsukura, and H. Ohno, IEEE Magn. Lett. **2**, 3000304 (2011).

[7]J. Z. Sun, R. P. Robertazzi, J. Nowak, P. L. Trouilloud, G. Hu, D. W. Abraham, M. C. Gaidis, S. L. Brown, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge, Phys. Rev. B **84**, 064413 (2011).

[8]J. Z. Sun, P. L. Trouilloud, M. J. Gajek, J. Nowak, R. P. Robertazzi, G. Hu, D. W. Abraham, M. C. Gaidis, S. L. Brown, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge, J. Appl. Phys. **111**, 07C711 (2012).

[9]H. Sato, M. Yamanouchi, K. Miura, S. Ikeda, R. Koizumi, F. Matsukura, and H. Ohno, IEEE Magn. Lett. **3**, 3000204 (2012).

[10]S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, Nat. Mater. **9**, 721 (2010).

[11]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, and D. C. Worledge, Appl. Phys. Lett. **100**, 132408 (2012).

[12]S. Yuasa (unpublished).

[13]G. Jan, Y.-J. Wang, T. Moriyama, Y.-J. Lee, M. Lin, T. Zhong, R.-Y. Tong, T. Torng, and P.-K. Wang, Appl. Phys. Exp. **5**, 093008 (2012).

[14]For devices of larger sizes compared to the exchange length, a curling mode equivalent to the formation of a vortex is often the energetically favored mode of reversal (Ref. 15). When the disk diameter is close to or shorter than the exchange length however, the curling mode becomes energetically less favorable as it would involve inhomogeneities at twice the spatial period as that of a single domain wall.

[15]R. Skomski, H.-P. Oepen, and J. Kirschner, Phys. Rev. B **58**, 3223 (1998).

[16]D. Craik, *Magnetism: Principles and Applications* (John Wiley & Sons, Chichester, England, 1995), p. 96.

[17]This sum rule is strictly true for three-dimensional ellipsoids as derived by, e.g., Stoner (Ref. 18).

[18]E. C. Stoner, Philos. Mag. Ser. 7 **36**, 803 (1945).

[19]G. T. Rado and R. J. Hicken, J. Appl. Phys. **63**, 3885 (1988).

[20]B. Hillebrands, Phys. Rev. B **41**, 530 (1990).

[21]J. Xiao and G. E. W. Bauer, Phys. Rev. Lett. **108**, 217204 (2012).

[22]C. Herring and C. Kittel, Phys. Rev. **81**, 869 (1951).

[23]There is a narrow parameter space of Q and Q_1, when $-1/2 < Q < 0$ and $1 < Q_1 + Q < 1 - Q - 2Q^2$ when the coefficient for k_x in Eq. (18) is negative, implying a reduction of finite wavelength spin-wave excitation by spin current, due to the competition of an in-plane interface anisotropy and that of a bulk perpendicular moment arrangement. In the ultrathin-film limit, which is the case one considers here, this parameter space is very narrow and is of little practical interest since the combined perpendicular anisotropy $Q + Q_1 < 1.125$ would severely limit the total PMA available.

[24]J. Z. Sun, Phys. Rev. B **62**, 570 (2000).