Importance Sampling for Thermally Induced Switching and Non-Switching Probabilities in Spin-Torque Magnetic Nanodevices

YiMing Yu, Cyrill B. Muratov, and Richard O. Moore

Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 USA

Spin-transfer torque magnetoresistive random access memory is a potentially transformative technology in the non-volatile memory market. Its viability depends, in part, on one's ability to predictably induce or prevent switching; however, thermal fluctuations cause small but important errors in both the writing and reading processes. Computing these very small probabilities for magnetic nanodevices using naive MC simulations is essentially impossible due to their slow statistical convergence, but variance reduction techniques can offer an effective way to improve their efficiency. Here, we provide an illustration of how importance sampling can be efficiently used to estimate low read and write soft error rates of macrospin and coupled-spin systems.

Index Terms—**Importance sampling, macrospin, micromagnetics, rare events, spintronics.**

I. INTRODUCTION

SPIN-TRANSFER torque magnetoresistive random access memory (STT-MRAM) has been proposed as a non-volatile replacement for random access memory that offers high-speed, low-power consumption, non-volatility, and unlimited endurance [1]–[5]. One of the primary obstacles to its widespread deployment is physical scaling due to an increased error rate that accompanies smaller volumes of storage cells. A memory device should switch quickly and reliably when switching is intended and otherwise maintain its current state. However, thermal fluctuations in the magnetization orientation can sometimes induce unwanted switching during either storage or an attempted read event, or failure to switch during an attempted write event. These contribute to the write soft error rate (WSER), read soft error rate (RSER), and retention failure rate [6]. The expected values of WSER and RSER in STT-MRAM should not exceed the order of 10^{-18} without error correction [2]. Due to the importance of these extremely small rates in quantifying the viability of experimental STT-MRAM configurations, analytical and computational techniques that facilitate their calculation are critically important.

One approach makes use of the Fokker–Planck equation (FPE) describing the time evolution of the switching probability [7]–[9]. In the macrospin approximation, treating each STT-MRAM bit as a single magnetic domain, the FPE can be solved directly or can be further approximated by the Brown–Kramers formula, which overestimates the RSER for short read times [6]. However, both the effects of spatial variations in the magnetization across a single memory cell and interactions between adjacent cells obviously cannot be captured in the macrospin approximation. These effects increase in importance as the size of each cell exceeds the scale of above 50 nm in lateral size and must, therefore, be taken into account to support development of magnetic nanodevices at this scale [10]. Direct numerical simulations of the FPE increase exponentially in computational cost as the dimension of the coupled system of macrospins is increased, and this high cost is exacerbated by the presence of boundary layers associated with the small size of thermal fluctuations. Efficient computational methods are therefore needed to provide a means of determining these small switching probabilities and rates.

The most common approach to computing switching probabilities uses sampling to provide an empirical estimate of the quantity of interest. However, the extremely low-switching probabilities and rates relevant to micromagnetic devices make naive Monte Carlo (MC) studies essentially impossible. A common approach to recover the tails of the probability distribution from MC simulations is via extrapolation (see [11]–[13]). However, this may introduce large uncontrolled inaccuracies due to the failure of the fitting form to capture the asymptotic behavior of the probability distribution in the small noise limit [14], [15]. Alternatively, variance reduction techniques such as importance splitting attempt to concentrate the samples generated on those with a higher likelihood of registering a rare event of interest [16]–[20].

Here, we demonstrate that accurate switching probabilities and error rates of STT-MRAM devices can be computed efficiently using another variance reduction technique known as importance sampling (IS) in an ensemble of biased MC simulations. We intentionally choose a particularly simple micromagnetic setting to clearly illustrate the IS methodology in the context of STT-MRAM modeling. The main point of this paper is to demonstrate that with the help of IS, the events with extremely low probability of occurrence can be accessed via direct MC sampling with an additional post-processing step. This together with the simplicity of its implementation could make the IS-based approach a powerful tool in assisting the design of the next generation of spintronic nanodevices.

Our paper is organized as follows. In Section II, we review IS and the strategies for optimal choices of bias. In Section III, we formulate the stochastic micromagnetic model for a single macrospin or a system of exchange coupled macrospins, and then show how to apply IS to sample switching probabilities in the macrospin approximation. In Section IV, we carry out IS simulations for a single macrospin and a coupled system of two identical macrospins with different choices of biases to demonstrate the power and efficiency of IS in the context of STT-MRAM applications. Finally, in Section V, we briefly summarize our findings.

II. IMPORTANCE SAMPLING

Suppose we wish to estimate the probability $P = \mathbb{E}_0[I(\omega)]$ of a system driven by random variable ω producing an event with indicator function $I(\omega)$. Here, $\mathbb{E}_0[\cdot]$ denotes the expected value with respect to the density $\rho_0(\omega)$, such that

$$P = \mathbb{E}_0[I(\omega)] = \int I(\omega)\rho_0(\omega)\,d\omega. \qquad (1)$$

A naive MC method uses M independent draws ω_i from ρ_0 to approximate P according to

$$P_{\mathrm{MC}} = \frac{1}{M}\sum_{i=1}^{M} I(\omega_i). \qquad (2)$$

The coefficient of variation (CV), which measures the statistical error of the sample, is given by

$$CV(P_{\mathrm{MC}}) = \frac{\sqrt{\mathrm{Var}(P_{\mathrm{MC}})}}{\mathbb{E}_0[P_{\mathrm{MC}}]} = \frac{1}{\sqrt{M}}\sqrt{\frac{1}{P_{\mathrm{MC}}} - 1}. \qquad (3)$$

For $P_{\mathrm{MC}} \ll 1$, smallness of $CV(P_{\mathrm{MC}})$ requires $M \gg 1/P_{\mathrm{MC}} \gg 1$. In the present case, this necessitates an impractically large number of simulations to produce a reasonable estimate of the probability. The idea behind IS is to sample ω from an alternative probability density $\rho_u(\omega)$ that depends on a bias u chosen to increase the likelihood of the event of interest. An unbiased estimator is then recovered by weighting each result according to

$$P_{\mathrm{IS}} = \frac{1}{M}\sum_{i=1}^{M} I(\omega_i)L(\omega_i), \qquad (4)$$

where $L(\omega) = \rho_0(\omega)/\rho_u(\omega)$ is called the likelihood ratio and assumes that $\rho_u(\omega) > 0$ whenever $\rho_0(\omega) > 0$. The resulting CV is then given by

$$\mathrm{CV}(P_{\mathrm{IS}}) = \frac{\sqrt{\mathrm{Var}(P_{\mathrm{IS}})}}{\mathbb{E}_u[P_{\mathrm{IS}}]} = \frac{1}{\sqrt{M}}\sqrt{\frac{\mathbb{E}_u[I(\omega)L^2(\omega)]}{\mathbb{E}_u[P_{\mathrm{IS}}]^2} - 1}, \qquad (5)$$

where $\mathbb{E}_u$ denotes the expectation with respect to ρ_u and $1/P \geq \mathbb{E}_u[I(\omega)L^2(\omega)]/P^2 \geq 1$. Expression (5) suggests that a "good" bias to use in IS keeps $\mathbb{E}_u[I(\omega)L^2(\omega)]/P^2$ close to 1.

In many cases of interest, including the one discussed here, the underlying physical model is described by a stochastic differential equation (SDE)

$$dX(t) = b(X(t))dt + \epsilon\sigma(X(t))dW(t), \quad X(0) = X_0, \qquad (6)$$

where $X(t) \in \mathbb{R}^N$, for some $N \geq 1$, is a randomly evolving state variable, $b(x)$ is its deterministic drift, $\sigma(x)$ is the noise strength matrix, dW is an infinitesimal increment of an N-dimensional Brownian motion, and $\epsilon > 0$ is the noise strength. In many situations of interest, the noise is weak: $\epsilon \ll 1$. If we wished to observe the state $X(t)$ exhibiting a behavior that is very rare for typical realizations of the Brownian motion $W(t)$, we would have to simulate this equation a very large number of times, even more so to produce an accurate probability estimate from these empirical observations. In the case of SDEs, the IS technique makes rare events happen more often by introducing a bias $\epsilon^{-1}u(X(t), t)dt$ to the mean of the noise increment dW, such that CV of the estimate P_{IS} is reduced at the same time [21], [22]. This provides new paths $\tilde{X}(t)$ that evolve according to

$$d\tilde{X} = \Big(b(\tilde{X}) + \sigma(\tilde{X})u(\tilde{X}, t)\Big)dt + \epsilon\sigma(\tilde{X})dW. \qquad (7)$$

By Girsanov's theorem [23], [24], for a time horizon $T > 0$, the likelihood ratio is given by

$$L = \exp\left(-\frac{1}{2\epsilon^2}\int_0^T |u(\tilde{X}, t)|^2 dt - \frac{1}{\epsilon}\int_0^T \langle u(\tilde{X}, t), dW(t)\rangle\right), \qquad (8)$$

where $\langle\cdot, \cdot\rangle$ stands for the Euclidean inner product in $\mathbb{R}^N$, $W(t)$ is the realization of the noise that produced $\tilde{X}(t)$, and the last integral is understood in the Itô sense. This expression may then be incorporated into the IS estimator (4) to recover an unbiased probability estimate. We emphasize that in the limit $M \to \infty$ one recovers from (4) the exact value of P for the original, unbiased process. The effect of good biasing is simply to concentrate the runs sampled in (4) on those with highest likelihood of activating the indicator function, thereby producing an accurate estimate of P with vastly fewer biased MC runs.

Over finite-time horizons, an effective bias $u = u^*$ can be obtained by minimizing the Freidlin–Wentzell large deviation action. Namely, for a given time horizon $T > 0$, current time $t < T$, current state x and the set of targeted outcomes $\mathcal{A}$ one looks for the minimizer $\phi_{t,x}^T(s)$ of the functional

$$S_T[\phi] = \int_t^T \frac{1}{2}|\sigma^{-1}(\phi(s))(\dot{\phi}(s) - b(\phi(s)))|^2 ds, \qquad (9)$$

where $\dot{\phi}(s) = d\phi(s)/ds$, over absolutely continuous paths $\phi : [t, T] \to \mathbb{R}^N$ satisfying $\phi(t) = x$ and $\phi(T) \in \mathcal{A}$ [25]. The finite-time bias function $u^* = u_T^*$ is then given by [21], [24], [26]

$$u_T^*(x, t) = \sigma^{-1}(x)\big(\dot{\phi}_{t,x}^T(t) - b(x)\big). \qquad (10)$$

Over infinite-time horizons, i.e., when $T \to \infty$, a convenient reparametrization allows the action in (9) to be minimized with respect to arclength rather than time. In this case, one can choose $u^* = u_\infty^*$, where $u_\infty^*(x)$ is obtained from the minimizer $\phi_x(\alpha)$ of the functional [25], [27]

$$S_\infty[\phi] = \int_0^1 (|\sigma^{-1}(\phi(\alpha))\phi'(\alpha)\|\sigma^{-1}(\phi(\alpha))b(\phi(\alpha))|$$
$$- \langle\sigma^{-1}(\phi(\alpha))\phi'(\alpha), \sigma^{-1}(\phi(\alpha))b(\phi(\alpha))\rangle)d\alpha \qquad (11)$$

among all absolutely continuous paths $\phi : [0, 1] \rightarrow \mathbb{R}^N$ satisfying $\phi(0) = x$ and $\phi(1) \in \mathcal{A}$. Here $\phi'(\alpha) = d\phi(\alpha)/d\alpha$. The infinite-time bias is then given by [21]

$$u_{\infty}^{*}(x) = \sigma^{-1}(x) \left(\frac{|\sigma^{-1}(x)b(x)|}{|\sigma^{-1}(x)\phi_x'(0)|} \phi_x'(0) - b(x) \right). \quad (12)$$

One of the strategies discussed in the following is the use of IS with infinite-time bias functions to obtain switching probabilities over finite-time horizons. This strategy is based on the observation that, when the characteristic speed obtained by dividing the domain radius by the time horizon is small relative to the maximum speed of the infinite-time minimizing path, the finite-time, and infinite-time minimizing paths are nearly identical outside of small neighborhoods around the dynamic fixed points. As will be seen in Section IV, this strategy is effective for intermediate times but does not correctly promote the long periods spent near the stable fixed point in the true dynamics. This manifests in reduced efficiency of the IS strategy in these cases. To address this phenomenon, we turn off the biasing for values of x within a diffusion length of the stable fixed point, which leads to a significant improvement of sampling efficiency.

III. MICROMAGNETIC FRAMEWORK AND THERMALLY INDUCED SWITCHING

We consider a region $\Omega \subset \mathbb{R}^3$ occupied by a ferromagnetic film with in-plane shape $D \subset \mathbb{R}^2$ and thickness d, i.e., $\Omega = D \times (0, d)$, characterized by saturation magnetization M_s, exchange stiffness A and an in-plane uniaxial anisotropy constant K_u, at temperature $k_B T$ in energy units. To model the free layer of an in-plane STT-MRAM cell, we use the stochastic Landau–Lifshitz–Gilbert equation [28], [29] for the unit magnetization vector $\mathbf{m} = (m_x, m_y, m_z)$, written in the Landau–Lifshitz form

$$\frac{\partial \mathbf{m}}{\partial t} = -\mathbf{m} \times \mathbf{h} - \alpha \mathbf{m} \times \mathbf{m} \times \mathbf{h} + \alpha \mathbf{m} \times \boldsymbol{\tau}_{\text{STT}} + \boldsymbol{\tau}_{\text{STT}}, \quad (13)$$

where time and lengths are measured in the units of $\tau_0 = (1+\alpha^2)/(\gamma \mu_0 M_s)$ and $l_{ex} = \sqrt{2 A/(\mu_0 M_s^2)}$, respectively, α is the Gilbert damping, γ is the gyromagnetic ratio, and $\boldsymbol{\tau}_{STT}$ is the spin-transfer torque. The effective field is given by

$$\mathbf{h} = -\frac{\delta E}{\delta \mathbf{m}} + \sqrt{\sigma} \, \boldsymbol{\eta}, \quad (14)$$

where

$$E[\mathbf{m}] = \frac{1}{2} \int_D \left(|\nabla \mathbf{m}|^2 + Q m_y^2 + m_z^2 \right) d^2 r \quad (15)$$

is the leading order thin-film micromagnetic energy measured in the units of $2Ad$, and $Q = 2K_u/(\mu_0 M_s^2)$ is the quality factor [30]. The thermal fluctuation term $\boldsymbol{\eta}(\mathbf{r}, t)$ is a delta-correlated, suitably regularized 3-D spatiotemporal Gaussian white noise [31], with noise strength

$$\sigma = \frac{\alpha k_B T}{Ad(1+\alpha^2)} \quad (16)$$

by the fluctuation–dissipation theorem [32], [33]. Equation (13) is to be interpreted in the Stratonovich sense, so that it preserves the norm constraint $|\mathbf{m}| = 1$ [28], [33].

The spin-transfer torque $\boldsymbol{\tau}_{\text{STT}}$ is given by

$$\boldsymbol{\tau}_{\text{STT}} = a_J \mathbf{m} \times \mathbf{m} \times \mathbf{m}_p + b_J \mathbf{m} \times \mathbf{m}_p, \quad (17)$$

where $a_J = -\eta j \hbar/(2de\mu_0 M_s^2)$ and $b_J = \beta a_J$ are dimensionless Slonczewski and fieldlike torque strengths [29]. Here, j is the density of electric current passing perpendicularly through the film, e is the elementary charge, $\eta \in (0, 1]$ is the spin polarization efficiency, β is the relative strength of the field-like spin torque, and $\mathbf{m}_p$ is the spin-polarization direction. In this paper, we consider $\mathbf{m}_p = (1, 0, 0)^T$, i.e., when the spin current is polarized along the easy axis in the film plane, as is the case in the basic in-plane spin valve [2], [34], [35].

The macrospin approximation assumes spatial uniformity across the ferromagnet, such that the first term in (15) is zero and (13) is an ordinary differential equation. In this case, $E(\mathbf{m}) = \frac{1}{2}(Q m_y^2 + m_z^2)S$, where S is the area of D in the units of l_{ex}^2, and the effective field is given by

$$\mathbf{h} = -S^{-1} \nabla_{\mathbf{m}} E + \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}} \, \dot{\mathbf{W}}(t), \quad (18)$$

where $\epsilon = k_B T/(2AdS)$, and $\mathbf{W}(t)$ is a 3-D Brownian motion. The noise coefficient $\sqrt{2\alpha\epsilon/(1+\alpha^2)}$ is consistent with the Gibbs distribution, in which ϵ plays the role of the dimensionless temperature.

When the dimensionless parameters satisfy

$$\alpha \sim 1 \text{ and } a_J \sim b_J \sim Q \ll 1, \quad (19)$$

i.e., in soft materials with relatively high damping and low spin torques, the magnetization is always constrained to lie almost entirely in the film plane [36]–[38]. In this case, the system (13) may be simplified to an equation for the angle θ such that $\mathbf{m} \simeq (\cos\theta, \sin\theta, 0)$ (see the Appendix)

$$\dot{\theta} = b(\theta) + \frac{1}{\sqrt{\Delta}} \dot{W}, \quad (20)$$

where

$$b(\theta) = (I_J - \cos(\theta)) \sin(\theta), \quad (21)$$

$I_J = b_J/Q$, $1/\Delta = 2\epsilon/Q$, and the unit of time is now $\tau_Q = \alpha/(\gamma \mu_0 M_s Q)$. We point out that even though (20) was obtained for an in-plane STT-MRAM cell, exactly the same equation arises in the modeling of perpendicular cells [6]. Therefore, a direct comparison with the results obtained by Butler *et al.* [6], who used the analysis of the FPE is also possible.

Equation (20) with $0 \leq I_J < 1$ will be used as the simplest example of a stochastic micromagnetic model with bistability, for which several IS biasing strategies will be illustrated. A more realistic model of an STT-MRAM cell would need to incorporate spatial heterogeneity within the cell, which may be captured by considering a system of N exchange-coupled macrospins associated with the magnetization in each of the polycrystalline grains. If each grain has a dimensionless area S_i, $1/\Delta_i = k_B T/(Ad S_i Q)$, and θ_i are such that the magnetization in each grain is $\mathbf{m}_i \simeq (\cos\theta_i, \sin\theta_i, 0)$, then (20) may be generalized to [39]–[41]

$$\dot{\theta}_i = b(\theta_i) + \sum_{j=1}^N a_{ij} S_i^{-1} \sin(\theta_j - \theta_i) + \frac{1}{\sqrt{\Delta_i}} \dot{W}_i, \quad (22)$$

using Heisenberg exchange with dimensionless strengths $a_{ij} = a_{ji} \geq 0$ for the interactions between the grains, with W_i being N uncorrelated Brownian motions. Note that this equation is a stochastic version of a gradient system governed by an *effective potential*

$$V_N(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} S_i \left(I_J \cos\theta_i + \frac{1}{2} \sin^2\theta_i \right)$$
$$- \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} a_{ij} \cos(\theta_i - \theta_j) \quad (23)$$

and obeying detailed balance. The coupling coefficients a_{ij} are non-zero only for the nearest neighbors and may, in principle, be determined from the geometric characteristics of the individual grains. For simplicity, in this paper, we will limit ourselves to the consideration of the case of two identical macrospins only, which may correspond, e.g., to exchange-coupled synthetic bilayers [42].

The single macrospin drift term defined in (21) has stable fixed points $\theta = 0$ and $\theta = \pm\pi$, separated by unstable fixed points $\theta = \pm\theta_J$, where $\theta_J = \arccos(I_J) \leq \pi/2$. Therefore, for a trajectory starting close to $\theta = 0$ at $t = 0$, a switching event for the time horizon $T > 0$ would be defined as one in which $|\theta|$ is close to π at $t = T$. For the purposes of this paper, we consider a switching event to have occurred if $|\theta(T_{sw})| = \pi/2$ for some $0 < T_{sw} \leq T$, starting with $\theta(0) = 0$, i.e., the macrospin changes its direction along the easy axis. Similarly, for the system of two-coupled macrospins governed by (22) and starting with $\theta_1(0) = \theta_2(0) = 0$, we define a switching event to have occurred if $\max(|\theta_1(T_{sw})|, |\theta_2(T_{sw})|) = \pi/2$ for some $0 < T_{sw} \leq T$, i.e., at least one macrospin changes its direction along the easy axis.

The exact finite-time switching probability for a single macrospin exhibited by (20) can be computed by solving the backward FPE for the probability $P_{sw}(\theta, t)$ that a trajectory starting at a given value of $\theta \in (-\pi/2, \pi/2)$ reached the value of $\theta = \pm\pi/2$ by time t [43]. This probability satisfies

$$\frac{\partial P_{sw}}{\partial t} = b(\theta) \frac{\partial P_{sw}}{\partial \theta} + \frac{1}{2\Delta} \frac{\partial^2 P_{sw}}{\partial \theta^2} \quad (24)$$

for $(\theta, t) \in (-\pi/2, \pi/2) \times (0, T)$, with initial and boundary conditions, respectively,

$$P_{sw}(\theta, 0) = 0, \quad P_{sw}(\pm\pi/2, t) = 1. \quad (25)$$

In particular, the probability of having switched by time T, given the initial state $\theta(0) = 0$ is equal to $P_{sw}(0, T)$.

As an alternative to solving (24), we incorporate IS in the estimation of P_{sw} by sampling controlled dynamics

$$\dot{\tilde{\theta}} = b(\tilde{\theta}) + u^* + \frac{1}{\sqrt{\Delta}} \dot{W}, \quad \tilde{\theta}(0) = 0 \quad (26)$$

for $t \in (0, T)$. For finite-time bias, we have $u^* = u_T^*$, where $u_T^*(\tilde{\theta}, t) = \dot{\theta}_{\tilde{\theta},t}^T(t) - b(\tilde{\theta})$ is the bias function obtained through (10) by interpreting (26) as an instance of (7) with $\epsilon = 1/\sqrt{\Delta}$, using the minimizer $\theta_{\tilde{\theta},t}^T(s)$ of the finite-time action

$$S_T[\theta] = \int_t^T \frac{1}{2} |\dot{\theta}(s) - b(\theta(s))|^2 ds \quad (27)$$

among all $\theta(s)$ with $\theta(t) = \tilde{\theta}$ and $|\theta(T)| = \pi/2$. For sufficiently large time horizons, we set instead $u^* = u_\infty^*$, where the ininfite-time bias u_∞^* is obtained from the minimizer of the infinite-time action S_∞. In the single macrospin case, the latter is simply given by a straight line segment, resulting in a particularly simple explicit form of the bias

$$u_\infty^*(\tilde{\theta}) = \begin{cases} -2b(\tilde{\theta}), & -\theta_J \leqslant \tilde{\theta} \leqslant \theta_J, \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

To account for switching events, we stop the trajectory at time $t = T_{sw}$ as soon as the switching criterion $|\tilde{\theta}(t)| = \pi/2$ is satisfied, or otherwise set $T_{sw} = T$. The likelihood ratio is recovered from (8), which in this case is

$$L = \exp\left(-\frac{\Delta}{2} \int_0^{T_{sw}} |u^*(t)|^2 dt - \sqrt{\Delta} \int_0^{T_{sw}} u^*(t) \, dW(t) \right), \quad (29)$$

where either $u^*(t) = u_T^*(\tilde{\theta}(t), t)$ or $u^*(t) = u_\infty^*(\tilde{\theta}(t))$, depending on whether we use the finite- or infinite-time bias function, respectively, and $\tilde{\theta}(t)$ is the solution of (26) with a particular realization $W(t)$ of the noise.

IV. SIMULATIONS

This section describes the results obtained from IS simulations of macrospin and coupled-spin systems using the physical parameters drawn from [6] for the purpose of comparison. All simulations use the Euler–Maruyama method with a fixed time step $\tau = 0.1$.

The IS results presented in the following are obtained using bias functions based on either finite- or infinite-time minimizers of the action given by (9). Infinite-time bias functions are based on (28) in the macrospin case and, therefore, do not require additional computation. Infinite-time bias functions for the coupled-spin system are obtained by minimizing the action in (9) through the geometric minimum action method (GMAM) with 50 gridpoints [27]. Finite-time bias functions are obtained by minimizing the action in (9) through a combination of Newton's method for the associated Euler–Lagrange equation and the improved adaptive minimum action method [44], with 500 gridpoints in the macrospin case and 100 in the coupled-spin case.

A. Single Macrospin

The discretized version of (26) reads explicitly

$$\tilde{\theta}^{k+1} = \tilde{\theta}^k + (b(\tilde{\theta}^k) + u^*(\tilde{\theta}^k, t_k))\tau + \frac{\sqrt{\tau}}{\sqrt{\Delta}} \xi_k, \quad \tilde{\theta}^0 = 0, \quad (30)$$

where $\tilde{\theta}^k = \tilde{\theta}(t_k)$, $t_k = k\tau$ for $k = 0, 1, \ldots, K$, and ξ_k are independent and drawn from the standard normal distribution. The value of K is chosen so that either $|\tilde{\theta}_k| < \pi/2$ for all $k < K < \lfloor T/\tau \rfloor$ and $|\tilde{\theta}^K| \geq \pi/2$, or $K = \lfloor T/\tau \rfloor$, i.e., we stop the simulation if a switching event occurs at time $t_k < T$.
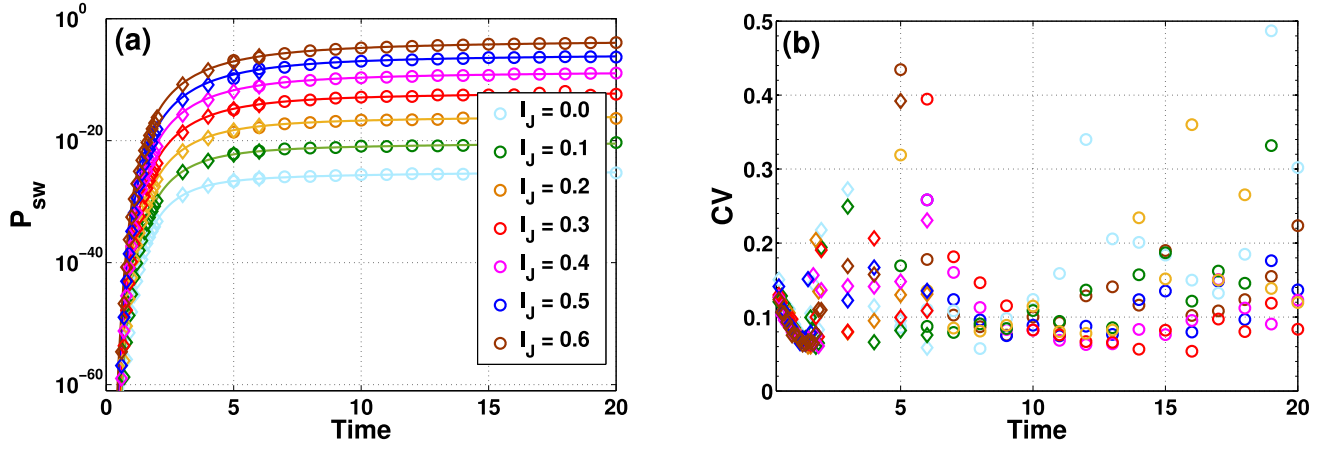
Fig. 1. IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$. (a) RSER versus reading pulse duration T and reading current I_J with thermal stability factor $\Delta = 60$. Solid lines denote numerical solutions of FPE (24). Open circles and open diamonds denote estimates generated by IS with finite-time bias functions and infinite-time bias functions, respectively, color coded by current amplitude. For $T \geq 5$, infinite-time bias functions are used; whereas for IS and for $T \leq 6$, finite-time bias functions are used. IS results at $T = 5$ and $T = 6$ obtained using the infinite- and finite-time bias functions are indistinguishable. (b) CV of the RSER.
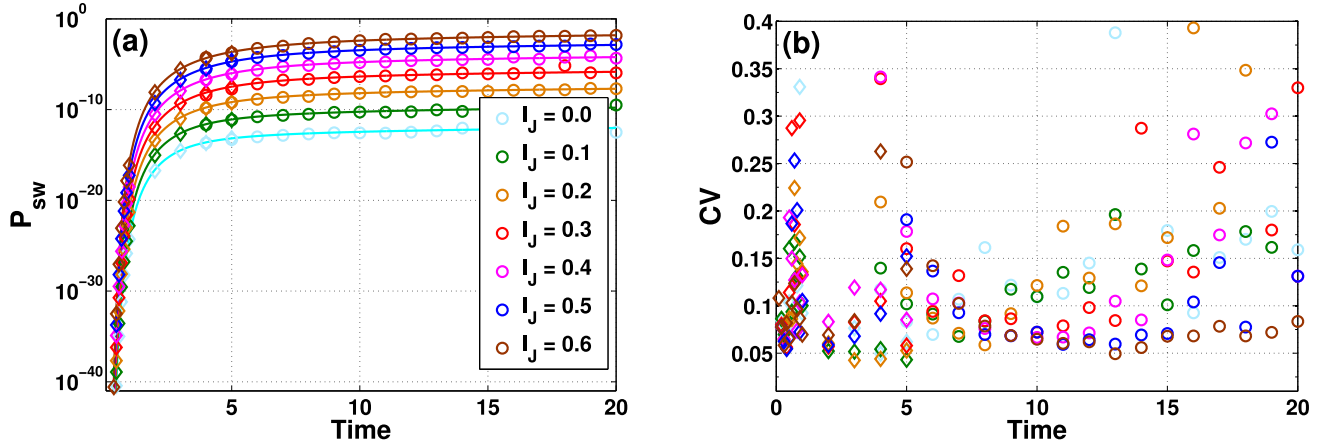

Fig. 2. Similar to Fig. 1, but with $\Delta = 30$. For $T \geq 4$, infinite-time bias functions are used, and for $T \leq 5$, finite-time bias functions are used. IS estimates at $T = 4$ and $T = 5$ obtained using infinite- and finite-time bias functions are indistinguishable. (a) RSER versus reading pulse duration T. (b) CV of the RSER.

The likelihood ratio corresponding to (29) is then given explicitly by

$$L = \exp\left(-\frac{\tau\,\Delta}{2}\sum_{k=0}^{K}|u^*(\tilde{\theta}^k, t_k)|^2 - \sqrt{\tau\,\Delta}\sum_{k=0}^{K}u^*(\tilde{\theta}^k, t_k)\,\xi_k\right). \quad (31)$$

The following figures are generated using switching probabilities and their coefficients of variation computed by applying formulas (4) and (5), respectively, to ensembles of M runs, where $I(\omega_i) = 1$ for runs that generate a switching event and $I(\omega_i) = 0$ otherwise.

Figs. 1 and 2 show the RSER as a function of time for seven values of I_J between 0 and 0.6 with thermal stability factors $\Delta = 60$ and $\Delta = 30$, respectively. Runs at both temperatures are included here to facilitate comparison with the results presented in [6]. Both show a comparison between numerical solutions of backward FPE (24) and IS simulations for the macrospin model. For $T \geq 5$ in Fig. 1 and $T \geq 4$ in Fig. 2, infinite-time bias functions are used for IS, while

for $T \leq 6$ in Fig. 1 and $T \leq 5$ in Fig. 2, finite-time bias functions are used. On the logarithmic scale, the agreement between the FPE and IS results is excellent throughout the range of times and currents, as can be seen from Figs. 1(a) and 2(a). In relative terms, for $M = 10^3$, the values of P_{sw} obtained using IS typically agree with the FPE results to within $10\% - 20\%$, while capturing correctly the magnitudes of P_{sw} that vary by many orders. Note that in this case, the IS sampling error dominates the discretization error of the Euler–Maruyama scheme, with the values of CV giving a good idea of the relative error for P_{sw}. We verified that increasing the sample size M decreases both the value of the CV and the relative error. For example, the relative error goes down to a few percent for $M = 10^5$ for most of the data points in Figs. 1 and 2. However, this unnecessary improvement in accuracy comes at the expense of a hundred-fold increase in the runtime. Finally, one can see that the IS results are internally consistent between the finite- and infinite-time bias functions used at $T = 5$ and $T = 6$ in Fig. 1, and used at $T = 4$ in Fig. 2.

Fig. 1(b) shows the CV for the IS estimates in Figs. 1(a) and Fig. 2(b) shows those for the estimates shown in Fig. 2(a). The CV values for the IS estimates range from approximately 0.05–0.5. We note that with an inappropriate choice of bias, the CV can be an imperfect measure of accuracy for MC estimates using variance reduction [45]. However, this is precluded by our choice of an asymptotically optimal bias function that is based on the large deviation theory [24], [46], as can also be seen from the excellent agreement with the solutions to FPE (24). The low CV values obtained for such extremely small probabilities with moderate sample size of $M = 10^3$ are, therefore, a clear demonstration of the efficiency of the bias functions used here.

In both figures, the CVs are observed to increase when T is decreased from $T = 10$ to $T = 5$, indicating that the infinite-time bias function used for these runs becomes progressively less efficient at capturing the switching events as the time horizon shrinks. The application of finite-time bias functions for smaller times lowers the CVs as expected. Furthermore, Figs. 1(b) and 2(b) show a similar pattern in the CVs generated by IS with infinite-time bias functions, where the CVs also increase for *large* times T, leading to a deterioration of the sampling accuracy. Fig. 3 illustrates this sudden increase in CV in the context of longer times, as well as a histogram of switching times, and the time evolution of the spread in values of likelihood ratio.

This decrease in efficiency of the infinite-time bias function for large but finite-time horizons is due to the fact that exits have a natural finite-time scale dictated by diffusion near the fixed points with the action minimizer bridging the gap between them. When the time horizon of the simulation is large relative to this time scale, it allows for exit events that hover near the stable fixed point before exiting just prior to the horizon time. These events occur with considerably higher likelihood under the *unbiased* dynamics than under the biased dynamics, leading to a very large likelihood ratio that causes them to dominate the CV computation. Since the diffusion time grows as the noise strength decreases, this phenomenon can be regarded as a finite-noise effect, and it, indeed, vanishes as $\Delta \to \infty$. We address this issue for finite noise by turning off the biasing near the stable fixed point, i.e., for $|\theta| < \theta_0$, with the results plotted for different values of θ_0 in Fig. 4. It is clearly seen that as θ_0 increases, the anomalous behavior for large horizon time is mitigated, at the expense of sampling efficiency for small horizon times.

B. Two-Coupled Identical Macrospins: Model

To demonstrate that the IS method can also be effective in coupled systems, we simulate two spins with identical volume and dynamics given by (22), which, in this case, is explicitly

$$\dot{\theta}_1 = c \sin(\theta_2 - \theta_1) + b(\theta_1) + \sqrt{\frac{2}{\Delta}} \dot{W}_1, \qquad (32)$$

$$\dot{\theta}_2 = c \sin(\theta_1 - \theta_2) + b(\theta_2) + \sqrt{\frac{2}{\Delta}} \dot{W}_2, \qquad (33)$$

where $c > 0$ is the ferromagnetic exchange coupling strength favoring parallel alignment of the two spins. The initial
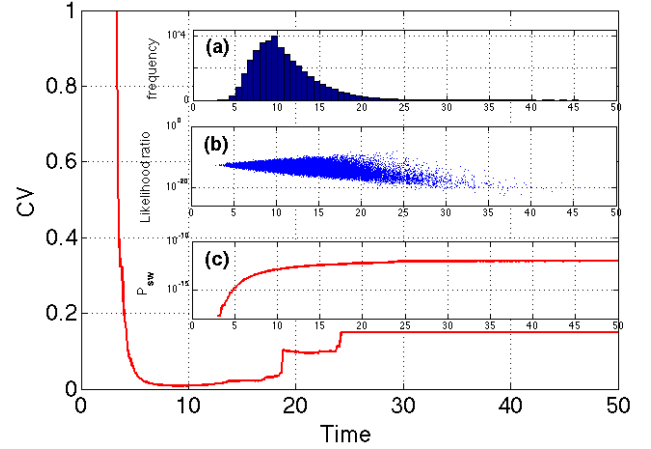

Fig. 3. CVs of IS estimators for $I_J = 0.3$ with infinite-time bias functions and sample size 10^5. Inset (a) is a histogram of exit times for the biased system. Inset (b) is the estimated likelihood ratio vs. time. Inset (c) is the probability of switching before time T. The sample size here is $M = 10^5$.
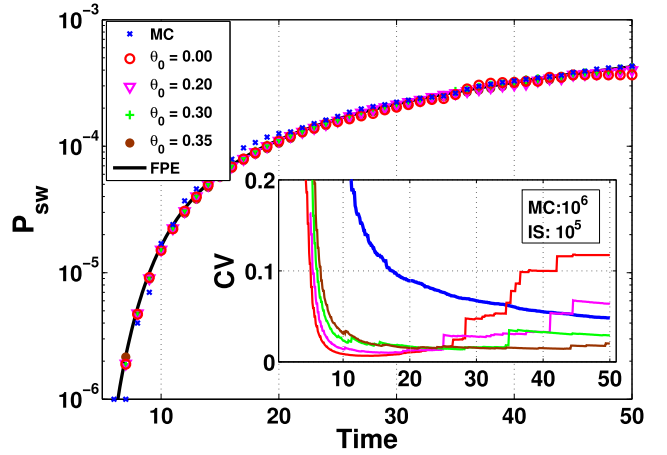

Fig. 4. Switching probabilities and CVs of IS estimators for $I_J = 0.6$ with infinite-time bias functions active outside of a region containing the stable fixed point, defined by $\theta_0 < |\theta| < \theta_J$, for different values of θ_0. The sample size is 10^5.

conditions are $\theta_1(0) = \theta_2(0) = 0$. Recall that as a switching criterion, we adopt that at least one of the angles reaches $\pi/2$ in absolute value.

At finite temperature, the rare events of switching for the coupled spin system occur along the maximum likelihood paths, which are also the minimum energy paths of the system. When the system undergoes a transition, it switches by coherent rotation for strongly coupled spins, asymmetric coherent rotation for weakly coupled spins, or single particle reversal for extremely weakly coupled spin [47]. More precisely, in the limit of infinite coupling strength c, the coupled spin system collapses to the macrospin model with $\theta_1 = \theta_2$, and the most probable path terminating at $\max(|\theta_1|, |\theta_2|) = \pi/2$ is identical to that of a single macrospin. As the coupling strength decreases, the dynamics of the coupled spin system changes significantly, ultimately leading to spins that evolve independently. Note that in N-dimensional coupled spin systems, the sequence of bifurcations from single macrospin dynamics to N-fold macrospin dynamics as c decreases from

infinity [48] further exacerbates the challenge in finding appropriate bias functions.

C. Two-Coupled Identical Macrospins: Biased Dynamics

The biased dynamics associated with (32) and (33) reads

$$\dot{\tilde{\theta}}_1 = c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) + b(\tilde{\theta}_1) + \sqrt{2}\,u_1^* + \sqrt{\frac{2}{\Delta}}\,\dot{W}_1, \quad (34)$$

$$\dot{\tilde{\theta}}_2 = c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) + b(\tilde{\theta}_2) + \sqrt{2}\,u_2^* + \sqrt{\frac{2}{\Delta}}\,\dot{W}_2. \quad (35)$$

In contrast to the single macrospin case, for two coupled macrospins, an exact analytical bias function is no longer available even for infinite-time biasing. Therefore, it is necessary to obtain numerical bias functions by minimizing the Freidlin–Wentzell action (9) with terminal condition $\max(|\theta_1(T_{\mathrm{sw}})|, |\theta_2(T_{\mathrm{sw}})|) = \pi/2$ for some $0 < T_{\mathrm{sw}} \leq T \leq \infty$.

For finite-time bias, the action functional S_T is given explicitly by

$$S_T[\theta_1, \theta_2] = \frac{1}{4}\int_t^T (\dot{\theta}_1 - c\sin(\theta_2 - \theta_1) - b(\theta_1))^2 ds$$

$$+ \frac{1}{4}\int_t^T (\dot{\theta}_2 - c\sin(\theta_1 - \theta_2) - b(\theta_2))^2 ds, \quad (36)$$

and the corresponding finite-time bias $(u_1^*, u_2^*) = (u_{T,1}^*, u_{T,2}^*)$ is

$$u_{T,1}^*(\tilde{\theta}_1, \tilde{\theta}_2, t) = \frac{1}{\sqrt{2}}\big(\dot{\theta}_{\tilde{\theta}_1,\tilde{\theta}_2,t,1}^T(t) - c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) - b(\tilde{\theta}_1)\big), \quad (37)$$

$$u_{T,2}^*(\tilde{\theta}_1, \tilde{\theta}_2, t) = \frac{1}{\sqrt{2}}\big(\dot{\theta}_{\tilde{\theta}_1,\tilde{\theta}_2,t,2}^T(t) - c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) - b(\tilde{\theta}_2)\big), \quad (38)$$

where $\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t}^T(s) = (\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,1}^T(s), \theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,2}^T(s))$ is the minimizer of S_T satisfying $\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t}^T(t) = (\tilde{\theta}_1, \tilde{\theta}_2)$ and $\max(|\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,1}^T(T)|, |\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,2}^T(T)|) = \pi/2$.

For infinite-time bias, we minimize

$$S_\infty[\theta_1, \theta_2] = \frac{1}{2}\int_0^1 \Big(\lambda(\theta_1, \theta_2)\sqrt{|\theta_1'|^2 + |\theta_2'|^2}$$

$$- (c\sin(\theta_2 - \theta_1) + b(\theta_1))\theta_1'$$

$$- (c\sin(\theta_1 - \theta_2) + b(\theta_2))\theta_2'\Big)\,ds, \quad (39)$$

where

$$\lambda(\theta_1, \theta_2) = [(c\sin(\theta_2 - \theta_1) + b(\theta_1))^2$$

$$+ (c\sin(\theta_1 - \theta_2) + b(\theta_2))^2]^{1/2}, \quad (40)$$

and express the bias as

$$u_{\infty,1}^*(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{\sqrt{2}} \times \left(\frac{\lambda(\tilde{\theta}_1, \tilde{\theta}_2)\theta_{\tilde{\theta}_1,\tilde{\theta}_2,1}'(0)}{\sqrt{|\theta_{\tilde{\theta}_1,\tilde{\theta}_2,1}'|^2 + |\theta_{\tilde{\theta}_1,\tilde{\theta}_2,2}'|^2}} \right.$$

$$\left. - c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) - b(\tilde{\theta}_1) \right), \quad (41)$$

$$u_{\infty,2}^*(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{\sqrt{2}} \times \left(\frac{\lambda(\tilde{\theta}_1, \tilde{\theta}_2)\theta_{\tilde{\theta}_1,\tilde{\theta}_2,2}'(0)}{\sqrt{|\theta_{\tilde{\theta}_1,\tilde{\theta}_2,1}'|^2 + |\theta_{\tilde{\theta}_1,\tilde{\theta}_2,2}'|^2}} \right.$$

$$\left. - c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) - b(\tilde{\theta}_2) \right), \quad (42)$$

where $\theta_{\tilde{\theta}_1,\tilde{\theta}_2} = (\theta_{\tilde{\theta}_1,\tilde{\theta}_2,1}, \theta_{\tilde{\theta}_1,\tilde{\theta}_2,2})$ is the minimizer of S_∞ with $\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t}(0) = (\tilde{\theta}_1, \tilde{\theta}_2)$ and $\max(|\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,1}(1)|, |\theta_{\tilde{\theta}_1,\tilde{\theta}_2,t,2}(1)|) = \pi/2$. Finally, the discretized version of the biased equations and the likelihood ratio are straightforward generalizations of (30) and (31).

Fig. 5(a) shows the IS estimates of switching probabilities obtained using finite-time bias functions and infinite-time bias functions with a suitable cutoff near the origin applied to the coupled system with non-dimensional temperature $\Delta = 60$ and coupling strength $c = 0.8$. The coupling strength $c = 0.8$ is an example of strong coupling, for which the spins rotate coherently along optimal switching paths. The open circles and open diamonds denote the estimates generated by IS using infinite-time and finite-time bias functions, respectively. With the sampling size of $M = 10^3$, the infinite-time bias functions allow us to sample switching probabilities for $T \geq 5$ with I_J ranging from 0 to 0.6. For $T \leq 4$, finite-time bias functions are used with IS. For infinite-time biasing, the bias is switched off, i.e., u_∞^* is set to zero, when the effective potential from (23) of a configuration $(\tilde{\theta}_1, \tilde{\theta}_2)$ falls below that of (θ_0, θ_0) with $\theta_0 \in [0, 0.3]$ chosen to minimize the CV. The accuracy of the obtained values of P_{sw} is expected to be similar to that obtained for the macrospin model in Section IV-A.

Fig. 5(b) shows the CVs becoming larger as time T decreases, indicating that the infinite-time bias functions become less efficient. The effectiveness of using finite-time bias functions is evident in the CV values in Fig. 5(b), where the CV values with open diamonds at time $T = 4$ are much smaller than the CV values with open circles at times $T = 4$ and $T = 5$. This improved effectiveness comes with a cost of computing updated finite-time bias functions at each time step.

Similar results for a weakly coupled system are shown in Fig. 6. Fig. 6(a) shows the switching probabilities as a function of time for different read currents with non-dimensional temperature $\Delta = 60$ and coupling strength $c = 0.2$. The coupling strength $c = 0.2$ is an example of weak coupling, for which the spins rotate incoherently along optimal switching paths. With a sampling size of $M = 10^3$, IS using infinite-time
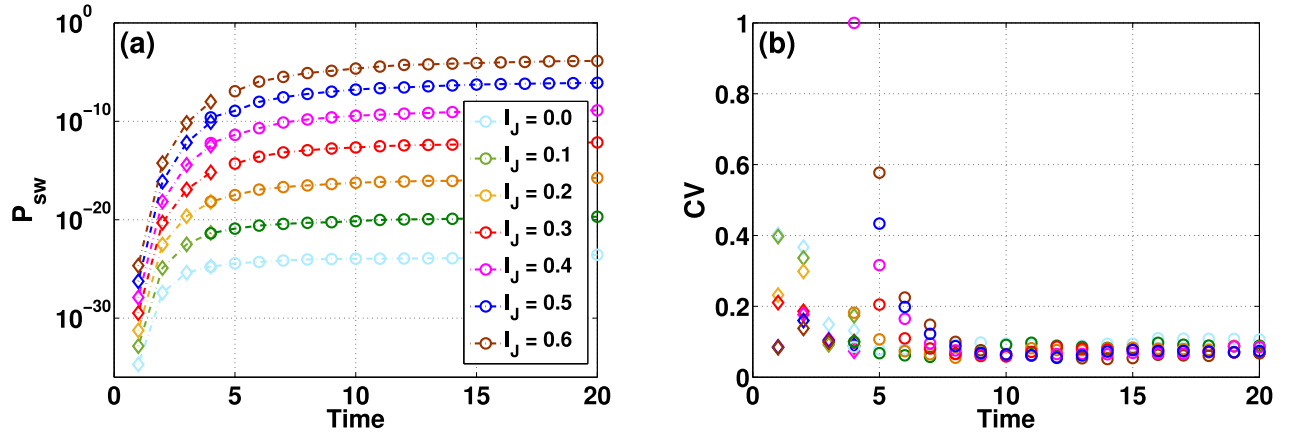
Fig. 5.　IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$ for a strongly coupled two-spin system. (a) RSER with applied currents ranging from 0 to 0.6 versus time T with thermal stability factor $\Delta = 60$ and coupling strength $c = 0.8$. (b) CVs of IS estimates in (a). Open circles and open diamonds denote the estimates generated by IS using infinite-time and finite-time bias functions, respectively. The colors correspond to different current amplitudes indicated in (a).
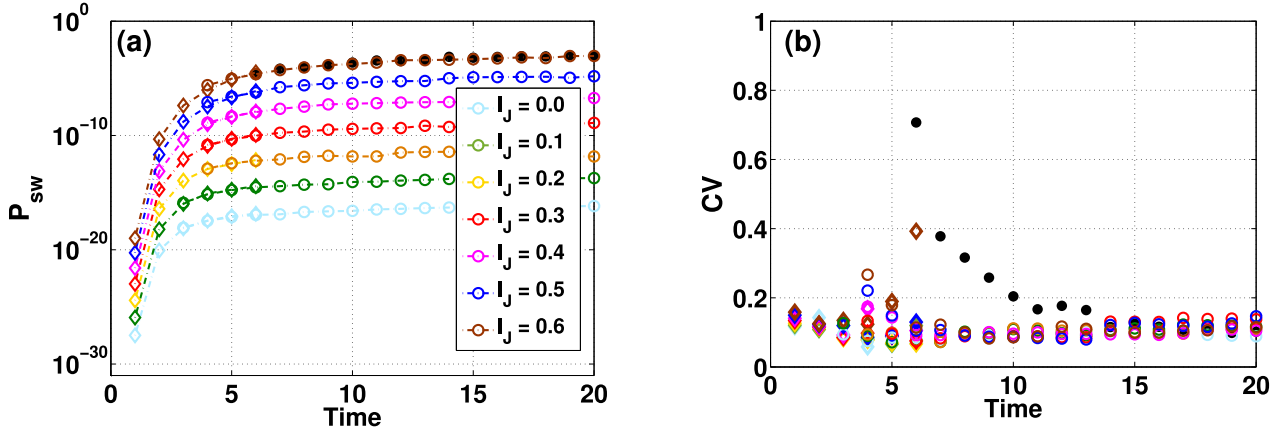

Fig. 6.　IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$ for a weakly coupled two-spin system. (a) RSER with applied currents ranging from 0 to 0.6 versus time T with thermal stability factor $\Delta = 60$ and coupling strength $c = 0.2$. (b) CVs of IS estimates in (a). Open circles and open diamonds denote the estimates generated by IS using infinite-time and finite-time bias functions, respectively, and black dots denote the estimates generated by naive MC simulations for $I_J = 0.6$ with sample size $M = 10^5$. The colors correspond to different current amplitudes indicated in (a).

bias functions allows us to sample switching probabilities for $T \geq 4$ with I_J ranging from 0 to 0.6. As shown in Figs. 5, 6(b) shows a decrease in efficiency of IS using infinite-time bias functions as T decreases.

Fig. 6(a) also shows the switching probabilities generated using naive MC simulations with a sample size of $M = 10^5$ for read current $I_J = 0.6$ at various pulse durations. Naive MC simulations at this sample size fail to accurately capture probabilities less than 10^{-5} while IS is able to estimate probabilities as low as 10^{-28}. This is reflected in Fig. 6(b) where the CV is seen to diverge as the probability estimate decreases. Finally, a few representative switching trajectories corresponding to the results in Figs. 5 and 6 are shown in Fig. 7. Fig. 8 shows a comparison between the CV for IS with sample size of 10^3 and the CV for naive MC with a sample size of 10^5 for $I_J = 0.6$. For time $6 \leq T \leq 15$, the CVs for naive MC exceed the CVs for IS, by a factor ranging from 1 to 6 as the value of T is decreased. This means that the number of IS samples required to achieve an estimate with the same or better accuracy than an estimate generated by naive MC

is smaller by several orders of magnitude. For example, for switching probability by time $T = 6$, where the CV ratio is about 6.2, at least 3.8×10^6 naive MC samples are required to achieve the same accuracy of an IS estimate generated using only $M = 10^3$. This contrast is even more stark for smaller applied currents with much lower associated switching probabilities, where the computational effort required by naive MC simulations is prohibitive. Sampling-based probability estimates are only available using IS at these parameter values.

D. Failure to Switch (WSER)

In this section, we estimate the WSER using MC and IS for both the macrospin model and two coupled-spin system, where the current is set sufficiently high as to drive the origin unstable, such that a switching event is expected to occur in the majority of random realizations. The two-spin state is considered to have switched when $\max(|\theta_1(T_{sw})|, |\theta_2(T_{sw})|) = \pi/2$ for some $0 < T_{sw} \leq T$.
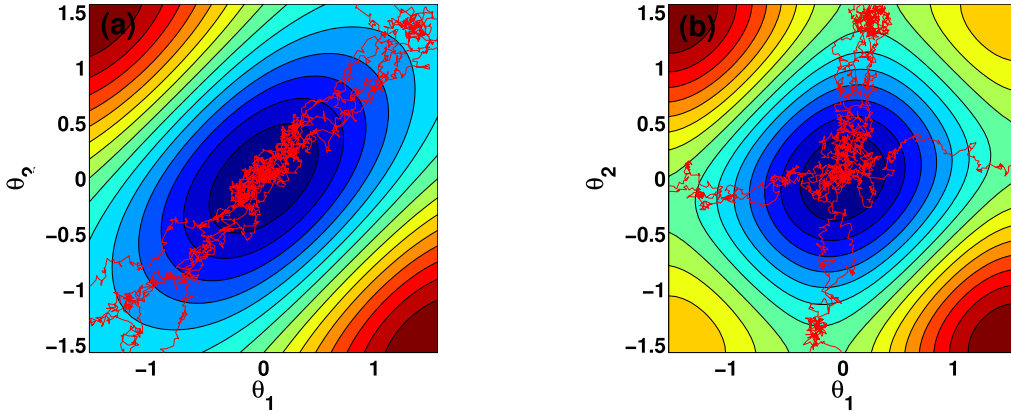
Fig. 7. Sample paths from IS for two-spin systems with current $I_J = 0.1$ and thermal stability factor $\Delta = 60$: (a) strongly coupled spins with $c = 0.8$ and (b) weakly coupled spins with $c = 0.2$. False color corresponds to the effective potential in (23).
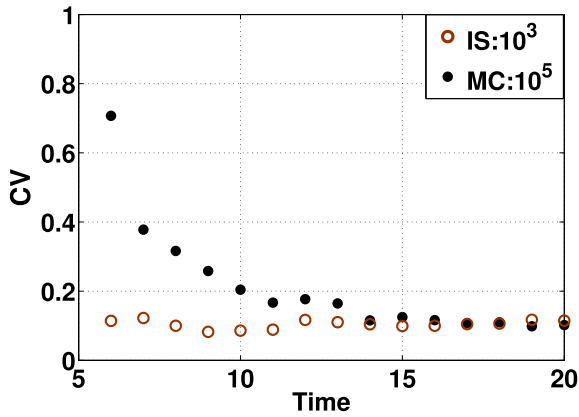

Fig. 8. CVs of estimates generated by IS (brown open circles) and naive MC (black dots) in Fig. 6.
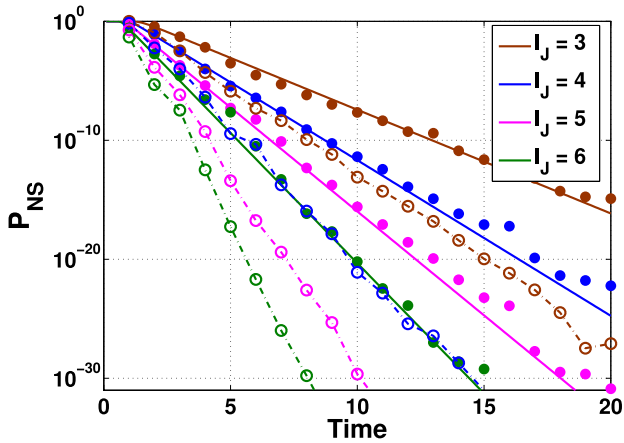

Fig. 9. Non-switching probability as a function of time horizon for several values of the writing current, using thermal stability factor $\Delta = 60$. The current I_J is identified by color and ranges from 3 to 6. Solid lines denote numerical solutions of the backward FPE for single macrospin, while the dashed lines are simply visual guides. Filled circles denote IS estimates of the non-switching probabilities for a single macrospin. Open circles denote IS estimates of non-switching probabilities for two coupled spins with $c = 0.2$.

Fig. 9 shows a comparison between the numerical solution of the FPE and simulation results using IS for a single macrospin. Independent realizations of the macrospin evolution were computed with the same initial conditions, currents $I_J = 3, 4, 5$, and 6, and independent thermal noise with $\Delta = 60$. The estimates obtained using $M = 10^4$ IS samples show good agreement with the numerical solution of the backward FPE for moderate and long times. For the two-spin system, 10^3 IS samples are used to estimate non-switching probabilities.

V. CONCLUSION

In conclusion, we have applied IS to estimate error rates in reading and writing spin-torque memory devices in the macrospin limit and for the case of two coupled spins, using a variety of applied currents and thermal stability factors. We have demonstrated how IS is able to compute probabilities well below those computable using naive MC simulations, while producing accurate estimates with improved efficiency in cases where the probability is computable using naive MC but still very small. Depending on the time horizon of the read or write event relative to the drift dynamics of the spin system, IS simulations using infinite- or finite-time bias functions are appropriate. Infinite-time bias functions require only moderate computational cost; however, the increased cost can be significant in computing finite-time bias functions. Further improvement of the infinite-time biasing can be achieved by introducing a threshold that turns on the bias only some distance away from the metastable equilibrium whose thermal stability is investigated.

While, in this paper, we intentionally restricted our attention to the particularly simple modeling setups involving either a single macrospin or two ferromagnetically coupled identical macrospins to make a clear illustration of the IS technique, we believe that IS can be used equally well for systems of many interacting macrospins of spatially nonuniform switching described by the stochastic Landau–Lifshitz–Gilbert equation. As we already noted, spatial nonuniformity starts to play an important role for noise-assisted magnetization reversal or failure for ferromagnetic elements exceeding a critical size of a few tens of nanometers. Here, the only limitation of the IS technique is the amount of computational resources that may be needed to compute the action minimizers for a good bias and to sample the biased trajectories sufficiently well.

Appendix

In this appendix, we derive a reduced model for the macrospin dynamics of the in-plane magnetization component in the regime of (19), which is done in the spirit of [36]–[38]. Our starting point is (13) for a macrospin **m**, in which **h** satisfies (18). Introducing cylindrical coordinates, we can write

$$\mathbf{m} = \left(\sqrt{1 - z^2}\,\cos\theta,\, \sqrt{1 - z^2}\,\sin\theta,\, z\right). \tag{43}$$

Then, changing the unit of time to τ_Q we arrive, after some algebra and a change from Stratonovich to Itô formulation, to the following Itô SDEs for $\theta(t)$ and $z(t)$:

$$\frac{Q(1+\alpha^2)}{\alpha}\dot{\theta} = -\frac{\partial E}{\partial z} - \frac{\alpha}{1-z^2}\frac{\partial E}{\partial \theta} + (a_J + \alpha b_J)\frac{\sin\theta}{\sqrt{1-z^2}}$$
$$+ (b_J - \alpha a_J)\frac{z\cos\theta}{\sqrt{1-z^2}} + Q\sqrt{\frac{1+\alpha^2}{\Delta(1-z^2)}}\,\dot{W}_1, \tag{44}$$

$$\frac{Q(1+\alpha^2)}{\alpha}\dot{z} = \frac{\partial E}{\partial \theta} - \alpha(1-z^2)\frac{\partial E}{\partial z} - \frac{Q\alpha z}{\Delta}$$
$$+ (a_J + \alpha b_J)z\sqrt{1-z^2}\,\cos\theta$$
$$- (b_J - \alpha a_J)\sqrt{1-z^2}\,\sin\theta$$
$$+ Q\sqrt{\Delta^{-1}(1+\alpha^2)(1-z^2)}\,\dot{W}_2, \tag{45}$$

where W_1 and W_2 are two uncorrelated Brownian motions and

$$\frac{\partial E}{\partial z} = z - Qz\sin^2\theta, \quad \frac{\partial E}{\partial \theta} = Q(1-z^2)\sin\theta\cos\theta. \tag{46}$$

We now assume that $Q \ll 1$, while $a_J = \beta^{-1}I_J Q$, $b_J = I_J Q$ and $\epsilon = Q/(2\Delta)$, with I_J, β and Δ of order unity. In this case, any deviations of z from zero are strongly suppressed. Therefore, linearizing the above-mentioned equations in z and introducing $\zeta = Q^{-1}z$, we obtain

$$\frac{Q(1+\alpha^2)}{\alpha}\dot{\theta} = -\frac{\partial E}{\partial z} - \alpha\frac{\partial E}{\partial \theta} + QI_J(\beta^{-1} + \alpha)\sin\theta$$
$$+ Q^2 I_J(1 - \alpha\beta^{-1})\zeta\cos\theta$$
$$+ Q\sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_1 \tag{47}$$

$$\frac{Q^2(1+\alpha^2)}{\alpha}\dot{\zeta} = \frac{\partial E}{\partial \theta} - \alpha\frac{\partial E}{\partial z} - \frac{Q^2\alpha\zeta}{\Delta}$$
$$+ Q^2 I_J(\beta^{-1} + \alpha)\zeta\cos\theta$$
$$- QI_J(1 - \alpha\beta^{-1})\sin\theta$$
$$+ Q\sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_2 \tag{48}$$

where

$$\frac{\partial E}{\partial z} = Q\zeta, \quad \frac{\partial E}{\partial \theta} = Q\sin\theta\cos\theta. \tag{49}$$

Substituting (49) into (47) and (48), to the leading order in $Q \ll 1$, we then arrive at

$$\frac{(1+\alpha^2)}{\alpha}\dot{\theta} = -\zeta - \alpha\sin\theta\cos\theta + I_J(\beta^{-1} + \alpha)\sin\theta$$
$$+ \sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_1 \tag{50}$$
$$0 = \sin\theta\cos\theta - \alpha\zeta - I_J(1 - \alpha\beta^{-1})\sin\theta$$
$$+ \sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_2. \tag{51}$$

Finally, solving for ζ in (51) and substituting it back into (50), with the help of the fact that $\alpha W_1 - W_2 = \sqrt{1 + \alpha^2}\,W$, where W is another Brownian motion, we obtain (20). It is interesting to note that only the contribution of the fieldlike spin torque appears in the reduced equation, while the contribution of the Slonczewski torque cancels out to the leading order.

References

[1] J.-G. Zhu, "Magnetoresistive random access memory: The path to competitiveness and scalability," *Proc. IEEE*, vol. 96, no. 11, pp. 1786–1798, Nov. 2008.

[2] A. V. Khvalkovskiy *et al.*, "Basic principles of STT-MRAM cell operation in memory arrays," *J. Phys. D, Appl. Phys.*, vol. 46, no. 7, p. 074001, 2013.

[3] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature Nanotechnol.*, vol. 10, no. 3, pp. 187–191, 2015.

[4] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive random access memory," *Proc. IEEE*, vol. 104, no. 10, pp. 1796–1830, Oct. 2016.

[5] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. N. Piramanayagam, "Spintronics based random access memory: A review," *Mater. Today*, vol. 20, no. 9, pp. 530–548, 2017.

[6] W. H. Butler *et al.*, "Switching distributions for perpendicular spin-torque devices within the macrospin approximation," *IEEE Trans. Magn.*, vol. 48, no. 12, pp. 4684–4700, Dec. 2012.

[7] Y. Xie, B. Behin-Aein, and A. W. Ghosh, "Fokker-Planck study of parameter dependence on write error slope in spin-torque switching," *IEEE Trans. Electron Devices*, vol. 64, no. 1, pp. 319–324, Jan. 2017.

[8] M. Tzoufras, "Switching probability of all-perpendicular spin valve nanopillars," *AIP Adv.*, vol. 8, no. 5, p. 056002, May 2018.

[9] Y. Xie, J. Ma, S. Ganguly, and A. W. Ghosh, "From materials to systems: A multiscale analysis of nanomagnetic switching," *J. Comput. Electron.*, vol. 16, no. 4, pp. 1201–1226, Dec. 2017.

[10] J. Z. Sun *et al.*, "Sullivan, W. J. Gallagher, and D. C. Worledge, "Effect of subvolume excitation and spin-torque efficiency on magnetic switching," *Phys. Rev. B, Condens. Matter*, vol. 84, no. 6, p. 064413, 2011.

[11] R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall, and A. B. Kos, "Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions," *J. Appl. Phys.*, vol. 109, no. 7, p. 073910, Apr. 2011.

[12] H. Zhao *et al.*, "Spin-torque driven switching probability density function asymmetry," *IEEE Trans. Magn.*, vol. 48, no. 11, pp. 3818–3820, Nov. 2012.

[13] Y. Shiota *et al.*, "Evaluation of write error rate for voltage-driven dynamic magnetization switching in magnetic tunnel junctions with perpendicular magnetization," *Appl. Phys. Exp.*, vol. 9, p. 013001, Jan. 2016.

[14] D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.

[15] R. O. Moore, G. Biondini, and W. L. Kath, "A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons," *SIAM Rev.*, vol. 50, p. 523, Jan. 2008.

[16] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge, MA, USA: Cambridge Univ. Press, 2000.

[17] T. Dean and P. Dupuis, "Splitting for rare event simulation: A large deviation approach to design and analysis," *Stochastic Process. Their Appl.*, vol. 119, no. 2, pp. 562–587, Feb. 2009.

[18] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method* (Wiley Series in Probability and Statistics), 3rd ed. Hoboken, NJ, USA: Wiley, 2017.

[19] U. Roy, T. Pramanik, L. F. Register, and S. K. Banerjee, "Write error rate of spin-transfer-torque random access memory including micromagnetic effects using rare event enhancement," *IEEE Trans. Magn.*, vol. 52, no. 10, pp. 1–6, Oct. 2016.

[20] T. Pramanik, U. Roy, P. Jadaun, L. F. Register, and S. K. Banerjee, "Write error rates of in-plane spin-transfer-torque random access memory calculated from rare-event enhanced micromagnetic simulations," *J. Magn. Magn. Mater.*, vol. 467, pp. 96–107, Dec. 2018.

[21] P. Dupuis and H. Kushner, "Stochastic systems with small noise, analysis and simulation; a phase locked loop example," *SIAM J. Appl. Math.*, vol. 47, no. 3, pp. 643–661, Jun. 1987.

[22] G. Rubino and B. Tuffin, *Rare Event Simulation using Monte Carlo Methods*. Hoboken, NJ, USA: Wiley, 2009.

[23] M. Cottrell, J.-C. Fort, and G. Malgouyres, "Large deviations and rare events in the study of stochastic algorithms," *IEEE Trans. Autom. Control*, vol. AC-28, no. 9, pp. 907–920, Sep. 1983.

[24] E. Vanden-Eijnden and J. Weare, "Rare event simulation of small noise diffusions," *Comm. Pure Appl. Math.*, vol. 65, no. 12, pp. 1770–1803, Dec. 2012.

[25] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Grundlehren der Mathematischen Wissenschaften). Berlin, Germany: Springer, 2012.

[26] P. Dupuis, K. Spiliopoulos, and H. Wang, "Importance sampling for multiscale diffusions," *Multiscale Model. Simul.*, vol. 10, no. 1, pp. 1–27, Jan. 2012. [Online]. Available: http://epubs.siam.org.libdb.njit.edu:8888/doi/abs/10.1137/110842545

[27] M. Heymann and E. Vanden-Eijnden, "The geometric minimum action method: A least action principle on the space of curves," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1052–1117, Aug. 2008.

[28] C. J. García-Cervera, "Numerical micromagnetics: A review," *Boletin Sociedad Espanola Matematica Aplicada*, vol. 39, pp. 103–135, 2007.

[29] A. Brataas, A. D. Kent, and H. Ohno, "Current-induced torques in magnetic materials," *Nature Mater.*, vol. 11, no. 5, pp. 372–381, 2012.

[30] G. Gioia and R. D. James, "Micromagnetics of very thin films," *Proc. Roy Soc. Lond. A*, vol. 453, pp. 213–223, Jan. 1997.

[31] M. Hairer. (2009). "An introduction to stochastic PDEs." [Online]. Available: https://arxiv.org/abs/0907.4178

[32] W. F. Brown, Jr., "Thermal fluctuations of a single-domain particle," *Phys. Rev.*, vol. 130, no. 5, pp. 1677–1686, Jun. 1963.

[33] J. L. García-Palacios and F. J. Lázaro, "Langevin-dynamics study of the dynamical properties of small magnetic particles," *Phys. Rev. B, Condens. Matter*, vol. 58, pp. 14937–14958, 1998.

[34] D. Pinna, A. D. Kent, and D. L. Stein, "Spin-transfer torque magnetization reversal in uniaxial nanomagnets with thermal noise," *J. Appl. Phys.*, vol. 114, no. 3, p. 033901, 2013.

[35] K. A. Newhall and E. Vanden-Eijnden, "Averaged equation for energy diffusion on a graph reveals bifurcation diagram and thermally assisted reversal times in spin-torque driven nanomagnets," *J. Appl. Phys.*, vol. 113, no. 18, pp. 184105-1–184105-12, May 2013.

[36] R. V. Kohn and V. Slastikov, "Effective dynamics for ferromagnetic thin films: A rigorous justification," *Proc. Roy. Soc. Lond. A*, vol. 461, pp. 143–154, Jan. 2005.

[37] C. J. Garciá-Cervera, "Effective dynamics for ferromagnetic thin films," *J. Appl. Phys.*, vol. 90, no. 1, pp. 370–374, 2001.

[38] R. G. Lund and G. D. Chaves-O, "Flynn, A. D. Kent, and C. B. Muratov, "Reduced model for precessional switching of thin-film nanomagnets under the influence of spin torque," *Phys. Rev. B*, vol. 94, p. 144425, Oct. 2016.

[39] J. Miles and B. Middleton, "A hierarchical micromagnetic model of longitudinal thin film recording media," *J. Magn. Magn. Mater.*, vol. 95, pp. 99–108, Apr. 1991.

[40] J. J. Miles, M. Wdowin, J. Oakley, and B. K. Middleton, "The effect of cluster size on thin film media noise," *IEEE Trans. Magn.*, vol. 31, no. 2, pp. 1013–1024, Mar. 1995.

[41] J. Fidler and T. Schrefl, "Micromagnetic modelling—the current state of the art," *J. Phys. D, Appl. Phys.*, vol. 33, pp. R135–R156, Aug. 2000.

[42] T. Taniguchi and H. Imamura, "Thermally assisted spin transfer torque switching in synthetic free layers," *Phys. Rev. B, Condens. Matter*, vol. 83, p. 054432, Feb. 2011.

[43] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer Series in Synergetics). Berlin, Germany: Springer-Verlag, 2004.

[44] Y. Sun and X. Zhou, "An improved adaptive minimum action method for the calculation of transition path in non-gradient systems," *Commun. Comput. Phys.*, vol. 24, pp. 44–68, Jul. 2018.

[45] P. Glasserman and Y. Wang, "Counterexamples in importance sampling for large deviations probabilities," *Ann. Appl. Prob.*, vol. 7, no. 3, pp. 731–746, Aug. 1997. [Online]. Available: https://projecteuclid.org/euclid.aoap/1034801251

[46] P. Dupuis and H. Wang, "Subsolutions of an Isaacs equation and efficient schemes for importance sampling," *Math. Oper. Res.*, vol. 32, no. 3, p. 723, Aug. 2007.

[47] W. Chen, S. Zhang, and N. H. Bertram, "Energy barriers for thermal reversal of interacting single domain particles," *J. Appl. Phys.*, vol. 71, no. 11, pp. 5579–5584, 1992.

[48] N. Berglund, B. Fernandez, and B. Gentz, "Metastability in interacting nonlinear stochastic differential equations: I. From weak coupling to synchronization," *Nonlinearity*, vol. 20, no. 11, p. 2551, 2007.