Advances in Magnetics

Switching Distributions for Perpendicular Spin-Torque Devices Within the Macrospin Approximation

W. H. Butler[1,2], Tim Mewes[1,2], Claudia K. A. Mewes[1,2], P. B. Visscher[1,2], William H. Rippard[3], Stephen E. Russek[3], and Ranko Heindl[3]

[1]Center for Materials for Information Technology, University of Alabama, Tuscaloosa, AL 35487 USA
[2]Department of Physics, University of Alabama, Tuscaloosa, AL 35487 USA
[3]National Institute of Standards and Technology, Boulder, CO 80305 USA

We model "soft" error rates for writing (WSER) and for reading (RSER) for spin-torque memory devices that have a free layer with easy axis perpendicular to the film plane by solving the Fokker-Planck equation for the probability distribution of the angle that the free layer magnetization makes with the normal to the plane of the film. We obtain: 1) an exact, closed form, analytical expression for the zero-temperature switching time as a function of initial angle; 2) an approximate analytical expression for the distribution function of the direction of the magnetization and the exponential decay of the WSER as a function of the time the current is applied; 3) comparison of the approximate analytical expressions for the distribution function and WSER to numerical solutions of the Fokker-Planck equation; 4) an approximate analytical expression for the distribution function and WSER for the case in which the pinned layer is not collinear with the perpendicular free layer; 5) an approximate analytical expression for the linear increase in RSER with current applied for reading; 6) comparison of the approximate analytical formula for the RSER to the numerical solution of the Fokker-Planck equation; and 7) confirmation of the accuracy of the Fokker-Planck solutions by comparison with results of direct simulation using the single-macrospin Landau-Lifshitz-Gilbert equations with a random fluctuating field in the short-time regime for which the latter is practical. We find that the WSER decays at long times as $\exp[-2(i-1)\tau]$ where the reduced time τ is related to the switching time, Gilbert damping and precession frequency through $\tau = \alpha\omega_0 t$, and the reduced current i is the ratio of the applied current to the critical current density for switching $i = \overline{I}/I_0$. This exponentially decaying tail in WSER is not easily reduced by tilting the pinned layer magnetization.

Index Terms—Error rate, Fokker-Planck, magnetic memory, spin torque, switching distribution.

I. INTRODUCTION

SPIN-POLARIZED electrical currents can transfer angular momentum between nanometer scale ferromagnetic electrodes separated by a nonmagnetic layer [1], [2]. The use of this effect to switch the direction of magnetization of a ferromagnetic layer as part of a magnetic memory device is of great current interest. In order to build a memory device one needs at least two distinct physical states that can be associated with the two logic states. In addition, one must have a means of switching the device between these states and a means of determining its state. A memory device is useful only if it switches (with very high probability) when switching is intended and does not switch (again with very high probability) when switching is not intended. In the most common type of spin-torque memory device, the two states are provided by the relative orientations of the direction of magnetization of two ferromagnetic layers, typically parallel and anti-parallel. The switching is achieved by the transfer of angular momentum carried by spin-polarized current, and the change in resistance between the parallel and anti-parallel states is used to determine the state of the memory device.

In this paper, we primarily consider spin-torque devices in which the magnetization of the two ferromagnetic layers is oriented perpendicular to the film plane in the quiescent state, i.e., in the absence of applied field or current. One layer is considered to have fixed magnetization, whereas the other layer's magnetization is free to process and switch. Such devices should switch at lower currents than devices (with equivalent thermal stability) for which the magnetization is in the plane of the layers in the quiescent state [3]. In Section V, we extend our treatment to include situations in which either the pinned or free layer is not perpendicular. At the present time, most of the effort aimed at practical spin-torque memory devices uses magnetic tunnel junctions because of the high spin polarization that can be achieved through the symmetry-based spin-filter effect [4]–[7] and because they offer the possibility of good impedance matching with a transistor that is used to select a particular device for reading or writing.

A desirable memory device should switch both quickly and reliably when switching is intended and it should not switch when switching is not intended, for example when current is applied to read the state of the device. Our concern will be the probability of switching events when switching is not intended [read soft-error rate (RSER)], or nonswitching events when switching is intended [write soft-error rate (WSER)].

A very large and rapidly growing literature on spin-torque switching, from the perspectives of both theory [1], [2], [8]–[25] and experiment [25]–[30], now exists. In the following, we will focus on the following aspects of the spin-torque

switching phenomenon: 1) Additional terms in the Landau-Lifshitz-Gilbert (LLG) equation that arise from Gilbert damping acting on the spin-torque; 2) exact integration of the zero temperature switching equation; 3) numerical solution to the Fokker-Planck equation giving the finite temperature distribution of switching elements; 4) approximate analytic solution to the Fokker-Planck equation valid for currents above the critical current for switching; 5) approximate analytic solution to the Fokker-Planck equation for the case of a pinned layer noncollinear with the perpendicular free layer or for a nonaxial magnetic field; 6) approximate analytic solution for the switching rate for currents below the critical current for switching; 7) demonstration that, for the purpose of determining switching distributions, the Fokker-Planck equation is equivalent to macrospin simulations that include a random thermal field, with the exception that the Fokker-Planck approach can be applied to determine switching probabilities that are extremely small or very close to unity, thus allowing the investigation of WSER and RSER; and 8) observation that, for cases with axial symmetry, the equations controlling switching via spin-polarized currents can be mapped onto mathematically equivalent equations for switching via an applied axial magnetic field.

A. LLG Equation for Film With Perpendicular Anisotropy

Consider a spin-torque transfer device consisting of pinned and free magnetic layers separated by a tunnel barrier. The lateral extent of the free layer in the x and y directions is assumed to be several times its thickness (in the z direction), but we require that all dimensions be smaller than the exchange length so that the approximation of a single direction for its magnetization is valid. The magnetization of the pinned layer is assumed to be fixed and parallel to the z direction. It is assumed that the fixed and pinned layers are separated by a tunnel barrier that eliminates any exchange coupling between the two layers.

Consider the evolution in time of the magnetization angular momentum $\mathbf{L}$ of the free layer (treated here as a macrospin, meaning that internal magnetic degrees of freedom are ignored) of a magnetic tunnel junction which is receiving electrons through a tunnel barrier from a pinned layer that has its magnetization pinned in the direction $\hat{\mathbf{m}}_p$. The rate of change of $\mathbf{L}$ can be written (in SI units) as [2], [31], [32], [50]

$$\frac{\partial \mathbf{L}}{\partial t} = -\mu_0 V \mathbf{M} \times \mathbf{H}_{\text{eff}} + \alpha \hat{\mathbf{m}} \times \frac{\partial \mathbf{L}}{\partial t} - \eta \frac{\hbar}{2} \frac{\overline{I}}{e} \hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p) + \frac{\hbar}{2e} f(\overline{I})(\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p) \quad (1.1)$$

where the first term on the right-hand side of (1.1) is the torque on the magnetic moment of the free layer, $V\mathbf{M}$ (V is the volume of the free layer and $\mathbf{M} = M\hat{\mathbf{m}}$ is its magnetization), due to the effective field, $\mathbf{H}_{\text{eff}}$; the second term is an empirical damping term designed to damp the precession of $\mathbf{L}$; and the third and fourth terms are spin-torque terms that arise from spin angular momentum carried by the electrons that tunnel through the barrier carrying angular momentum from the pinned layer to the free layer or vice versa. Each of these terms will be briefly discussed in turn.

The magnetic moment and angular momentum are related through the gyromagnetic ratio γ by $\mathbf{L} = M V / \gamma$, so the first term causes the angular momentum to precess in a direction perpendicular to both the effective field and the angular momentum. The effective field arises from magnetic anisotropy, demagnetizing effects, and any external field. We shall initially assume that all three of these contributions are perpendicular to the plane of the film. This assumption significantly simplifies the analysis. The additional assumption that $\hat{\mathbf{m}}_p$ is also perpendicular to the plane of the film is sufficient to provide the problem with axial symmetry. The requirement for axial symmetry and collinearity of $\hat{\mathbf{m}}_p$ and the external field with the film normal will be relaxed in Section V. To avoid ambiguity, we note that the gyromagnetic ratio γ, as used in this paper, is in units of *angular* frequency per Tesla, i.e., $\gamma \approx 1.76 \times 10^{11}$ radians/(Ts).

The second term is perpendicular to both the angular momentum and its precessional motion. This term drains energy from the precessing free layer moment and eventually causes the moment and angular momentum vector to align with $\mathbf{H}_{\text{eff}}$. Note that any term that causes torque (i.e., the third and fourth terms as well as the first) is expected to contribute to the damping because the semi-empirical Gilbert damping term is proportional to dL/dt, regardless of its origin.

The torque described by the third term can be understood heuristically as follows: $\overline{I}/e$ is the number per second of spin-polarized electrons entering the free layer when moments of the free and pinned layers are perpendicular. Each of these electrons carries angular momentum $\hbar/2$. The parameter η is given in terms of the currents for collinear orientation of the moments of the pinned and free layers by [10], [14], [17]

$$\eta = \frac{I_{++} + I_{+-} - I_{--} - I_{-+}}{I_{++} + I_{+-} + I_{--} + I_{-+}}. \quad (1.2)$$

Here, I_{++} and I_{--} are the majority and minority currents respectively that would flow for parallel alignment of the layers, while I_{+-} and I_{-+} are the majority and minority currents (from the perspective of the pinned layer), respectively, for anti-parallel alignment of the two layers. It is important to note that $\overline{I}$ and η depend on the applied voltage but are independent of the relative orientation of the free and pinned magnetic layers. The angle-dependent term $\hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p)$ accounts for the fact that only the transverse part of the incoming angular momentum can be absorbed by the free layer. In a spherical coordinate system in which the polar angle is measured from the normal to the plane and $\hat{\mathbf{m}}_p$ is also in this direction, this torque is in a direction to increase or decrease the polar angle. For brevity and convenience we refer to this term as the Slonczewski spin-torque term.

The torque described by the fourth term is often described as the "field-like" spin torque and cannot be simply expressed in terms of the electron particle current [17], [18]. The nomenclature "field-like" is based on the fact that as long as $f(\overline{I})$ does not depend on the angle between $\hat{\mathbf{m}}$ and $\hat{\mathbf{m}}_p$ this term can be treated as an additional contribution to the effective field, primarily increasing or decreasing the rate of precession, but also changing the effective energy function, i.e., the effective energy as a function of polar angle that determines the switching rate.

The LLG (1.1) can be written in the classical Landau-Lifshitz form by converting to an equation for $\mathbf{M}$ using $\mathbf{L} = \mathbf{M}V/\gamma$ and

then inserting the terms on the right-hand side of the equation for $\partial \mathbf{M}/\partial t$ and simplifying

$$(1+\alpha^2)\frac{\partial \mathbf{M}}{\partial t} = -\gamma\mu_0 M_S \hat{\mathbf{m}} \times \mathbf{H}_{\text{eff}} - \frac{\gamma\eta\overline{I}\hbar}{2\mathrm{e}V}\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)$$
$$+\frac{\gamma\hbar f(\overline{I})}{2\mathrm{e}V}(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p) - \alpha\gamma\mu_0 M_S \hat{\mathbf{m}}\times\hat{\mathbf{m}}\times\mathbf{H}_{\text{eff}}$$
$$+\frac{\alpha\gamma\eta\overline{I}\hbar}{2\mathrm{e}V}(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)$$
$$+\frac{\alpha\gamma\hbar f(\overline{I})}{2\mathrm{e}V}\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p). \tag{1.3}$$

Written this way, we can see that the effect of the Gilbert damping term $\alpha\hat{\mathbf{m}} \times (d\mathbf{L}/dt)$ is to renormalize the rate of precession by a factor $1 + \alpha^2$ (e.g., by effectively reducing the gyromagnetic ratio) and to add a term, $-\alpha\gamma\mu_0 M_S\hat{\mathbf{m}}\times\hat{\mathbf{m}}\times\mathbf{H}_{\text{eff}}$, which yields a contribution to $\partial\mathbf{M}/\partial t$ in the direction of the polar angle that acts to return the magnetization towards $\mathbf{H}_{\text{eff}}$. In addition to these two well-known effects, the damping also acts on the spin-torque terms so that the Slonczewski spin-torque term generates a field-like torque and the field-like torque term generates a torque in the same direction as the Slonczewski term. These terms are smaller in magnitude by a factor of α compared to spin-torque terms that generated them. If α is sufficiently small they may be neglected.

We obtain the effective field $\mathbf{H}_{\text{eff}}$ using

$$\mu_0\mathbf{H}_{\text{eff}} = -\left(\frac{\partial E}{\partial \mathbf{M}}\right) = -\left(\hat{x}\frac{\partial E}{\partial M_x} + \hat{y}\frac{\partial E}{\partial M_y} + \hat{z}\frac{\partial E}{\partial M_z}\right) \tag{1.4}$$

where E is the magnetic energy per unit volume of the free layer. If the free layer has no in-plane anisotropy, E can be written as

$$E = \frac{1}{2}\mu_0 N_{zz}M_z^2 + \left(K_U^B + \frac{K_U^S}{t}\right)\left(1 - \frac{M_z^2}{M_s^2}\right) \\ -\mu_0 M_z H_{z-\text{ext}} \tag{1.5}$$

where the first and second terms come from demagnetization and anisotropy, respectively. N_{zz} is the zz component of the demagnetization tensor. The anisotropy may arise from bulk magnetocrystalline anisotropy (K_U^B) or from surface anisotropy (K_U^S/t). Here, t represents the thickness of the magnetic free layer. The third term comes from an external applied magnetic field. In this paper, in order to preserve axial symmetry, we initially restrict our treatment to the case in which any external field is perpendicular to the layers. In fact, our primary interest is in RSER and WSER with no external field present. The external field is included here partly for completeness, but more importantly, because it allows us to make a connection with previous work by Brown [33]–[35] and others [36]–[40]. In Section V, we treat the more general field configuration in which a component is in the plane of the layers.

In the macrospin model, the magnitude of the magnetic moment of the free layer is assumed to be constant in time. Thus, the important quantity is its direction, which we will describe in spherical polar coordinates, i.e., $\hat{\mathbf{m}} = \sin\theta(\cos\phi\hat{\mathbf{x}} + \sin\phi\hat{\mathbf{y}}) +$

$\cos\theta\hat{\mathbf{z}}$. The magnetic energy per unit volume (relative to its value for $\theta = 0$), expressed in spherical polar coordinates, is

$$E = K_U^{\text{eff}}\sin^2\theta - \mu_0 M_S H_{z-\text{ext}}(\cos\theta - 1) \tag{1.6}$$

where $K_U^{\text{eff}} = K_U^B + K_U^S/t - (1/2)\mu_0 N_{zz}M_S^2$. This energy expression leads to an effective magnetic field

$$\mathbf{H}_{\text{eff}} = \left(H_K^{\text{eff}}\cos\theta + H_{z-\text{ext}}\right)\hat{z} = H_K^{\text{eff}}(\cos\theta + h)\hat{z} \tag{1.7}$$

where $H_K^{eff} = 2K_U^{\text{eff}}/\mu_0 M_s$ and $h = H_{z-\text{ext}}/H_K^{\text{eff}}$. H_K^{eff} is called the switching field because a field of this magnitude applied along the easy axis (perpendicular to the layers) will cause the free layer to switch. In the following, we shall suppress the superscript on H_K. It will be understood that H_K is the effective anisotropy field resulting from both anisotropy and demagnetization.

II. EQUATIONS OF MOTION

Substitution of (1.7) for $\mathbf{H}_{\text{eff}}$ in (1.3) yields

$$(1+\alpha^2)\frac{\partial\hat{\mathbf{m}}}{\partial t} = -\gamma\mu_0 H_K\left[\cos\theta + h + \alpha^2\overline{i} + \frac{\alpha}{\eta}\frac{f(\overline{I})}{I_0}\right]\hat{\mathbf{m}}\times\hat{\mathbf{z}}$$
$$-\alpha\gamma\mu_0 H_K\left[\cos\theta + h - \overline{i} + \frac{\alpha}{\eta}\frac{f(\overline{I})}{I_0}\right]\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}}). \tag{2.1}$$

Here, we have introduced a reduced current $\overline{i} = \overline{I}/I_0$, where I_0 is defined by

$$I_0 = \frac{\alpha}{\eta}\frac{2\mathrm{e}}{\hbar}\mu_0 H_K M_S V \tag{2.2}$$

and is the critical current for switching via the first spin-torque term. We have also assumed that the pinned layer magnetization is along $\hat{\mathbf{z}}$. The terms in (2.1), proportional to $\hat{\mathbf{m}}\times\hat{\mathbf{z}}$, yield a torque in the azimuthal (ϕ) direction and contribute to precession around the $\hat{\mathbf{z}}$ axis. On the other hand, the terms proportional to $\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})$ generate a torque that changes the polar angle and are responsible for spin-torque switching.

In terms of θ and ϕ, the LLG equation becomes

$$(1+\alpha^2)\frac{d\phi}{dt} = \omega_0\left(\cos\theta + h + \alpha^2\overline{i} + \frac{\alpha}{\eta}\frac{f(\overline{I})}{I_0}\right) \tag{2.3}$$

$$(1+\alpha^2)\frac{d\theta}{dt} = \alpha\omega_0\sin\theta\left(\overline{i} - h - \frac{\alpha}{\eta}\frac{f(\overline{I})}{I_0} - \cos\theta\right) \tag{2.4}$$

where $\omega_0 = \gamma\mu_0 H_K$. Equation (2.3) for ϕ simply describes the precession of the magnetic moment around the $\hat{\mathbf{z}}$ axis. We will not consider it further except to note that the field-like term (and to a lesser extent the Sloncewski-torque, since typically $\alpha \ll 1$) yields a current or bias dependence to the FMR frequency that might be observable. Equation (2.4) for θ describes the variation of the polar angle as a competition among the Slonczeweski spin-torque term (proportional to $\overline{i}$), external field (proportional to h), field-like spin-torque term and the damping term (proportional to $\cos\theta$). Depending on their signs, the first three either increase or decrease θ, whereas the third always decreases it (if $\theta < \pi/2$). Note that the field-like term in (2.4) can be viewed

either as a modification to i or a modification to h. Physically, modification of h may seem more appealing since this term originates from the change in damping due to the change in the rate of precession, similar to that caused by an applied field. Practically, however, it may be more convenient to include the field-like term with the Slonczewski term because both are controlled by the electrical current or bias.

We can simplify our equation even more if we measure time in units of $(1 + \alpha^2)/(\alpha\omega_0)$. Thus, if our unit of time is that required for $(1 + \alpha^2)/(2\pi\alpha)$ precessional orbits (in the absence of applied fields or currents) we have

$$\frac{d\theta}{d\tau} = (i - h - \cos\theta)\sin\theta \qquad (2.5)$$

where

$$\tau = \frac{\alpha\gamma\mu_0 H_K}{1 + \alpha^2}t \qquad (2.6)$$

and

$$i = \left(1 - \frac{\alpha}{\eta}\frac{f(\overline{I})}{\overline{I}}\right)\frac{\overline{I}}{\frac{\alpha}{\eta}\frac{2e}{\hbar}\mu_0 H_K M_S V}. \qquad (2.7)$$

Since $f(\overline{I})/\overline{I}$ may not be independent of $\overline{I}$, i may not be precisely proportional to $\overline{I}$. We shall nevertheless, for simplicity, refer to i as the "reduced current".

The critical current for switching is an important quantity for spin-torque devices because the magnitude of the current that must be supplied determines the size of the transistor needed to supply the current and ultimately the energy required for switching. Another important quantity is the thermal stability factor Δ, defined as the energy barrier for switching divided by the thermal energy $k_B T$

$$\Delta = \frac{K_U^{\text{eff}}V}{k_B T} = \frac{\mu_0 H_K M_S V}{2k_B T}. \qquad (2.8)$$

The thermal stability factor determines the rate at which thermal fluctuations cause an element to switch. To avoid thermally induced switching (over practically relevant periods), Δ should be greater than ~ 50. The critical current for switching is proportional to the thermal stability factor

$$I_0 = \frac{\alpha}{\eta}\frac{2e}{\hbar}2\Delta k_B T \approx \frac{\alpha}{\eta}1.5\ \text{mA}. \qquad (2.9)$$

The approximate estimate for the critical current in (2.9) assumes $\Delta = 60$.

A. Deterministic Switching Time

In the limit of $\theta \to 0$, (2.5) becomes

$$\frac{d\theta}{d\tau} = (i - h - 1)\theta \qquad (2.10)$$

with solution

$$\theta = \theta_0 \exp\left[(i - h - 1)\tau\right]. \qquad (2.11)$$

The deterministic switching time in this approximation is [41]

$$\tau_{sw} = \frac{\ln\left(\frac{\pi}{2\theta_0}\right)}{i - h - 1}. \qquad (2.12)$$

This expression is based on the notion that the element switches when θ equals $\pi/2$.

Equation (2.5) can also be solved without making the $\theta \to 0$ approximation. Substituting $z = \cos\theta$ into (2.5), we have

$$\frac{dz}{d\tau} = -(i - h - z)(1 - z^2) \qquad (2.13)$$

which can be integrated to give

$$\tau(z) = \\ \frac{\frac{1}{2}(i-h+1)\ln\left(\frac{1-z}{1-z_0}\right) + \frac{1}{2}(i-h-1)\ln\left(\frac{1+z_0}{1+z}\right) + \ln\left(\frac{i-h-z_0}{i-h-z}\right)}{(i-h)^2 - 1} \qquad (2.14)$$

where z_0 is the value of $\cos\theta$ at $\tau = 0$. When θ reaches $\pi/2$, $\cos\theta$ changes sign so that the damping term changes from impeding switching by pushing θ back towards zero to assisting switching by pushing θ towards π, thus it is reasonable to assume that the element switches when $\theta = \pi/2$ or $\cos\theta = 0$, yielding

$$\tau_{sw} = \\ \frac{-\frac{1}{2}(i-h+1)\ln(1-z_0) + \frac{1}{2}(i-h-1)\ln(1+z_0) + \ln\left(1 - \frac{z_0}{i-h}\right)}{(i-h)^2 - 1}. \qquad (2.15)$$

III. FOKKER-PLANCK EQUATION

Equation (2.15) gives the time required for an element that has its moment pointing at an angle $\theta_0 = \cos^{-1}z_0$ at $\tau = 0$ to switch. It implicitly assumes that the motion is deterministic, i.e., that neither the initial displacement angle θ_0, nor the trajectory in θ, ϕ space are affected by thermal fluctuations. In actuality, of course, thermal fluctuations will have important effects upon the switching. We are primarily interested in the probability of rare events, namely switching that occurs when not intended (when current is applied for reading the state of the system) or switching that does *not* occur when switching is intended (when current is applied for switching). One approach to this problem is to solve the Landau-Lifshitz-Slonczewski equation [see (1.1)] with an additional, random thermal field many times while recording whether the magnetization has switched. We will use this approach to validate our results based on the Fokker-Planck method described in the following.

We derive and solve a Fokker-Planck equation [33], [42], [43] for the probability distribution of the angle θ as a function of τ. We define $\rho(\theta, \tau)$ to be the probability that the magnetization

is pointing in direction θ relative to the film normal at time τ. $\rho(\theta, \tau)$ is normalized so that

$$\int_0^\pi \rho(\theta, \tau) \sin \theta d\theta = 1 \qquad (3.1)$$

for all τ. The Fokker-Planck equation is based on the continuity equation for this probability density

$$\frac{\partial \rho(\theta, \tau)}{\partial \tau} = -\nabla \cdot \mathbf{J}(\theta, \tau) = -\frac{1}{\sin \theta} \frac{\partial (\sin \theta J_\theta(\theta, \tau))}{\partial \theta} \quad (3.2)$$

which simply states that the rate of change of the probability density at angle θ is equal to the net rate at which probability density flows in. The current in probability density consists of a flow term and a diffusion term

$$J_\theta(\theta, \tau) = \rho(\theta, \tau) \frac{\partial \theta}{\partial \tau} - D \frac{\partial \rho(\theta, \tau)}{\partial \theta} \qquad (3.3)$$

where D is the diffusion coefficient and $\partial \theta / \partial \tau$ is given by (2.5). Combining (3.2) and (3.3) yields

$$\frac{\partial \rho(\theta, \tau)}{\partial \tau} = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin^2 \theta (i - h - \cos \theta) \rho(\theta, \tau) \right.$$
$$\left. - D \sin \theta \frac{\partial \rho(\theta, \tau)}{\partial \theta} \right). \quad (3.4)$$

The diffusion constant D can be determined from the equilibrium condition in the absence of an applied electrical current or field, because in this case (3.4) reduces to

$$\frac{\partial \rho_{\text{eq}}(\theta)}{\partial \theta} = \frac{-1}{D} \sin \theta \cos \theta \rho_{\text{eq}}(\theta) \qquad (3.5)$$

which has the solution

$$\rho_{\text{eq}}(\theta) = A \exp \left(\frac{-\sin^2 \theta}{2D} \right). \qquad (3.6)$$

On the other hand, the equilibrium distribution should be Maxwell-Boltzmann, which implies

$$\rho_{\text{eq}}(\theta) \propto \exp \left[-\frac{E(\theta)}{k_B T} \right] = \exp[-\Delta \sin^2 \theta]. \qquad (3.7)$$

Thus, $D = 1/2\Delta$, so the Fokker-Planck equation is

$$\frac{\partial \rho(\theta, \tau)}{\partial \tau} = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin^2 \theta (i - h - \cos \theta) \rho(\theta, \tau) \right.$$
$$\left. - \frac{\sin \theta}{2\Delta} \frac{\partial \rho(\theta, \tau)}{\partial \theta} \right). \quad (3.8)$$

It can also be expressed in terms of $z = \hat{m}_z = \cos \theta$ as

$$\frac{\partial \rho(z, \tau)}{\partial \tau} = \frac{\partial}{\partial z} \left[(i - h - z)(1 - z^2) \rho + \frac{(1 - z^2)}{2\Delta} \frac{\partial \rho}{\partial z} \right]. \quad (3.9)$$

Equation (3.9) is equivalent, aside from the term i in $(i - h - z)$, to an expression derived by Brown [33] for the similar reversal of a macrospin by a magnetic field.

A. Approximate Analytical Solutions

We can obtain an approximate analytical solution to (3.8) in the limit in which θ is small. In this limit, (3.8) becomes

$$\frac{\partial \rho(\theta, \tau)}{\partial \tau} \approx -\frac{1}{\theta} \frac{\partial}{\partial \theta} \left(\theta^2 (i - h - 1) \rho(\theta, \tau) - \frac{\theta}{2\Delta} \frac{\partial \rho(\theta, \tau)}{\partial \theta} \right). \quad (3.10)$$

To solve this equation, we will use the ansatz $\rho(\theta, \tau) = A(\tau) \exp(-\theta^2 / W(\tau))$. $A(\tau)$ and $W(\tau)$ can be related through the normalization condition

$$1 = \int_0^\pi \sin \theta d\theta \rho(\theta, \tau) \approx \int_0^\pi \theta d\theta A(\tau) \exp \left(-\frac{\theta^2}{W(\tau)} \right)$$

$$\approx \frac{1}{2} A(\tau) W(\tau) \quad (3.11)$$

which implies that $A(\tau) = 2/W(\tau)$. Substitution of this ansatz into (3.10) yields the following equation for $W(\tau)$:

$$\frac{\partial W(\tau)}{\partial \tau} = 2\nu W(\tau) + \frac{2}{\Delta} \qquad (3.12)$$

where $\nu = i - h - 1$.

The solution to (3.12) subject to the boundary condition $W(\tau) \to (1/\Delta_0)$ for $\tau \to 0$ (where the initial thermal stability factor Δ_0 is possibly different from that Δ, during the reversal process, e.g., due to heating, field induced strains, etc.) is

$$W(\tau) = \frac{\exp(2\tau\nu)}{\Delta_0} + \frac{\exp(2\tau\nu) - 1}{\Delta\nu} \qquad (3.13)$$

so that the distribution function is approximated as

$$\rho(\theta, \tau) = \frac{2}{W(\tau)} \exp \left(-\frac{\theta^2}{W(\tau)} \right) \qquad (3.14)$$

with $W(\tau)$ given by (3.13).

Note that (3.13) predicts that at large times ($2\nu\tau \gg 1$) the distribution will become independent of θ and decay as

$$\rho \to \frac{2}{W(\tau)} = \frac{2\nu\Delta\Delta_0}{\nu\Delta + \Delta_0} \exp[-2\tau\nu] \to \frac{2\Delta\nu}{\nu + 1} \exp[-2\tau\nu] \quad (3.15)$$

where the last form assumes $\Delta = \Delta_0$.

The nonswitched fraction in this approximation, calculated by assuming that an element switches at $\theta = \pi/2$, is

$$P_{NS} \approx \int_0^{\frac{1}{2}\pi} \theta d\theta \frac{2}{W} \exp \left[-\frac{\theta^2}{W} \right] = 1 - \exp \left(-\frac{\pi^2}{4W(\tau)} \right). \quad (3.16)$$

The accuracy of this approximation is tested in the next section by comparison to numerical solutions of the Fokker-Planck equation that do not invoke the small-angle approximation.

IV. NUMERICAL AND APPROXIMATE ANALYTICAL SOLUTIONS TO AXIALLY SYMMETRIC FOKKER-PLANCK

Equation (3.9) for $h = 0$ is solved on the line $-1 \leq z \leq 1$, where $z = \cos \theta$. The usual starting distribution is
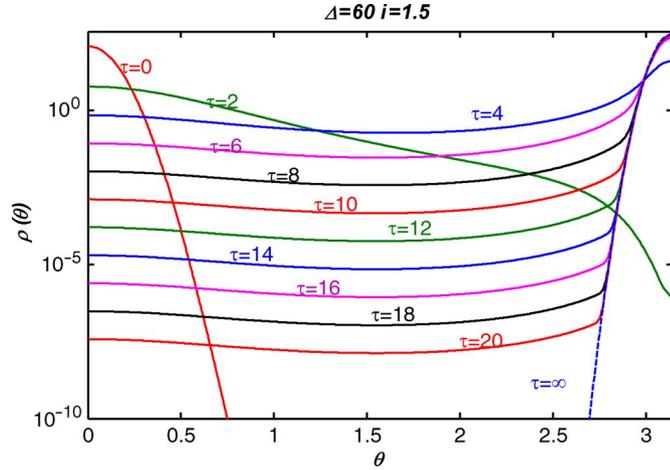
Fig. 1. Probability distribution for different reduced times (numerical solution to equation (3.8) or (3.9) for $\Delta = 60$ and $i = 1.5$).

$$\rho_{\text{initial}}(\theta) = \frac{\exp(-\Delta \sin^2 \theta)\Theta\left(\frac{\pi}{2} - \theta\right)}{\int\limits_0^{\frac{\pi}{2}} \sin \theta \exp(-\Delta \sin^2 \theta)d\theta} \qquad (4.1)$$

or

$$\rho_{\text{initial}}(x) = \frac{\exp\left(-\Delta(1 - z^2)\right)\Theta(x)}{\int\limits_0^1 \exp\left(-\Delta(1 - z^2)\right)dx} \qquad (4.2)$$

where the Heaviside function $\Theta((\pi/2)-\theta)$ constrains the initial distribution to the well near $\theta = 0$. Evolving (3.9) forward in time gives the curves shown in Fig. 1, for the case $\Delta = \Delta_0 = 60$ and $i = 1.5$. Initially, the probability is confined to the well at $\theta = 0$. By reduced time $\tau \sim 4$, however, the distribution on a semi-log plot becomes relatively flat [as predicted by (3.15)], except for a sharp peak in the final state at the right. The flat value decreases exponentially as the probability flows into the final state near $\theta = \pi$. At long times, the distribution function becomes

$$\rho(\theta, \infty) = Ne^{-\Delta(\sin^2 \theta + 2i \cos \theta)} \qquad (4.3)$$

where N is a normalization factor. This function gives the parabola labeled $\tau = \infty$ on the right side of Fig. 1. It provides a particular solution to (3.8), for which $\partial\rho(\theta, \tau)/\partial\tau = 0$. Equation (4.3) is easily understood from (2.5), which implies that the spin-torque current, insofar as it enters the equation for $d\theta/d\tau$, acts like an additional axial magnetic field. Thus the Fokker-Planck equation for perpendicular spin-torque systems is equivalent to the 1-D diffusion equation for a particle in a potential of form

$$E(\theta)V = K_U V \left[\sin^2 \theta + 2(i - h)(\cos \theta - 1)\right]. \qquad (4.4)$$

The $\tau \to \infty$ solution (4.3) is the spin-torque analog of an expression derived by Brown [33] for the steady state solution to the Fokker-Planck equation for a Stoner-Wohlfarth element in an easy axis field.

Fig. 2 compares the numerical solution shown in Fig. 1 with the approximate solution (3.13), which certainly is not valid in
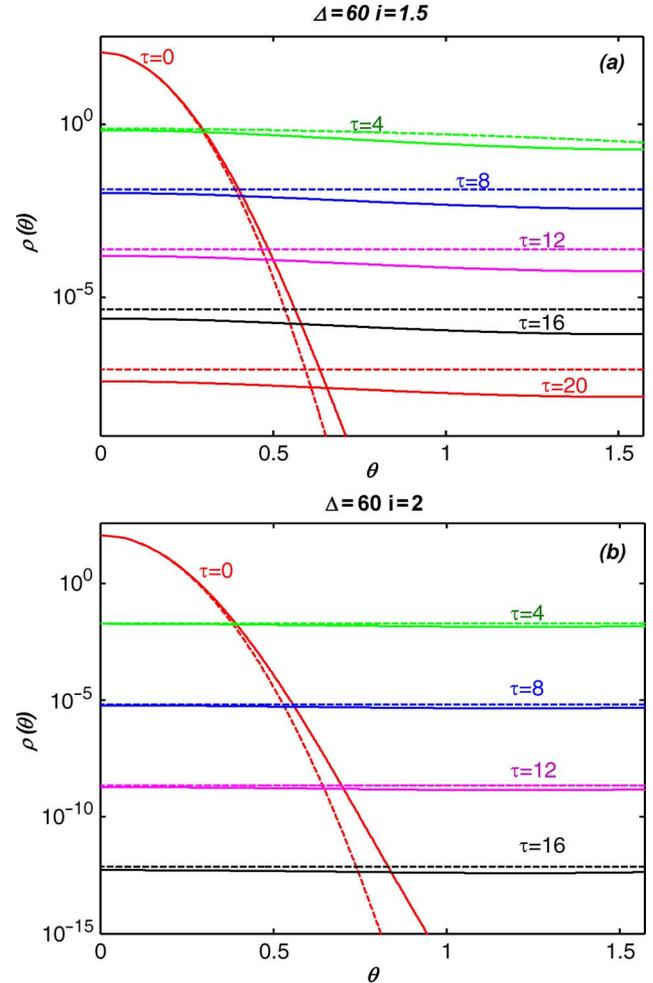
Fig. 2. Probability distribution within the well centered at $\theta = 0$ for reduced currents [see (2.7)] (a) $i = 1.5$ and for (b) $i = 2$. Dashed lines are approximate solution of (3.13). For both calculations, thermal stability factor is $\Delta = 60$. Reduced time [see (2.6)] is labeled by τ.

the well centered at $\theta = \pi$, since it is based on an expansion around the bottom of the well at $\theta = 0$. For this reason, we show in Fig. 2 only the part of the distribution between $\theta = 0$ and $\theta = \pi/2$. The surprising prediction [see (3.15)] that the distribution becomes approximately independent of θ for large values of $(i - 1)\tau$ is obvious in both the numerical solutions and the approximate analytic solutions. The fraction of ρ that has not switched can be obtained by integrating over this part of the probability distribution function.

The approximate analytical solution becomes more accurate as the reduced current i is increased, as can be seen by comparing Fig. 2(a) with (b). Note that the approximate solution consistently overestimates the probability distribution in the well at $\theta = 0$. This is understandable because the actual equation of motion for θ, $d\theta/d\tau = (i - h - \cos \theta)\sin \theta$ is approximated by setting $\cos \theta$ to 1, which significantly underestimates the relative amount by which the spin-torque term exceeds the damping term, especially when $i \approx 1$. Thus, the frequency of write errors is less than would be predicted by the approximate analytical model, especially for low values of $i - 1$.

Fig. 3. Distribution functions for thermal stability factor $\Delta = 30$ for reduced time ranging from 0 to 20. Time is measured in units of FMR period divided by damping parameter [see (2.6)].

Fig. 5. Nonswitched probability, WSER using a thermal stability factor of $\Delta = 60$ as a function of time for several values of the reduced current i, measured in units of I_0, critical current for switching. Time is measured in units of FMR period divided by 2π times damping parameter. Smooth solid curves are solutions of Fokker-Planck equation. Jagged lines composed of discrete points are a result of Landau-Lifshitz simulations of a macrospin including a thermal field [44] as described in the text.

A. Effect of Thermal Stability Factor

Fig. 3 shows the time evolution of the distribution function for a thermal stability factor $\Delta = 30$, one half the value used in the comparable calculation shown in Fig. 1. The initial and final distribution functions are noticeably broader as would be expected. Somewhat surprisingly, the effect of this large reduction in Δ on the distribution functions at intermediate times is relatively modest. There is a relatively small (on this exponential scale) shift downward in the distribution function. The spacing of the curves for different times hardly changes. The implication of this result is that soft error rates for writing associated with nonswitching events should be relatively insensitive to the thermal stability factor.

Fig. 4(a) shows the WSER as a function of time for five values of i with Δ fixed at 60 (solid lines). Similar results are presented for $\Delta = 30$ in Fig. 4(b). For comparison, we also show in both figures the WSER calculated from the analytic approximation (3.16). The analytic approximation overestimates P_{ns}, especially for low values of $i - 1$, but accurately represents the exponential decay which is well described by $\exp[-2\tau(i-1)]$. It can be seen that Δ has only a small effect on the WSER.

In Fig. 5, we show a comparison between the numerical solution of the Fokker-Planck equation and macrospin simulations including a thermal field [44]. Both calculations were carried out for $\Delta = 60$ and for $i = 2, 3, 4, 5$ and 6. In the simulations, for each value of current i, 10 000 realizations of the system evolution were computed with independent initial conditions and thermal noise. The probability shown in Fig. 5 is proportional to the number of these that remain unswitched at a particular time—thus the discrete downward steps visible in the lower part of the figure occur each time one or more of the realizations is observed to switch.

The fact that we found no statistically significant difference between the two approaches is consistent with their mathematical equivalence and supports our assertion that Brown's

Fig. 4. Nonswitched probability, soft-error rate for writing using a thermal stability factor of (a) $\Delta = 60$ and (b) $\Delta = 30$ as a function of time for several values of the reduced current. Current i is measured in units of I_0, critical current for switching and time is measured in units of FMR period divided by 2π times damping parameter. Solid lines are solutions of the Fokker-Planck equation. Dashed lines represent approximate analytic solution [see (3.16)].

Fig. 6. Dashed lines are nonswitched probability as a function of reduced time for the case that initial angle is $\cos^{-1}(0.9) \sim 26°$. Solid lines are numerical solutions to Fokker-Planck equation for an initial thermal distribution. Panels (a) and (b) show effect of an initial canting angle for thermal stability factors of 60 and 20, respectively.

derivation of the Fokker-Planck equation [33] from the stochastic Langevin equation for the analogous problem in which the switching is induced by a magnetic field has been successfully generalized to include current induced spin torques.

B. Effect of Initial Canting Angle

Since we speculate that the exponentially decaying tail in the nonswitched probability arises from the fact that the spin torque vanishes at $\theta = 0$, we investigated the effect of an initial canting angle on the switching probability. Fig. 6 shows the nonswitched probability as a function of the reduced time for three values of the reduced current calculated with two types of initial distribution. In one case (dashed lines) the initial distribution is very tightly concentrated at a particular value of θ_0 ($\cos \theta_0 = 0.9$). In the other case (solid lines) the initial distribution is a thermal distribution centered at the origin similar to the calculations shown in Fig. 4. From the figure, it can be seen that the nonswitched fraction is significantly reduced if the initial distribution is canted, especially if Δ and $i - 1$ are both large; however, the exponential long-time tail persists with the same decay constant $2(i - 1)$.

For this example, the initial distribution had a maximum value of $\cos \theta$ of 0.905 and a minimum value of 0.895. For small values of $i - 1$, the decrease in the P_{ns} caused by an initial canting angle is smaller than for larger values, because there is a smaller spin-torque driving term so that the system spends
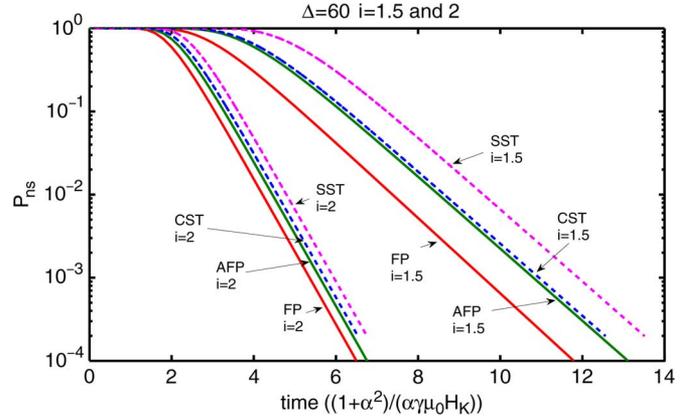
Fig. 7. Comparison of four approximations for time evolution of $P_{ns}(\tau)$, probability that the element has not switched after reduced time τ. FP and AFP indicate, respectively, the numerical Fokker-Planck solution and approximate analytical solution, based on (3.16). Curves SST and CST indicate, respectively, the estimation of switching time from initial probability density using deterministic expressions for switching time given by (2.12) and (2.15).

more time at smaller angles allowing the diffusion term to establish a population at $\theta = 0$. This is illustrated in Fig. 6(b) which shows that an initial canting angle has a much smaller effect for $\Delta = 20$ compared to the case of $\Delta = 60$ [see Fig. 6(a)] because of the larger diffusion term.

C. Comparison With Sun Switching Time Ansatz

An exponential tail in the nonswitched probability was obtained previously by Sun *et al.* [47] and by He *et al.* [48] who postulated that the nonswitched probability could be related to the initial thermal distribution through the relation (2.12) which approximately and deterministically relates the initial angle to the time to switch. Thus, if one assumes that those (and only those) systems that have an initial angle greater than θ_0 will have switched in time, $\tau_{sw} = \ln(\pi/2\theta_0)/\nu$ (where we again use $\nu = i - h - 1$), one can estimate the nonswitched fraction at time τ by integrating over the initial probability distribution from $\theta = 0$ to $\theta = \theta_0(\tau_{sw})$

$$P_{ns}(\tau) = \int_0^{\frac{\pi}{2} \exp[-\tau\nu]} \rho_0(\theta) \sin\theta \, d\theta. \qquad (4.5)$$

This approximation is shown in Fig. 7 as the dashed curve labeled Sun Switching Time (SST) Approximation. An improvement on this approximation can be made by using the exact expression for the switching time given in (2.14), which is shown in Fig. 7 as the dashed curve marked Corrected Switching Time (CST) Approximation

$$P_{ns}(\tau_{sw}(z_0)) = \int_{z_0(\tau_{sw})}^1 \rho_0(z) \, dz \qquad (4.6)$$

where $\tau_{sw}(z_0)$ is given by (2.15). This latter curve is slightly greater than the approximate solution to the Fokker-Planck equation that we derived as (3.16) and show as the solid curve labeled AFP. The full solutions to the Fokker-Planck equation are given by the solid lines labeled FP.

The asymptotic forms (for large values of τ) for the SST and CST approximations for P_{ns} are easily obtained since an analytic expression for (4.6) can be obtained for $z_0 \approx 1$

$$P_{ns}^{SST} \to \frac{\Delta \pi^2}{4} \exp[-2\tau\nu] \qquad (4.7)$$

and

$$P_{ns}^{CST} \to \Delta 4^{\frac{\nu+1}{\nu+2}} \left(\frac{\nu+2}{\nu+1}\right)^{\frac{2}{\nu}} \exp[-2\tau\nu]. \qquad (4.8)$$

These may be compared with the nonswitched fraction obtained from the approximate analytic solution [from (3.15)]

$$P_{ns}^{AFP} \to \frac{\Delta 2\nu}{\nu+1} \exp[-2\nu\tau]. \qquad (4.9)$$

Comparing the three approximate solutions among themselves and with the numerical Fokker-Planck solution, we can remark that the "SST" and the "AFP" results are similar in that both are based on a linear approximation for θ. Thus, both approximate $\sin\theta$ as θ and $\cos\theta$ as 1. They differ in that the SST approximation neglects diffusion during the switching process. The "CST" and the full Fokker-Planck solution do not make the linear approximation, but CST like SST neglects diffusion during the switching process. Thus, SST and CST become more accurate approximations to the Fokker-Planck (FP) solution as Δ becomes larger. It is clear that the WSER is less for CST than for SST and is less for FP than for AFP because the linear approximation reduces the increase of polar angle with time in (2.5). It is also clear that the WSER is less for AFP than SST and is less for FP than for CST, because diffusion during the switching process will reduce the probability of finding a system at the "stagnation point" at $\theta = 0$ where the switching torque vanishes.

Somewhat counterintuitively, the linear approximations (SST and AFP) become more accurate as the overdrive $\nu = i - 1$ becomes larger. This is due to the fact that for larger ν, the difference between $i - \cos\theta$ and $i - 1$ in (2.5) becomes relatively less significant and because the flow term in the Fokker-Planck equation (3.4) becomes relatively more important compared to the diffusion term.

Physically, the expressions (4.5) and (4.6), which are based on deterministic trajectories from the initial distribution, are equivalent to solving the Fokker-Planck equation without the diffusion term (i.e., assuming zero temperature) starting from an initial finite temperature distribution determined by Δ. Thus the distribution function corresponding to the SST approximation can be obtained from (3.13) and (3.14) with the substitutions $\Delta_0 \to \Delta$ and $\Delta \to \infty$ in (3.13)

$$\rho(\theta, \tau) = 2\Delta \exp(-2\nu\tau) \exp\left(-\Delta\theta^2 \exp(-2\nu\tau)\right). \quad (4.10)$$

This result is easily confirmed because the assumed deterministic relation between $\theta(\tau)$ and the initial $\theta(\tau = 0) = \theta_0$ given by (2.11), allows us to relate the distribution function at $\tau = 0$ to that at any τ through

$$\rho(\theta, \tau)\theta d\theta = \rho_0(\theta_0, 0)\theta_0 d\theta_0 \qquad (4.11)$$

so that in this approximation, using $\theta_0^2 = \theta^2 \exp[-2\nu\tau]$ [from (2.12)] we have

$$\rho(\theta, \tau) = \rho_0(\theta_0[\theta, \tau]) \frac{d\theta_0^2}{d\theta^2}$$
$$= 2\Delta \exp\left[-\Delta\theta^2 \exp(-2\nu\tau)\right] \exp(-2\nu\tau) \quad (4.12)$$

which is identical to (4.10). More generally, it can be shown using (4.12) that in the deterministic or "no-diffusion" approximation (e.g., SST or CST), if the initial distribution is proportional to θ^n for small values of θ, the unswitched fraction will decay at long times as $\exp[-\nu(2+n)\tau]$. It is clear, however, both from the calculations shown in Fig. 6 and from the analytical approximations of Section V as follows, that even if the initial distribution vanishes near $\theta = 0$, the Fokker-Planck solutions give a nonswitched fraction that decays at long times as $\exp[-2\nu\tau]$. If the SST approximation were applied for the situation described by Fig. 6 with an initial canting angle of 0.451 radians, it would yield a step function for P_{ns} which would drop abruptly from 1 to 0 at $\tau \approx 1.25/\nu = 1.25/(i-1)$.

V. PERTURBING WITH IN-PLANE COMPONENTS OF PINNED-LAYER MAGNETIZATION OR EXTERNAL FIELD: STAGNATION POINTS

Fig. 6 shows that the WSER (probability of not switching) can be decreased by moving the initial distribution away from $\theta = 0$. The reason for this is that there is no spin-transfer torque when the magnetization of the free layer is collinear with that of the pinned layer—switching must wait for a thermal fluctuation. We call such a free-layer direction a "stagnation point". Clearly, the precession term $\sim \hat{m} \times \mathbf{H}_{\text{eff}}$ and the damping term $\sim \hat{m} \times (\hat{m} \times \mathbf{H}_{\text{eff}})$ vanish for $\hat{m}$ along $\mathbf{H}_{\text{eff}}$, and when $\hat{m}_p$ is along $\mathbf{H}_{eff}$, the spin torque term $\sim \hat{m} \times (\hat{m} \times \hat{m}_p)$ does as well. But if the pinned layer magnetization is not along the easy axis (i.e., $\mathbf{H}_{\text{eff}}$) the stagnation points of the terms are different, and we might hope that the total torque does not vanish anywhere, eliminating this problem of slow switching.

However, there is a well-known theorem of topology that eliminates this hope. The "Hairy Ball Theorem", proved by Brouwer in 1912 [49], states that any continuous tangent field on an even-dimensional sphere (such as the 2-D sphere on which $\hat{m}$ lies) must vanish at some point on the sphere. Each of the torque terms that enter (1.3) are tangent fields because they are perpendicular to $\hat{m}$. They are also continuous (in fact infinitely differentiable), which explains why each has a stagnation point—but their sum is also a continuous tangent field and therefore has a stagnation point. The theorem guarantees only the existence of the stagnation point, but we will see visually how it arises and where on the sphere it occurs in Fig. 8.

A. WSER for Tilted Pinned Layer From Linearized Fokker-Planck Equation

To calculate the WSER for the case of the pinned layer magnetization not collinear with the easy direction of the free layer or an in-plane component of the magnetic field, we will generalize the linear approximation discussed in Section IV previously, by breaking the rotational symmetry that allowed us to describe the switching process using only the polar angle θ. In
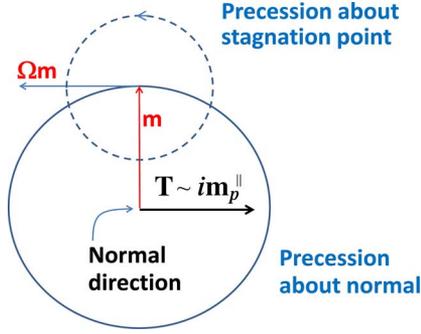
Fig. 8. Stagnation point for case of low damping ($\nu \ll \Omega$) and $F = h_{\parallel} = 0$.

case there is any concern that our linearization process may generate the stagnation point or put it in the wrong place, we provide the (nonlinearized) equations that determine the position of the stagnation point in Appendix A.

We linearize about the normal direction ($\hat{\mathbf{z}}$), working only with the transverse component of the unit vector $\hat{\mathbf{m}}$, which we denote by $\mathbf{m}^{\parallel}$

$$\hat{\mathbf{m}} = \mathbf{m}^{\parallel} + \hat{\mathbf{z}}\cos\theta \approx \mathbf{m}^{\parallel} + \hat{\mathbf{z}}; \quad \left(\mathbf{m}^{\parallel}\cdot\hat{\mathbf{z}} = 0, \quad |\mathbf{m}^{\parallel}| \ll 1\right) \tag{5.1}$$

where the error in replacing $\cos\theta$ by 1 is second order. Explicitly, $\mathbf{m}^{\parallel}$ can be written in terms of the polar and azimuthal angles by $\mathbf{m}^{\parallel} = \theta(\cos\phi\hat{\mathbf{x}} + \sin\phi\hat{\mathbf{y}})$, to linear order (neglecting terms of order θ^2 and higher). Effectively, we are approximating the region around the "north pole" of the $\hat{\mathbf{m}}$-sphere as a plane.

We allow a tilted pinned magnetization direction, which we write as

$$\mathbf{m}_p = \mathbf{m}_p^{\parallel} + m_{pz}\hat{\mathbf{z}}. \tag{5.2}$$

It will *not* be necessary to linearize in $\mathbf{m}_p^{\parallel}$; the tilt angle can be as large as desired. We also introduce a dimensionless in-plane field $\mathbf{h}_{\parallel}$ (whose magnitude is the tilt angle in radians) defined by

$$\mathbf{H}_{\text{eff}} = H_K(1+h)\hat{\mathbf{z}} + H_K\mathbf{h}_{\parallel}. \tag{5.3}$$

The approximation that θ is small so that we can neglect terms proportional to θ^2 and higher also allows us to approximate $\hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \hat{\mathbf{z}}) = \mathbf{m}^{\parallel}$ and $\hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \mathbf{h}_{\parallel}) = -\mathbf{h}_{\parallel}$ so that the Landau-Lifshitz-Gilbert-Slonczewski equation (1.3) becomes

$$\begin{aligned}(1+\alpha^2)\frac{\partial\mathbf{M}}{\partial t} = &-\gamma\mu_0 M_s H_K(1+h)\mathbf{m}^{\parallel}\times\hat{\mathbf{z}}\\&-\alpha\gamma\mu_0 M_s H_K(1+h)\mathbf{m}^{\parallel} - \alpha\gamma\mu_0 M_s H_K\mathbf{h}_{\parallel}\\&-\gamma\mu_0 M_s H_K\hat{\mathbf{z}}\\&\times\mathbf{h}_{\parallel} - \frac{\gamma\hbar}{2eV}(\eta\bar{I}-\alpha f)\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)\\&+\frac{\gamma\hbar}{2eV}(f+\alpha\eta\bar{I})(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p).\end{aligned} \tag{5.4}$$

Note that in the absence of field tilt and spin torque, the first term describes precession about the z-axis with a frequency $\gamma\mu_0 H_K(1+h)/(1+\alpha^2)$ and the second describes dissipation at a rate α times this frequency. The change of time variable

to τ (2.6) removes this rate and leaves us with a dimensionless equation

$$\begin{aligned}\frac{\partial\hat{\mathbf{m}}}{\partial\tau} = &-\alpha^{-1}(1+h)\mathbf{m}^{\parallel}\times\hat{\mathbf{z}} - (1+h)\mathbf{m}^{\parallel} - \mathbf{h}_{\parallel} - \alpha^{-1}\hat{\mathbf{z}}\times\mathbf{h}_{\parallel}\\&-\frac{1}{\alpha\gamma\mu_0 M_s H_K}\frac{\gamma\hbar}{2eV}(\eta\bar{I}-\alpha f)\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)\\&+\frac{1}{\alpha\gamma\mu_0 M_s H_K}\frac{\gamma\hbar}{2eV}(f+\alpha\eta\bar{I})(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p).\end{aligned} \tag{5.5}$$

The coefficient of the triple product will be recognized as the dimensionless current i defined in (2.7)

$$i = \frac{1}{\alpha\gamma\mu_0 M_s H_K}\frac{\gamma\hbar}{2eV}(\eta\bar{I} - \alpha f). \tag{5.6}$$

The last term is the field-like part of the current (with a small correction proportional to $\alpha\bar{I}$). We will denote its coefficient by

$$F = \frac{1}{\alpha\gamma\mu_0 M_s H_K}\frac{\gamma\hbar}{2eV}(f + \alpha\eta\bar{I}). \tag{5.7}$$

The cross products involving $\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p$ and $\hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p)$ in (5.5) can be expanded using $\hat{\mathbf{m}} \approx \mathbf{m}^{\parallel} + \hat{\mathbf{z}}$ and dropping terms higher than linear in θ (or $\mathbf{m}^{\parallel}$)

$$\hat{\mathbf{m}} \times \hat{\mathbf{m}}_p = \mathbf{m}\times\mathbf{m}_p^{\parallel} + m_{pz}\mathbf{m}^{\parallel}\times\hat{\mathbf{z}} + \hat{\mathbf{z}}\times\mathbf{m}_p^{\parallel}$$
$$\hat{\mathbf{m}} \times (\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p) = m_{pz}\mathbf{m}^{\parallel} - \mathbf{m}_p^{\parallel} + \hat{\mathbf{z}}\left(\mathbf{m}^{\parallel}\cdot\mathbf{m}_p^{\parallel}\right). \tag{5.8}$$

Finally, we omit out-of-plane contributions to $d\hat{\mathbf{m}}/d\tau$ because changes in $(\hat{\mathbf{m}})_z = \cos\theta$ would not be consistent with our linear approximation for θ. These approximations yield the following Landau-Lifshitz equation for the in-plane linearized magnetization

$$\frac{\partial\mathbf{m}^{\parallel}}{\partial\tau} = -\Omega\mathbf{m}^{\parallel}\times\hat{\mathbf{z}} + \nu\mathbf{m}^{\parallel} + \mathbf{T} \tag{5.9}$$

where

$$\Omega = \alpha^{-1}(1+h) - F \tag{5.10}$$

and

$$\nu = -1 - im_{pz} - h. \tag{5.11}$$

The sign of ν is chosen so that it is positive for switching. For a perpendicular pinned layer ($m_{pz} = -1$) and in the absence of applied axial field ($h = 0$), we have $\nu = i - 1$. This is sometimes called the "overdrive" since it is the relative amount of current in excess of the critical current. All of the azimuthal-symmetry-breaking effects are contained in the 2-D "tilt" vector $\mathbf{T}$

$$\mathbf{T} = i\mathbf{m}_p^{\parallel} + F\hat{\mathbf{z}}\times\mathbf{m}_p^{\parallel} - \alpha^{-1}\mathbf{h}_{\parallel}. \tag{5.12}$$

The Fokker-Planck equation for the linearized magnetization is

$$\frac{\partial\rho(\mathbf{m}^{\parallel},\tau)}{\partial\tau} = -\nabla\cdot\mathbf{J}(\mathbf{m}^{\parallel},\tau) \tag{5.13}$$

where the divergence is respect to the vector $\mathbf{m}^{\parallel}$, and the probability current is

$$\mathbf{J} = \rho \frac{d\mathbf{m}^{\parallel}}{d\tau} - D\nabla\rho. \tag{5.14}$$

To solve this equation, we generalize our ansatz (Section III) to a Gaussian distribution with a drifting center $\mathbf{m_d}$ and a width parameter W

$$\rho(\mathbf{m}^{\parallel}, \tau) = A(\tau)\exp\left[-\frac{\left(\mathbf{m}^{\parallel} - \mathbf{m}_d(\tau)\right)^2}{W(\tau)}\right]. \tag{5.15}$$

Computing the necessary derivatives with respect to the 2-D vector $\mathbf{m}^{\parallel}$

$$\nabla\rho = -\frac{2(\mathbf{m}^{\parallel} - \mathbf{m}_d)}{W}\rho \tag{5.16}$$

$$\nabla^2\rho = 4\frac{(\mathbf{m}^{\parallel} - \mathbf{m}_d)^2 - W}{W^2}\rho \tag{5.17}$$

$$\nabla\cdot(\rho\mathbf{m}^{\parallel}) = 2\frac{W - \mathbf{m}^{\parallel}\cdot(\mathbf{m}^{\parallel} - \mathbf{m}_d)}{W}\rho \tag{5.18}$$

$$\nabla\cdot(\rho\mathbf{m}^{\parallel}\times\hat{\mathbf{z}}) = 2\frac{\rho}{W}(\hat{\mathbf{z}}\times\mathbf{m}_d)\cdot(\mathbf{m}^{\parallel} - \mathbf{m}_d) \tag{5.19}$$

and substituting into the Landau-Lifshitz equation (5.4) gives an expression with various powers of $(\mathbf{m}^{\parallel} - \mathbf{m}_d)$. The coefficients of each power must match, giving for powers 0, 1, and 2 respectively

$$\frac{dA}{d\tau} = -\left(2\nu + \frac{4D}{W}\right)A \tag{5.20}$$

$$\frac{d\mathbf{m}_d}{d\tau} = -\Omega\mathbf{m}_d\times\hat{\mathbf{z}} + \nu\mathbf{m}_d + \mathbf{T} \tag{5.21}$$

and

$$\frac{dW}{d\tau} = 2\nu W + 4D. \tag{5.22}$$

Note that (5.21) is exactly the deterministic part of the linearized LLG equation (5.9).

As in Section III, we can relate D to the temperature by insisting that in equilibrium ($\nu = -1$, due to damping) the Boltzmann distribution ($W = 1/\Delta$) is a time-independent solution

$$D = \frac{1}{2\Delta}. \tag{5.23}$$

As in (3.13), it is convenient to make a distinction between the initial thermal distribution, with thermal stability factor, Δ_0, and the thermal stability factor during switching Δ. Then solving (5.22) with the initial condition $W(0) = 1/\Delta_0$ gives an increasing width

$$W(\tau) = \frac{1}{\Delta_0}e^{2\nu t} + \frac{1}{\nu\Delta}(e^{2\nu t} - 1). \tag{5.24}$$

Normalization of the probability (5.15) requires that

$$A(\tau) = \frac{1}{\pi W(\tau)} \tag{5.25}$$

which is easily seen to be consistent with (5.20) and with (3.11) when allowance is made for the factor of 2π needed to account for integration over the azimuthal angle as well as the polar angle in the present context.

To understand the motion of the drifting center, $\mathbf{m_d}$ (5.21), note that if there is no tilt ($\mathbf{T} = 0$), $\mathbf{m}^{\parallel}$ simply precesses about the origin in a circle of some radius m (Fig. 8) with frequency Ω. The velocity at the top of the orbit is $d\mathbf{m}^{\parallel}/d\tau = -\Omega\mathbf{m}^{\parallel}\times\hat{\mathbf{z}}$ with magnitude $\Omega|\mathbf{m}^{\parallel}|$.

We can use Fig. 8 to visualize why there is a stagnation point—for simplicity, let us ignore damping, which is smaller by a factor of α. If we add a tilt $\mathbf{T}$, pointing to the right, (5.21) indicates that this adds $\mathbf{T}$ to the velocity, exactly canceling the precession velocity if $m = T/\Omega$. Thus this is the stagnation point $\mathbf{m}_s$ at which the magnetization will remain stationary, and $m_s \equiv |\mathbf{m}_s| = T/\Omega$. If the tilt arises from a pinned-layer tilt, $m_s = im_p^{\parallel}/\Omega \sim i\alpha m_p^{\parallel}$, so $m_s \ll m_p^{\parallel}$; the stagnation point tilt is much less than the pinned-layer tilt. It is not hard to see that if $\mathbf{m}^{\parallel}$ starts somewhere else, it will precess around this stagnation point (dashed circle). If we now take damping into account, we can find the stagnation point by setting the torque (5.9) equal to zero and solving for $\mathbf{m}^{\parallel}$: the stagnation point is slightly modified to

$$\mathbf{m}_s = -\frac{\Omega\mathbf{T}\times\hat{\mathbf{z}} + \nu\mathbf{T}}{\Omega^2 + \nu^2}. \tag{5.26}$$

The displacement of the drifting center $\mathbf{m}_d$ from the stagnation point, $\boldsymbol{\delta} \equiv \mathbf{m}_d - \mathbf{m}_s$ obeys [from (5.21)] the simple equation

$$\frac{d\boldsymbol{\delta}}{d\tau} = -\Omega\boldsymbol{\delta}\times\hat{\mathbf{z}} + \nu\boldsymbol{\delta} \tag{5.27}$$

which describes a vector that rotates at the rate Ω and expands at the rate ν

$$\boldsymbol{\delta} = e^{\nu\tau}\delta_0\hat{\mathbf{n}} \tag{5.28}$$

where

$$\hat{\mathbf{n}} \equiv (\cos(\Omega\tau + \xi), \sin(\Omega\tau + \xi), 0) \tag{5.29}$$

is a unit vector rotating about the z axis, $\delta_0 \equiv |\boldsymbol{\delta}(\tau = 0)| = |\mathbf{m}_d(0) - \mathbf{m}_s|$ is the initial displacement from the stagnation point, and the initial phase angle is

$$\xi \equiv \tan^{-1}\left[\frac{\delta_y(0)}{\delta_x(0)}\right] = \tan^{-1}\left[\frac{m_{dy}(0) - m_{sy}}{m_{dx}(0) - m_{sx}}\right]. \tag{5.30}$$

Thus the drifting center spirals outward from the stagnation point

$$\mathbf{m}_d(\tau) = \mathbf{m}_s + e^{\nu\tau}\delta_0\hat{\mathbf{n}}. \tag{5.31}$$

To describe the case of a canted pinned layer and an initially perpendicular free layer, we would take initial position of the drifting center $\mathbf{m}_d(0)$ to be at the origin, so $\delta_0 = \mathbf{m}_s$. However, these equations can also describe the linear approximation to the case of an initially canted free layer, in which case δ_0 is the canting angle. Our final exact result for the probability density (within this linear approximation for $\mathbf{m}^{\parallel}$) is given by (5.15), together with (5.24), (5.25), and (5.31).

To estimate the nonswitching probability $P_{NS}(\tau)$, we choose an angle θ_{sw} (measured from the stagnation point, which is typically almost the same as measuring it from the origin) beyond which we will assume switching becomes inevitable (in Section IV, we took this to be $\pi/2$). After most systems have switched (i.e., in the long time tail of P_{NS}) we can neglect 1 in comparison with the exponential growth factor $e^{2\nu\pi}$, so the width becomes

$$W(\tau) \approx \left(\frac{1}{\Delta_0} + \frac{1}{\nu\Delta} \right) e^{2\nu\tau}. \qquad (5.32)$$

Because this is much larger than the switching angle $\theta_{\rm sw}$, the probability density is nearly constant over the switching circle, and we may approximate the integral by the value at $\mathbf{m}_s$ times the area $\pi\theta_{\rm sw}^2$ of the circle. From (5.31), the $(\mathbf{m}^{\|} - \mathbf{m}_d(\tau))^2$ in (5.15) is $\delta_0^2 e^{2\nu\tau}$, giving a general result for the nonswitching probability at long times

$$P_{NS}(\tau) \approx \frac{\theta_{sw}^2}{\frac{1}{\Delta_0} + \frac{1}{\nu\Delta}} e^{-2\nu\tau} \exp\left[-\frac{\delta_0^2}{\frac{1}{\Delta_0} + \frac{1}{\nu\Delta}} \right]. \qquad (5.33)$$

B. Canted Pinned Layer

In the case of a pinned magnetization tilt $m_P^{\|}$, the initial displacement from the stagnation point is given by

$$\delta_0^2 = m_s^2 = \frac{T^2}{(\Omega^2 + \nu^2)} \approx \frac{\left(i m_p^{\|}\right)^2}{\Omega^2}. \qquad (5.34)$$

If we assume (as previously) that $\Delta = \Delta_0$, the long time WSER becomes

$$P_{NS}(\tau) \approx \theta_{sw}^2 \Delta \frac{\nu}{\nu+1} e^{-2\nu\tau} \exp\left[-\Delta \frac{\nu}{\nu+1} m_s^2 \right]. \qquad (5.35)$$

Note that this differs from our earlier analytic result for the axially symmetrical case (tilt $\mathbf{T} = 0$) only by the last exponential factor $\exp[-\Delta(\nu/\nu + 1)m_s^2]$. We can relate this to a simpler estimate obtained by considering only the effect of shifting the stagnation point away from the origin. By this estimate we simply gain a factor of the initial thermal probability of being at the stagnation point, the Boltzmann factor, $\exp[-\Delta_0 m_s^2]$. In the limit of high overdrive ν, when diffusion is negligible, this is exactly correct. The additional term $1/(\nu\Delta)$ in (5.33) increases the nonswitching probability by taking into account diffusion back to the stagnation point of systems that do not start there. In general, as discussed previously in Section IV, we expect a linearized theory to work best for high overdrive. Fortunately, to get a reasonably low nonswitching probability it is normally necessary to use a fairly high overdrive.

C. Canted Initial Magnetization

As an additional test of the linearized theory, we can compare to numerical solutions of the Fokker-Planck equation for the case of a definite initial angle. Then the drifting center is initially at $\mathbf{m}_d(0) = (\theta_0, 0, 0)$ and the spread of the initial angle (proportional to $1/\Delta_0$) is zero. In this case, $\mathbf{m}_s = 0$, because we are not considering a canted free layer. Thus (5.33) gives the

long time tail of the nonswitching probability, with $\delta_0 = \theta_0$ and $1/\Delta_0 = 0$

$$P_{NS}(\tau) \approx \theta_{sw}^2 \nu\Delta e^{-2\nu\tau} \exp\left[-\nu\Delta\theta_0^2 \right]. \qquad (5.36)$$

However, in this case we can get more than just the long time tail. Using equations (5.15), (5.24), and (5.25) with $1/\Delta_0 \to 0$, we have

$$\rho(\mathbf{m}^{\|}, \tau) = \frac{1}{\pi W} \exp\left[-\frac{\left(\mathbf{m}^{\|} - \mathbf{m}_d\right)^2}{W} \right] \qquad (5.37)$$

with

$$W = \frac{(e^{2\nu\tau} - 1)}{\nu\Delta} \qquad (5.38)$$

where we now include the "−1" since we no longer assume τ is large. We can evaluate $\left(\mathbf{m}^{\|} - \mathbf{m}_d\right)^2$ using (5.31) with $\mathbf{m}_s = 0$ as

$$
\begin{aligned}
(\mathbf{m}^{\|} - \mathbf{m}_d)^2 &= m^{\|2} + m_d^2 - 2\mathbf{m}^{\|} \cdot \mathbf{m}_d \\
&= \theta^2 + \theta_0^2 e^{2\nu\tau} - 2\theta\theta_0 e^{\nu\tau} \cos(\phi - \phi_0 - \Omega\tau).
\end{aligned} \qquad (5.39)
$$

Thus the distribution function in our linear approximation is

$$\rho(\theta, \phi, \tau) = \frac{e^{-\frac{\theta_0^2 e^{2\nu\tau}}{W}} e^{-\frac{\theta^2}{W}}}{\pi W} \exp\left[\frac{2\theta\theta_0 e^{\nu\tau}}{W} \cos(\phi - \phi_0 - \Omega\tau) \right]. \qquad (5.40)$$

We estimate the nonswitched fraction as before by integrating over the upper-half of the sphere. The integral over the azimuthal angle gives a modified Bessel function

$$\oint \exp(x \cos\phi) d\phi = 2\pi I_0(x) \qquad (5.41)$$

which allows us to write the distribution as a function of time and polar angle as

$$\rho(\theta, \tau) = \frac{2}{W} e^{-\frac{\theta_0^2 e^{2\nu\tau}}{W}} e^{-\frac{\theta^2}{W}} I_0\left(\frac{2\theta\theta_0 e^{\nu\tau}}{W} \right). \qquad (5.42)$$

The nonswitched fraction is given by

$$
\begin{aligned}
P_{NS} &= e^{-\frac{\theta_0^2 e^{2\nu\tau}}{W}} \frac{2}{W} \int_0^{\frac{\pi}{2}} e^{-\frac{\theta^2}{W}} I_0\left(\frac{2\theta\theta_0 e^{\nu\tau}}{W} \right) \theta d\theta \\
&= e^{-\frac{\theta_0^2 e^{2\nu\tau}}{W}} \int_0^{\frac{\pi^2}{4W}} e^{-x} I_0\left(2\theta_0 e^{\nu\tau} \sqrt{\frac{x}{W}} \right) dx.
\end{aligned} \qquad (5.43)
$$

This approximation is plotted as the dashed lines in Fig. 9 and may be compared to the solid lines which are solutions to the Fokker-Planck equation without linearization. At long times, $1/W \to \nu\Delta \exp(-2\nu\tau)$, which restricts the integral at the right in (5.43) to the region around $x = 0$ where the integrand is unity, and (5.43) reduces to our previous expression (5.36) (with the switching angle θ_{sw} set to $\pi/2$) for the long time tail. This approximation is shown as the dotted line in Fig. 9.

The key practical question is whether the stagnation point can in fact be shifted out (or mostly out) of the initial thermal prob-
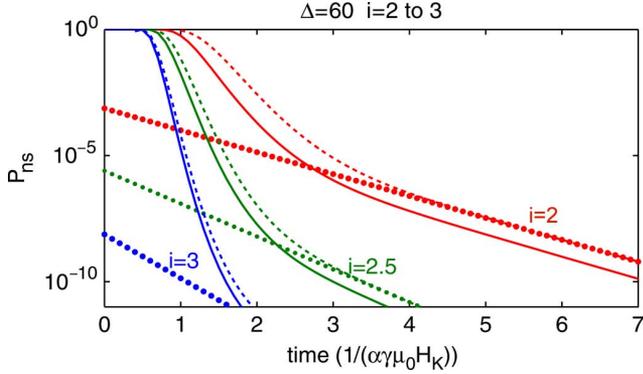
Fig. 9. Nonswitched fraction for reduced current, i = 2, 2.5, and 3 (ν = 1, 1.5, 2). Solid lines are numerical Fokker-Planck solutions; dashed lines are obtained from linear theory (5.43). Dotted lines are long time approximation to linear theory (5.36).

ability distribution. Simplifying to the case $h = 0$ and dropping terms of relative size α or F, the tilt-dependent exponential factor in (5.35), is

$$\exp\left[-i(i-1)m_p^{\|2}\alpha^2\Delta\right]. \tag{5.44}$$

Even if the tilt angle is large ($m_p^{\|} \sim 1$) and estimating $i \sim 2$, $\alpha \sim 0.02$, even if Δ is as high as 60, the gain factor is still negligible, $\exp(-0.024)$. We conclude that it is very difficult to shift the stagnation point by tilting the pinned magnetization. It appears from (5.44) that this is because of the factor α^2, which may seem surprising because the stagnation point shift m_s arises from the balance between the tilted pinned magnetization spin torque and the precession term, neither of which involves the damping α. However, this is because we have incorporated the factor of α into the dimensionless current i; each factor $i\alpha$ in (5.44) is actually just proportional to the physical current. It is the smallness of the physical current that makes the gain small. Imposing an in-plane magnetic field may be more promising—in the context of (5.12), it is enhanced by a factor $1/\alpha$. A static bias field does not help, of course, since it will shift the initial distribution to the stagnation point and we lose the enhancement factor involving m_s [the last exponential in (5.33)]. However, if we turn on the field when we turn on the current, (5.35) is valid with $m_s \approx h_{\|}$, and there is an improvement factor in P_{NS} of

$$\exp\left[-\Delta\frac{\nu}{\nu+1}h_{\|}^2\right] \tag{5.45}$$

which can be substantial—if we set the tilt angle of the total field $h_{\|}$ to 0.45, with $\Delta = 60$ and $\nu = 1$ the factor becomes $e^{-6} \sim 0.0025$.

VI. SOFT ERROR RATES FOR READ DISTURB

Fig. 10 shows the distribution function calculated for $\Delta = 60$ and for $i = 0.5$ as a function of time. It is assumed that a current is applied for reading in order to determine whether the device is in a low (typically parallel moments of free and pinned layers) or a high (anti-parallel moments) resistance state. The applied current has a nonzero probability of causing or assisting a switching event. Spin-torque memory devices must be capable
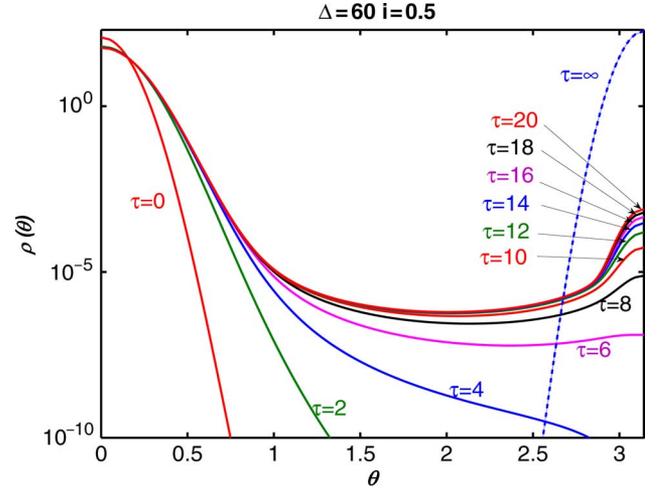
Fig. 10. Probability distribution as a function of time for $\Delta = 60$ and $i = 0.5$.

of switching from anti-parallel to parallel and from parallel to anti-parallel when information is being stored. Typically, the spin-torque efficiency is somewhat higher for anti-parallel to parallel switching than for parallel to anti-parallel [10], [45], [46], thus anti-parallel to parallel switching is usually easier. For reading one has the freedom to choose the current direction and thus one can use the current direction that stabilizes the anti-parallel configuration, i.e., electrons flow from free layer to pinned layer, thereby minimizing the probability of accidental switching into the parallel state.

Fig. 11(a) and (b) shows the calculated read soft error rate (RSER) for $\Delta = 60$ ($\Delta = 30$) and values of the reduced current between $i = 0.1$ and $i = 0.6$. The solid lines represent the numerical solutions to the Fokker-Planck equation. The dashed lines represent an approximate solution to the Fokker-Planck equation obtained by Brown for the case of a Stoner-Wohlfarth particle in an external magnetic field. We can apply his result to our case because of the correspondence between the spin-torque current and the applied magnetic field established in (2.5).

The switching probability is observed to increase very rapidly initially, and then, after $\tau \approx 5$, the rate of increase slows and the switching probability is observed to increase linearly at a rate that increases with i. This can be understood in terms of a rapid equilibration within the well at $\theta = 0$ for $\tau <\sim 5$ followed by Brown-Kramers hopping over the energy barrier. The initial equilibration occurs because the effective energy function within the well changes when the current is applied. The Brown-Kramers approximation [33]–[35] to the switching probability is linear in τ

$$P_{sw} = \tau\sqrt{\frac{\Delta}{\pi}}\left(1-(i-h)^2\right)\left\{(1-i+h)\exp\left[-\Delta(1-i+h)^2\right]\right.$$
$$\left.+(1+i-h)\exp\left[-\Delta(1+i-h)^2\right]\right\}. \tag{5.46}$$

The generalization of Brown's original derivation to include spin-torque is given in Appendix B. It can be seen from Figs. 9 and 10 that (5.46) provides an upper limit for the switching probability and that it significantly overestimates the switching probability for reduced times less than ~ 10. Such times may be of interest for spin-torque devices, for example if $\alpha = 0.01$ and $\theta_0 H_K = 1$ T, this would correspond to ~ 6 ns.
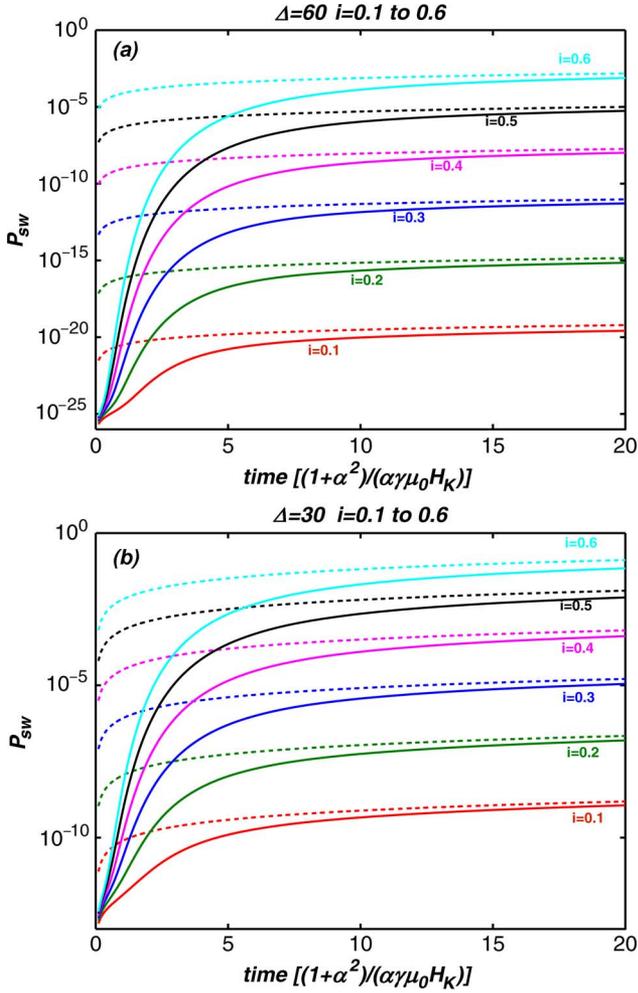
Fig. 11. Calculated RSER for (a) $\Delta = 60$ and (b) $\Delta = 30$. Solid lines are solutions to Fokker-Planck equation. Dashed lines were calculated using (5.46).

VII. CONCLUSION

In summary, we investigated spin-torque switching for devices in which the magnetization of the pinned and free layer is perpendicular to the plane of the layers, treating the free layer as a macrospin. Our investigation emphasized the time-dependent probability for not switching when the applied current exceeds the critical current for switching and the probability for switching when the applied current is significantly below the critical current for switching. The former case determines the WSER and the latter case determines the RSER. The Fokker-Planck equation was solved for the switching distribution in the presence of random thermal fluctuations. Results were presented in terms of reduced currents and reduced time so that effectively all relevant cases are represented in the figures. We also provide approximate analytical formulas that can be used to estimate the RSER and WSERs. For axially symmetric macrospins we find that the spin-polarized current enters the Fokker-Planck equation in essentially the same way as an axial magnetic field, allowing previous results derived for magnetic field induced switching to be used to describe current induced switching. For a perpendicular pinned layer and a free layer with net perpendicular anisotropy we find that the WSER decays at long times as $\exp(-2\nu\tau)$, where ν is defined in term of the critical current

for switching, I_0 as $\nu = (I - I_0)/I_0$, and $\tau = \alpha\gamma\mu_0 H_K t$. We relate this long time decay to the lack of spin-torque when the pinned and free layer magnetizations are collinear (stagnation point). We also investigated the effects of a canted pinned layer and found that it shifted the position of the stagnation point away from the perpendicular direction, but the amount of this shift is smaller by a factor of approximately α (Gilbert damping rate) than the tilt of the pinned layer. Thus the canted pinned layer only increases the long time tail decay rate by a small amount [see (5.44)].

To investigate the RSER, we solved the Fokker-Planck equation for values of the current that were less than the critical current for switching. We also generalized the Brown-Kramers low-field expression for field switching to include the case of current induced switching. We found that it overestimates the RSER for very short read times.

APPENDIX A
POSITION OF STAGNATION POINT

In this appendix, we prove that there is a stagnation point even if we do not linearize the LL equation. A stagnation point exists if, for some magnetization direction $\hat{\mathbf{m}}$, the torque in (1.3) vanishes; using the notation of (2.1), this gives

$$-\cos\theta\,\hat{\mathbf{m}}\times\hat{\mathbf{z}} - \alpha\cos\theta\,\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}}) + i\alpha\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p) = 0 \tag{A.1}$$

(note that in (2.1) this $\hat{\mathbf{m}}_p$ was replaced by $\hat{\mathbf{z}}$; here we assume an arbitrary pinned-layer direction.) For simplicity, we have omitted the external field, the field like term, and higher order terms in α. They can be included at the cost of a slightly more complicated notation. This equation involves three vectors, $\hat{\mathbf{m}}\times\hat{\mathbf{z}}$, $\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})$, and $\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)$. The first two of these vectors are orthogonal to each other and all three are orthogonal to $\hat{\mathbf{m}}$. Thus the 2-D space orthogonal to $\hat{\mathbf{m}}$ will be spanned by the two orthogonal vectors, $\hat{\mathbf{m}}\times\hat{\mathbf{z}}$ and $\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})$. This allows us to express the third vector in terms of the first two as

$$[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)] = a\hat{\mathbf{m}}\times\hat{\mathbf{z}} + b[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})]. \tag{A.2}$$

Since

$$(\hat{\mathbf{m}}\times\hat{\mathbf{z}})\cdot[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)] = \hat{\mathbf{m}}_p\cdot(\hat{\mathbf{m}}\times\hat{\mathbf{z}}) \text{ and}$$
$$[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})]\cdot[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{m}}_p)] = \hat{\mathbf{m}}_p\cdot[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})] \tag{A.3}$$

we have

$$a = \frac{\hat{\mathbf{m}}_p\cdot(\hat{\mathbf{m}}\times\hat{\mathbf{z}})}{\sin^2\theta} \quad \text{and} \quad b = \frac{\hat{\mathbf{m}}_p\cdot[\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})]}{\sin^2\theta}. \tag{A.4}$$

Thus a stagnation point will exist if, for some $\hat{\mathbf{m}}$,

$$\left[ia - \frac{1}{\alpha}\cos\theta\right](\hat{\mathbf{m}}\times\hat{\mathbf{z}}) + [ib - \cos\theta][\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}})] = 0 \tag{A.5}$$

which implies $\cos\theta = \alpha ia = ib$.

Expressing $\hat{\mathbf{m}}$ in terms of polar coordinates, we have

$$\hat{\mathbf{m}} = \sin\theta(\cos\phi\hat{\mathbf{x}} + \sin\phi\hat{\mathbf{y}}) + \cos\theta\hat{\mathbf{z}}$$
$$\hat{\mathbf{m}}\times\hat{\mathbf{z}} = \sin\theta(\sin\phi\hat{\mathbf{x}} - \cos\phi\hat{\mathbf{y}})$$
$$\hat{\mathbf{m}}\times(\hat{\mathbf{m}}\times\hat{\mathbf{z}}) = \cos\theta\sin\theta(\cos\phi\hat{\mathbf{x}} + \sin\phi\hat{\mathbf{y}}) - \sin^2\theta\hat{\mathbf{z}}. \tag{A.6}$$

Without loss of generality we can assume that the in-plane component of $\hat{\mathbf{m}}_p$ is in the $\hat{\mathbf{x}}$ direction, so that

$$\hat{\mathbf{m}}_p = \cos\eta\hat{\mathbf{z}} + \sin\eta\hat{\mathbf{x}}. \tag{A.7}$$

Then, we have

$$a = \frac{\sin\phi\sin\eta}{\sin\theta}; \quad b = \frac{-\cos\eta\sin\theta + \cos\theta\cos\phi\sin\eta}{\sin\theta} \tag{A.8}$$

which leads to

$$\sin\theta\cos\theta = i\alpha\sin\phi\sin\eta$$
$$\sin\theta\cos\theta = i(\cos\theta\cos\phi\sin\eta - \sin\theta\cos\eta). \tag{A.9}$$

These two equations determine θ_0 and ϕ_0, the direction of $\hat{\mathbf{m}}$, for which the torque vanishes.

If the pinned layer magnetization is completely in-plane ($\sin\eta = 1$, $\cos\eta = 0$), the (A.9) for θ_0 and ϕ_0 become particularly simple

$$\sin\theta_0 = i\cos\phi_0; \quad \cos\theta_0 = \alpha\tan\phi_0 \tag{A.10}$$

with solution

$$\cos\phi_0 = \frac{1}{i\sqrt{2}}\sqrt{(1+\alpha^2) - \sqrt{(1+\alpha^2)^2 - 4i^2\alpha^2}}$$
$$= \alpha + \frac{1}{2}(i^2-1)\alpha^3 + \cdots$$
$$\sin\theta_0 = \frac{1}{\sqrt{2}}\sqrt{(1+\alpha^2) - \sqrt{(1+\alpha^2)^2 - 4i^2\alpha^2}}$$
$$= i\alpha\left[1 + \frac{1}{2}(i^2-1)\alpha^2 + \cdots\right]\left(\eta = \frac{\pi}{2}\right). \tag{A.11}$$

For a general angle η, we can take advantage of the fact that $i\alpha$ is normally $\ll 1$, which allows us to approximate (A.9) by

$$\theta \approx i\alpha\sin\phi\sin\eta = i(\sin\eta\cos\phi - \theta\cos\eta) \tag{A.12}$$

which leads to

$$\phi_0 \approx \frac{\pi}{2} - \alpha(1 + i\cos\eta); \quad \theta_0 \approx i\alpha\sin\eta. \tag{A.13}$$

This is consistent with the solution obtained in Section V. Numerical solutions of (A.9) were well represented by the linearized estimate (A.13) for reasonable values of i and α.

APPENDIX B
DERIVATION OF GENERALIZED BROWN-KRAMERS HIGH-ENERGY BARRIER FORMULA IN PRESENCE OF SPIN TORQUE

In this appendix, we present an approximate solution to the Fokker-Planck equation for the case in which the current is less than the critical current for switching. The approximation is valid in a limit in which the distribution function has come to quasi-equilibrium within each of two minima in the effective energy, but the total density in the two wells is not in equilibrium. The development follows closely Brown's treatment [33] of the analogous problem for field induced switching.
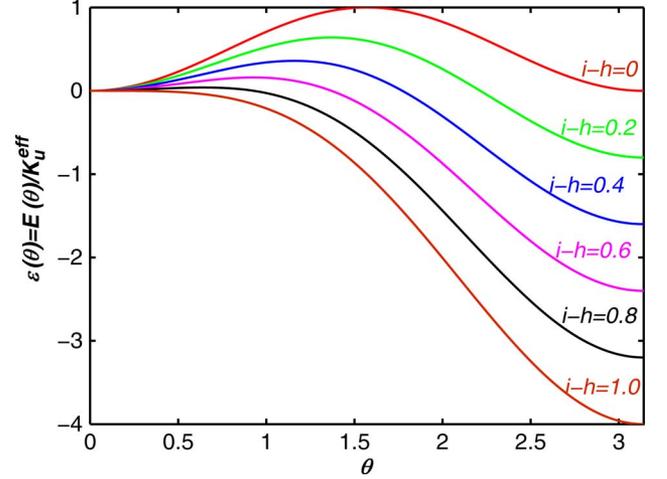
Fig. 12. Effective energy function for values of $i - h$ between 0 and 1.

The Fokker-Planck equation is given by (3.8)

$$\frac{\partial\rho(\theta,\tau)}{\partial\tau} = -\nabla\cdot\mathbf{J}(\theta,\tau) \tag{B.1}$$

where

$$J_\theta(\theta,\tau) = \left[\sin\theta(i - h - \cos\theta)\rho(\theta,\tau) - \frac{1}{2\Delta}\frac{\partial\rho(\theta,\tau)}{\partial\theta}\right]. \tag{B.2}$$

For very long times the system comes to equilibrium; $\partial\rho(\theta,\tau)/\partial\tau = 0$, and $J_\theta(\theta,\tau) = 0$. Using the latter result we have $(1/\rho(\theta,\infty))(\partial\rho(\theta,\infty)/\partial\theta) = 2\Delta\sin\theta(i - h - \cos\theta)$. Integrating from $\theta = 0$ to θ yields

$$\ln\rho(\theta,\infty) - \ln\rho(0,\infty)$$
$$= 2\Delta\int_0^\theta d\theta\sin\theta(i - h - \cos\theta)$$
$$= -\Delta\left[\sin^2\theta + 2(i - h)(\cos\theta - 1)\right] \tag{B.3}$$

or

$$\rho(\theta,\infty) = \rho(0,\infty)\exp\left\{-\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\}. \tag{B.4}$$

According to this result, the equilibrium distribution is appropriate to an effective energy function that has been modified from $E(\theta)/K_U^{\text{eff}} \equiv \varepsilon(\theta) = \sin^2\theta$ to $\varepsilon_{eff}(\theta) = \sin^2\theta + 2(i - h)(\cos\theta - 1)$ by the spin polarized current and the applied field. Fig. 12 has minima at $\theta = 0$ and $\theta = \pi$ and a maximum at $\cos\theta = i - h$ if $i - h < 1$.

We now consider the case for which there is a maximum in the effective energy function ($i - h < 1$) and the current pulse has been applied long enough for local equilibrium to be established in the vicinity of the bottom of the two wells, but there has not been sufficient time for the distribution to be equilibrated between the wells. If we assume that the effective energy function in the vicinity of each minimum is given by (B.4), we can relate the total probability of being in either well to the value of

the distribution functions at the minima. For the well at $\theta = 0$, we have

$$n_0 = \rho_0 \int_0^{\theta_0} \sin\theta d\theta \exp\left\{-\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\}$$
(B.5)

and for the well at $\theta = \pi$

$$n_\pi = \rho_\pi \int_{\pi-\theta_\pi}^{\pi} \sin\theta d\theta \exp\left\{-\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right.\right.$$
$$\left.\left. +4(i-h)\right]\right\}. \quad \text{(B.6)}$$

Here, θ_0 should be large enough and θ_π small enough that almost all of the distribution on the respective sides of the maximum are contained within the ranges $0 \le \theta \le \theta_0$ and $\theta_\pi \le \theta \le \pi$, respectively. We also require that θ_0 be less than the angle for which the effective energy is a maximum and θ_π be greater $(\theta_0 < \cos^{-1}(i-h) < \theta_\pi)$.

We can approximate the integrals in (B.5) and (B.6) by expanding the effective energy about the respective minima to order θ^2, approximating $\sin\theta$ by θ and extending the range of integration to infinity, yielding

$$n_0 = \frac{\rho_0}{2\Delta(1-(i-h))} \qquad n_\pi = \frac{\rho_\pi}{2\Delta(1+(i-h))}. \quad \text{(B.7)}$$

Because equilibrium has not yet been established between the wells, there must be a current that flows between them. If we assume that there is no appreciable accumulation of the particles in the region between θ_0 and θ_π, we can calculate the current flowing between the wells. We concentrate on the interval $\theta_0 < \theta < \theta_\pi$. By assumption, $\partial\rho(\theta,\tau)/\partial\tau = 0$ for this region. This implies that $J_\theta(\theta,\tau)\sin\theta$ is constant in this region

$$J_\theta(\theta,\tau)\sin\theta = \sin\theta\left[\sin\theta(i-h-\cos\theta)\rho(\theta,\tau) - \frac{1}{2\Delta}\frac{\partial\rho(\theta,\tau)}{\partial\theta}\right]$$
$$= C. \quad \text{(B.8)}$$

This may be written as

$$\left[2\Delta\sin\theta(i-h-\cos\theta)\rho(\theta,\tau) - \frac{\partial\rho(\theta,\tau)}{\partial\theta}\right] = 2\Delta\frac{C}{\sin\theta}. \quad \text{(B.9)}$$

Multiplication of both sides by $\exp[\Delta\varepsilon_{\text{eff}}(\theta)]$ allows us to integrate the left-hand side between θ_0 and θ_π

$$\frac{\partial}{\partial\theta}\left[\rho(\theta)\exp\left\{\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\}\right]$$
$$= -2\Delta\frac{C}{\sin\theta}\exp\left\{\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\} \quad \text{(B.10)}$$

or

$$\left[\rho(\theta)\exp\left\{\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\}\right]_{\theta_0}^{\theta_\pi}$$
$$= -2\Delta C \int_{\theta_0}^{\theta_\pi} \frac{d\theta}{\sin\theta}\exp\left\{\Delta\left[\sin^2\theta + 2(i-h)(\cos\theta - 1)\right]\right\}. \quad \text{(B.11)}$$

Using the assumed variations of ρ within the two wells, we have

$$\rho_\pi - \rho_0 = -2\Delta C \int_{\theta_0}^{\theta_\pi} \frac{d\theta}{\sin\theta}\exp\left\{\Delta\left[\sin^2\theta\right.\right.$$
$$\left.\left. +2(i-h)(\cos\theta - 1)\right]\right\}. \quad \text{(B.12)}$$

The argument of the exponential has a maximum at $\cos\theta_m = (i-h)$. Expanding the argument about this maximum and approximating $\sin\theta$ by its value at the maximum, we have

$$C = \frac{(\rho_0 - \rho_\pi)}{2\Delta}\sqrt{\frac{\Delta}{\pi}}\left(1 - (i-h)^2\right)\exp\left[-\Delta(i-h-1)^2\right]. \quad \text{(B.13)}$$

Using (B.7), we have

$$C = \left[n_0(1 - i + h) - n_\pi(1 + i - h)\right]\sqrt{\frac{\Delta}{\pi}}\left(1 - (i-h)^2\right)$$
$$\exp\left[-\Delta(i-h-1)^2\right]. \quad \text{(B.14)}$$

For the case where n_π is still very small, the switched fraction will be given by

$$P_{sw} = \tau\sqrt{\frac{\Delta}{\pi}}(1 - i + h))^2(1 + i - h))\exp\left[-\Delta(1 - i + h))^2\right]. \quad \text{(B.15)}$$

This may be written as

$$\ln\left(\frac{P_{sw}}{\tau}\right) - \ln\left(\sqrt{\frac{\Delta}{\pi}}\right) = \ln\left[(1 - i + h)^2(1 + i - h)\right]$$
$$-\Delta(1 - i + h)^2. \quad \text{(B.16)}$$

REFERENCES

[1] J. C. Slonczewski, *Phys. Rev. B*, vol. 39, p. 6995, 1989.
[2] J. C. Slonczewski, *J. Magn. Magn. Mater.*, vol. 159, p. L1, 1996.
[3] D. C. Worledge, G. Hu, D. W. Abraham, J. Z. Sun, P. L. Trouilloud, J. Nowak, S. Brown, M. C. Gaidis, E. J. O'Sullivan, and R. P. Robertazzi, *Appl. Phys. Lett.*, vol. 98, p. 022501, 2011.
[4] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, *Phys. Rev. B*, vol. 63, p. 054416, 2001.
[5] W. H. Butler, "Tunneling magnetoresistance from a symmetry filtering effect," *Sci. Technol. Advanced Mater.*, vol. 9, p. 014106, 2008.
[6] X.-G. Zhang and W. H. Butler, *Phys. Rev. B*, vol. 70, p. 172407, 2004.
[7] X.-G. Zhang and W. H. Butler, *J. Phys. Condens. Mater.*, vol. 15, pp. R1–R37, 2003.
[8] J. Slonczewski, *J. Magn. Magn. Mater.*, vol. 195, p. L261, 1999.
[9] J. Slonczewski, *J. Magn. Magn. Mater.*, vol. 247, p. 324, 2002.
[10] J. C. Slonczewski, *Phys. Rev. B*, vol. 71, p. 024411, 2005.
[11] A. Manchon and J. C. Slonczewski, *Phys. Rev. B*, vol. 73, p. 184419, 2006.
[12] L. Berger, *Phys. Rev. B*, vol. 54, p. 9353, 1996.

[13] D. M. Edwards, F. Federici, J. Mathon, and A. Umerski, *Phys. Rev. B*, vol. 71, p. 054407, 2005.

[14] I. Theodonis, N. Kioussis, A. Kalitsov, M. Chshiev, and W. H. Butler, *Phys. Rev. Lett.*, vol. 97, p. 237205, 2006.

[15] A. Kalitsov, I. Theodonnis, N. Kioussis, M. Chshiev, W. H. Butler, and A. Vedyayev, *J. Appl. Phys.*, vol. 99, p. 08G501, 2006.

[16] M. Chshiev, I. Theodonis, A. Kalitsov, N. Kioussis, and W. H. Butler, "Voltage dependence of spin transfer torque in magnetic tunnel junctions," *IEEE Trans. Magn.*, vol. 44, no. 1, p. 2543, Jan. 2008.

[17] A. Kalitsov, M. Chshiev, I. Theodonis, N. Kioussis, and W. H. Butler, *Phys. Rev. B*, vol. 79, p. 174416, 2009.

[18] Y.-H. Tang, N. Kioussis, A. Kalitsov, W. H. Butler, and R. Car, *Phys. Rev. Lett.*, vol. 103, p. 057206, 2009.

[19] Y.-H. Tang, N. Kioussis, A. Kalitsov, W. H. Butler, and R. Car, *Phys. Rev. B*, vol. 81, p. 054437, 2010.

[20] Y.-H. Tang, N. Kioussis, A. Kalitsov, W. H. Butler, and R. Car, *J. Phys. Conf. Series*, vol. 200, p. 062033, 2010.

[21] Z. Li and S. Zhang, *Phys. Rev. B*, vol. 68, p. 024404, 2003.

[22] Z. Li and S. Zhang, *Phys. Rev. B*, vol. 69, p. 134416, 2004.

[23] S. Zhang and Z. Li, *Phys. Rev. Lett.*, vol. 93, p. 127204, 2004.

[24] Z. Li, S. Zhang, Z. Diao, Y. Ding, X. Tang, D. M. Apalkov, Z. Yang, K. Kawabata, and Y. Huai, *Phys. Rev. Lett.*, vol. 100, p. 246602, 2008.

[25] R. Heindl, W. H. Rippard, S. E. Russek, and A. B. Kos, *Phys. Rev. B*, vol. 83, p. 055430, 2011.

[26] I. N. Krivorotov, N. C. Emley, J. C. Sankey, S. I. Kiselev, D. C. Ralph, and R. A. Buhrman, *Science*, vol. 307, p. 228, 2005.

[27] Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchura, E. Chen, L.-C. Wang, and Y. Huang, *J. Phys.: Condensed Matter*, vol. 19, p. 165209, 2007.

[28] H. Kubota, A. Fukushima, K. Yakushiji, T. Nagahama, S. Yuasa, K. Ando, H. Maehara, Y. Nagamine, K. Tsunekawa, D. D. Djayaprawira, N. Watanabe, and Y. Suzki, *Nature Phys.*, vol. 4, p. 37, 2008.

[29] C. Wang, Y.-T. Cui, J. Z. Sun, J. A. Katine, R. A. Buhrman, and D. C. Ralph, *Phys. Rev. B*, vol. 79, p. 224416, 2009.

[30] M. J. Jung, S. Park, S.-Y. You, and S. Yuasa, *Phys. Rev. B*, vol. 81, p. 134419, 2010.

[31] L. D. Landau and E. M. Lifshitz, *Phys. Z. Sowjet.*, vol. 8, pp. 153–169, 1935.

[32] T. L. Gilbert, *IEEE Trans. Magn.*, vol. 40, pp. 3443–3448, 2004.

[33] W. F. Brown, *Phys. Rev.*, vol. 130, pp. 1677–1686, 1963.

[34] W. F. Brown, *Physica*, vol. 86–88B, p. 1423, 1977.

[35] W. F. Brown, *IEEE Trans. Magn.*, vol. Mag-15, no. 6, p. 1196, Dec. 1979.

[36] A. Aharoni, *Phys, Rev.*, vol. 135, pp. A447–A449, 1964.

[37] A. Aharoni, *Phys. Rev.*, vol. 177, pp. 793–796, 1969.

[38] A. Aharoni, *Phys. Rev. B*, vol. 46, p. 5434, 1992.

[39] P. J. Cregg, D. S. F. Crothers, and A. W. Wickstead, *J. Appl. Phys.*, vol. 76, pp. 4900–4902, 1994.

[40] D. A. Garanin, *Phys. Rev. E*, vol. 54, pp. 3250–3255, 1996.

[41] J. Z. Sun, *Phys. Rev. B*, vol. 62, p. 570, 2000.

[42] D. M. Apalkov and P. B. Visscher, *Phys. Rev. B*, vol. 72, p. 180405(R), 2005.

[43] G. Bertotti, I. E. Mayergoyz, and C. Serpico, *Nonlinear Magnetization Dynamics in Nanosystems*. Amsterdam, The Netherlands: Elsevier, 2009.

[44] C. K. A. Mewes and T. Mewes, Matlab Based Micromagnetics Code $\mathbb{M}^3$ [Online]. Available: http://bama.ua.edu/~tmewes/Mcube/Mcube.shtml

[45] C. Wang, Y.-T. Cui, J. Z. Sun, J. A. Katine, R. A. Buhrman, and D. C. Ralph, *Phys. Rev. B*, vol. 79, p. 224416, 2009.

[46] M. H. Jung, S. Park, C.-Y. You, and S. Yuasa, *Phys. Rev. B*, vol. 81, p. 134419, 2010.

[47] J. Z. Sun, T. S. Kuan, J. A. Katine, and R. H. Koch, in *Proc. SPIE*, 2004, vol. 5359, p. 445.

[48] J. He, J. Z. Sun, and Z. Zhang, *J. Appl. Phys.*, vol. 101, p. 09A501, 2007.

[49] L. E. J. Brouwer, "Über Abbildung von Mannigfaltigkeiten," *Mathematische Annalen*, vol. 71, no. 1, pp. 97–115, 1912.

[50] L. D. Landau, *Collected Papers*, D. ter Haar, Ed. New York: Gordon and Breach, 1967, p. 101.