Spring 2019

Rare event sampling in applied stochastic dynamical systems

YiMing Yu
New Jersey Institute of Technology

ABSTRACT

RARE EVENT SAMPLING IN APPLIED STOCHASTIC DYNAMICAL SYSTEMS

by
YiMing Yu

Predicting rare events is a challenging problem in many complex systems arising in physics, chemistry, biology, and materials science. Simulating rare events is often prohibitive in such systems due to their high dimensionality and the numerical cost of their simulation, yet analytical expressions for rare event probabilities are usually not available. This dissertation tackles the problem of approximation of the probability of rare catastrophic events in optical communication systems and spin-torque magnetic nanodevices. With the application of the geometric minimum action method, the probability of pulse position shifts or other parameter changes in a model of an actively mode-locked laser subject to noise from amplified spontaneous emission can be quantified. Similarly, by applying importance sampling with biasing functions motivated by optimal control, read and write soft error rates of macrospin and coupled-spin systems of spin-torque magnetic nanodevices can be efficiently estimated.

RARE EVENT SAMPLING IN APPLIED STOCHASTIC DYNAMICAL SYSTEMS

by
YiMing Yu

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
and Rutgers, The State University of New Jersey–Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers–Newark

May 2019

APPROVAL PAGE

RARE EVENT SAMPLING IN APPLIED STOCHASTIC DYNAMICAL SYSTEMS

YiMing Yu

Dr. Richard O. Moore, Dissertation Co-Advisor — Date
Associate Professor of Mathematical Sciences, New Jersey Institute of Technology

---

Dr. Cyrill B. Muratov, Dissertation Co-Advisor — Date
Professor of Mathematical Sciences, New Jersey Institute of Technology

---

Dr. David J. Horntrop, Committee Member — Date
Associate Professor of Mathematical Sciences, New Jersey Institute of Technology

---

Dr. Roy H. Goodman, Committee Member — Date
Associate Professor of Mathematical Sciences, New Jersey Institute of Technology

---

Dr. Tobias Schäfer, Committee Member — Date
Professor of Mathematics, City University of New York, New York

BIOGRAPHICAL SKETCH

Author: YiMing Yu

Degree: Doctor of Philosophy

Date: May 2019

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2019
- Bachelor of Science in Applied Mathematics,
  New York City College of Technology, Brooklyn, NY, 2013

Major: Applied Mathematics

Presentations and Publications:

Y. Yu, C. B. Muratov, and R. O. Moore, "Importance sampling for thermally induced switching and non-switching probabilities in spin-torque magnetic nanodevices," *Magnetics, IEEE Transactions on. (to appear).*

Y. Yu and R. O. Moore, "Quasi-stable dynamics of a mode-locked laser," *Proceedings of 13th International Conference on Mathematical and Numerical Aspects of Wave Propagation*, Minneapolis, MN, May 2017.

Y. Yu and R. O. Moore, "Biased Monte Carlo simulations to compute phase slip probabilities in a mode-locked laser model," *SIAM Conference on Nonlinear Waves and Coherent Structures*, Philadelphia, PA, August 2016.

Y. Yu and R. O. Moore, "Biased Monte Carlo simulations to compute phase slip probabilities in a mode-locked laser model," *Gene Golub SIAM Summer School*, Philadelphia, PA, August 2016.

This dissertation is dedicated to my wife and my family.

ACKNOWLEDGMENT

I want to thank my advisors Dr. Richard O. Moore and Dr. Cyrill B. Muratov for teaching and mentoring. Dr. Moore has been very supportive and patient, allowing me to explore different ideas in research and different options in my career, and Dr. Muratov has been very encouraging and enthusiastic to pursue exciting ideas. Our collaboration on our first paper has provided me a unique, invaluable experience on thoughts and process of collaborative research throughout many discussions and commitment. It would not have been possible to finish this dissertation without both of their guidance, valuable feedback, and instruction.

I next thank Dr. David J. Horntrop, Dr. Roy H. Goodman, and Dr. Tobias Schäfer for being my dissertation committee.

I would like to thank the Department of Mathematical Science of NJIT for giving me the opportunity to pursue a Ph.D. degree at such a prestigious university and for understanding the need of its graduate students. In particular, the department continuously supported me while I was away from NJIT during my last year.

Many thanks go to graduate students Emel Khan, Szu-Pei Fu, Micheal Pedeult, Pejman Sannaei, Xieyang Jia, Li Yu, Rui Cao, Ruihua, Alyssa loving and Vanessa Rivera Quinones for their friendship and encouragement.

This research is supported, in part, by the National Science Foundation via grant DMS-1313687.

Finally, I want to thank my mother and my sisters for their endless support and understanding. I also want to thank my beloved wife Marissa Loving Yu for her conditional love, endless support, and having faith in me. We might live in two different states, but we went through countless nights via Skype. She would come to visit whenever she got chances. I thank all the time and energy that she spent for us in our long-distance relationship to make it last.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Rare but important events arise due to the impact of noise on complex systems in physics, chemistry, biology, materials science, optical devices, and networks. Examples are thermal switching in micromagnetic devices, phase slips in coherent optical systems, conformational changes of biomolecules, chemical reactions, and dislocation dynamics in crystalline solids. Simulating these rare events is often prohibitive due to computational cost since the physical system usually involves high dimension, nonlinearity, and disparity of time scales. Thus, analysis and computations must concentrate on the most critical noise-induced events. In many cases, the noisy complex dynamical system can be modeled as a system of stochastic ordinary differential equations, in which the noise is modeled by Brownian motion. Many tools exist to assist in the computation of rare events in these cases, including computational tool development arising from the application of large deviation theory.

The second chapter is devoted to reviewing these tools with emphasis on large deviation theory and importance sampling. Large deviation theory is a tool to analyze the maximum likelihood transition path from one configuration to another configuration of these systems. Simulation with the guidance of large deviation theory can be an efficient way to approximate the probability of rare events.

The third chapter is devoted to studying rare events in examples from nonlinear optical communications, such as a fiber-based optical communication system and an actively mode-locked laser. The primary source of noise is amplified spontaneous emissions. To investigate the quasi-stability induced by the noise of an optical system, the infinite-dimensional model is reduced to a finite-dimensional system of stochastic ordinary differential equations and numerically verified for agreement. The quasi-stable state is computed by evaluating the action functional via the geometric

minimum action method (gMAM). This computation shows how and to what extent noise effectively destabilizes the system, as demonstrated by depicting a region of quasi-stability in its parameter space similar to a stability diagram for a deterministic system.

The fourth chapter is devoted to studying rare events in spin-torque magnetic nanodevices. 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 small probabilities for magnetic nanodevices using naive Monte Carlo simulations is virtually 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.

CHAPTER 2

TECHNIQUES FOR RARE EVENTS

The need to compute rare event probabilities is often driven by performance requirement. For example, reliable operation for a working memory requires a write error rate in a magnetic RAM less than 10^{-9} in the presence of an error correction code in the chip or less than 10^{-18} without an error correction code. In other settings, the importance of rare events stems from their potentially catastrophic impact. Examples include major earthquakes, tsunamis, hurricanes, financial market crash [2], and population collapses [3].

In many cases, the complex system can be modeled by stochastic differential equations (SDEs) subject to random noise. It is the noise that induces transitions between basins of attraction of stable fixed points in the deterministic dynamical system. This often implies the presence of at least two time-scales in these perturbed systems: the time-scale of the deterministic dynamics and the time-scales between the transitions of rare events. The former is determined by eigenvalues of the system; Jacobian evaluated at the fixed point, while the latter scales inverse exponentially with the noise strength. This will be discussed more in section (2.2).

A traditional method to quantify the effect of random noise is direct simulation of the stochastic dynamical system in Monte Carlo (MC) simulations [4, 5]. When the amplitude of the random noise is small, naive or unbiased MC simulations require a prohibitively large number of runs to obtain a sufficient number of the events of interest to provide reasonable statistical accuracy. This can be addressed through the use of importance sampling (IS) or biased MC simulation. As we mention before, Freidlin-Wentzell large deviation theory provides a framework for understanding the transitions of the rare events from one configuration to another configuration

of the underlying system [6–9], which provides useful information to construct an appropriate biasing function.

2.1 Monte Carlo Simulation and Importance Sampling

Let us start with the basic concepts of the Monte Carlo simulation. 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 $I(\omega)$ of $\mathcal{A}$, that is, $I(\omega) = 1$ if $\omega \in \mathcal{A}$ and $I(\omega) = 0$ otherwise. Here, $\mathbb{E}_0[\cdot]$ denotes the expected value with respect to the density $\rho_0(\omega)$, such that

$$P = \mathbb{E}_0[I(\omega)] = \int I(\omega)\rho_0(\omega)\, d\omega. \tag{2.1}$$

The strong law of large numbers states that if $\omega_1, \omega_2, ..., \omega_M$ are independent and identically distributed (i.i.d.) draws from the probability density ρ_0, then for any function $f(\omega)$ with finite average, meaning that $\mathbb{E}_0[f(\omega)] < \infty$, the sample average $\frac{1}{M}\sum_{i=1}^{M} f(\omega_i)$ converges to the true mean $\mathbb{E}_0[f(\omega)]$ as $M \to \infty$ with probability one [10]. Thus, a naive Monte Carlo method uses M independent draws ω_i from distribution ρ_0 to approximate P according to

$$P_{\mathrm{MC}} = \frac{1}{M}\sum_{i=1}^{M} I(\omega_i), \tag{2.2}$$

which is an unbiased estimator of P with mean

$$\mathbb{E}_0[P_{\mathrm{MC}}] = \mathbb{E}_0\left[\frac{1}{M}\sum_{i=1}^{M} I(\omega_i)\right] = \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}_0[I(\omega_i)] = P, \tag{2.3}$$

and variance

$$\mathrm{Var}[P_{\mathrm{MC}}] = \mathrm{Var}\left[\frac{1}{\mathrm{M}}\sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{I}(\omega_{\mathrm{i}})\right] = \frac{1}{\mathrm{M}^2}\sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{Var}[\mathrm{I}(\omega_{\mathrm{i}})] = \frac{1}{\mathrm{M}}\mathrm{P}(1-\mathrm{P}). \tag{2.4}$$

According to the central limit theorem, the MC estimator P_{MC} is distributed approximately as a normal distribution with mean $\mathbb{E}_0[P_{\mathrm{MC}}]$ and variance $\mathrm{Var}[P_{\mathrm{MC}}]$, as M goes to infinity [10].

The coefficient of variation (CV), which measures the relative statistical error of the estimator P_{MC}, is given by

$$CV(P_{\mathrm{MC}}) = \frac{\sqrt{\mathrm{Var}(P_{\mathrm{MC}})}}{\mathbb{E}_0[P_{\mathrm{MC}}]} = \frac{1}{\sqrt{M}}\sqrt{\frac{1}{P} - 1} \approx \frac{1}{\sqrt{M}}\sqrt{\frac{1}{P_{\mathrm{MC}}} - 1}. \tag{2.5}$$

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 function u chosen to increase the likelihood of the event of interest with the observation of changing the measure

$$P = \int I(\omega)\rho_0(\omega)\, d\omega = \int I(\omega)\frac{\rho_0(\omega)}{\rho_u(\omega)}\rho_u(\omega)\, d\omega = \mathbb{E}_u[I(\omega)L(\omega)], \tag{2.6}$$

where $\mathbb{E}_u[\cdot]$ denotes the expected value with respect to the density $\rho_u(\omega)$, and $L(\omega) = \rho_0(\omega)/\rho_u(\omega)$ is called the likelihood ratio, provided that $\rho_u(\omega) > 0$ whenever $\rho_0(\omega) > 0$. Similarly to the construction of naive MC estimator, the unbiased IS estimator is then recovered by weighting each result according to

$$P_{\mathrm{IS}} = \frac{1}{M}\sum_{i=1}^{M} I(\omega_i)L(\omega_i), \tag{2.7}$$

with mean

$$\mathbb{E}_u[P_{\mathrm{IS}}] = \mathbb{E}_u\left[\frac{1}{M}\sum_{i=1}^{M} I(\omega_i)L(\omega_i)\right] = \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}_u[I(\omega_i)L(\omega_i)] = P \tag{2.8}$$

and variance

$$\begin{aligned}\mathrm{Var}[P_{\mathrm{IS}}] &= \mathrm{Var}\left[\frac{1}{M}\sum_{i=1}^{M} I(\omega_i)L(\omega_i)\right] \\ &= \frac{1}{M^2}\sum_{i=1}^{M}\mathrm{Var}[I(\omega_i)L(\omega_i)] \\ &= \frac{1}{M}(\mathbb{E}_u[I(\omega_i)L^2(\omega_i)] - P^2). \end{aligned} \tag{2.9}$$

The resulting CV is then given by

$$CV(P_{\mathrm{IS}}) = \frac{\sqrt{\mathrm{Var}(P_{\mathrm{IS}})}}{\mathbb{E}_u[P_{\mathrm{IS}}]} = \frac{1}{\sqrt{M}}\sqrt{\frac{\mathbb{E}_u[I(\omega)L^2(\omega)]}{\mathbb{E}_u[I(\omega)L(\omega)]^2} - 1}, \tag{2.10}$$

where $1/P \geq \mathbb{E}_u[I(\omega)L^2(\omega)]/P^2 \geq 1$.

Expression (2.10) suggests that a "good" bias to use in IS keeps $\mathbb{E}_u[I(\omega)L^2(\omega)]$ close to P^2. It is easy to show that the optimal choice for $\rho_u(\omega)$ is given by

$$\rho_u(\omega) = \frac{I(\omega)\rho_0(\omega)}{P}, \tag{2.11}$$

which gives that $\mathrm{Var}[P_{IS}] = 0$. This indicates that there exists an optimal change of measure that leads to a zero-variance estimator. In practice, this is impossible since the equation (2.11) requires the knowledge of the probability of rare events P. The result also indicates that we should choose ρ_u as close to equation (2.11) as possible.

2.2 Importance Sampling for Stochastic Differential Equations

In many cases of interest, the underlying physical model is described by an Itô stochastic differential equation

$$dX(t) = b(X(t))dt + \epsilon\sigma(X(t))dW(t), \quad X(0) = x_0, \tag{2.12}$$

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 diffusion coefficient matrix, dW is an infinitesimal

increment of an N-dimensional Brownian motion, and $\epsilon > 0$ is noise strength. We focus on system (2.12) above whose deterministic part is autonomous, i.e, it does not explicitly, depend on time t. 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 to obverse a single event of interest, even more so to produce an accurate probability estimate from these empirical observations.

Likelihood ratio for SDEs. In the case of SDEs, the IS technique makes rare events happen more often by introducing a bias to the mean of the noise increment dW, such that the coefficient of variation of the estimate P_{IS} is reduced at the same time [8, 11]. This provides new paths $\tilde{X}(t)$ that evolve according to

$$d\tilde{X} = \Big(b(\tilde{X}) + \sigma(\tilde{X})u(\tilde{X}, t)\Big)\, dt + \epsilon\sigma(\tilde{X})d\tilde{W}, \tag{2.13}$$

where we have chosen a form of bias term

$$dW = d\tilde{W} + \epsilon^{-1}udt, \tag{2.14}$$

suggestive of optimal control. By Girsanov's theorem [6, 7], for a time horizon $T > 0$, the likelihood ratio is then given by

$$L = \exp\left(-\frac{1}{2\epsilon^2}\int_0^T |u(\tilde{X}, t)|^2 dt - \frac{1}{\epsilon}\int_0^T \langle u(\tilde{X}, t), dW(t)\rangle\right), \tag{2.15}$$

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 (2.7) to recover an unbiased probability estimate.

Large deviation theory. The theory of large deviations has applications in many fields, such as physics, chemistry, and biology. It provides a mathematically rigorous

framework to estimate the probability of rare events and the rate of occurrence of the rare events, and to identify the maximum likelihood pathway of the rare events.

The large deviation theory gives a description of the behavior of a trajectory solution of equation (2.12) to a function $\phi(t)$, which is given by

$$P\left\{\sup_{0\leq t\leq T}|X(t)-\phi(t)|\leq\delta|X(0)=x_0\right\}\sim\exp(-S_T(\phi)/\epsilon^2), \tag{2.16}$$

for ϵ and δ sufficiently small. The action functional is defined by

$$S_T[\phi]=\int_0^T L(\phi(t),\dot{\phi}(t))dt, \tag{2.17}$$

where $\dot{\phi}(t)=d\phi(t)/dt$, over absolutely continuous paths $\phi:[0,T]\to\mathbb{R}^N$ satisfying $\phi(0)=x$ and $\phi(T)\in\mathcal{A}$ [12]. This functional has a Lagrangian and associated Hamiltonian are given by

$$L(x,y)=\frac{1}{2}|y-b(x)|^2_{a(x)}, \tag{2.18}$$

and

$$H(x,\theta)=\langle b(x),\theta\rangle+\frac{1}{2}\langle\theta,a(x)\theta\rangle, \tag{2.19}$$

where $\langle u,v\rangle_{a(x)}=\langle u,a^{-1}(x)v\rangle$, $|u|_{a(x)}=\langle u,u\rangle^{1/2}_{a(x)}$, and $a(x)=\sigma(x)\sigma^T(x)$, for any $u,v,x\in\mathbb{R}^N$ [7].

Equation (2.16) indicates that the probability density distribution of $X(t)$ is concentrated in the neighborhood of $\phi(t)$ and that the probabilities of paths that are not in the neighborhood of $\phi(t)$ are exponentially small. It provides an estimate of the probability of a given solution of equation (2.12). Moreover, minimizing the functional under endpoint constraints provides an estimate of the probability that is associated with transition from one state to the another state. For instance, if $\mathcal{A}$ is a

Borel subset of $\mathbb{R}^N$, we have

$$\lim_{\epsilon\to 0}\epsilon^2\log P\left\{X(T)\in\mathcal{A}|X(0)=x_0\right\}\sim -\inf_{\substack{\phi(0)=x,\\ \phi(T)=\mathcal{A}}} S_T[\phi], \tag{2.20}$$

whose minimizer gives us the most probable path for the transition of the trajectory starting from $X(0)=x_0$ and ending at $X(T)\in\mathcal{A}$ [7, 13].

Another important concept in large deviation theory is the quasi-potential, which is defined by

$$V(x_0,x_T)=\inf_{T>0}\inf_{\substack{\phi(0)=x_0,\\ \phi(T)=x_T}} S_T[\phi]. \tag{2.21}$$

The quasi-potential allows us to understand the stochastic dynamical system over an exponentially long or infinite time interval. For example, consider O to be an asymptotically stable fixed point of system (2.12) with its domain of attraction $D\in\mathbb{R}^N$ having a smooth boundary ∂D. Then the mean first exit time $E[\tau]$ is given by

$$\lim_{\epsilon\to 0}\epsilon^2\ln E[\tau]\sim\min_{y\in\partial D} V(O,y). \tag{2.22}$$

Minimization of variance. To minimize the error of the estimator (2.10), we need to control the ratio

$$\begin{aligned} R(P_{IS}) &= \frac{\mathbb{E}_u[I(\omega)L^2(\omega)]}{\mathbb{E}_u[I(\omega)L(\omega)]^2} \\ &= \frac{\mathbb{E}_u[I(\omega)L^2(\omega)]}{\mathbb{E}_0[I(\omega)]^2}. \end{aligned} \tag{2.23}$$

Numerical and theoretical evidence show that equation (2.10) has asymptotically bounded relative error in the limit as ϵ goes to zero, provided that the boundary of $\mathcal{A}$ is smooth [7].

Jensen's inequality implies that $R(P_{IS})\geq 1$, and in the context of SDEs with vanishing noise, we have

$$\lim_{\epsilon\to 0}\epsilon^2\log\mathbb{E}_u[I(\omega)L^2(\omega)]\geq 2\lim_{\epsilon\to 0}\epsilon^2\log\mathbb{E}_0[I(\omega)]. \tag{2.24}$$

In our interest of rare events for SDEs, the mean $\mathbb{E}_0[I(\omega)]$ denotes the probability $P(X(T) \in \mathcal{A}|X(0) = x_0)$, where the $X(t)$ are solutions of equation (2.12) with initial point at $X(0) = x_0$.

To find the density of X, let us look at a discrete version of equation (2.12) given by

$$X^{i+1} = X^i + b(X^i)\Delta t + \epsilon\sigma(X^i)\Delta W^i, \quad X^0 = x_0, \tag{2.25}$$

for $i = 0, 1, ..., K$ and $\Delta t = T/K$. Here $X^i = X(i\Delta t)$ and the $\Delta W^i s$ are i.i.d Gaussian random variables with mean 0 and variance Δt. The conditional density of X^{i+1} given X^i is normal with mean $X^i + b(X^i)\Delta t$ and variance $\epsilon^2\sigma^T(X^i)\sigma(X^i)\Delta t$, meaning that

$$\begin{aligned} p(X^{i+1}|X^i) &= \frac{\exp(-\frac{1}{2\epsilon^2\Delta t}(X^{i+1} - X^i - b(X^i)\Delta t)^T a^{-1}(X^i)(X^{i+1} - X^i - b(X^i)\Delta t))}{\sqrt{(2\pi)^N\epsilon^2|a(X^i)|\Delta t}} \\ &= \frac{\exp(-\frac{1}{2\epsilon^2}|\sigma^{-1}(X^i)(\frac{X^{i+1}-X^i}{\Delta t} - b(X^i))|^2\Delta t)}{\sqrt{(2\pi)^N\epsilon^2|a(X^i)|\Delta t}}, \end{aligned} \tag{2.26}$$

where $a(X) = \sigma^T(X)\sigma(X)$ and $|a(X)|$ is the determinant of $a(X)$. Thus the density for a discrete path $\{X_i\}_{i=0,1,...K}$ is given by

$$\begin{aligned} p(X^N|X^0) &= \prod_{i=0}^{K-1} p(X^{i+1}|X^i) \\ &= \frac{\exp(-\frac{1}{2\epsilon^2}\sum_{i=0}^{K-1}|\sigma^{-1}(X^i)(\frac{X^{i+1}-X^i}{\Delta t} - b(X^i))|^2\Delta t)}{\prod_{i=0}^{K-1}\sqrt{(2\pi)^N\epsilon^2|a(X^i)|\Delta t}}. \end{aligned} \tag{2.27}$$

The exponential term in equation (2.27) converges to integral

$$\frac{\Delta t}{2\epsilon^2}\sum_{i=0}^{K-1}\left|\sigma^{-1}(X^i)\left(\frac{X^{i+1}-X^i}{\Delta t} - b(X^i)\right)\right|^2 \to \frac{1}{2\epsilon^2}\int_0^T \left|\sigma^{-1}(X(t))\left(\dot{X}(t) - b(X(t))\right)\right|^2 dt, \tag{2.28}$$

as $\Delta t \to 0$. Therefore, the density for a continuous path $X(t)$ is given by

$$p_0(X(T)|X(0) = x_0) = C\exp\left(-\frac{1}{\epsilon^2}S_T[X]\right), \tag{2.29}$$

where C is a normalizing constant. As a result, the probability for $X(T) \in \mathcal{A}$ given $X(0) = x_0$ is

$$P(X(T) \in \mathcal{A}|X(0) = x_0) = C \int_{X(t):X(T)\in A} \exp\left(-\frac{1}{\epsilon^2} S_T[X]\right) dX(t). \tag{2.30}$$

Recall that to calculate the asymptotic approximation for integrals of the form

$$\int_a^b f(t) \exp(-g(t)/\epsilon) dt, \tag{2.31}$$

one uses Laplace's method as $\epsilon \to 0$. If there is a unique minimum $g(t_0)$ in the interval $a \leq t_0 \leq b$ and $f(t)$ is continuous and positive, then the major contribution to this integral is given by the neighborhood of t_0. In analogy with the Laplace's method, if a minimum of $S_T[X]$ connected with the process $X(t)$ starting at $X(0) = x_0$ and terminal at $X(T) \in \mathcal{A}$ exists, then we have

$$P(X(T) \in \mathcal{A}|X(0) = x_0) \sim C \exp\left(-\frac{1}{\epsilon^2} \inf_{\substack{\phi(0)=x_0, \\ \phi(T)\in\mathcal{A}}} S_T[\phi]\right), \tag{2.32}$$

as $\epsilon \to 0$. Applying Laplace's principle to equation (2.29) [14], we have

$$\lim_{\epsilon\to 0} \epsilon^2 \log \mathbb{E}_0[I(\omega)] \sim - \inf_{\substack{\phi(0)=x_0, \\ \phi(T)\in\mathcal{A}}} S_T[\phi]. \tag{2.33}$$

Meanwhile, applying Varadhan's lemma [14], we have

$$\lim_{\epsilon\to 0} \epsilon^2 \log \mathbb{E}_u[I(\omega)L^2(\omega)] \sim - \inf_{\substack{\phi(0)=x_0, \\ \phi(T)\in\mathcal{A}}} \left(\tilde{S}_T[\phi] + \int_0^T |u(\phi, s)|^2 ds\right) \tag{2.34}$$

where the Freidlin-Wentzell large deviation action is defined by

$$\tilde{S}_T[\phi] = \int_0^T \frac{1}{2} |\sigma^{-1}(\phi(s))(\dot{\phi}(s) - b(\phi(s))) - u(\phi, s)|^2 ds, \tag{2.35}$$

By choosing $u = \phi(s)(\dot{\phi}(s) - b(\phi(s)))$, equations (2.34) becomes

$$\lim_{\epsilon \to 0} \epsilon^2 \log \mathbb{E}_u[I(\omega)L^2(\omega)] \sim - \inf_{\substack{\phi(0)=x_0, \\ \phi(T)=\mathcal{A}}} \left(\tilde{S}_T[\phi] + \int_0^T |u(\phi, s)|^2 ds \right)$$

$$= -2 \inf_{\substack{\phi(0)=x_0, \\ \phi(T)=\mathcal{A}}} S_T[\phi]. \tag{2.36}$$

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 $\phi^T_{t,x}(s)$ of the functional

$$S_T[\phi] = \int_t^T \frac{1}{2} |\sigma^{-1}(\phi(s))(\dot{\phi}(s) - b(\phi(s)))|^2 ds, \tag{2.37}$$

where $\dot{\phi}(s) = d\phi(s)/ds$, over absolutely continuous paths $\phi : [t, T] \to \mathbb{R}^N$ satisfying $\phi(t) = x$ and $\phi(T) \in \mathcal{A}$ [12]. The finite-time bias function $u^* = u^*_T$ is then given by [7, 9, 11]

$$u^*_T(x, t) = \sigma^{-1}(x)(\dot{\phi}^T_{t,x}(t) - b(x)). \tag{2.38}$$

Over infinite-time horizons, i.e., when $T \to \infty$, a convenient reparametrization allows the action in equation (2.37) 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 [12, 13]

$$S_\infty[\phi] = \int_0^1 \big(|\sigma^{-1}(\phi(\alpha))\phi'(\alpha)| \, |\sigma^{-1}(\phi(\alpha))b(\phi(\alpha))| \\ - \langle \sigma^{-1}(\phi(\alpha))\phi'(\alpha), \sigma^{-1}(\phi(\alpha))b(\phi(\alpha)) \rangle \big) d\alpha, \tag{2.39}$$

among all absolutely continuous paths $\phi : [0, 1] \to \mathbb{R}^N$ 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 [11]

$$u^*_\infty(x) = \sigma^{-1}(x) \left(\frac{|\sigma^{-1}(x)b(x)|}{|\sigma^{-1}(x)\phi'_x(0)|} \phi'_x(0) - b(x) \right). \tag{2.40}$$

One of the strategies discussed below 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 Chapter 4 Section 4.3, this strategy is effective for intermediate times but does not correctly recover 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.

2.3 Minimization of Action

2.3.1 Minimization of action with fixed endpoints

The calculation of the minimizer of the action functional (2.17) and the quasi-potential (2.21) is critical to the application of large deviation theory. Here I will describe the minimization problems for each case and their solutions in terms of a Euler-Lagrange equation with suitable boundary conditions.

Finite-time horizons. The minimization problem of the action functional over finite-time horizons is given by

$$\inf_{\phi} S_T[\phi], \tag{2.41}$$

subject to $\phi(0) = x_0$ and $\phi(T) = x_T$. The minimizer satisfies the Euler-Lagrange equations yielding the Hamiltonian system of ODEs,

$$\begin{aligned} \dot{\phi} &= H_\theta(\phi, \theta) \\ \dot{\theta} &= -H_x(\phi, \theta), \end{aligned} \tag{2.42}$$

subject to boundary conditions $\phi(0) = x_0$ and $\phi(T) = x_T$. The simplest way to minimize the action functional perhaps is the shooting method. In practice, this method become inefficient when the dimension of the system increases. Other methods such as the minimum action method (MAM) [13] and the improved adaptive minimum action method (aMAM) [15] are introduced to solve the equivalent system obtained by differentiating ϕ twice, ie,

$$\begin{aligned} \ddot{\phi} &= H_{\theta x}(\phi, \theta)\dot{\phi} + H_{\theta\theta}\dot{\theta} \\ &= H_{\theta x}\dot{\phi} - H_{\theta\theta}H_x. \end{aligned} \tag{2.43}$$

Infinite-time horizons. Minimizers of the action functional over infinite-time horizons are given by

$$V(x_0, x_T) = \inf_{T>0} \inf_{\phi} S_T[\phi], \tag{2.44}$$

subject to $\phi(0) = x_0$ and $\phi(T) = x_T$.

The shooting method, MAM, and improved aMAM are often not suitable for this double minimization problem (2.44) over T and ϕ, since they discretize the time interval $[0, T]$. If the prescribed starting point x_0 and terminal point x_T are fixed points of the deterministic system (2.12), the minimizer takes infinite time, i.e $T \to \infty$. The geometric minimum action method overcomes this problem of infinite time interval through a geometric reformation of the quasi-potential (2.44).

Using equations (2.17) and (2.18), the minimization problem (2.44) reduces to

$$V(x_0, x_T) = \inf_{T>0} \inf_{\substack{\phi(0)=x_0,\\ \phi(T)=x_T}} \frac{1}{2} \int_0^T |\dot{\phi}(t) - b(\phi(t))|^2_{a(\phi)} dt. \tag{2.45}$$

Expanding the inner product inside the integral and applying the inequality $|u|^2_{a(x)} + |v|^2_{a(x)} \geq 2|u|_{a(x)}|v|_{a(x)}$, equation (2.45) can be reduced to

$$V(x_0, x_T) \geq \inf_{T>0} \inf_{\substack{\phi(0)=x_0,\\ \phi(T)=x_T}} \int_0^T |\dot{\phi}(t)|_{a(\phi)}|b(\phi(t))|_{a(x)} - \langle \dot{\phi}(t), b(\phi(t))\rangle dt \tag{2.46}$$

$$= 2 \inf_{T>0} \inf_{\substack{\phi(0)=x_0,\\ \phi(T)=x_T}} \int_0^T |\dot{\phi}(t)|_{a(\phi)}|b(\phi(t))|_{a(x)} \sin^2(\frac{1}{2}\eta(t)) dt, \tag{2.47}$$

where $\eta(t)$ is the angle between $\dot{\phi}(t)$ and $b(\phi(t))$ in the inner product induced by $\langle \cdot, \cdot \rangle_{a(\phi)}$. The equality between equations (2.45) and (2.46) can be achieved when $|\dot{\phi}(t)|_{a(\phi)} = |b(\phi(t))|_{a(x)}$. As a result, the minimization problem (2.44) has a geometric expression as

$$V(x_0, x_T) = 2 \inf_{\gamma} \int_\gamma |b|_a \sin^2(\frac{1}{2}\eta) ds, \tag{2.48}$$

where the ds is the arclength element along the curve γ, η is the angle between γ and b at location s along the curve, and the infimum is taken in the space of curves that starting at point x_0 and ending at point x_T.

Thus, we can introduce a parametrization of the curve $\gamma = \{\phi(\alpha) : \alpha \in [0, 1]\}$, with $|\phi'(\alpha)| =$ constant, where $\phi'(\alpha)$ is the derivative with respect to α such that

$$\dot{\phi} = \lambda \phi', \tag{2.49}$$

$$\ddot{\phi} = \lambda^2 \phi'' + \lambda \lambda' \phi', \tag{2.50}$$

and the action functional becomes

$$S_\infty[\phi] = \int_0^1 \frac{1}{\lambda} L(\phi(\alpha), \lambda \phi'(\alpha)) d\alpha, \tag{2.51}$$

or

$$S_\infty[\phi] = \int_0^1 \langle \phi', \theta(\phi, \phi') \rangle d\alpha, \tag{2.52}$$

where $\lambda = \lambda(\phi, \phi')$ [13] .

In the case of SDEs, one can show that θ is given by

$$\theta(x, y) = a^{-1}(x)\big(\lambda y - b(x)\big), \tag{2.53}$$

and

$$\lambda(x, y) = \frac{|b(x)|_a}{|y|_a}. \tag{2.54}$$

Using equations (2.49) and (2.50), equation (2.43) can be rewritten as

$$\lambda^2 \phi'' - \lambda H_{\theta x} \phi' + H_{\theta\theta} H_x + \lambda \lambda' \phi' = 0 \tag{2.55}$$

with boundary conditions $\phi(0) = x_0, \phi(1) = x_T$ and subject to the constraint $|\phi'(\alpha)|$ is constant. Equation (2.55) can be solved by a relaxation method with artificial time τ, i.e, the PDE

$$\dot{\phi} = \lambda^2 \phi'' - \lambda H_{\theta x} \phi' + H_{\theta\theta} H_x + \lambda \lambda' \phi' + \mu \phi', \tag{2.56}$$

where $\phi(\tau, 0) = x_0,\ \phi(\tau, 1) = x_T,$ and $\phi(0, \alpha) = \phi^0(\alpha)$. Here $\tau \geq 0,\ \dot{\phi} = \partial\phi/\partial\tau,$ and $\phi(0, \alpha) = \phi^0(\alpha)$ is the initial condition, which must satisfy the constraint that $|\phi'(\tau, \cdot)|$ is constant.

Our implementation of the gMAM uses the relaxation method with discretization of the Euler-Lagrange equation and uses an interpolation re-parametrization step to enforce the constraint on the parametrization of ϕ. We discretized $\phi(\tau, \alpha)$ in τ and α and denote $\phi_i^k = \phi(k\Delta\tau, i\Delta\alpha),\ k \in \mathbb{N}_0,\ i = 0, 1, ..., N$ where $\Delta\tau$ is the artificial time step and $\Delta\alpha = 1/N$ if the curve is discretized into $N + 1$ points.

Algorithm 1 GMAM

1: At the kth iteration, given $\{\phi_i^k\}_{i=0,...,N}$, compute $\phi_i'^k = (\phi_{i+1}^k - \phi_{i-1}^k)/(2/N)$, $\theta_i^k = \theta_i^k(\phi_i^k, \phi_i'^k)$ using equation (2.53), $\lambda_i^k = \lambda_i^k(\phi_i^k, \phi_i'^k)$ using equation (2.54), and $\lambda_i'^k = (\lambda_{i+1}^k - \lambda_{i-1}^k)/(2/N)$ for $i = 1, ..., N-1$, where $\lambda_0^k = 3(\lambda_1^k - \lambda_2^k) + \lambda_3^k$ and $\lambda_N^k = 3(\lambda_1^{N-1} - \lambda_2^{N-2}) + \lambda_{N-3}^k$.

2: Let $\{\tilde{\phi}_i\}_{i=0,...,N}$ be the solution of the linear system, for $i = 1, ..., N-1$

$$\frac{\tilde{\phi}_i - \phi_i^k}{\Delta\tau} = (\lambda_i^k)^2 \frac{\tilde{\phi}_{i+1} - 2\tilde{\phi}_i + \tilde{\phi}_{i-1}}{1/N^2} - \lambda_i^k H_{\theta x} \phi_i'^k + H_{\theta\theta} H_x + \lambda_i^k \lambda_i'^k \phi_i'^k \qquad (2.57)$$

with boundary conditions $\tilde{\phi}_0 = x_0$ and $\tilde{\phi}_N = x_T$. Note that $H_{\theta\theta}$, $H_{\theta x}$, and H_x are evaluated at (ϕ_i^k, θ_i^k).

3: Given $\{\tilde{\phi}_i\}_{i=0,...,N}$, find $\{\phi_i^{k+1}\}_{i=0,...,N}$ that satisfies the prescribed constraint $|\phi'(\tau, \cdot)| = cst$.

4: Repeat step 1-4 until stoping criterion is satisfied.

5: Action is given by

$$S = \frac{1}{2N} \sum_i^N \left(\langle \phi_i'^k, \theta_i'^k \rangle + \langle \phi_{i-1}'^k, \theta_{i-1}'^k \rangle \right). \qquad (2.58)$$

2.3.2 Minimization of action with variable endpoints

The gMAM with fixed endpoints gives the most likely path for a transition between two defined points, in many cases of interest, e.g, exit problem, the right end criterion is a set with dimension higher than zero, i.e, $\phi(T) \in \mathcal{A}$ where $\mathcal{A}$ is a Borel subset of $\mathbb{R}^N$. Define a function g such that $g(x) = 0$ for all $x \in \mathcal{A}$. The equation (2.45) becomes

$$\begin{aligned}
\inf V(x_0, y) &= \inf_{y\in\mathcal{A}} \inf_{T>0} \inf_{\substack{\phi(0)=x_0,\\ \phi(T)=y}} \frac{1}{2}\int_0^T |\dot{\phi}(t) - b(\phi(t))|^2_{a(\phi)} dt \\
&= \inf_{y\in\mathcal{A}} \inf_{\substack{\phi(0)=x_0,\\ \phi(1)=y}} \int_0^1 \langle \phi', \theta(\phi, \phi')\rangle d\alpha \qquad (2.59) \\
&= \inf_{\substack{\phi(0)=x_0\\ g(\phi(1))=0}} \int_0^1 \tilde{L}(\phi, \phi') d\alpha + \nu g(\phi(1)), \qquad (2.60)
\end{aligned}$$

where $\tilde{L} = \langle \phi', \theta(\phi, \phi')\rangle$. In the equation (2.60), the endpoint constraint $g(\phi(1)) = 0$ is enforced by a Lagrange multiplier term $\nu g(\phi(1))$ for $\nu \geq 0$. One can show that the minimizer of equation (2.60) satisfies the following equation [16]:

$$\dot{\phi} = \lambda^2 \phi'' - \lambda H_{\theta x}\phi' + H_{\theta\theta} H_x + \lambda\lambda'\phi' \qquad (2.61)$$

$$\dot{\phi}(\tau, 1) = -\theta(\phi(\tau, 1), \phi'(\tau, 1)) + \nu \nabla g(\phi(\tau, 1)), \qquad (2.62)$$

$$\phi(\tau, 0) = x_0, \ \phi(0, \alpha) = \phi^0(\alpha), \qquad (2.63)$$

where $\tau \geq 0$ and ν is given by

$$\nu = \frac{\langle \nabla g(\phi(\tau, 1)), \theta(\phi(\tau, 1), \phi'(\tau, 1))\rangle}{|\nabla g(\phi(\tau, 1))|^2}. \qquad (2.64)$$

As a result, we can modify the gMAM algorithm by updating

$$\tilde{\phi}_N = \phi_N - \Delta\tau(\theta^k_N + \nu^k \nabla g(\phi_N)). \qquad (2.65)$$

Notice that a simple modification of equation (2.57) in Step 3 of the gMAM algorithm can set some of the variables at the endpoint ϕ_N free by updating them

with

$$\frac{\tilde{\phi}_{Nj} - \phi^k_{Nj}}{\Delta\tau} = b_j(\phi^k_i), \tag{2.66}$$

for some $j \in [0, 1, ..., N]$.

For example, here we consider a two-dimensional system

$$\dot{x}_1 = c\sin(x_2 - x_1) + (I_J - \cos(x_1))\sin(x_1) + \sqrt{\frac{2}{\Delta}}\dot{W}_1, \tag{2.67}$$

$$\dot{x}_2 = c\sin(x_1 - x_2) + (I_J - \cos(x_2))\sin(x_2) + \sqrt{\frac{2}{\Delta}}\dot{W}_2, \tag{2.68}$$

where c, I_J and Δ are constant, and W_1 and W_2 are Wiener processes. The starting point of the system is $(0.01, 0.01)$ and we define a rare event to have occurred if $\max(|x_1(t)|, |x_2(t)|) = \pi/2$ for any $t > 0$. Let the drift vector be

$$b(x_1, x_2) = \begin{pmatrix} c\sin(x_2 - x_1) + (I_J - \cos(x_1))\sin(x_1) \\ c\sin(x_1 - x_2) + (I_J - \cos(x_2))\sin(x_2) \end{pmatrix} \tag{2.69}$$

Now let us apply the gMAM to this system. Since the diffusion matrix is the identity, we have

$$\theta = \lambda\phi' - b(\phi) \text{ and } \lambda = \frac{|b(\phi)|}{|\phi'|}. \tag{2.70}$$

We also have

$$H_x = (\nabla b)^T(\lambda\phi' - b(\phi)), H_{\theta\theta} = I, \text{ and } H_{x\theta} = \nabla b. \tag{2.71}$$

Since the exit criterion is $\max(|x_1(t)|, |x_2(t)|) = \pi/2$ for any $t > 0$, we have

$$g(x_1, x_2) = \max(|x_1|, |x_2|) - \pi/2. \tag{2.72}$$

In particular, if we consider that an exit will occur at $x_1 = \pi/2$, we have $g(x_1, x_2) = x_1 - \pi/2$ and $\nu = \theta_1 = (\lambda\phi' - b(\phi))_1$. Thus equation (2.61) can be written explicitly

as

$$\dot{\phi} = \lambda^2 \phi'' - \nabla b^T b + \lambda \lambda' \phi' \tag{2.73}$$

$$\dot{\phi}_2(\tau, 1) = -(\lambda \phi' - b(\phi))_2, \tag{2.74}$$

$$\phi(\tau, 0) = x_0, \ \ \phi(0, \alpha) = \phi^0(\alpha), \tag{2.75}$$

where $x_0 = (0.01, 0.01)$.

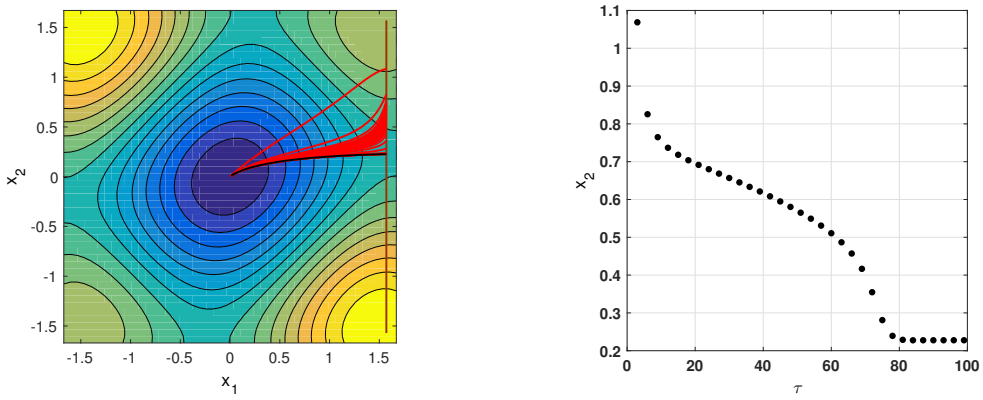

Figure 2.1 The minimum action path starting from (0.01,0.01) with variable endpoint along the line $x_1 = \pi/2$ (left). The red curves are intermediate action paths generated by equation (2.73), while the black curves is minimum action paths starting at (0.01,0.01) and ending at $(\pi/2, 0.228)$. Convergence of endpoint x_2 over iterations(right).

Figure 2.1 shows the action path generated by gMAM and the convergence of the endpoint variable x_2. The red curves on the left are intermediate action paths generated by gMAM algorithm, while the black curve is final minimum action path staring from $(0.01, 0.01)$ and ending at $(\pi/2, 0.0228)$. The graph on the right of Figure 2.1 shows that the endpoint variable x_2 converges to 0.228 after $\tau \geq 80$.

CHAPTER 3

PARAMETER EXCURSIONS IN MODE-LOCKED LASERS

3.1 Introduction

3.1.1 Optical fiber communication systems

A rapidly increasing demand for bandwidth delivered to internet services (e.g. streaming video) has led to the widespread deployment of optical fiber communication systems [17]. Optical fiber is a cylindrical tube of fused silica, which confines light to its interior using a latent gradient in the refractive index or a more complex microstructure with advantageous waveguide properties.

Modern optical fiber communication involves sending pulses of light through an optical fiber to transfer information. Optical fibers are multi-mode optical fibers or single-mode optical fibers. Basic multi-mode optical fiber has a core greater than 50 micrometers in diameter and admits many simultaneously propagating optical modes. Single-mode optical fiber has an inner core less than 10 micrometers in diameter and only admits one propagating spatial mode. Most optical fiber deployed is single-mode. Optical fiber communication systems consist of an optical transmitter to convert electrical signals to optical signals, a cable made of bundles of optical fibers, optical amplifiers to compensate for attenuation of pulses of light during transmission, and an optical receiver to convert optical signals into electrical signals that can be read.

3.1.2 Mode-locked laser basics

The most common type of laser consists of a gain medium to amplify light and an optical cavity to provide feedback. A gain medium is a material with excited atoms that amplify light through the process of stimulated emission. The optical cavity is a pair of mirrors on the ends of the gain medium. One of the mirrors is an output coupler and is partially transparent. Light can bounce back and forth between the

two mirrors, with a small fraction of the incident light passing through the output coupler. Depending on the design of the cavity, the light experience signification diffraction upon each round trip. The laser device is also called a laser oscillator.

Mode-locking is a technique by which a laser can generate coherent pulses of light. In a simple laser, each of the standing waves, known as longitudinal modes of the cavity, oscillates independently. In a mode-locked laser, each mode operates with a fixed phase between it and the other modes. The modes of the laser periodically interfere with each other to produce a short intense pulse of light. The periodic interference is produced by the presence of an element in the cavity that distinguishes mode-lock laser from other lasers. Active mode-locking uses an externally driven modulation while passive mode-locking uses passive element such as a saturable absorber.

A mode-locked laser can generate ultra-short optical pulses on the order of tens of picoseconds to femtoseconds. The pulse circulating in the cavity of a mode-locked laser can be thought of as a dissipative soliton, where the dissipation is due to loss of energy through interaction with cavity elements. The loss is compensated by gain. The mode-locked laser is an excellent system for experimental studies of soliton dynamics. The evolution of a pulse can be monitored since the pulse is emitted once per round trip.

In the context of a periodically amplified pulse of light either propagating through an optical fiber or circulating in the cavity of the laser, a primary noise source is amplified spontaneous emission (ASE) noise, which always accompanies stimulated emission. The ASE perturbs the pulses of light in ways that we will characterize using parametrized representation of the pulses. The noise is often modeled as effectively white in both space and time solitons with the fact that the pulse has a substantially smaller bandwidth than the noise source.

3.2 Model Including Loss and Filtering

In the cavity of an optical fiber laser with an amplifier, the average round-trip propagation of optical pulse is governed by a perturbed NLS equation:

$$iu_z + du_{tt} + c|u|^2u = -\alpha iu + i\kappa u_{tt} + i\epsilon\eta(z,t)\,, \tag{3.1}$$

where the first two terms on the right hand side are fiber loss and filtering. Here $u(t,z)$ is the complex electromagnetic field envelope, z is the dimensionless distance obtained from the number of round trips, and t is the normalized time, which is a spacelike variable in this formulation. The second and third terms on the left reflect dispersion and self-phase modulation, respectively. The scalar complex random variable $\eta(z,t)$ has identically independently distributed real and imaginary components each distributed as mean-zero Gaussian white noise, delta-correlated in z and t such that

$$\langle \eta(z,t), \eta^*(z',t') \rangle = 2\delta(z-z')\delta(t-t'), \tag{3.2}$$

and

$$\langle \eta(z,t), \eta(z',t') \rangle = 0. \tag{3.3}$$

As alluded to above, equation (3.2) is a mathematical idealization, using the fact that the noise bandwidth is very large relative to the bandwidth of optical pulses in practice [18, 19]. Although only the noise with same spectral range as the soliton itself can impact the soliton directly, the idealization can lead to error in important physical cases.

The simple finite-dimensional reduction is given by

$$\tilde{u}(z,t) = a(z)f(t/T(z))\,\mathrm{e}^{i\phi(z)+i\mu(z)t^2}, \tag{3.4}$$

where f is chosen to be Gaussian to recover an exact soliton in the linear case. This pulse is characterized by its amplitude a, width T, phase ϕ, and "chirp" (a quadratic variation in phase) μ.

Equation (3.1) includes terms that are in the form of a variational gradient and terms that are not variational. We now use a variational technique to reduce the dynamics (3.1) to the approximate 4-dimensional manifold suggested by the parameters in equation (3.4). The Lagrangian density $\mathcal{L}(u, u^*)$ is given by

$$\mathcal{L}(u, u^*) = \mathrm{Im}(u u_z^*) - d|u_t|^2 + \frac{c}{2}|u|^4 \tag{3.5}$$

Integrating the Lagrangian evaluated with equation (3.4) yields average Lagrangian

$$L(a, T, \mu, \lambda) = \int_{-\infty}^{\infty} \mathcal{L}(\tilde{u}, \tilde{u}^*)\, dt. \tag{3.6}$$

To include non-variational terms, we match the non-variational terms by taking variational derivative of L with these terms project onto the tangent manifold of equation (3.4), which is given by

$$\frac{\partial L}{\partial p_j} - \frac{d}{dz}\frac{\partial L}{\partial \dot{p}_j} = 2\mathrm{Re}\int_{-\infty}^{\infty} \mathcal{L} i \left(-\alpha\tilde{u} + \kappa\tilde{u}_{tt} + \epsilon\eta(z,t)\right) \frac{\partial \tilde{u}^*}{\partial p_j}\, dt. \tag{3.7}$$

3.2.1 Derivation of the SDEs

With unit L_2 norm, we choose for f a Gaussian of the form

$$f(x) = \tilde{c}\, \mathrm{e}^{-x^2/2}, \qquad \tilde{c}^2 = 1/\sqrt{\pi}. \tag{3.8}$$

After some algebra, we obtain the averaged Lagrangian

$$L = -\dot{\phi}\, a^2\, T - \frac{1}{2}\dot{\mu}\, a^2\, T^3 - \frac{1}{2} d\, a^2 \left(\frac{1}{T} + 4\mu^2 T^3\right) + \frac{c\, a^4}{2\sqrt{2\pi}} T. \tag{3.9}$$

As an aside, we note that the Euler-Lagrange equations associated with equation (3.9) provide a system that has been well-studied in the context of dispersion-compensated

solitons [20];

$$\begin{aligned}
\dot{T} &= 4d\mu T\,, \\
\dot{\mu} &= \frac{d}{T^4} - 4d\mu^2 - \frac{c}{2\sqrt{2\pi}}\frac{N^2}{T^3} \text{ and,} \\
\dot{\phi} &= -\frac{d}{T^2} + \frac{5c}{4\sqrt{2\pi}}\frac{N^2}{T}
\end{aligned}$$

supplemented by the conservation law $a^2T = N^2$. These reflect the integral nonlinear Schrödinger equation, albeit with a Gaussian pulse form rather than a classic NLS soliton.

Turning to the non-variational terms, we note that the noise contributions take the form

$$\frac{\partial L}{\partial p_j} - \frac{d}{dz}\frac{\partial L}{\partial \dot{p}_j} = 2\epsilon \mathrm{Re}\int i\eta(z,t)\frac{\partial \tilde{u}^*}{\partial p_j}dt, \tag{3.10}$$

$$= -2\epsilon \mathrm{Im}\int \eta(z,t)\frac{\partial \tilde{u}^*}{\partial p_j}dt, \tag{3.11}$$

$$= 2\epsilon\gamma_i(z). \tag{3.12}$$

The covariance of $\gamma_i(z)$ is given by

$$\langle \gamma_i(z), \gamma_i(z')\rangle = \langle \mathrm{Im}\int \eta(z,t)\frac{\partial \tilde{u}^*}{\partial p_j}dt, \mathrm{Im}\int \eta(z',t')\frac{\partial \tilde{u}}{\partial p_j}dt'\rangle, \tag{3.13}$$

$$= \int \frac{\partial \tilde{u}^*}{\partial p_j}\frac{\partial \tilde{u}}{\partial p_j}dt\delta(z-z'). \tag{3.14}$$

From equation (3.4), we have

$$\frac{\partial \tilde{u}^*}{\partial a} = f(t/T(z))\,\mathrm{e}^{-i\phi(z)-i\mu(z)t^2}, \tag{3.15}$$

$$\frac{\partial \tilde{u}^*}{\partial T} = a(z)\frac{t^2}{T^3}f(t/T(z))\,\mathrm{e}^{-i\phi(z)-i\mu(z)t^2}, \tag{3.16}$$

$$\frac{\partial \tilde{u}^*}{\partial \phi} = -ia(z)f(t/T(z))\,\mathrm{e}^{-i\phi(z)-i\mu(z)t^2}, \tag{3.17}$$

$$\frac{\partial \tilde{u}^*}{\partial \mu} = -it^2a(z)f(t/T(z))\,\mathrm{e}^{-i\phi(z)-i\mu(z)t^2}. \tag{3.18}$$

Thus the covariances among the $\gamma_i(z),$ for $i = a, T, \mu, \phi$ are given by

$$\langle \gamma_a(z), \gamma_a(z') \rangle = \int f^2(t/T)\, dt \delta(z - z') = T\delta(z - z'), \tag{3.19}$$

$$\langle \gamma_T(z), \gamma_T(z') \rangle = \int a^2(z) \frac{t^4}{T^6} f^2(t/T)\, dt \delta(z - z') = \frac{3a^2}{4T}\delta(z - z'), \tag{3.20}$$

$$\langle \gamma_\mu(z), \gamma_\mu(z') \rangle = \int a^2(z) t^4 f^2(t/T)\, dt \delta(z - z') = \frac{3}{4} a^2 T^5 \delta(z - z'), \tag{3.21}$$

$$\langle \gamma_\phi(z), \gamma_\phi(z') \rangle = \int a^2(z) f^2(t/T)\, dt \delta(z - z') = a^2 T \delta(z - z'), \tag{3.22}$$

$$\langle \gamma_a(z), \gamma_T(z') \rangle = \frac{a}{2}\delta(z - z'), \tag{3.23}$$

$$\langle \gamma_\mu(z), \gamma_\phi(z') \rangle = \frac{a^2 T^3}{2}\delta(z - z'). \tag{3.24}$$

Combining all the equations from equation (3.10) to equation (3.24), the ODEs for (a, T, μ, ϕ) are given by

$$-2\dot{\phi}\, a\, T - \dot{\mu}\, a\, T^3 - d\, a \left(\frac{1}{T} + 4\mu^2 T^3\right) + \frac{2}{\sqrt{2\pi}} c\, a^3 T = 2\epsilon\gamma_a(z) \tag{3.25}$$

$$-\dot{\phi}\, a^2 - \frac{3}{2}\dot{\mu}\, a^2 T^2 - \frac{1}{2} d\, a^2 \left(-\frac{1}{T^2} + 12\mu^2 T^2\right) + \frac{c\, a^4}{2\sqrt{2\pi}} = 4\kappa\, a^2 \mu + 2\epsilon\gamma_T(z) \tag{3.26}$$

$$-4d\, a^2\, \mu\, T^3 + \frac{1}{2}\frac{d}{dz}\left(a^2 T^3\right) = -\alpha a^2 T^3 + 2\kappa a^2 (\frac{T}{4} - 3\mu^2 T^5) + 2\epsilon\gamma_\mu(z) \tag{3.27}$$

$$\frac{d}{dz}\left(a^2 T\right) = -\kappa a^2 \frac{1}{T} - 4\kappa\mu^2 T^3 a^2 - 2\alpha a^2 T + 2\epsilon\gamma_\phi(z), \tag{3.28}$$

which can be simplified to

$$\dot{a} = -2ad\mu - \alpha a - \frac{\kappa a}{T^2} - \frac{\epsilon}{aT^3}\gamma_\mu(z) + \frac{3\epsilon}{2aT}\gamma_\phi(z), \tag{3.29}$$

$$\dot{T} = 4d\mu T + \frac{\kappa}{T} - 4\kappa\mu^2 T^3 + \frac{2\epsilon}{a^2 T^2}\gamma_\mu(z) - \frac{\epsilon}{a^2}\gamma_\phi(z), \tag{3.30}$$

$$\dot{\mu} = \frac{d}{T^4} - 4d\mu^2 - \frac{ca^2}{2\sqrt{2\pi}T^2} - \frac{4\kappa\mu}{T^2} - \frac{2\epsilon}{a^2 T^2}\gamma_T(z) + \frac{\epsilon}{aT^3}\gamma_a(z), \tag{3.31}$$

$$\dot{\phi} = -\frac{d}{T^2} + \frac{5ca^2}{4\sqrt{2\pi}} + 2\kappa\mu - \frac{3\epsilon}{2aT}\gamma_a(z) + \frac{\epsilon}{a^2}\gamma_T(z), \tag{3.32}$$

where the covariance for each $\gamma_i(z), i = a, T, \mu, \phi$ is given by

$$\langle \gamma_a(z), \gamma_a(z') \rangle = \int f^2(t/T)\, dt \delta(z - z') = T\delta(z - z'), \tag{3.33}$$

$$\langle \gamma_T(z), \gamma_T(z') \rangle = \int a^2(z) \frac{t^4}{T^6} f^2(t/T)\, dt \delta(z - z') = \frac{3a^2}{4T}\delta(z - z'), \tag{3.34}$$

$$\langle \gamma_\mu(z), \gamma_\mu(z') \rangle = \int a^2(z) t^4 f^2(t/T)\, dt \delta(z - z') = \frac{3}{4} a^2 T^5 \delta(z - z'), \tag{3.35}$$

$$\langle \gamma_\phi(z), \gamma_\phi(z') \rangle = \int a^2(z) f^2(t/T)\, dt \delta(z - z') = a^2 T \delta(z - z'), \tag{3.36}$$

$$\langle \gamma_a(z), \gamma_T(z') \rangle = \frac{a}{2}\delta(z - z'), \tag{3.37}$$

$$\langle \gamma_\mu(z), \gamma_\phi(z') \rangle = \frac{a^2 T^3}{2}\delta(z - z'). \tag{3.38}$$

Thus the covariance matrix, $\Sigma = \sigma\sigma^T$, is given by

$$\Sigma = \begin{bmatrix} \frac{3}{2T} & -\frac{1}{a} & 0 & 0 \\ -\frac{1}{a} & \frac{2T}{a^2} & 0 & 0 \\ 0 & 0 & \frac{2}{a^2T^5} & -\frac{1}{a^2T^3} \\ 0 & 0 & -\frac{1}{a^2T^3} & \frac{3}{2a^2T} \end{bmatrix}. \tag{3.39}$$

Using Cholesky factorization, we obtain

$$\sigma = \begin{bmatrix} \sqrt{\frac{3}{2T}} & 0 & 0 & 0 \\ -\frac{1}{a}\sqrt{\frac{2T}{3}} & \sqrt{\frac{4T}{3a^2}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{2}{a^2T^5}} & 0 \\ 0 & 0 & -\frac{1}{a\sqrt{2T}} & \sqrt{\frac{1}{a^2T}} \end{bmatrix}. \tag{3.40}$$

Thus the reduced ODEs of $\mathbf{x} = (a, T, \mu, \phi)$ are given by

$$\dot{\mathbf{x}} = F(\mathbf{x}) + \epsilon\sigma(x)\dot{\mathbf{w}} \tag{3.41}$$

where

$$F(\mathbf{x}) = \begin{pmatrix} -2ad\mu - \alpha a - \frac{\kappa a}{T^2} \\ 4d\mu T + \frac{\kappa}{T} - 4\kappa\mu^2 T^3 \\ \frac{d}{T^4} - 4d\mu^2 - \frac{ca^2}{2\sqrt{2\pi}T^2} - \frac{4\kappa\mu}{T^2} \\ -\frac{d}{T^2} + \frac{5ca^2}{4\sqrt{2\pi}} + 2\kappa\mu \end{pmatrix} \tag{3.42}$$

and $\sigma(\mathbf{x})$ is defined as matrix (3.40) with $\mathbf{w} = [W_\mu, W_\phi, W_a, W_T]$.

3.2.2 Verification of reduced system

In this section, we compare the PDE system (3.1) and its reduced ODE system (3.40)-(3.42) numerically for different values of parameters. The ODE system is solved using a fourth order Runge-Kutta method with $\Delta z = 0.01$ and the PDE system is solved using a pseudospectral method in t with N Fourier modes and a fourth-order Runge-Kutta method with $\Delta z = 0.01$.

We will first compare the pathwise solutions of the PDE system and the ODE system without noise. Then we will compare their solutions with noise both pathwise and in the sense of distribution.

Pathwise comparison of the deterministic PDE and ODEs. In order to compare the solutions between the deterministic PDE and ODEs, we extract the parameters (a, T, μ, ϕ) from the numerical solutions of the deterministic PDE using the following identities

$$a^2 T = \int_{-\infty}^{\infty} |u|^2 dt, \tag{3.43}$$

$$a^2 \frac{T^3}{2} = \int_{-\infty}^{\infty} t^2 |u|^2 dt, \tag{3.44}$$

$$a^2 (\frac{1}{2T} + 2\mu^2 T^3) = \int_{-\infty}^{\infty} |u_t|^2 dt, \tag{3.45}$$

$$a^2 T\phi = \int_{-\infty}^{\infty} \arctan(\frac{\Im(u \exp(-i\mu t^2))}{\Re(u \exp(-i\mu t^2))}) |u|^2 dt. \tag{3.46}$$

In practice, the lower limit and the upper limit of the integrals are replaced by $-L/2$ and $L/2$ and the integrals are approximated by the composite trapezoidal rule.

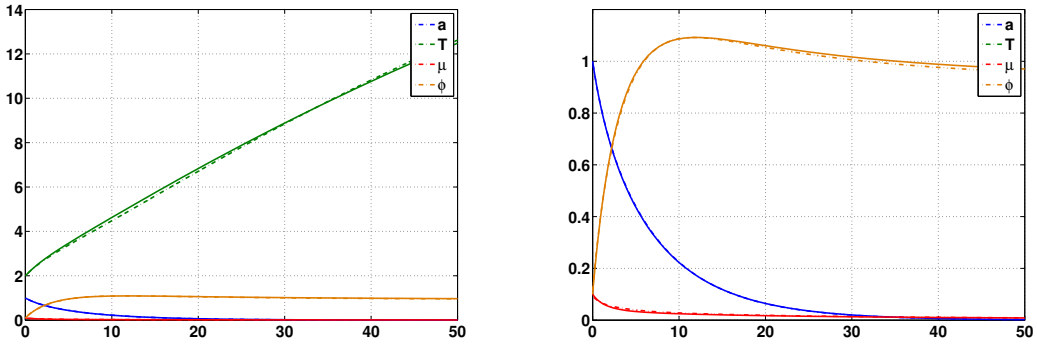

Figure 3.1 Pathwise comparison of the deterministic PDE and ODEs. The dashed lines denote the PDE solution and the solid lines denote the ODE solutions.

Figure 3.1 shows the pathwise comparison of the deterministic PDE solution and the ODE solution. The initial point of the path starts at $(1, 2, 0.1, 0.1)$ with parameters $\alpha = \kappa = 0.1, d = 1/2$ and $c = 1$. We see good agreement between the two solutions. The evolutions of the parameters for amplitude a, phase ϕ, and chirp μ for both systems are almost the same. There is a small difference between the evolutions of the width T.

To have a better understanding for which parameters loss α and filtering κ the two models agree well, we compare their solutions for different values of α and κ with fixed $d = 1/2$ and $c = 1$.

Figure 3.2 shows the pathwise comparison of the deterministic PDE solution and the ODE solution. The initial point of the path starts at $(1, 2, 0.1, 0.1)$ with parameter $\kappa = 0.1$ and different values of α, ranging from 0 to 0.1. The results indicate that the two models agree better with stronger fiber loss.

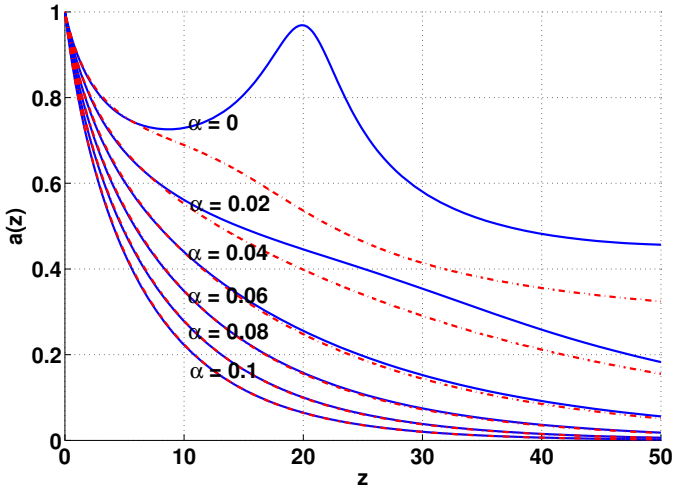

Figure 3.2 Effect of fiber loss on agreement between the PDE and ODE models. Dashed lines denote the PDE solution, and solids lines denote the ODE solution.

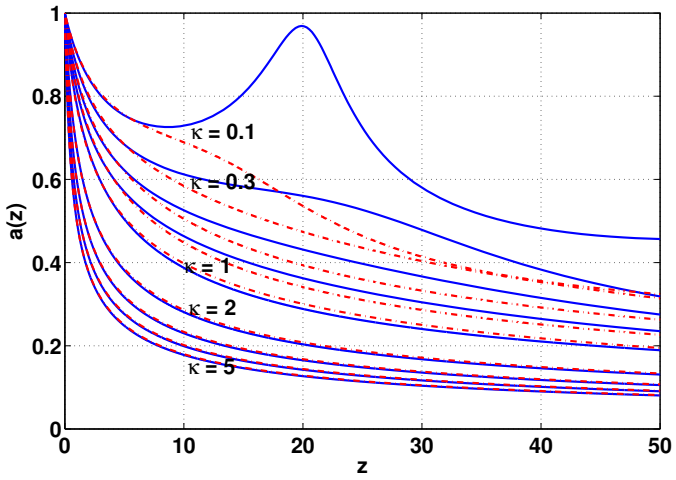

Figure 3.3 Effect of filtering on agreement between the PDE and ODE models. Dashed lines denote the PDE solution, and solids lines denote the ODE solution.

Figure 3.3 shows the pathwise comparison of the deterministic PDE solution and the ODE solution. The initial point of the path starts at $(1, 2, 0.1, 0.1)$ with parameters $\alpha = 0$ and different values of κ, ranging from 0.1 to 5. The results indicate that the two models agree better with stronger filtering.

Figure 3.1 and Figure 3.3 show that the agreement for the evolution of amplitude between the deterministic PDE system and the deterministic ODE system improves as α and κ increases.

Pathwise comparison of the stochastic PDE and ODEs. In order to compare the stochastic PDE and ODEs, we first generate a random process for the PDE system then project the noise $\eta(z,t)$ onto the corresponding derivative of the ansatz using equations (3.10). The data below is generated with $\alpha = 0.1, \kappa = 1, d = 1/2, L = 100$ and $c = 1$, with initial point $(0.2675, 2.8915, 0.0218, 0.2648)$. The discretization in t of the PDE uses $N = 412$. The L_2 error is defined by

$$L_2(u_p, u_o) = \int_{-L/2}^{L/2} |u_p - u_o|^2 dt, \tag{3.47}$$

where u_p is the solution of the PDE system and u_o is the solution of the ODE system.

Figure 3.4 compares the deterministic PDE and ODE systems. The parameters and initial point are chosen to have a small L_2 error, which is our starting point to include the effect of noise of both systems. The results in Figure 3.5 are generated using the same parameters and initial point as Figure 3.4, except including the noise with strength $\epsilon = 10^{-5}$. Figure 3.5 shows that the evolutions of all the characteristic parameters of both systems agree well, except for the phase of ϕ. After $z = 30$, the phase of the solution from the PDE model decreases to 0 very fast while the phase of the ODE remains between 0.25 and 0.3. A similar trend also appears in the evolution of the graph of L_2, where the error stops decreasing. Interestingly, even though the difference in the evolution of T becomes larger after $z = 40$, their fluctuations are very similar. Despite the difference between the frequencies of both models, their solitons at the end of the simulation when $z = 50$ stay the same shape and at the same center.

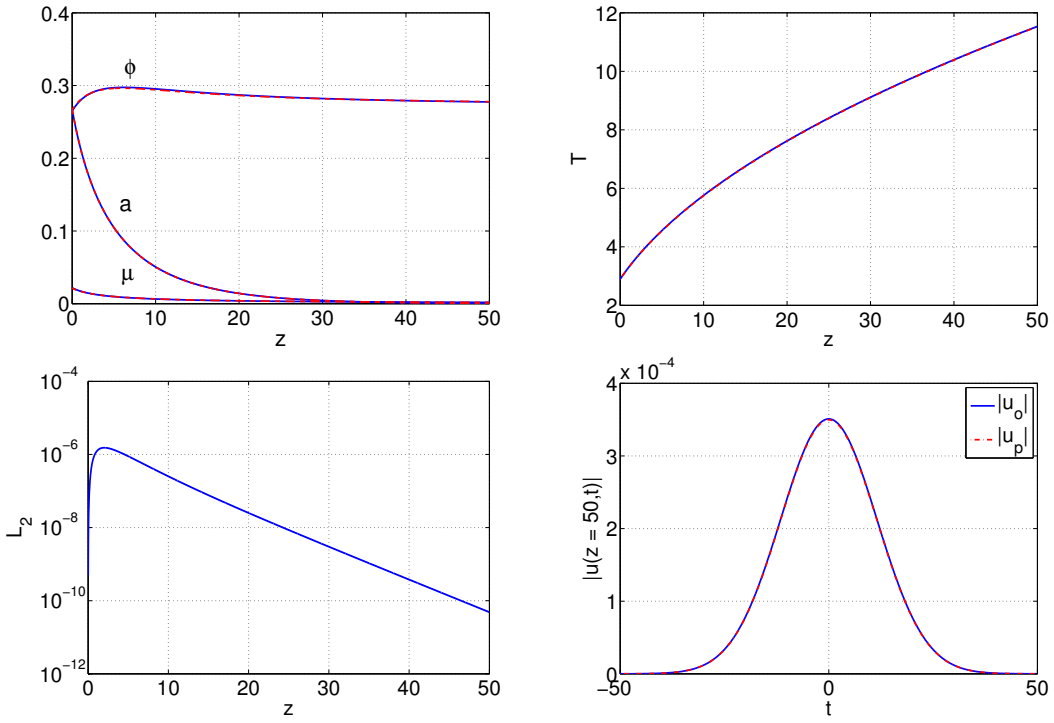

Figure 3.4 Pathwise comparison of the deterministic PDE and ODEs. Dashed lines denote the PDE solution and solids lines denote the ODE solutions. The bottom left figure shows the pulse at $z = 50$.

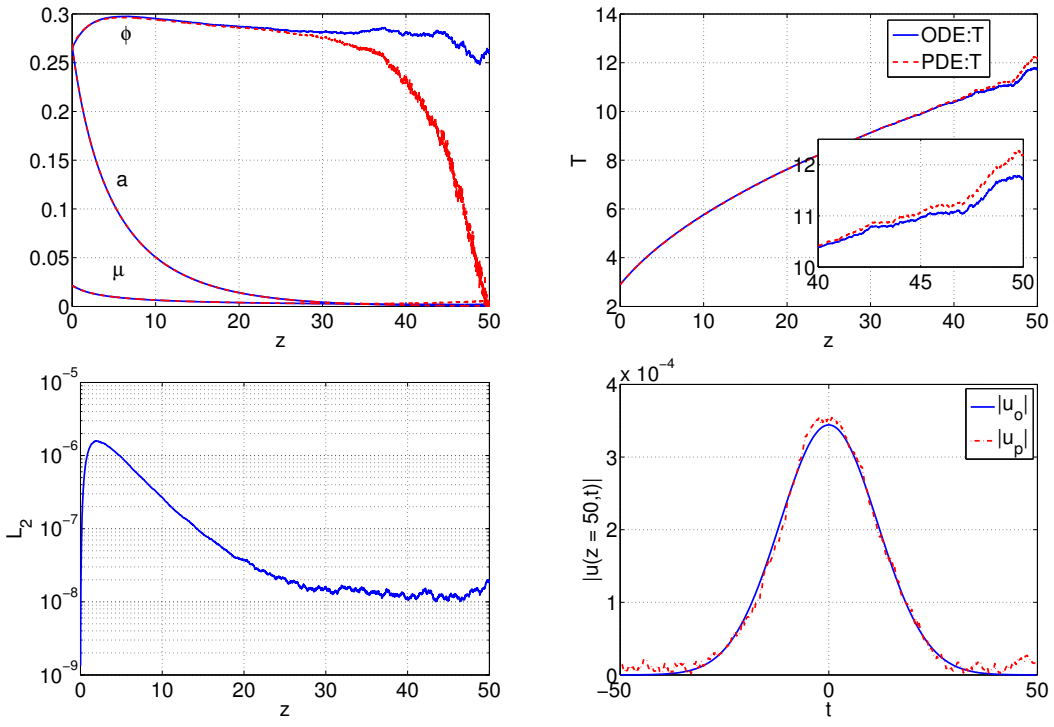

Figure 3.5 Pathwise comparison of the stochastic PDE and ODEs with $\epsilon = 10^{-5}$. Dash lines denote PDE solution and solid lines denote the ODE solution.

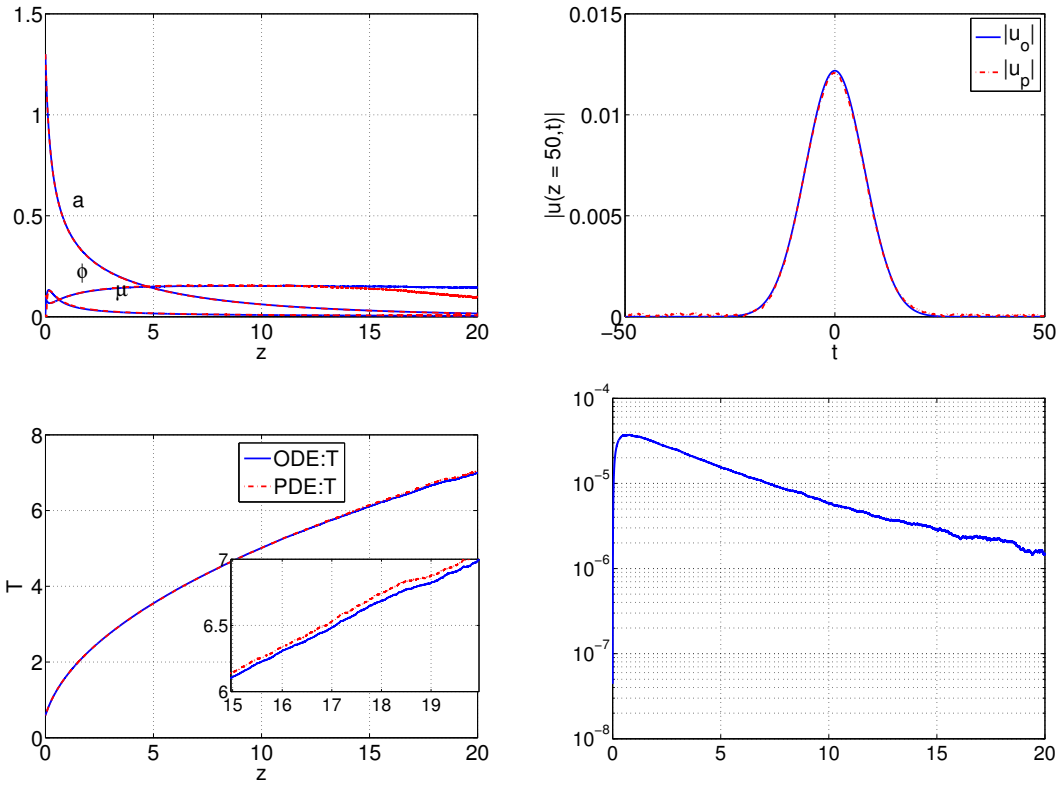

Figure 3.6 Pairwise comparison between stochastic PDE and its reduced stochastic ODEs with $\epsilon = 10^{-4}, N = 10^3,$ and $\Delta z = 0.0013$.

Figure 3.6 compares the two models with a larger initial value of $a(0) = 1.3$ As a result, the two systems experience a stronger nonlinear effect. The results in this figure are generated with $\alpha = 0.1, \kappa = 1, d = 1/2, L = 100$ and $c = 1$ with initial point $(1.3, 1/1.3^2, 0.01, 0.1)$. The discretization in t of the PDE uses $N = 412$.

Simulations of stochastic PDE and stochastic ODEs for comparison in the sense of distributions. To compare the stochastic PDE and stochastic ODEs, we study the time evolution of the variances of the four parameters. The simulation results are generated with $\alpha = 0.1, \kappa = 1, d = 1/2, L = 100,$ and $c = 1,$ and with initial point $(0.2675, 2.8915, 0.0218, 0.2648)$ and $\epsilon = 10^{-6}$. The discretization in t of the PDE uses $N = 412$.

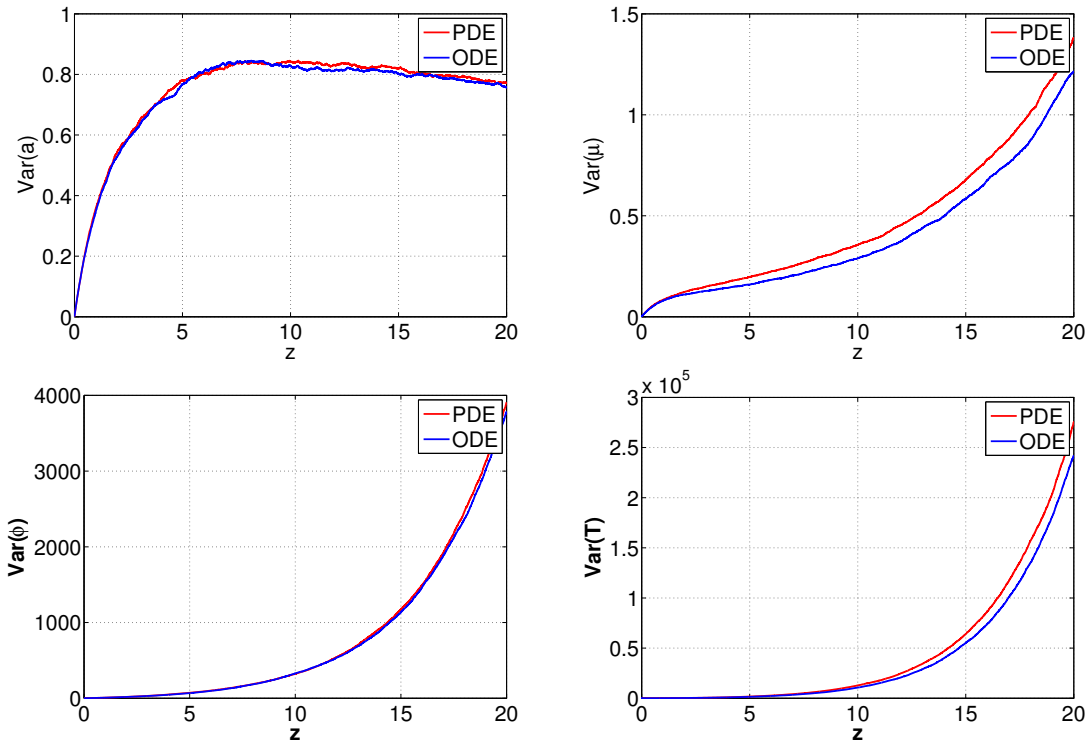

Figure 3.7 Evolution of the variances for each characteristic parameter.

Figure 3.7 shows the time evolution of the variances for each parameter of the stochastic PDE model and the stochastic ODE model. The variances of the parameters for both models stay close over time, but there is a significant difference between the variance of the width T for the PDE model and that of the ODE model. The evolution of the width T for the stochastic model has a larger variance than that of the stochastic ODE model. We can observe a similar pattern for the variance of μ.

3.2.3 Application of the gMAM with variable endpoint

Before applying the gMAM to compute optimal trajectories in our reduced stochastic ODE system, we first run a test to ensure that the path computed by gMAM recovers a deterministic trajectory when one exists in the admissible set corresponding to the applied boundary conditions. The trajectory is first obtained numerically with

starting point $(1.3, 1/1.3^2, 0.01, 0.1)$ at $z = 0$, and the computed end point at $z = 20$ found to be $(0.0161, 7.0721, 0.0047, 0.1462)$. The parameters for the ODE system are given by $\alpha = 0.1, \kappa = 1, d = 1/2, L = 100$,and $c = 1$. A fixed initial point and a varied terminal point for the gMAM are $(1.3, 1/1.3^2, 0.01, 0.1)$ and $(0.0161, T, 0.0047, 0.1462)$, where $T = 2$ in used as an initial value. The value for T from the gMAM path is 7.0739, agreeing well with the precomputed action (See Figure 3.8) .

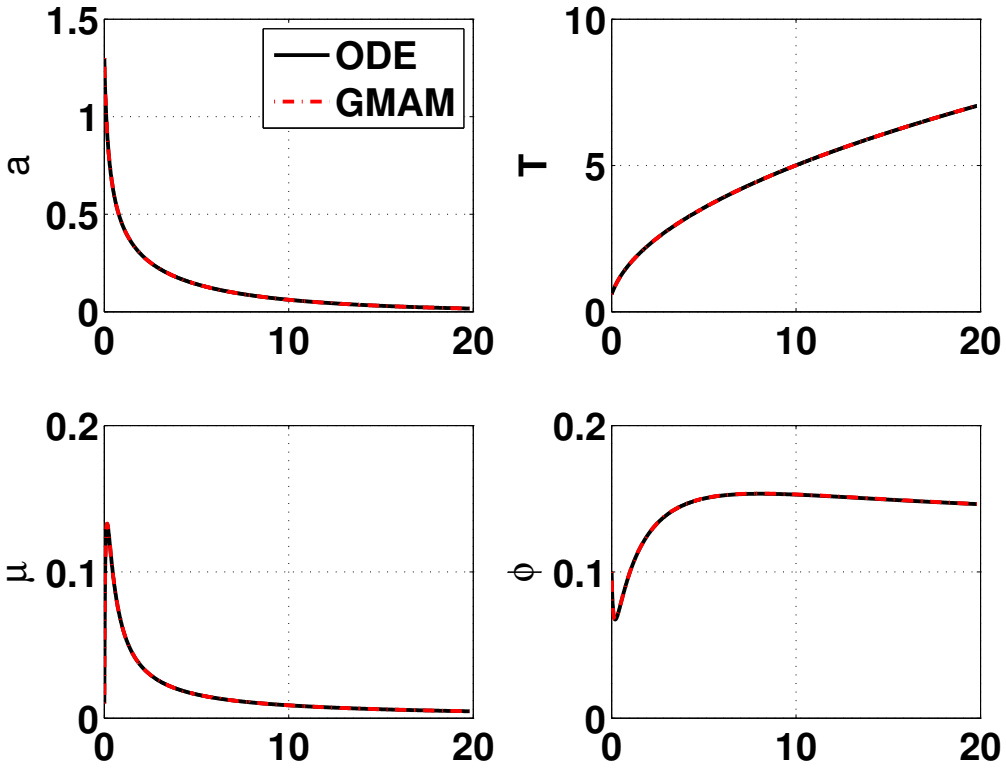

Figure 3.8 Comparison of gMAM path and deterministic ODE path.

Figure 3.8 compares the gMAM path with the deterministic ODE path. Black lines denote the deterministic trajectory of the ODEs. Red dashed lines denotes the gMAM path. The figures show that the gMAM path is indistinguishable from the deterministic trajectory of the ODEs.

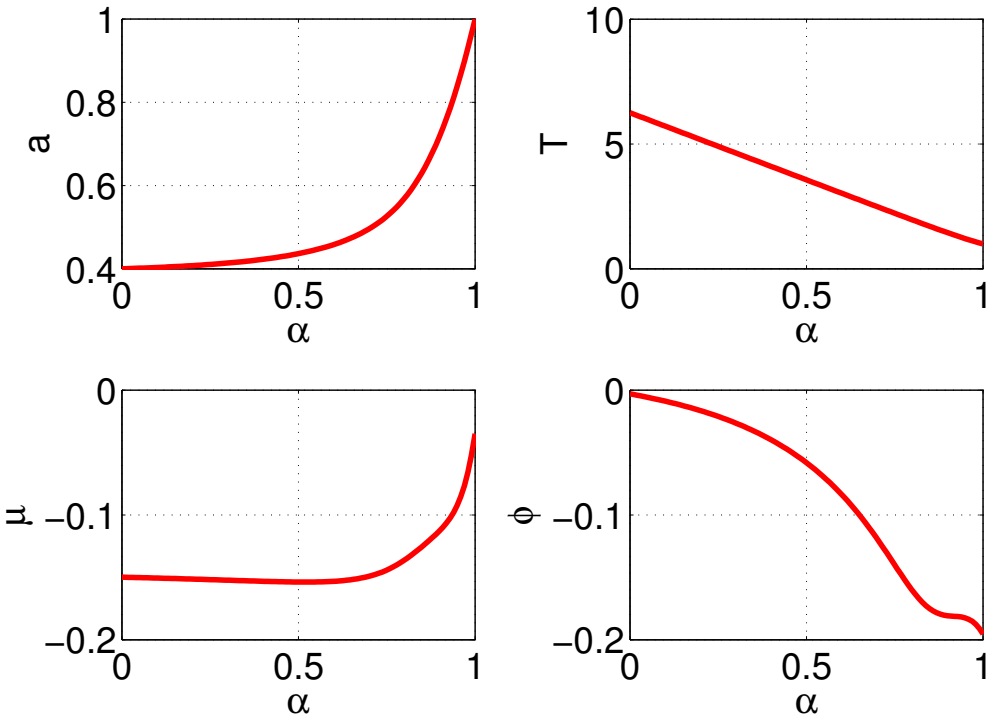

Figure 3.9 The computed starting point is $(0.4, 6.25, -0.1499, -0.003)$ and the ending point is $(1, 1, -0.03, -0.1954)$. The action is 0.443.

In Figure 3.9, we consider a situation where the amplitude of the soliton is below 0.5 such as $a = 0.4$, and we are interested in the transition of the amplitude from $a = 0.4$ to $a = 1$. As the amplitude transits from $a = 0.4$ to $a = 1$, the width T decreases but system does not follow the conservation law a^2T = constant (See Figure 3.10).

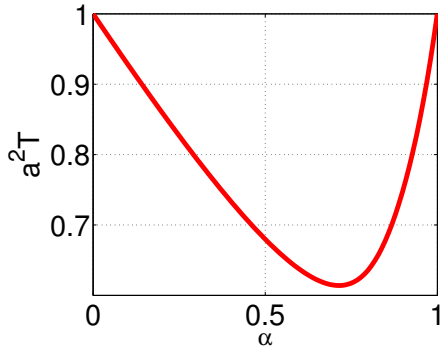

Figure 3.10 The system does not follow the conservation law.

3.3 A Mode-Locked Laser with Active Feedback

We now study the quasi-stable dynamics of a mode-locked laser with active feedback and noise due to amplified spontaneous emission. We show that, in a distinguished small-noise limit, an effective boundary can be drawn in parameter space for quasi-stability that is distinct from the deterministic stability boundary. We consider the probability that a mode-locked laser with active feedback will experience a transition between stable equilibria in a potential well when subjected to amplified spontaneous emission noise generated by the gain medium. To investigate the influence of noise on quasi-stability, we reduce the infinite-dimensional model to a finite-dimensional system of stochastic ordinary differential equations and compute the quasi-stable state by evaluating the action functional via the geometric minimum action method. This computation shows how and to what extent noise effectively destabilizes the system, producing a region of quasi-stability in its parameter space that is smaller than that of the deterministic system. We also estimate the probability of pulse position shifts using importance sampling.

We consider a mode-locked fiber laser with a nonlinear gain element that controls the pulse amplitude and active phase modulation that controls the pulse position. It is also natural to consider the dynamical system evolving in time, so we write in in a form commonly used by mathematicians,

$$\begin{aligned} iu_t + \frac{1}{2}u_{xx} + |u|^2u &= -b\cos(\omega x)u - ic_1u \\ &+ ic_2u_{xx} + id_1|u|^2u - id_2|u|^4u + i\epsilon\eta(x,t), \end{aligned} \tag{3.48}$$

where $u(x,t)$ is the electric field envelope, u_{xx} represents filtering, $\cos(\omega x)u$ is the active phase modulation, and $-ic_1u + id_1|u|^2u - id_2|u|^4u$ represents linear loss and saturable gain. The noise process $\eta(x,t)$ is assumed to derive from spontaneous emission noise and assumed to be mean-zero Gaussian white noise, delta-correlated in x and t [1, 21]. Equation (3.48) with trivial right-hand side has a soliton solution

of the following form:

$$u_s(x,t) = A(t)\,\mathrm{sech}[A(t)(x - X(t))]\exp[i\phi], \tag{3.49}$$

where $\phi(x,t) = \varphi(t) + (x - X(t))\Omega(t)$. The parameters A, X, φ, and Ω represent amplitude, position, phase and frequency, respectively.

3.3.1 A variational approach for model reduction

After application of the same variational technique discussed in section 3.2 , we obtain a set of four stochastic ordinary differential equations(SODE) for the four soliton parameters:

$$dU = F(U)dt + \epsilon\sigma(U)dW, \tag{3.50}$$

where $U = (A, \Omega, X)^T$,

$$F(U) = \begin{pmatrix} -2c_1u_1 + (\frac{4}{3}d_1 - \frac{2}{3}c_2)u_1^3 - \frac{16}{15}d_2u_1^5 - 2c_2u_1u_2^2 \\ -\frac{4}{3}c_2u_1^2u_2 - \frac{\pi bw^2}{2u_1^3}\,\mathrm{csch}(\frac{\pi w}{2u_1})\sin(wu_3) \\ u_2 \end{pmatrix}, \tag{3.51}$$

and

$$\sigma(U) = \begin{pmatrix} \sqrt{u_1} & 0 & 0 \\ -\frac{u_2}{\sqrt{u_1}} & \sqrt{\frac{u_1}{3} + \frac{u_2^2}{u_1}} & 0 \\ 0 & 0 & \frac{\pi}{\sqrt{12u_1^3}} \end{pmatrix}. \tag{3.52}$$

The phase evolution is not included in the above dynamical system since it does not affect the dynamics. It is governed by

$$d\varphi = \frac{1}{2}(u_1^2 + u_2^2)dt - \frac{\pi\omega b}{u_1^3}\cos(\omega u_3)\,\mathrm{csch}(\frac{\pi\omega}{2u_1})dt$$
$$+ \frac{\pi^2\omega^2 b}{4u_1^4}\cos(\omega u_3)\,\mathrm{csch}(\frac{\pi\omega}{2u_1})\coth(\frac{\pi\omega}{2u_1})dt$$
$$- \epsilon\sqrt{\frac{12+\pi^2}{36u_1}}dW_4 + \epsilon\frac{\pi u_2}{u_1\sqrt{12u_1}}dW_3. \tag{3.53}$$

3.3.2 Linearization and stability

Considering $\epsilon = 0$, the fixed points of $F(U)$ are

$$U_n = (A_{0\pm}, 0, n\pi/\omega),$$

$n \in \mathbb{Z}$ and

$$A_{0\pm}^2 = \frac{5}{16d_2}\left[2d_1 - c_2 \pm \sqrt{(2d_1 - c_2)^2 - \frac{96}{5}c_2 d_2}\right],$$

provided $(2d_1 - c_2)^2 - \frac{96}{5}c_2 d_2 > 0$. Recalling that all the physical parameters of equation (3.48) are positive, the fixed points with A_{0+} are stable if n is even, otherwise they are unstable. The fixed points with A_{0-} are always saddles. The stable fixed points are nodes if $(\frac{4}{3}c_2 A_{0+}^2)^2 > \frac{2\pi b\omega^3}{A_{0+}^3}\,\mathrm{csch}(\frac{\pi\omega}{2A_{0+}})$; otherwise, they are spirals. The eigenvalues of the fixed points are given by

$$\lambda_{U_1} = 8c_1 - \frac{4}{3}(2d_1 - c_2)A_{0+}^2,$$
$$\lambda_{U_2} = \frac{1}{2}(M_{22} - \sqrt{M_{22}^2 + 4M_{23}}),$$
$$\lambda_{U_3} = \frac{1}{2}(M_{22} + \sqrt{M_{22}^2 + 4M_{23}}),$$

where $M_{22} = -4/3c_2 A_{0+}^2$ and

$$M_{23} = (-1)^{n+1}\pi b\omega^3/(2A_{0+}^3)\,\mathrm{csch}(\pi\omega/(2A_{0+})).$$

3.3.3 Paths for amplitude drop-outs and pulse position shifts

Although equation (3.48) has a nonlinear gain element to control the pulse amplitude and an active phase modulation to control the pulse position, it suffers from amplitude drop-outs and pulse position shifts due to ASE noise.

The parameters defined in Table 3.1 are used to compute actions and minimum action paths in Figure 3.11, and are used for MC simulation in Figure 3.12.

Table 3.1 Parameters for an Actively Mode-locked Fiber Laser Model

Parameter	ω	c_1	c_2	d_1	d_2	b
Value	4	1	1	5	1.5	4.45

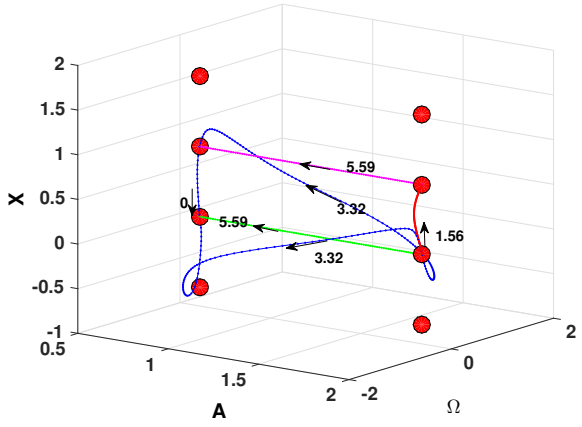

Figure 3.11 Transition paths with different starting points and different endpoints and their corresponding actions.

Figure 3.11 shows different transition paths with their corresponding action. For any trajectory starting at stable fixed point $(A_{0+}, 0, 0)$, the most likely transition is the minimum action path connecting $(A_{0+}, 0, 0)$ and $(A_{0+}, 0, \pm\pi/\omega)$. In other words, an error occurs mostly likely due to pulse position shifts rather than amplitude drop-outs.

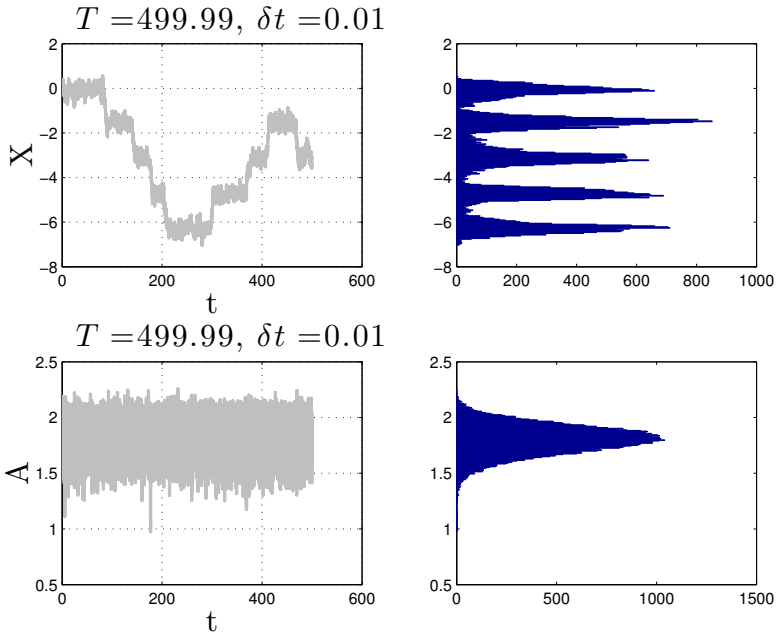

Figure 3.12 MC simulation of stochastic ODEs using parameters defined in Table 3.1.

Figure 3.12 shows that the system suffers from position shifts more often than amplitude drop-outs. The figures on the left are the time evaluations of amplitude and position. The figures on the right are the histograms of amplitude and position over all time steps. The MC simulation uses the parameters defined in Table 3.1 with noise strength $\epsilon = 0.65$, which is chosen to be large so that the rare events occur more often. During the simulation, the amplitude stays close to A_0^+, while the position spends most of its time near a stable equilibrium with brief transits between them.

3.3.4 Quasi-stability and large deviation theory

To examine the effect of noise on quasi-stability of the mode-locked laser we focus on the active feedback parameters (ω, b) and we fix the other parameters. The fixed points with even n are stable in the first quadrant of parameter plane (ω,b). Notice that when $b \to 0$ we have $\lambda_{U_2} = \lambda_{U_3} \to 0$. Meanwhile as $\omega \to \infty, U_n =$

$(A_{0+}, 0, n\pi/\omega) \to (A_{0+}, 0, 0)$. These limits confirm the intuition that the susceptibility of the state to undergo a transition is affected by b and ω.

In presence of noise $\epsilon > 0$, any trajectory that starts at stable fixed point $(A_{0+}, 0, 0)$, will almost surely exit the basin of attraction and enter the basin of attraction of $(A_{0+}, 0, \pm 2\pi/\omega)$. Large deviation theory [12] states that the probability of exiting the basin of attraction G from a stable fixed point before finite T is given by

$$\lim_{\epsilon \to 0} \epsilon^2 \ln P(t \leq T) \sim - \inf_{\phi(T) \notin G} S_T(\phi) \tag{3.54}$$

and that the mean first exit time τ is given by

$$\lim_{\epsilon \to 0} \epsilon^2 \ln E\tau \sim \inf_{\phi(\tau) \notin G} S_\infty(\phi) \tag{3.55}$$

where

$$S_T(\phi) = \frac{1}{2} \int_0^T |\sigma^{-1}(\phi)(\dot{\phi} - F(\phi))|^2 dt. \tag{3.56}$$

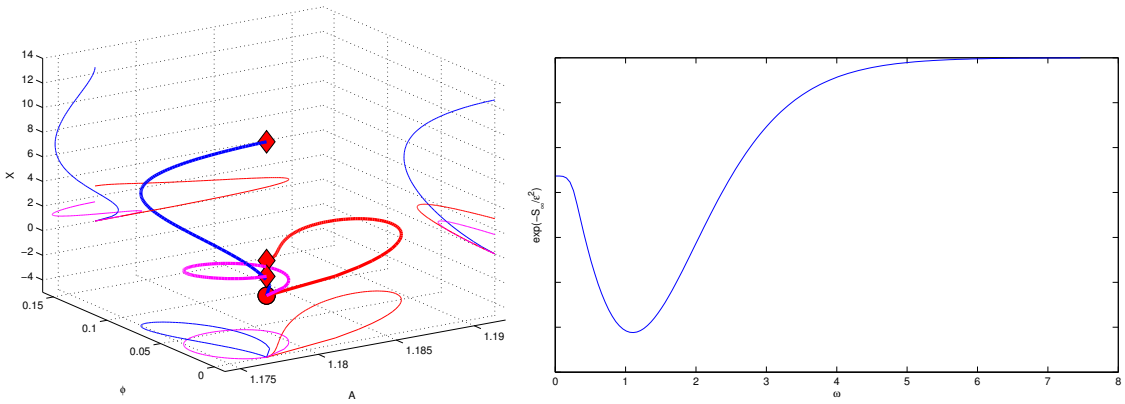

Figure 3.13 Three different optimal paths corresponding to three different values of ω (left). There is a unique ω that minimizes the transition rate, i.e., maximizes the mean exit time (right).

Equations (3.54) and (3.55) suggest that as $T \to \infty$, $P(t \leq T) \to 1$ almost surely. Meanwhile, for any T and $\epsilon = 0$, the transition event does not occur. By

choosing $T_\epsilon = \exp(A/\epsilon^2)(B + O(\epsilon))$, where $A, B > 0$ and $\tau \sim \nu \exp(S/\epsilon^2)$, the probability of an exit before T_ϵ is given by

$$\begin{aligned} P &\sim 1 - \exp(-T_\epsilon/\tau) \\ &= 1 - \exp(-\exp((A-S)/\epsilon^2)\frac{B}{\nu}). \end{aligned}$$

If $A = S$, we have that P remains bounded away from 0 and 1 as $\epsilon \to 0$. If $A > S$, we have $P \to 1$ as $\epsilon \to 0$.If $A < S$, we have $P \to 0$ as $\epsilon \to 0$. Thus, in a manner related to the finite-temperature astroids studied for magnetic materials in [22], this distinguished limit of vanishing noise strength provides a quasi-stable region with a boundary inside the deterministic stability boundary.

The physical parameters defined in Table 3.2 are used to compute the action in Figure 3.14, as well as the probability for pulse position shifts in section (3.3.5).

Table 3.2 Typical Parameters for an Actively Mode-locked Fiber Laser Model [1]

Parameter	Value	Units
ω	$2\pi/25$	ps^{-1}
b	0.01	km^{-1}
c_1	0.01	km^{-1}
c_2	0.002	ps^2/km
d_1	0.034	$(W/km)^{-1}$
d_2	0.02	$(W^2/km)^{-1}$

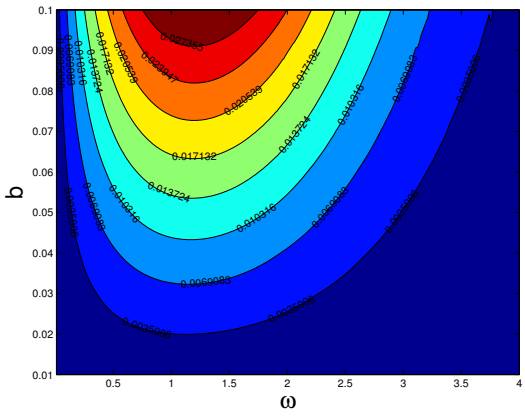

Figure 3.14 Contours of the action functional in the active feedback parameter space of (b, ω).

Figure 3.14 shows contours of the minimum action along paths connecting starting point $(A_{0+}, 0, 0)$ and ending point at $(A_{0+}, 0, \pi/\omega)$ for various values of amplitude and frequency of the active modulation. The larger the value of the action, the more difficult it is for the system to undergo a transition from the basin of attraction of $(A_{0+}, 0, 0)$ to the basin of attraction of $(A_{0+}, 0, 2\pi/\omega)$ or $(A_{0+}, 0, -2\pi/\omega)$.

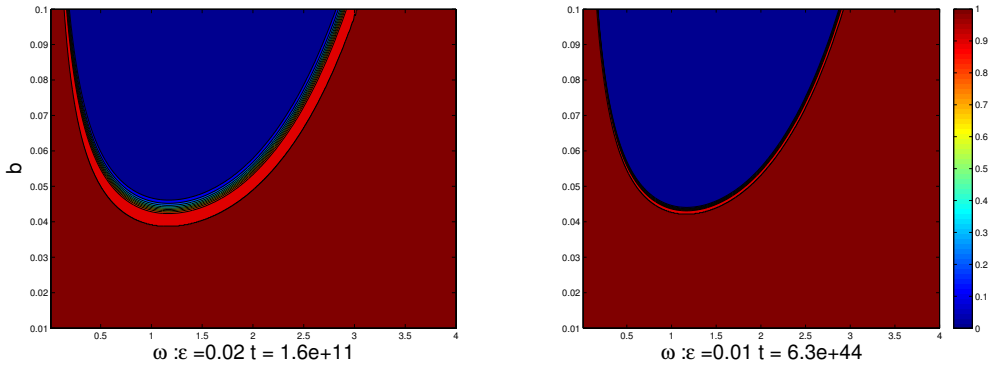

Figure 3.15 Quasi-stable region of exit probability before T_ϵ with $A = 0.010316$. For small ϵ, the transition region shrinks to a narrow band, which converges to one of contours in Figure 3.14.

Figure 3.15 shows the quasi-stable region in the parameter space of (b, ω). The parameter space can be divided into two major regions and a narrow region

in terms of the probability of exit before T_ϵ. The blue region indicates vanishingly small probability of exit before T_ϵ, while the red region indicates overwhelmingly large probability of exit before T_ϵ. The probability of exit before T_ϵ in the exponentially narrow region connecting the red region and the blue region is between 0 and 1. By choosing T_ϵ to be the order of the mean first exit time of the system as ϵ goes to zero, the narrow region becomes one of the contours in Figure 3.14.

3.3.5 Importance sampling for probability of pulse position shifts

Here we demonstrate the application of importance sampling to compute the probability $P(t \leq T)$ of pulse position shifts before time T. In our simulation, all the trajectories start at an stable fixed point $(A_{0+}, 0, 0)$ and errors occur due to pulse position shifts whenever a trajectory exits the basin of attraction of the stable fixed point $(A_{0+}, 0, 0)$ and enters the basin attraction of $(A_{0+}, 0, \pm 2\pi/\omega)$ before time T. The exit criterion is determined by two planes, which are spanned by the two stable directions of the two different saddle points $(A_{0+}, 0, \pm\pi/\omega)$ (See figure (3.16)).

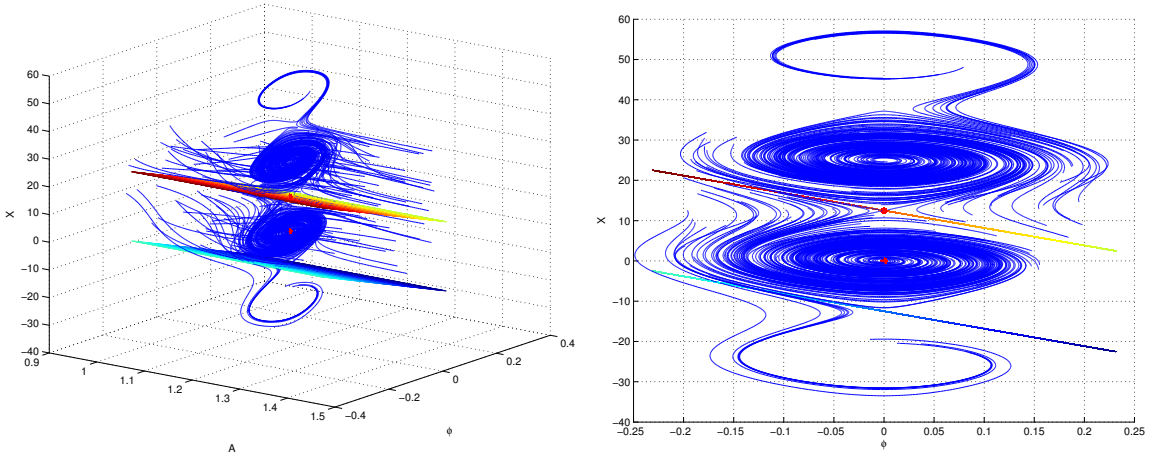

Figure 3.16 The blue curves are trajectories with random initial points. The red dots are fixed points of the system. The two tangent planes are spanned by two stable directions of two different saddle points.

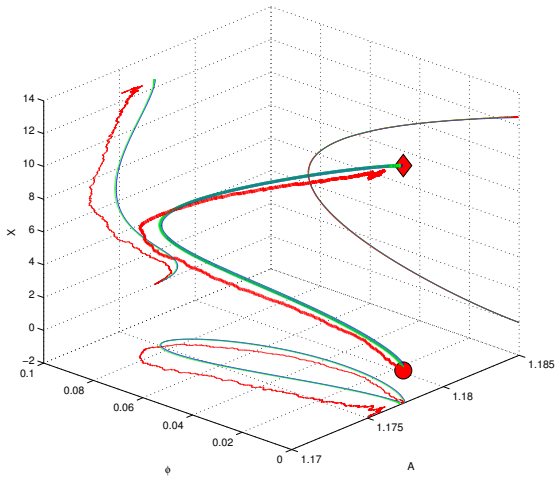

Figure 3.17 Mean exit paths from importance sampling.

Figure (3.17) shows the mean exit paths from importance sampling. The red curve and the green curve are the mean exit paths from IS with $\epsilon = 10^{-3}$ and $\epsilon = 10^{-4}$, respectively. The blue curve is the minimum action path connecting the stable fixed point $(A_{0+}, 0, 0)$ and the saddle point $(A_{0+}, 0, \pi/\omega)$.

Table 3.3 Probability of Pulse Position Shifts before $T = 180$ Generated by IS with Sample Size of 10^5

ϵ	$P_{\text{IS}}(t \leq 180)$	CV
0.05	9.50×10^{-1}	1.61×10^{-3}
0.02	2.79×10^{-1}	2.44×10^{-2}
0.01	1.46×10^{-3}	2.68×10^{-2}
0.008	4.95×10^{-5}	1.96×10^{-2}
0.006	3.95×10^{-8}	3.07×10^{-2}
0.004	6.56×10^{-17}	2.30×10^{-2}
0.002	1.75×10^{-63}	1.03×10^{-1}

Table 3.3 shows the finite probability of pulse position shifts before $T = 180$ for differential values of ϵ. With a sample size of 10^5, IS allows us to sample probability

as low as 10^{-63} for $\epsilon = 0.002$. For ϵ between 0.02 and 0.04, the coefficient of variations are around 10^{-2}, which imply the estimators achieve relatively high accuracy.

Table 3.4 Probability of Pulse Position Shifts before $T = 180$ Generated by MC with $\epsilon = 0.1$ and Sample Size of 10^5

ϵ	$P_{\mathrm{MC}}(t \leq 180)$	CV
0.05	9.52×10^{-1}	7.13×10^{-4}
0.02	2.77×10^{-1}	5.10×10^{-3}
0.01	2.83×10^{-3}	5.94×10^{-2}
0.008	1.85×10^{-4}	2.35×10^{-1}

Table 3.4 shows the exit probability $P_{\mathrm{MC}}(t \leq 180)$ estimated by MC simulation for different values of ϵ. For $\epsilon = 0.05$ and $\epsilon = 0.02$, the results estimated by IS agree well with the results estimated by MC simulations. For small values of ϵ such as 0.01 and 0.008, their agreement becomes less satisfactory.

CHAPTER 4

IMPORTANCE SAMPLING FOR MAGNETIC NANODEVICES

4.1 Introduction

Spin-transfer torque magnetoresistive random access memory (STT-MRAM) has been proposed as a replacement for random access memory that offers high speed, low power consumption, non-volatility and unlimited endurance [23–27]. 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. But 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 [28]. The expected values of WSER and RSER in STT-MRAM should not exceed the order of 10^{-18} without error correction [24]. 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 [29–31]. 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 [28]. 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 [32]. 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.

Although the most common approach to computing switching probabilities uses sampling to provide an empirical estimate of the quantity of interest, the extremely low switching probabilities and rates relevant to micromagnetic devices make naive Monte Carlo studies essentially impossible. A common approach to recover the tails of the probability distribution from Monte Carlo simulations is via extrapolation (see, e.g., [33–35]). 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 [36, 37]. 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 [5, 38–41].

Here, we demonstrate that accurate switching probabilities and error rates of STT-MRAM devices can be computed efficiently using importance sampling in an ensemble of biased Monte Carlo simulations (See Section 2.1). 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 chapter is to demonstrate that with the help of IS the events with extremely low probability of occurrence can be accessed via direct Monte Carlo 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.

This chapter is organized as follows. In Section 4.2 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 4.3, we carry out IS simulations for a single macrospin and a coupled system of two indentical macrospins with different choices of biases to demonstrate the power and efficiency of IS in the context of STT-MRAM applications. Finally, in Section 4.4 we briefly summarize our findings.

4.2 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. To model the free layer of an in-plane STT-MRAM cell, we use the stochastic Landau-Lifshitz-Gilbert equation [42, 43] for the unit magnetization vector $\mathbf{m} = (m_x, m_y, m_z)$,

$$\frac{\partial \mathbf{m}}{\partial t} = -\gamma\mu_0 \mathbf{m} \times \mathbf{H}_{\text{eff}} + \alpha \mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} + \boldsymbol{\tau}_{\text{STT}}, \tag{4.1}$$

where γ is the gyromagnetic ratio, $\mathbf{H}_{\text{eff}}$ is the (dimensional) effective magnetic field, α is the Gilbert damping parameter, and $\boldsymbol{\tau}_{\text{STT}}$ is the contribution from spin-transfer torque. The effective field $\mathbf{H}_{\text{eff}}$ is defined as the negative gradient of an energy functional plus a stochastic term,

$$\mathbf{H}_{\text{eff}} = -\frac{1}{\mu_0 M_s} \frac{\delta E}{\delta \mathbf{m}} + \sqrt{\sigma}\boldsymbol{\eta}, \tag{4.2}$$

where for now we consider only the exchange (i.e., spatial coupling) and crystalline anisotropy contributions given by

$$E[\mathbf{m}] = \int_\Omega d^3r \left[A|\nabla \mathbf{m}|^2 + K_u \Phi(\mathbf{m}) \right], \tag{4.3}$$

and

$$\Phi(\mathbf{m}) = m_y^2, \tag{4.4}$$

where A is the exchange stiffness constant, K_u is a in-plane uniaxial anisotropy constant, μ_0 is the permeability of the vacuum, and σ is determined by the fluctuation-dissipation theorem:

$$\sigma = \frac{2\alpha k_B T}{\gamma \mu_0^2 M_s}. \tag{4.5}$$

The parameter k_B is the Boltzmann constant ($k_B = 1.38054 \times 10^{-23}$ Joules/degree) and T is the absolute temperature. The random term in $\mathbf{H}_{\text{eff}}$ is assumed to be space-time white noise, with independent components and

$$\begin{aligned} \langle \boldsymbol{\eta}^\dagger(\mathbf{x}, t), \boldsymbol{\eta}(\mathbf{x}', t') \rangle &= \delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \\ &= \delta(\mathbf{r} - \mathbf{r}')\delta(x_3 - x_3')\delta(t - t'), \end{aligned} \tag{4.6}$$

where $\mathbf{r} \in D$ and $x_3 \in (0, d)$. The spin-transfer torque $\boldsymbol{\tau}_{\text{STT}}$ is given by

$$\boldsymbol{\tau}_{\text{STT}} = -\eta \frac{\hbar}{2} \frac{j}{de} \frac{\gamma}{M_s} \mathbf{m} \times (\mathbf{m} \times \mathbf{m}_p) - \beta \eta \frac{\hbar}{2} \frac{j}{de} \frac{\gamma}{M_s} \mathbf{m} \times \mathbf{m}_p, \tag{4.7}$$

where j is the density of current passing perpendicularly through the film, e is the elementary charge (positive), $\eta \in (0, 1]$ 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.

To obtain the Landau-Lifshitz form of the LLGS equation, we resubstitute $\partial\mathbf{m}/\partial t$ and note that

$$\mathbf{m}\times\left(\mathbf{m}\times\frac{\partial\mathbf{m}}{\partial t}\right)=\left(\mathbf{m}\cdot\frac{\partial\mathbf{m}}{\partial t}\right)\mathbf{m}-|\mathbf{m}|^2\frac{\partial\mathbf{m}}{\partial t}=-\frac{\partial\mathbf{m}}{\partial t}, \tag{4.8}$$

giving

$$\frac{\partial\mathbf{m}}{\partial t}=-\gamma\mu_0\mathbf{m}\times\mathbf{H}_{\text{eff}}+\alpha\mathbf{m}\times(-\gamma\mu_0\mathbf{m}\times\mathbf{H}_{\text{eff}}+\boldsymbol{\tau}_{\text{STT}})+\boldsymbol{\tau}_{\text{STT}}+\alpha^2\mathbf{m}\times(\mathbf{m}\times\frac{\partial\mathbf{m}}{\partial t}),$$

which can be simplified to

$$(1+\alpha^2)\frac{\partial\mathbf{m}}{\partial t}=-\gamma\mu_0\mathbf{m}\times\mathbf{H}_{\text{eff}}-\gamma\mu_0\alpha\mathbf{m}\times(\mathbf{m}\times\mathbf{H}_{\text{eff}})+\alpha\mathbf{m}\times\boldsymbol{\tau}_{\text{STT}}+\boldsymbol{\tau}_{\text{STT}}. \tag{4.9}$$

Nondimensionalization: To nondimensionalize the system (4.9), we first define the following characteristic scales:

$$l_{ex}=\sqrt{\frac{2A}{\mu_0M_s^2}}\text{ and }t_s=\frac{1+\alpha^2}{\gamma\mu_0M_s}. \tag{4.10}$$

We also define dimensionless variables $\tilde{\mathbf{r}}=\mathbf{r}/l_{ex}$, $\tilde{x}_3=x_3/d$, $\tilde{t}=t/t_s$, $\tilde{\mathbf{H}}_{\text{eff}}=\mathbf{H}_{\text{eff}}/M_s$, $\tilde{E}=E/(\mu_0M_s^2dl_{ex}^2)$, and $\tilde{\boldsymbol{\tau}}_{\text{STT}}=\boldsymbol{\tau}_{\text{STT}}/(\gamma\mu_0M_s)$ to obtain,

$$\begin{aligned}\frac{\partial\mathbf{m}}{\partial\tilde{t}}&=\frac{1+\alpha^2}{\gamma\mu_0M_s}\frac{\partial\mathbf{m}}{\partial t}\\&=-\mathbf{m}\times\frac{\mathbf{H}_{\text{eff}}}{M_s}-\alpha\mathbf{m}\times(\mathbf{m}\times\frac{\mathbf{H}_{\text{eff}}}{M_s})+\alpha\mathbf{m}\times\frac{\boldsymbol{\tau}_{\text{STT}}}{\gamma\mu_0M_s}+\frac{\boldsymbol{\tau}_{\text{STT}}}{\gamma\mu_0M_s}\\&=-\mathbf{m}\times\tilde{\mathbf{H}}_{\text{eff}}-\alpha\mathbf{m}\times(\mathbf{m}\times\tilde{\mathbf{H}}_{\text{eff}})+\alpha\mathbf{m}\times\tilde{\boldsymbol{\tau}}_{\text{STT}}+\tilde{\boldsymbol{\tau}}_{\text{STT}}\end{aligned} \tag{4.11}$$

where

$$\tilde{\mathbf{H}}_{\text{eff}}=-\frac{\delta\tilde{E}}{\delta\mathbf{m}}+\frac{\sqrt{\sigma}}{M_s}\boldsymbol{\eta}(\mathbf{x},t). \tag{4.12}$$

with

$$\sigma=\frac{2\alpha k_BT}{\gamma\mu_0^2M_s}. \tag{4.13}$$

Since we are interested in thin films [44–46], we assume that $\mathbf{m}$ is independent of the film thickness, we have

$$\begin{aligned}\tilde{E} \simeq & \frac{E}{(\mu_0 M_s^2 d l_{ex}^2)} \\ = & \frac{d}{(M_s^2 \mu_0 d l_{ex}^2)} \int_{\tilde{D}} l_{ex}^2 d^2\tilde{r} \left[\frac{A}{l_{ex}^2} |\tilde{\nabla}\mathbf{m}|^2 + K_u \Phi(\mathbf{m}) \right] \\ = & \int_{\tilde{D}} d^2\tilde{r} \left[\frac{A}{\mu_0 M_s^2 l_{ex}^2} |\tilde{\nabla}\mathbf{m}|^2 + \frac{K_u}{\mu_0 M_s^2} \Phi(\mathbf{m}) \right] \\ = & \frac{1}{2} \int_{\tilde{D}} d^2\tilde{r} \left[|\tilde{\nabla}\mathbf{m}|^2 + Q\Phi(\mathbf{m}) \right], \end{aligned} \tag{4.14}$$

where $Q = 2K_u/(\mu_0 M_s^2)$.

Notice that

$$\begin{aligned}\langle \boldsymbol{\eta}^\dagger(\mathbf{x}, t), \boldsymbol{\eta}(\mathbf{x}', t') \rangle &= \delta(\mathbf{x} - \mathbf{x}')\delta(t - t') \\ &= \delta(\mathbf{r} - \mathbf{r}')\delta(x_3 - x_3')\delta(t - t') \\ &= \delta(l_{ex}(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'))\delta(d(\tilde{x}_3 - \tilde{x}_3'))\delta(t_s(\tilde{t} - \tilde{t}')) \\ &= \frac{\mu_0 \gamma M_s}{d l_{ex}^2 (1 + \alpha^2)} \delta(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}')\delta(\tilde{t} - \tilde{t}') \\ &= \frac{\mu_0 \gamma M_s}{d l_{ex}^2 (1 + \alpha^2)} \langle \tilde{\boldsymbol{\eta}}^\dagger(\tilde{\mathbf{x}}, \tilde{t}), \tilde{\boldsymbol{\eta}}(\tilde{\mathbf{x}}', \tilde{t}') \rangle. \end{aligned} \tag{4.15}$$

Thus we have

$$\tilde{\mathbf{H}}_{\text{eff}} = -\frac{\delta \tilde{E}}{\delta \mathbf{m}} + \sqrt{\tilde{\sigma}} \tilde{\boldsymbol{\eta}}, \tag{4.16}$$

and $\tilde{\boldsymbol{\eta}}$ is delta-correlated in dimensionless space in 3D and time with

$$\tilde{\sigma} = \frac{2\alpha k_B T}{\mu_0 M_s^2 d l_{ex}^2 (1 + \alpha^2)} = \frac{\alpha k_B T}{A d (1 + \alpha^2)}. \tag{4.17}$$

Considering that $\mathbf{m}$ is independent of the film thickness, the noise along the thickness of the film is projected by

$$\tilde{\boldsymbol{\eta}}(\tilde{\mathbf{r}}, \tilde{t}) = \int_0^1 \tilde{\boldsymbol{\eta}}(\tilde{\mathbf{r}}, \tilde{x}_3, \tilde{t}) d\tilde{x}_3, \tag{4.18}$$

where now $\tilde{\boldsymbol{\eta}}(\mathbf{r}, t)$ is delta-correlated in 2-D space and time.

After dropping tildes and redefining $\mathbf{h} = \mathbf{H}_{\text{eff}}$,

$$\frac{\partial \mathbf{m}}{\partial t} = -\mathbf{m} \times \mathbf{h} - \alpha \mathbf{m} \times (\mathbf{m} \times \mathbf{h}) + \alpha \mathbf{m} \times \boldsymbol{\tau}_{\text{STT}} + \boldsymbol{\tau}_{\text{STT}}, \tag{4.19}$$

where

$$\mathbf{h} = -\frac{\delta E}{\delta \mathbf{m}} + \sqrt{\sigma}\boldsymbol{\eta}, \tag{4.20}$$

with

$$E = \frac{1}{2} \int_D d^2 r \left[|\nabla \mathbf{m}|^2 + Q\Phi(\mathbf{m}) \right]. \tag{4.21}$$

Now, $\boldsymbol{\eta}(\mathbf{r}, t)$ is delta-correlated in dimensionless 2-D space and time and

$$\sigma = \frac{2\alpha k_B T}{\mu_0 M_s^2 d l_{ex}^2 (1 + \alpha^2)} = \frac{\alpha k_B T}{A d (1 + \alpha^2)}. \tag{4.22}$$

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

$$\boldsymbol{\tau}_{\text{STT}} = a_J \mathbf{m} \times (\mathbf{m} \times \mathbf{m}_p) + b_J \mathbf{m} \times \mathbf{m}_p. \tag{4.23}$$

where $a_J = -\eta j \hbar / (2de\mu_0 M_s^2)$ and $b_J = \beta a_J$ are dimensionless Slonczewski and field-like torque strengths [43]. 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 Chapter 4, 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 [24, 47, 48].

Micromagnetic with stray field. A ferromagnetic thin film also generates a magnetic field, called stray field. In the limit $d \to 0$, it can be showed that the limiting

energy is completely local, where the magnetostatics equation disappears [49]. The resulting dimensionless energy given by

$$E[\mathbf{m}] = \frac{1}{2}\int_D \left(|\nabla \mathbf{m}|^2 + Qm_y^2 + m_z^2\right)\, d^2r \tag{4.24}$$

is the leading order thin film micromagnetic energy measured in the units of $2Ad$.

The macrospin approximation assumes spatial uniformity across the ferromagnet, such that the first term in equation (4.24) is zero and equation (4.19) is an ordinary differential equation. In this case $E(\mathbf{m}) = \frac{1}{2}(Qm_y^2 + m_z^2)S$, where S is the area of D in the units of l_{ex}^2, and the effective field is given by

$$\mathbf{h} = -S^{-1}\nabla_{\mathbf{m}} E + \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}\,\dot{\boldsymbol{W}}(t), \tag{4.25}$$

where $\epsilon = k_B T/(2AdS)$, and $\boldsymbol{W}(t)$ is a three-dimensional 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

$$\alpha \sim 1 \text{ and } a_J \sim b_J \sim Q \ll 1, \tag{4.26}$$

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 [44–46]. In this case the system (1) may be simplified to an equation for the angle θ such that $\mathbf{m} \simeq (\cos\theta, \sin\theta, 0)$ (see Appendix):

$$\dot{\theta} = b(\theta) + \frac{1}{\sqrt{\Delta}}\dot{W}, \tag{4.27}$$

where

$$b(\theta) = (I_J - \cos(\theta))\sin(\theta), \tag{4.28}$$

$I_J = b_J/Q$, $1/\Delta = 2\epsilon/Q$, and the unit of time is now $\tau_Q = \alpha/(\gamma\mu_0 M_s Q)$. We point out that even though equation (4.27) was obtained for an in-plane STT-MRAM cell, exactly the same equation arises in the modeling of perpendicular cells [28]. Therefore, a direct comparison with the results obtained by Butler et al. [28], who used the analysis of the Fokker-Planck equation is also possible.

Equation (4.27) 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, $1/\Delta_i = k_B T/(AdS_i Q)$, and θ_i are such that the magnetization in each grain is $\mathbf{m}_i \simeq (\cos\theta_i, \sin\theta_i, 0)$, then equation (4.27) may be generalized to [50–52]

$$\dot{\theta}_i = b(\theta_i) + \sum_{j=1}^{N} a_{ij} S_i^{-1} \sin(\theta_j - \theta_i) + \frac{1}{\sqrt{\Delta_i}} \dot{W}_i, \tag{4.29}$$

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*

$$V_N(\theta_1, \ldots, \theta_N) \;=\; \sum_{i=1}^{N} S_i \left(I_J \cos\theta_i + \tfrac{1}{2}\sin^2\theta_i \right) \;-\; \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} a_{ij} \cos(\theta_i \,-\, \theta_j), \tag{4.30}$$

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 chapter 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 [53].

The single macrospin drift term defined in equation (4.28) 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 chapter , 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 equation (4.29) 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 equation (4.27) can be computed by solving the backward Fokker-Planck equation for the probability $P_{\text{sw}}(\theta, t)$ 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 [54]. This probability satisfies

$$\frac{\partial P_{\text{sw}}}{\partial t} = b(\theta)\frac{\partial P_{\text{sw}}}{\partial \theta} + \frac{1}{2\Delta}\frac{\partial^2 P_{\text{sw}}}{\partial \theta^2}, \tag{4.31}$$

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

$$P_{\text{sw}}(\theta, 0) = 0, \qquad P_{\text{sw}}(\pm\pi/2, t) = 1, \tag{4.32}$$

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

As an alternative to solving equation (4.31), we incorporate IS in the estimation of P_{sw} by sampling controlled dynamics

$$\dot{\tilde{\theta}} = b(\tilde{\theta}) + u^* + \frac{1}{\sqrt{\Delta}}\dot{W}, \qquad \tilde{\theta}(0) = 0, \tag{4.33}$$

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 equation (2.38) by interpreting equation (4.33) as an instance of equation (2.13) with $\epsilon = 1/\sqrt{\Delta}$, using the minimizer $\theta_{\tilde{\theta},t}^T(s)$ of the finite-time action

$$S_T[\theta] = \int_t^T \frac{1}{2}|\dot{\theta}(s) - b(\theta(s))|^2 ds, \tag{4.34}$$

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:

$$u_\infty^*(\tilde{\theta}) = \begin{cases} -2b(\tilde{\theta}), & -\theta_J \leqslant \tilde{\theta} \leqslant \theta_J, \\ 0, & \text{otherwise.} \end{cases} \tag{4.35}$$

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 equation (2.15), which in this case is

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

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 equation (4.33) with a particular realization $W(t)$ of the noise.

4.3 Simulations

The following sections describe results obtained from IS simulations of macrospin and coupled-spin systems using physical parameters drawn from reference [28] for the purpose of comparison. All simulations use the Euler-Maruyama method with a fixed time step $\tau = 0.1$.

The IS results presented below are obtained using bias functions based on either finite- or infinite-time minimizers of the action given by equation (2.37). Infinite-time bias functions are based on equation (4.35) 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 equation (2.37) through the geometric minimum action method (GMAM) with 50 gridpoints [13]. Finite-time bias functions are obtained by minimizing the action in equation (2.37) through a combination of Newton's method for the associated Euler-Lagrange equation and the improved adaptive minimum action method [15], with 500 gridpoints in the macrospin case and 100 in the coupled-spin case.

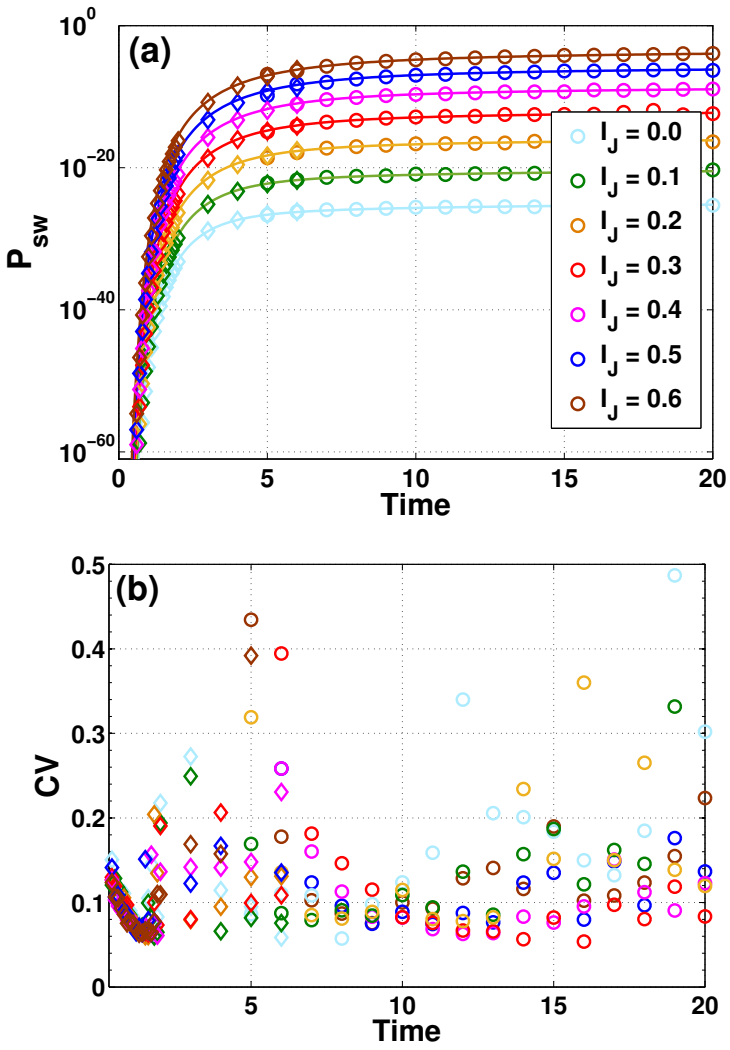

Figure 4.1 IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$. (a) RSER vs. reading pulse duration T and reading current I_J with thermal stability factor $\Delta = 60$. Solid lines denote numerical solutions of FPE (4.31). Open circles and open diamonds denote 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, while for IS and for $T \leq 6$ finite-time bias functions are used. IS results at $T = 5$ and $T = 6$ obtained using infinite- and finite-time bias functions are indistinguishable. (b) CV of the RSER.

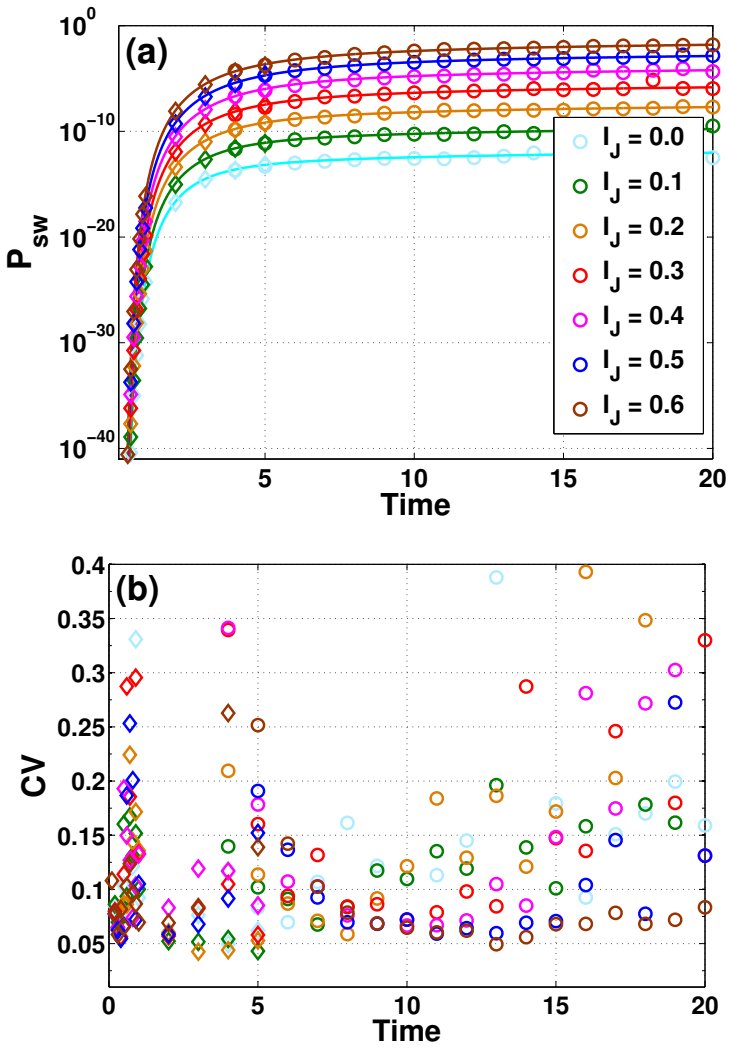

Figure 4.2 Similar to Figure 4.1, but with $\Delta = 30$. For $T \geq 4$ infinite-time bias functions are used, and for $T \leq 5$ finite-time bias functions are used. IS estimates at $T = 4$ and $T = 5$ obtained using infinite- and finite-time bias functions are indistinguishable.

4.3.1 Single macrospin

The discretized version of equation (4.33) reads explicitly

$$\tilde{\theta}^{k+1} = \tilde{\theta}^k + (b(\tilde{\theta}^k) + u^*(\tilde{\theta}^k, t_k))\tau + \frac{\sqrt{\tau}}{\sqrt{\Delta}}\xi_k, \quad \tilde{\theta}^0 = 0, \tag{4.36}$$

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 $t_k < T$. The likelihood ratio corresponding to equation (4.36) is then given explicitly by

$$L = \exp\left(-\frac{\tau\Delta}{2} \sum_{k=0}^{K} |u^*(\tilde{\theta}^k, t_k)|^2 - \sqrt{\tau\Delta} \sum_{k=0}^{K} u^*(\tilde{\theta}^k, t_k)\, \xi_k \right). \tag{4.37}$$

The figures below are generated using switching probabilities and their coefficients of variation computed by applying formulas (2.7) and (2.10), respectively, to ensembles of M runs, where $I(\omega_i) = 1$ for runs that generate a switching event and $I(\omega_i) = 0$ otherwise.

Figures 4.1 and 4.2 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 reference [28]. Both show a comparison between numerical solutions of backward FPE (4.31) and IS simulations for the macrospin model. For $T \geq 5$ in Figure 4.1 and $T \geq 4$ in Figure 4.2, infinite-time bias functions are used for IS while for $T \leq 6$ in Figure 4.1 and $T \leq 5$ in Figure 4.2, 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 Figures 4.1(a) and 4.2(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 per cent for $M = 10^5$ for most of the data points in Figures 4.1 and 4.2. 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 Figure 4.1, and used at $T = 4$ in Figure 4.2.

Figure 4.1 (b) shows the CV for the IS estimates in Figure 4.1 (a), and Figure 4.2 (b) shows those for the estimates shown in Figure 4.2 (a). The CV values for the IS estimates range from approximately 0.05 to 0.5. We note that with an inappropriate choice of bias the CV can be an imperfect measure of accuracy for Monte Carlo estimates using variance reduction [55]. However, this is precluded by our choice of an asymptotically optimal bias function that is based on large deviation theory [7, 56], as can also be seen from the excellent agreement with the solutions to FPE (4.31). 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, Figures 4.1 (b) and 4.2 (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. Figure 4.3 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 there exists 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 Figure 4.4. 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.

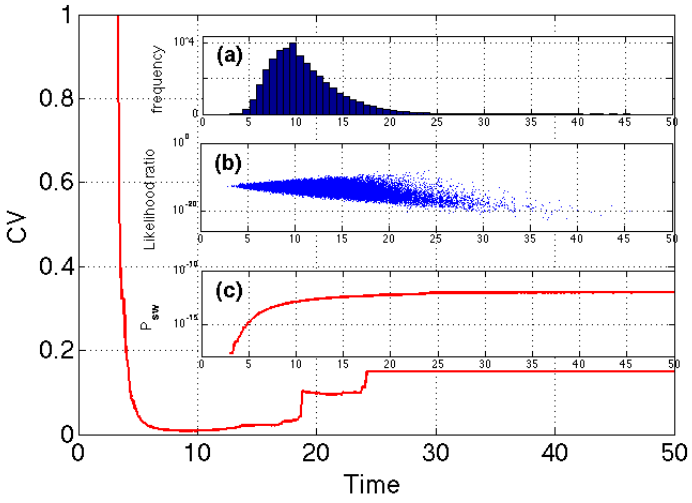

Figure 4.3 CVs of IS estimators for $I_J = 0.3$ with infinite-time bias functions and sample size 10^5. Inset (a) is a histogram of exit times for the biased system. Inset (b) is the estimated likelihood ratio vs. time. Inset (c) is the probability of switching before time T. The sample size here is $M = 10^5$.

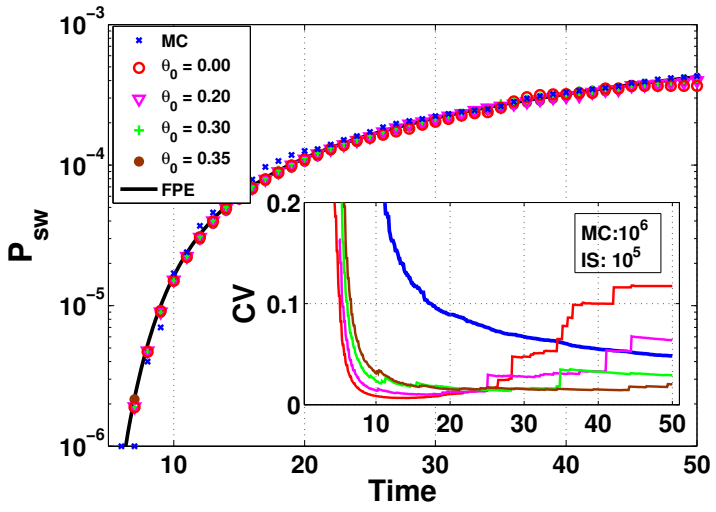

Figure 4.4 Switching probabilities and CVs of IS estimators for $I_J = 0.6$ with infinite-time bias functions active outside of a region containing the stable fixed point, defined by $\theta_0 < |\theta| < \theta_J$, for different values of θ_0. The sample size is 10^5.

4.3.2 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 equation (4.29), which in this case is explicitly

$$\dot{\theta}_1 = c\sin(\theta_2 - \theta_1) + b(\theta_1) + \sqrt{\frac{2}{\Delta}}\dot{W}_1, \tag{4.38}$$

$$\dot{\theta}_2 = c\sin(\theta_1 - \theta_2) + b(\theta_2) + \sqrt{\frac{2}{\Delta}}\dot{W}_2, \tag{4.39}$$

where $c > 0$ is the ferromagnetic exchange coupling strength favoring parallel alignment of the two spins. The initial 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 [57]. 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 [58] further exacerbates the challenge in finding appropriate bias functions.

4.3.3 Two coupled identical macrospins: biased dynamics

The biased dynamics associated with equation (4.38) and equation (4.39) reads

$$\dot{\tilde{\theta}}_1 = c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) + b(\tilde{\theta}_1) + \sqrt{2}\,u_1^* + \sqrt{\frac{2}{\Delta}}\dot{W}_1, \tag{4.40}$$

$$\dot{\tilde{\theta}}_2 = c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) + b(\tilde{\theta}_2) + \sqrt{2}\,u_2^* + \sqrt{\frac{2}{\Delta}}\dot{W}_2. \tag{4.41}$$

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 (2.37) with terminal condition $\max(|\theta_1(T_{sw})|, |\theta_2(T_{sw})|) = \pi/2$ for some $0 < T_{sw} \leq T \leq \infty$.

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

$$S_T[\theta_1, \theta_2] = \frac{1}{4}\int_t^T \Big(\dot{\theta}_1 - c\sin(\theta_2 - \theta_1) - b(\theta_1)\Big)^2 ds + \frac{1}{4}\int_t^T \Big(\dot{\theta}_2 - c\sin(\theta_1 - \theta_2) - b(\theta_2)\Big)^2 ds, \tag{4.42}$$

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

$$u_{T,1}^*(\tilde{\theta}_1, \tilde{\theta}_2, t) = \frac{1}{\sqrt{2}}(\dot{\theta}^T_{\tilde{\theta}_1,\tilde{\theta}_2,t,1}(t) - c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) - b(\tilde{\theta}_1)), \tag{4.43}$$

$$u_{T,2}^*(\tilde{\theta}_1, \tilde{\theta}_2, t) = \frac{1}{\sqrt{2}}(\dot{\theta}^T_{\tilde{\theta}_1,\tilde{\theta}_2,t,2}(t) - c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) - b(\tilde{\theta}_2)), \tag{4.44}$$

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

For infinite-time bias, we minimize

$$S_\infty[\theta_1, \theta_2] = \frac{1}{2}\int_0^1 \Bigg(\lambda(\theta_1, \theta_2)\sqrt{|\theta_1'|^2 + |\theta_2'|^2} \\ - (c\sin(\theta_2 - \theta_1) + b(\theta_1))\theta_1' - (c\sin(\theta_1 - \theta_2) + b(\theta_2))\theta_2'\Bigg)ds, \tag{4.45}$$

where

$$\lambda(\theta_1, \theta_2) = \Big[(c\sin(\theta_2 - \theta_1) + b(\theta_1))^2 + (c\sin(\theta_1 - \theta_2) + b(\theta_2))^2\Big]^{1/2}, \tag{4.46}$$

and express the bias as

$$u_{\infty,1}^*(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{\sqrt{2}}\left(\frac{\lambda(\tilde{\theta}_1, \tilde{\theta}_2)\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,1}(0)}{\sqrt{|\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,1}|^2 + |\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,2}|^2}} - c\sin(\tilde{\theta}_2 - \tilde{\theta}_1) - b(\tilde{\theta}_1)\right), \tag{4.47}$$

$$u_{\infty,2}^*(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{\sqrt{2}}\left(\frac{\lambda(\tilde{\theta}_1, \tilde{\theta}_2)\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,2}(0)}{\sqrt{|\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,1}|^2 + |\theta'_{\tilde{\theta}_1,\tilde{\theta}_2,2}|^2}} - c\sin(\tilde{\theta}_1 - \tilde{\theta}_2) - b(\tilde{\theta}_2)\right), \tag{4.48}$$

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 equation (4.36) and equation (4.37).

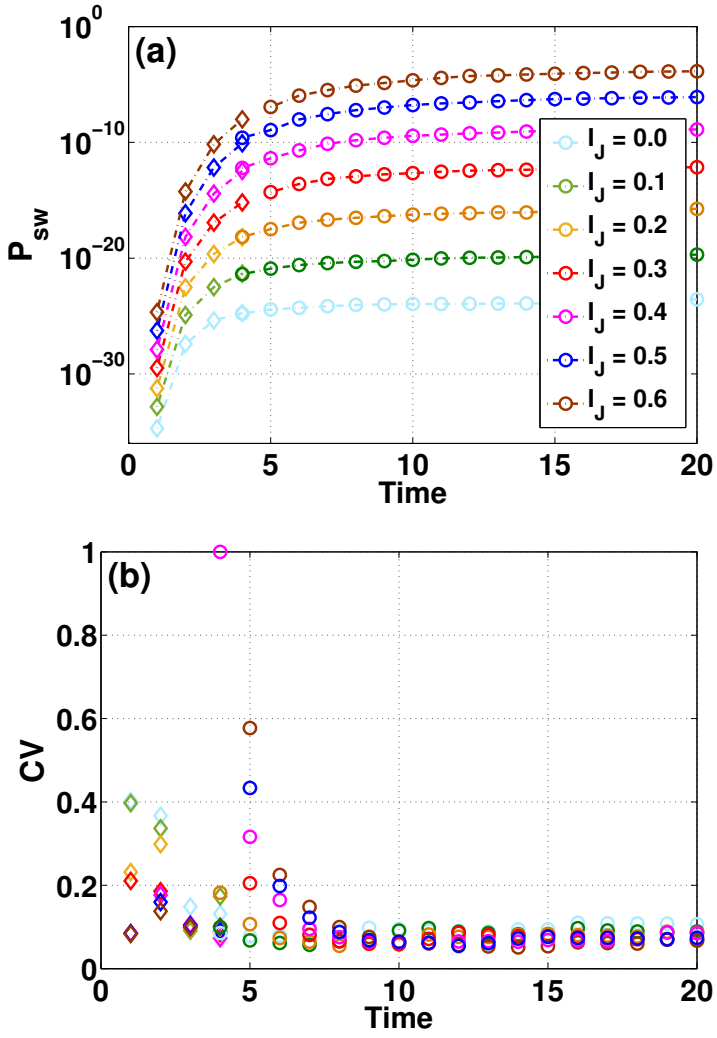

Figure 4.5 IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$ for a strongly coupled two-spin system. (a) RSER with applied currents ranging from 0 to 0.6 versus time T with thermal stability factor $\Delta = 60$ and coupling strength $c = 0.8$. (b) CVs of IS estimates in (a). In both panels, open circles and open diamonds denote estimates generated by IS using infinite-time and finite-time bias functions, respectively. The colors correspond to different current amplitudes indicated in (a).

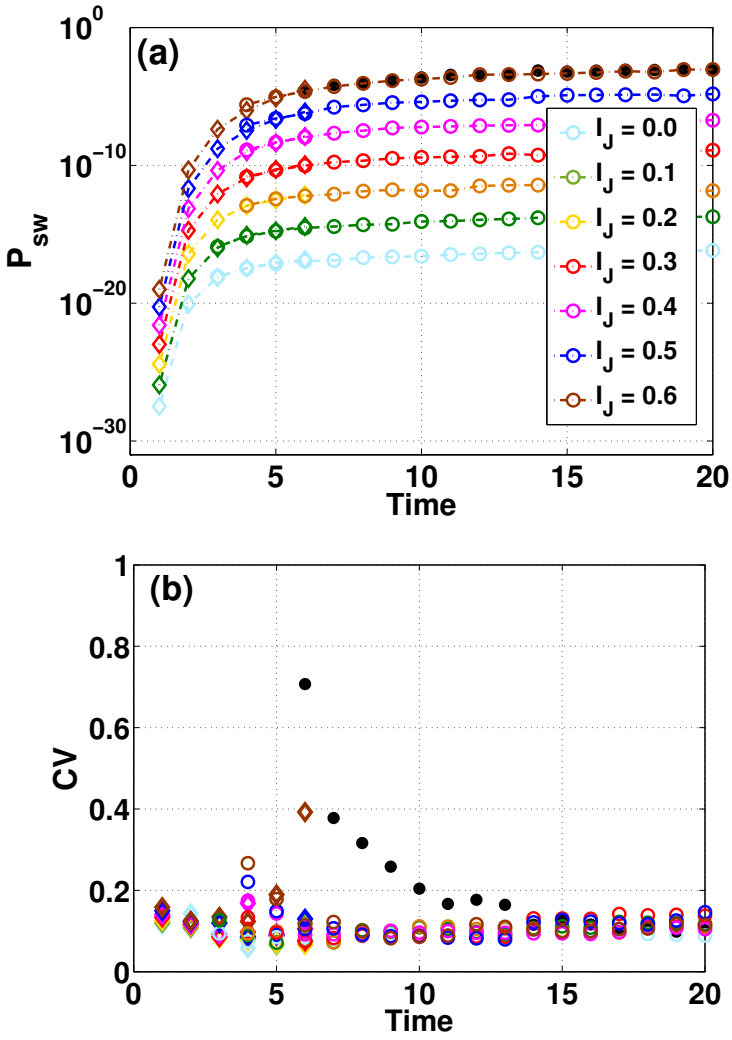

Figure 4.6 IS estimate for switching probability (RSER) and CV of the RSER with sample size $M = 10^3$ for a weakly coupled two-spin system. (a) RSER with applied currents ranging from 0 to 0.6 versus time T with thermal stability factor $\Delta = 60$ and coupling strength $c = 0.2$. (b) CVs of IS estimates in (a). In both panels, open circles and open diamonds denote estimates generated by IS using infinite-time and finite-time bias functions, respectively, and black dots denote 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).

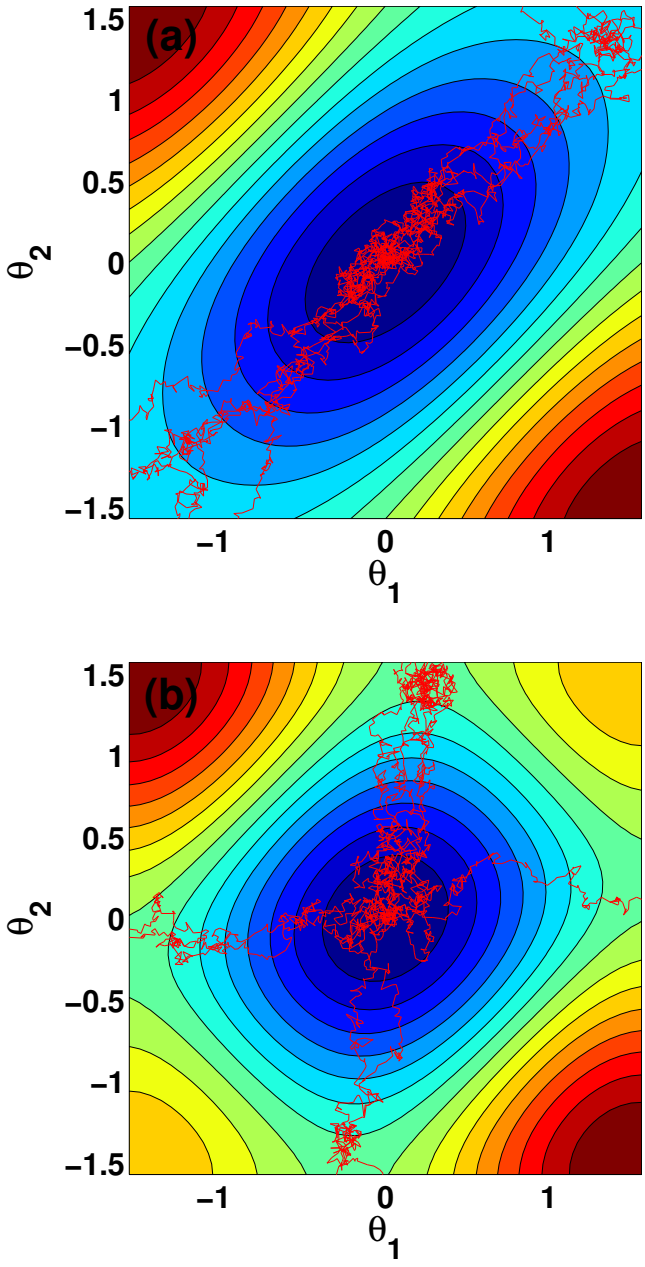

Figure 4.7 Sample paths from IS for two-spin systems with current $I_J = 0.1$ and thermal stability factor $\Delta = 60$: strongly coupled spins with $c = 0.8$ (a), weakly coupled spins with $c = 0.2$ (b). False color corresponds to the effective potential in (4.30).

Figure 4.5 (a) shows 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 estimates generated by IS using infinite-time and finite-time bias functions, respectively. With 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 equation (4.30) 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.

Figure 4.5 (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 Figure 4.5 (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 Figure 4.6. Figure 4.6 (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 bias functions allows us to sample switching probabilities for $T \geq 4$ with I_J ranging from 0 to 0.6. As in Figure 4.5, Figure 4.6 (b) shows a decrease in efficiency of IS using infinite-time bias functions as T decreases.

Figure 4.6 (a) also shows 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 Figure 4.6 (b) where the CV is seen to diverge as the probability estimate decreases. Finally, a few representative switching trajectories corresponding to the results in Figures 4.5 and 4.6 are shown in Figure 4.7.

Figure 4.7 (a) shows the sample paths generated by IS for a strongly coupled spins. Since the two spins are strongly coupled, the spins switch by coherent rotation. In Figure 4.7 (b), the sample paths generated by IS for a weakly coupled spins and the spins switch by asymmetric coherent rotation.

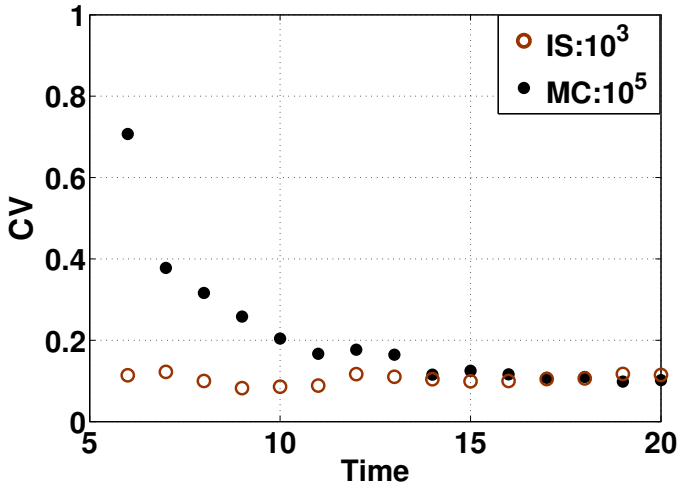

Figure 4.8 CVs of estimates generated by IS (brown open circles) and naive MC (black dots) in Figure 4.6.

Figure 4.8 shows a comparison between the CV for IS with sample size of 10^3 and the CV for naive MC with 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, al 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.

4.3.4 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 $0 < T_{sw} \leq T$.

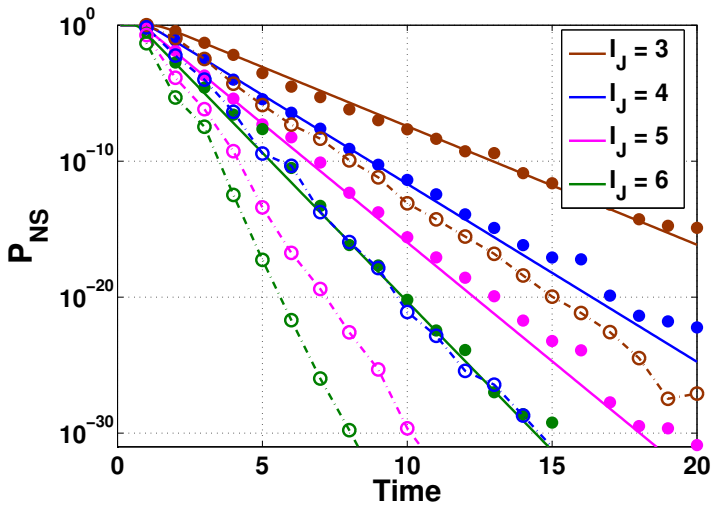

Figure 4.9 Non-switching probability as a function of time horizon for several values of the writing current, using thermal stability factor $\Delta = 60$. The current I_J is identified by color and ranges from 3 to 6. Solid lines denote numerical solutions of the backward Fokker-Planck equation for single macrospin, while 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 $c = 0.2$.

Figure 4.9 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.

4.4 Summary

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 chapter 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 non-uniform 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.

CHAPTER 5

CONCLUSIONS

This work is devoted to quantifying the probabilities of rare events in stochastic system. In particular, we estimate the probabilities of parameter excursions in optical communication systems and the probabilities of read error rates and write error rates for a spin-transfer torque nano-devices. In both settings, the first step is to reduce infinite-dimensional stochastic PDEs to finite-dimensional stochastic ODEs. In the case of the laser model, this is accomplished using a variational reduction. In the case of the Landau-Lifshitz equation, this is accomplished using the macrospin approximation.

Using the numerical computation tool gMAM, the Freidlin-Wentzell minimum action was obtained for amplitude drop-outs and position shifts in the finite-dimensional optical systems we considered, providing estimates for the probability of these events occurring. By analogy with the concept of finite-temperature single-pulse astroids in micromagnetics, we explored the quasi-stability of the system in its parameter space and identified for which parameters the system was most stable. Furthermore, the application of IS with bias functions computed by gMAM with fixed endpoints allowed a numerical estimation of the probabilities above.

We extended the application of gMAM and IS to the estimation of read error rates and write error rates for macrospin spin systems modeling STT-MRAM devices. Throughout this dissertation, we have intentionally selected simple models to focus on the feasibility of using IS to estimate switching probabilities. One of the strategies to obtain switching probabilities over finite time horizons is the use of IS with infinite-time bias functions. This strategy is useful for intermediate times but does not correctly promote the long periods spent near the stable fixed point in the actual dynamics. This phenomenon can be addressed by turning off the biasing while

the current state is within a diffusion length of the stable fixed point, which leads to significant improvement of the sampling efficiency. Another strategy to obtain switching probabilities for short times is the use of IS with finite-time bias functions. We also extended these strategies to estimate switching probabilities using IS applied to a two coupled-spin system. The infinite-time bias functions are computed by gMAM with variable endpoints, while the finite-time bias functions are computed by MAM with variable endpoints. The switching probability is underestimated if the bias function is based on gMAM or MAM with fixed endpoint at the saddle point.

The critical component of these IS strategies is the approximation of the bias functions. As we have presented, these importance sampling strategies are readily adaptable to more complex ODE systems, such as a system with many interacting macrospins. The only limitation of these strategies is the amount of computational effort needed to obtain the bias function. The IS strategies can also be extended to stochastic PDEs such as the nonlinear Schrödinger equations and the Landau-Lifshitz equation. Ongoing work includes the implementation of gMAM or MAM for these systems with variable terminal conditions and the investigation of effective numerical schemes to obtain bias functions.

APPENDIX

DERIVATION OF MACROSPIN AND COUPLED SPINS

This appendix presents the derivation of the macrospin approximation for the stochastic Landau-Lifshitz equation with the spin-transfer torque and the derivation of the coupled spin system for the coupled stochastic Landau-Lifshitz equation with the spin-transfer torque.

The macrospin model assumes that the magnetic element is so small that the exchange energy dominates and the magnetization is almost uniform throughout the element. The standard stochastic Landau-Lifshitz equation with the spin-transfer torque is a stochastic ODE :

$$\frac{\partial \mathbf{m}}{\partial t} = -\mathbf{m} \times \mathbf{h} - \alpha \mathbf{m} \times (\mathbf{m} \times \mathbf{h}) + \alpha \mathbf{m} \times \boldsymbol{\tau}_{\mathrm{STT}} + \boldsymbol{\tau}_{\mathrm{STT}}, \tag{1}$$

where $\mathbf{m}(t) = (m_x(t), m_y(t), m_z(t))$ is a unit vector whose direction is parallel to the magnetization of a magnetic system with constant length M_s. The effective field is given by

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

where $\dot{\boldsymbol{W}}(t)$ is homogenized by

$$\dot{\boldsymbol{W}}(t) = \frac{1}{\sqrt{S}} \int_D \boldsymbol{\eta}(\mathbf{r}, t) d^2 r, \tag{3}$$

with S stands for the area of D in the units of l_{ex}^2. The dimensionless energy density is given by

$$\frac{E}{S} = \frac{Q}{2} m_y^2 + \frac{1}{2} m_z^2. \tag{4}$$

$\dot{\boldsymbol{W}}(t)$ is uncorrelated, independent, Gaussian white noise. The noise coefficient, $\sqrt{2\alpha\epsilon/(1+\alpha^2)}$, is consistent with the Gibbs distribution and

$$\epsilon = k_B T/(\mu_0 M_s^2 dl_{ex}^2 S) = k_B T/(2AdS)$$

is the dimensionless temperature. The unit of time is $(1+\alpha^2)/(\gamma\mu_0 M_S)$. Equation (1) is to be interpreted in the Stratonovich sense so that this equation preserve the norm constraint, $|\mathbf{m}| = \mathbf{1}$. Here we consider $\mathbf{m}_p = (1,0,0)^T$. a_J and b_J are constant.

Rewriting the stochastic Landau-Lifshitz equation. To convert equation (1) into cylindrical coordinates, we first rewrite it as following

$$\frac{d\mathbf{m}}{dt} = \mathbf{b}(\mathbf{m}) + \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}} G(\mathbf{m}) \circ \dot{\mathbf{W}}, \tag{5}$$

where the drift term is

$$\begin{aligned} b(\mathbf{m}) =& -\mathbf{m}\times\mathbf{h} - \alpha\mathbf{m}\times(\mathbf{m}\times\mathbf{h}) + a_J\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_p) - a_J\alpha\mathbf{m}\times\mathbf{m}_p \qquad (6)\\ &+ b_J\mathbf{m}\times\mathbf{m}_p + \alpha b_J\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_p)\\ =& -\mathbf{m}\times\mathbf{h} - \alpha\mathbf{m}\times(\mathbf{m}\times\mathbf{h}) + (a_J+\alpha b_J)\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_p)\\ &+ (b_J - \alpha a_J)\mathbf{m}\times\mathbf{m}_p \qquad (7)\\ =& \begin{pmatrix} h_2 m_z - h_3 m_y \\ h_3 m_x - h_1 m_z \\ h_1 m_y - h_2 m_x \end{pmatrix} + \alpha \begin{pmatrix} h_1(m_y^2+m_z^2) - h_2 m_x m_y - h_3 m_x m_z \\ h_2(m_x^2+m_z^2) - h_1 m_x m_y, -h_3 m_y m_z \\ h_3(m_x^2+m_y^2) - h_1 m_x m_z - h_2 m_y m_z \end{pmatrix}\\ &+ (a_J+\alpha b_J)\begin{pmatrix} -m_z^2 - m_y^2 \\ m_x m_y \\ m_x m_z \end{pmatrix} + (b_J - \alpha a_J)\begin{pmatrix} 0 \\ m_z \\ -m_y \end{pmatrix}, \end{aligned}$$

and the diffusion matrix is

$$\mathbf{G}(\mathbf{m}) = \begin{pmatrix} \alpha(m_y^2 + m_z^2) & m_z - \alpha m_x m_y & -m_y - \alpha m_x m_z \\ -m_z - \alpha m_x m_y & \alpha(m_x^2 + m_z^2) & m_x - \alpha m_y m_z \\ m_y - \alpha m_x m_z & -m_x - \alpha m_y m_z & \alpha(m_x^2 + m_y^2), \end{pmatrix}$$

where $\mathbf{h} = -\nabla_{\mathbf{m}} E$.

Cylindrical coordinates. We will show that under the transformation

$$\mathbf{m} = (\cos(\theta)\sqrt{1-z^2}, \sin(\theta)\sqrt{1-z^2}, z) \tag{8}$$

equation (5) can be written as the Itô stochastic differential differential equation

$$\begin{aligned} \dot{\theta} =& -\frac{\partial E}{\partial z} - \frac{\alpha}{1-z^2}\frac{\partial E}{\partial \theta} + \frac{\sqrt{2\epsilon\alpha}}{\sqrt{1-z^2}}\dot{W}_1 \\ &+ (a_J + \alpha b_J)\frac{\sin(\theta)}{\sqrt{1-z^2}} + (b_J - \alpha a_J)\frac{z\cos(\theta)}{\sqrt{1-z^2}}, \end{aligned} \tag{9}$$

$$\begin{aligned} \dot{z} =& \frac{\partial E}{\partial \theta} - \alpha(1-z^2)\frac{\partial E}{\partial z} - 2\epsilon\alpha z + \sqrt{2\epsilon\alpha}\sqrt{1-z^2}\dot{W}_2 \\ &+ (a_J + \alpha b_J)(z\cos(\theta)\sqrt{1-z^2}) - (b_J - \alpha a_J)\sin(\theta)\sqrt{1-z^2} \end{aligned} \tag{10}$$

where

$$\frac{\partial E}{\partial z} = z - Q\sin(\theta)^2 z, \tag{11}$$

$$\frac{\partial E}{\partial \theta} = Q\sin(\theta)\cos(\theta)(1-z^2). \tag{12}$$

Transformation. Let us define

$$\theta = \tan^{-1}(\frac{m_y}{m_x}) \tag{13}$$

$$z = m_z. \tag{14}$$

Since equation(5) is interpreted in the Strotonovich sense, ordinary chain rule applies to Stratonovich calculus. We have

$$
\begin{aligned}
\dot{\theta} &= \frac{\dot{m}_y m_x - \dot{m}_x m_y}{m_x^2 + m_y^2} \\
&= \frac{1}{m_x^2 + m_y^2}(b_2 m_x - b_1 m_y) \\
&+ \frac{1}{m_x^2 + m_y^2}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}((m_x \mathbf{G}_{2,1} - m_y \mathbf{G}_{1,1}) \circ \dot{W}_1 + (m_x \mathbf{G}_{2,2} - m_y \mathbf{G}_{1,2}) \circ \dot{W}_2 \\
&+ (m_x \mathbf{G}_{2,3} - m_y \mathbf{G}_{1,3}) \circ \dot{W}_3) \\
&= \frac{1}{m_x^2 + m_y^2}(b_2 m_x - b_1 m_y) \\
&+ \frac{1}{m_x^2 + m_y^2}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}((-m_x m_z - \alpha m_y) \circ \dot{W}_1 + (-m_y m_z + \alpha m_x) \circ \dot{W}_2 \\
&+ (m_x^2 + m_y^2) \circ \dot{W}_3) \\
&= \frac{1}{m_x^2 + m_y^2}(b_2 m_x - b_1 m_y) \\
&+ \frac{1}{\sqrt{1-z^2}}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}((-z\cos(\theta) - \alpha\sin(\theta)) \circ \dot{W}_1 + (-z\sin(\theta) + \alpha\cos(\theta)) \circ \dot{W}_2) \\
&+ \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}\dot{W}_3
\end{aligned} \tag{15}
$$

To simply the deterministic part, we need to use the following equations:

$$
\begin{aligned}
\frac{\partial E}{\partial \theta} &= \frac{\partial E}{\partial m_x}\frac{\partial m_x}{\partial \theta} + \frac{\partial E}{\partial m_y}\frac{\partial m_y}{\partial \theta} + \frac{\partial E}{\partial m_z}\frac{\partial m_z}{\partial \theta} \\
&= -\frac{\partial E}{\partial m_x} m_y + \frac{\partial E}{\partial m_y} m_x \\
&= h_1 m_y - h_2 m_x,
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial z} &= \frac{\partial E}{\partial m_x}\frac{\partial m_x}{\partial z} + \frac{\partial E}{\partial m_y}\frac{\partial m_y}{\partial z} + \frac{\partial E}{\partial m_z}\frac{\partial m_z}{\partial z} \\
&= \frac{\partial E}{\partial m_x}\frac{-z\cos(\theta)}{\sqrt{1-z^2}} + \frac{\partial E}{\partial m_y}\frac{-z\sin(\theta)}{\sqrt{1-z^2}} + \frac{\partial E}{\partial m_z} \\
&= h_1 \frac{m_z m_x}{m_x^2 + m_y^2} + h_2 \frac{m_z m_y}{m_x^2 + m_y^2} - h_3.
\end{aligned} \tag{17}
$$

Thus the deterministic part of θ dynamic can be simplified as following:

$$\begin{aligned}
&\frac{1}{m_x^2+m_y^2}(b_2 m_x - b_1 m_y) = \\
&\frac{1}{m_x^2+m_y^2}(h_3(m_x^2+m_y^2) - h_1 m_x m_z - h_2 m_z m_y + \alpha(h_2 m_x - h_1 m_y) \\
&+ (a_J + \alpha b_J) m_y + (b_J - a_J \alpha) m_x m_z) \\
&= -\frac{\partial E}{\partial z} - \frac{\alpha}{(1-z^2)}\frac{\partial E}{\partial \theta} + (a_J + \alpha b_J)\frac{\sin(\theta)}{\sqrt{1-z^2}} + (b_J - \alpha a_J)\frac{z\cos(\theta)}{\sqrt{1-z^2}} \qquad (18)
\end{aligned}$$

As a result, the dynamic for θ is given by

$$\begin{aligned}
\dot{\theta} = &-\frac{\partial E}{\partial z} - \frac{\alpha}{(1-z^2)}\frac{\partial E}{\partial \theta} + (a_J + \alpha b_J)\frac{\sin(\theta)}{\sqrt{1-z^2}} + (b_J - \alpha a_J)\frac{z\cos(\theta)}{\sqrt{1-z^2}} \\
&+ \frac{1}{\sqrt{1-z^2}}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}((-z\cos(\theta) - \alpha\sin(\theta)) \circ \dot{W}_1 \\
&+ (-z\sin(\theta) + \alpha\cos(\theta)) \circ \dot{W}_2) + \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}\dot{W}_3
\end{aligned}$$

We have

$$\begin{aligned}
\dot{z} &= \dot{m}_z \\
&= b_3 + \mathbf{G}_{3,1} \circ \dot{W}_1 + \mathbf{G}_{3,2} \circ \dot{W}_2 + \mathbf{G}_{3,3} \circ \dot{W}_3 \\
&= h_1 m_y - h_2 m_x - \alpha(h_1 m_x m_z + h_2 m_y m_z - h_3(m_x^2 + m_y^2))) \\
&+ (a_J + \alpha b_J) m_x m_z - (b_J - \alpha a_J) m_y \\
&+ \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}((m_y - \alpha m_x m_z) \circ \dot{W}_1 + (-m_x - \alpha m_y m_z) \circ \dot{W}_2 + (\alpha(m_x^2 + m_y^2)) \circ \dot{W}_3) \\
=&\frac{\partial E}{\partial \theta} - \alpha(1-z^2)\frac{\partial E}{\partial z} \\
&+ (a_J + \alpha b_J) z\cos(\theta)\sqrt{1-z^2} - (b_J - \alpha a_J)\sin(\theta)\sqrt{1-z^2} \\
&+ \sqrt{1-z^2}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}(\sin(\theta) - \alpha\cos(\theta)z) \circ \dot{W}_1 \\
&+ \sqrt{1-z^2}\sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}(-\cos(\theta) - \alpha\sin(\theta)z) \circ \dot{W}_2 + \sqrt{\frac{2\alpha\epsilon}{1+\alpha^2}}(\alpha(1-z^2)) \circ \dot{W}_3
\end{aligned}$$

The diffusion matrix is given by

$$\sigma = \begin{pmatrix} (-z\cos(\theta) - \alpha\sin(\theta))/\sqrt{1-z^2} & (-z\sin(\theta) + \alpha\cos(\theta))/\sqrt{1-z^2} & 1 \\ \sqrt{1-z^2}(\sin(\theta) - \alpha\cos(\theta)z) & \sqrt{1-z^2}(-\cos(\theta) - \alpha\sin(\theta)z) & \alpha(1-z^2) \end{pmatrix}$$

with

$$\sigma\sigma^T = \begin{pmatrix} (1+\alpha^2)/1 - z^2 & 0 \\ 0 & (1+\alpha^2)(1-z^2) \end{pmatrix}$$

Compute Itô correction. Here we convert the system into Itô form. The Itô correction term for θ is given by

$$\begin{aligned} &\frac{\alpha\epsilon}{1+\alpha^2}\left(\frac{\partial\sigma_{1,1}}{\partial\theta}\sigma_{1,1} + \frac{\partial\sigma_{1,1}}{\partial z}\sigma_{2,1} + \frac{\partial\sigma_{1,2}}{\partial\theta}\sigma_{1,2} + \frac{\partial\sigma_{1,2}}{\partial z}\sigma_{2,2}\right) \\ &= 0 \end{aligned} \tag{19}$$

The itô correction term for z is given by

$$\begin{aligned} &\frac{\alpha\epsilon}{1+\alpha^2}\left(\frac{\partial\sigma_{2,1}}{\partial\theta}\sigma_{1,1} + \frac{\partial\sigma_{2,1}}{\partial z}\sigma_{2,1} + \frac{\partial\sigma_{2,2}}{\partial\theta}\sigma_{1,2} + \frac{\partial\sigma_{2,2}}{\partial z}\sigma_{2,2} + \frac{\partial\sigma_{3,2}}{\partial\theta}\sigma_{1,3} + \frac{\partial\sigma_{3,2}}{\partial z}\sigma_{2,3}\right) \\ &= -\frac{\alpha\epsilon}{1+\alpha^2}2(\alpha^2+1)z = -2\alpha\epsilon z. \end{aligned} \tag{20}$$

We also want to reduce the three-dimensional noise into two-dimensional noise. Observing that $\sigma\sigma^T$ is a diagonal matrix, we redefine

$$\sigma = \sqrt{1+\alpha^2}\begin{pmatrix} 1/\sqrt{1-z^2} & 0 \\ 0 & \sqrt{1-z^2} \end{pmatrix}$$

Putting everything together, we can see that equation (5) can be written as the Itô stochastic differential differential equation

$$\begin{aligned}\dot{\theta} =& -\frac{\partial E}{\partial z} - \frac{\alpha}{1-z^2}\frac{\partial E}{\partial \theta} + \frac{\sqrt{2\epsilon\alpha}}{\sqrt{1-z^2}}\dot{W}_1 \\ &+ (a_J + \alpha b_J)\frac{\sin(\theta)}{\sqrt{1-z^2}} + (b_J - \alpha a_J)\frac{z\cos(\theta)}{\sqrt{1-z^2}}, \end{aligned} \tag{21}$$

$$\begin{aligned}\dot{z} =& \frac{\partial E}{\partial \theta} - \alpha(1-z^2)\frac{\partial E}{\partial z} - 2\epsilon\alpha z + \sqrt{2\epsilon\alpha}\sqrt{1-z^2}\dot{W}_2 \\ &+ (a_J + \alpha b_J)(z\cos(\theta)\sqrt{1-z^2}) - (b_J - \alpha a_J)\sin(\theta)\sqrt{1-z^2}, \end{aligned} \tag{22}$$

where

$$\frac{\partial E}{\partial z} = z - Q\sin(\theta)^2 z, \tag{23}$$

$$\frac{\partial E}{\partial \theta} = Q\sin(\theta)\cos(\theta)(1-z^2). \tag{24}$$

Limit equation for θ. We derive a reduced model for the macrospin dynamics of the in-plane magnetization component in the regime of equation (4.26), which is done in the spirit of [44–46]. Our starting point is equation (1) for a macrospin $\mathbf{m}$, in which $\mathbf{h}$ satisfies equation (4.25). Introducing cylindrical coordinates, we can write

$$\mathbf{m} = \left(\sqrt{1-z^2}\,\cos\theta, \sqrt{1-z^2}\,\sin\theta, z\right). \tag{25}$$

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 $z(t)$:

$$\begin{aligned}\frac{Q(1+\alpha^2)}{\alpha}\dot{\theta} &= -\frac{\partial E}{\partial z} - \frac{\alpha}{1-z^2}\frac{\partial E}{\partial \theta} + (a_J + \alpha b_J)\frac{\sin\theta}{\sqrt{1-z^2}} \\ &\quad + (b_J - \alpha a_J)\frac{z\cos\theta}{\sqrt{1-z^2}} + Q\sqrt{\frac{1+\alpha^2}{\Delta(1-z^2)}}\,\dot{W}_1, \qquad (26)\\ \frac{Q(1+\alpha^2)}{\alpha}\dot{z} &= \frac{\partial E}{\partial \theta} - \alpha(1-z^2)\frac{\partial E}{\partial z} - \frac{Q\alpha z}{\Delta} \\ &\quad + (a_J + \alpha b_J) z\sqrt{1-z^2}\,\cos\theta \\ &\quad - (b_J - \alpha a_J)\sqrt{1-z^2}\,\sin\theta \\ &\quad + Q\sqrt{\Delta^{-1}(1+\alpha^2)(1-z^2)}\,\dot{W}_2, \qquad (27)\end{aligned}$$

where W_1 and W_2 are two uncorrelated Brownian motions and

$$\frac{\partial E}{\partial z} = z - Qz\sin^2\theta, \quad \frac{\partial E}{\partial \theta} = Q(1-z^2)\,\sin\theta\cos\theta. \qquad (28)$$

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 equations in z and introducing $\zeta = Q^{-1} z$, we obtain

$$\begin{aligned}\frac{Q(1+\alpha^2)}{\alpha}\dot{\theta} &= -\frac{\partial E}{\partial z} - \alpha\frac{\partial E}{\partial \theta} + QI_J(\beta^{-1} + \alpha)\sin\theta \\ &\quad + Q^2 I_J(1 - \alpha\beta^{-1})\zeta\cos\theta \\ &\quad + Q\sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_1, \qquad (29)\\ \frac{Q^2(1+\alpha^2)}{\alpha}\dot{\zeta} &= \frac{\partial E}{\partial \theta} - \alpha\frac{\partial E}{\partial z} - \frac{Q^2\alpha\zeta}{\Delta} \\ &\quad + Q^2 I_J(\beta^{-1} + \alpha)\zeta\cos\theta \\ &\quad - QI_J(1 - \alpha\beta^{-1})\sin\theta \\ &\quad + Q\sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_2, \qquad (30)\end{aligned}$$

where

$$\frac{\partial E}{\partial z} = Q\zeta, \qquad \frac{\partial E}{\partial \theta} = Q \sin\theta \cos\theta. \tag{31}$$

Substituting equation (31) into equations (29) and (30), to the leading order in $Q \ll 1$ we then arrive at

$$\begin{aligned} \frac{(1+\alpha^2)}{\alpha}\dot{\theta} &= -\zeta - \alpha \sin\theta\cos\theta + I_J(\beta^{-1}+\alpha)\sin\theta \\ &\quad + \sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_1, \end{aligned} \tag{32}$$

$$\begin{aligned} 0 &= \sin\theta\cos\theta - \alpha\zeta - I_J(1-\alpha\beta^{-1})\sin\theta \\ &\quad + \sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_2, \end{aligned} \tag{33}$$

Finally, solving for ζ in equation (33) and substituting it back into equation (32), 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 equation (4.27). It is interesting to note that only the contribution of the field-like spin torque appears in the reduced equation, while the contribution of the Slonczewski torque cancels out to the leading order.

Let us consider N coupled spins with identical volumes that are coupled via Heisenberg hamiltonians

$$E_{ex} = -\sum_{i,j} a_{ij}\mathbf{m}_i \cdot \mathbf{m}_j, \tag{34}$$

with a total energy that is given by

$$E = -\sum_{i,j} a_{ij}\mathbf{m}_i \cdot \mathbf{m}_j + \sum_i S_i\left(\frac{Q}{2}m_{i,y}^2 + \frac{1}{2}m_{i,3}^2\right), \tag{35}$$

where a_{ij} is the exchange parameter between moments $\mathbf{m}_i$ and $\mathbf{m}_j$.

The dynamics of the two spins are governed by a system of standard stochastic Landau-Lifshitz equation with the spin-transfer torque:

$$\frac{\partial \mathbf{m}_i}{\partial t} = -\mathbf{m}_i \times \mathbf{h}_i - \alpha \mathbf{m}_i \times (\mathbf{m}_i \times \mathbf{h}_i) + \alpha \mathbf{m}_i \times \boldsymbol{\tau}^i_{\mathrm{STT}} + \boldsymbol{\tau}^i_{\mathrm{STT}} \tag{36}$$

The effective fields are given by

$$\mathbf{h}_i = -\frac{\delta E}{\delta \mathbf{m}_i} + \sqrt{\frac{2\alpha\epsilon_i}{1+\alpha^2}} \boldsymbol{W}_i(t). \tag{37}$$

where $i = 1, 2$. Let us denote

$$\mathbf{m}_i = (\cos(\theta_i)\sqrt{1-z_i^2}, \sin(\theta_i)\sqrt{1-z_i^2}, z_i),\ i = 1, 2,$$

and the dynamic for spin $\mathbf{m}_1$ is given by

$$\begin{aligned}
\dot{\theta_1} =& -\frac{\partial E}{\partial z_1} - \frac{\alpha}{1-z_1^2}\frac{\partial E}{\partial \theta_1} + \frac{\sqrt{\epsilon_i \alpha}}{\sqrt{1-z_1^2}}\dot{W}_1 \\
&+ (a_J + \alpha b_J)\frac{\sin(\theta_1)}{\sqrt{1-z_1^2}} + (b_J - \alpha a_J)\frac{z_1\cos(\theta_1)}{\sqrt{1-z_1^2}}, \\
\dot{z_1} =& \frac{\partial E}{\partial \theta_1} - \alpha(1-z_1^2)\frac{\partial E}{\partial z_1} - 2\epsilon\alpha z_1 + \sqrt{\epsilon_i\alpha}\sqrt{1-z_1^2}\dot{W}_2 \\
&+ (a_J + \alpha b_J) z_1 \cos(\theta_1)\sqrt{1-z_1^2} \\
&- (b_J - \alpha a_J)\sin(\theta_1)\sqrt{1-z_1^2},
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
\frac{\partial E}{\partial z_1} &= z_1 - Q\sin(\theta_1)^2 z_1 \\
&\quad + a_{12}\frac{m_z}{m_x^2 + m_y^2}(m_{2,x}m_{1,x} + m_{1,y}m_{2,y} - m_{2,z}) \\
&= z_1 - Q\sin(\theta_1)^2 z_1 + a_{12} z_1 \frac{\sqrt{1-z_2^2}}{\sqrt{1-z_1^2}}\cos(\theta_1 - \theta_2) - a_{12} z_2, \\
\frac{\partial E}{\partial \theta_1} &= Q\sin(\theta_1)\cos(\theta_1)(1-z_1^2) + a_{12}(m_{2,x}m_{1,y} - m_{1,x}m_{2,y}) \\
&= Q\sin(\theta_1)\cos(\theta_1)(1-z_1^2) + a_{12}\sin(\theta_1 - \theta_2)\sqrt{(1-z_1^2)(1-z_2^2)}.
\end{aligned}$$

We now assume that $Q \ll 1$, while $a_{12} = cQ$, $a_J = \beta^{-1} I_J Q$, $b_J = I_J Q$ and $\epsilon_i = Q/(2\Delta_i)$, with c, I_J, β and Δ of order unity. In this case any deviations of z_i from zero are strongly suppressed. Therefore, linearizing the above equations in z_i, introducing $\zeta_i = Q^{-1} z_i$, and rescaling time by $(\alpha + \frac{1}{\alpha})$, we obtain

$$\begin{aligned}
\frac{Q(1+\alpha^2)}{\alpha}\dot{\theta_1} &= -\frac{\partial E}{\partial z_1} - \alpha\frac{\partial E}{\partial \theta_1} + QI_J(\beta^{-1}+\alpha)\sin\theta_1 \\
&\quad + Q^2 I_J(1-\alpha\beta^{-1})\zeta_1\cos\theta_1 \\
&\quad + Q\sqrt{\Delta^{-1}(1+\alpha^2)}\,\dot{W}_1, \qquad (39) \\
\frac{Q^2(1+\alpha^2)}{\alpha}\dot{\zeta_1} &= \frac{\partial E}{\partial \theta_1} - \alpha\frac{\partial E}{\partial z_1} - \frac{Q^2\alpha\zeta_1}{\Delta_1} \\
&\quad + Q^2 I_J(\beta^{-1}+\alpha)\zeta_1\cos\theta_1 \\
&\quad - QI_J(1-\alpha\beta^{-1})\sin\theta_1 \\
&\quad + Q\sqrt{\Delta_1^{-1}(1+\alpha^2)}\,\dot{W}_2, \qquad (40)
\end{aligned}$$

where

$$\frac{\partial E}{\partial z_1} = Q\zeta_1, \qquad \frac{\partial E}{\partial \theta_1} = Q\sin\theta_1\cos\theta_1 + cQ\sin(\theta_1-\theta_2). \qquad (41)$$

Substituting equation (31) into equations (39) and (40), to the leading order in $Q \ll 1$ we then arrive at

$$\begin{aligned}
\frac{(1+\alpha^2)}{\alpha}\dot{\theta_1} &= -\zeta_1 - \alpha(\sin\theta_1\cos\theta_1 + c\sin(\theta_1-\theta_2)) \\
&\quad + I_J(\beta^{-1}+\alpha)\sin\theta_1 \\
&\quad + \sqrt{\Delta_1^{-1}(1+\alpha^2)}\,\dot{W}_1, \qquad (42) \\
0 &= \sin\theta_1\cos\theta_1 + c\sin(\theta_1-\theta_2) - \alpha\zeta_1 \\
&\quad - I_J(1-\alpha\beta^{-1})\sin\theta_1 \\
&\quad + \sqrt{\Delta_1^{-1}(1+\alpha^2)}\,\dot{W}_2, \qquad (43)
\end{aligned}$$

Now again, solving for ζ_1 in equation (43) and substituting it back into equation (42), we obtain

$$\dot{\theta_1} = c\sin(\theta_2 - \theta_1) + (I_J - \cos(\theta_1))\sin(\theta_1) + \sqrt{\frac{1}{\Delta_1}}\dot{W}_1, \tag{44}$$

The derivation of the dynamics for $\mathbf{m}_2$ or θ_2 is the same for θ_1. The coupled system is given by

$$\dot{\theta_1} = c\sin(\theta_2 - \theta_1) + (I_J - \cos(\theta_1))\sin(\theta_1) + \sqrt{\frac{1}{\Delta_1}}\dot{W}_1 \tag{45}$$

$$\dot{\theta_2} = c\sin(\theta_1 - \theta_2) + (I_J - \cos(\theta_2))\sin(\theta_2) + \sqrt{\frac{1}{\Delta_2}}\dot{W}_2. \tag{46}$$

This can be generalized to N coupled spin system [50–52]

$$\dot{\theta}_i = b(\theta_i) + \sum_{j=1}^{N} a_{ij} S_i^{-1} \sin(\theta_j - \theta_i) + \frac{1}{\sqrt{\Delta_i}}\,\dot{W}_i, \tag{47}$$

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.

BIBLIOGRAPHY

[1] G. M. Donovan and W. L. Kath, "Rare event simulation of the performance of an actively mode-locked fiber laser model," in *Conf. Lasers and Electro-Opt.*, 2007.

[2] V. Linetsky, "The path integral approach to financial modeling and options pricing," *Comput. Econ.*, no. 11, pp. 129–163, 1998.

[3] M. Bachar, J. J. Batzel, and S. Ditlevsen, *Stochastic Biomathematical Models: with Applications to Neuronal Modeling*, ser. 2524-6771. Heidelberg, Germany: Springer, 2013, vol. 2058, no. 1.

[4] N. Metropolis and S. Ulam, "The Monte Carlo method," *J. Amer. Statist. Assoc.*, vol. 44, no. 247, pp. 335–341, 1949.

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

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

[7] 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.

[8] G. Rubino and B. Tuffin, *Rare Event Simulation Using Monte Carlo Methods*. Hoboken, NJ: Wiley Publishing, Mar. 2009.

[9] P. Dupuis, K. Spiliopoulos, and H. Wang, "Importance sampling for multiscale diffusions," *Multiscale Model. Simul.*, vol. 10, no. 1, pp. 1–27, Jan. 2012.

[10] S. Ross, *A First Course in Probability*, 8th ed. Upper Saddle River, NJ: Prentice Hall, 2009.

[11] 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.

[12] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, ser. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

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

[14] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, ser. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 38.

[15] 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, 2018.

[16] M. Heymann, "The geometric minimum action method: a least action principle on the space of curves," PhD dissertation, Dept. Math., N.Y. Univ., New York, NY, 2007.

[17] F. Idachaba, D. U. I., and O. Evwieroghene, "Future trends in fiber optics communication," *Lecture Notes Eng. Comput. Sci.*, vol. 1, pp. 438–442, Jul. 2014.

[18] W. H. Louisell, A. Yariv, and A. E. Siegman, "Quantum fluctuations and noise in parametric processes. i," *Phys. Rev.*, vol. 124, pp. 1646–1654, Dec. 1961.

[19] C. R. Giles and E. Desurvire, "Modeling erbium-doped fiber amplifiers," *J. Lightw. Technol.*, vol. 9, no. 2, pp. 271–283, Feb. 1991.

[20] S. Turitsyn, I. Gabitov, E. Laedke, V. Mezentsev, S. Musher, E. Shapiro, T. Schäfer, and K. Spatschek, "Variational approach to optical pulse propagation in dispersion compensated transmission systems," *Opt. Commun.*, vol. 1, no. 117-135, 1998.

[21] R. O. Moore, "Trade-off between linewidth and slip rate in a mode-locked laser model," *Opt. Lett.*, vol. 39, no. 10, pp. 3042–3045, May. 2014.

[22] R. Kohn, M. Reznikoff, and E. Vanden-Eijnden, "Magnetic elements at finite temperature and large deviation theory," *J. Nonlinear Sci.*, vol. 15, no. 4, pp. 223–253, Aug. 2005.

[23] J.-G. Zhu, "Magnetoresistive random access memory: The path to competitiveness and scalability," *Proc. IEEE*, vol. 96, pp. 1786 –1798, Nov. 2008.

[24] A. V. Khvalkovskiy *et al.*, "Basic principles of STT-MRAM cell operation in memory arrays," *J. Phys. D: Appl. Phys.*, vol. 46, p. 074001, Feb. 2013.

[25] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature Nanotech.*, vol. 10, pp. 187–191, Mar. 2015.

[26] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive random access memory," *Proc. IEEE*, vol. 104, pp. 1796–1830, Oct. 2016.

[27] 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. Nov., pp. 530–548, Nov. 2017.

[28] 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.

[29] 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.

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

[31] 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, pp. 1201–1226, Dec. 2017.

[32] J. Z. Sun, R. P. Robertazzi, J. Nowak, P. L. Trouilloud, G. Hu, D. W. Abraham, M. C. Gaidis, S. L. Brown, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge, "Effect of subvolume excitation and spin-torque efficiency on magnetic switching," *Phys. Rev. B*, vol. 84, p. 064413, Aug. 2011.

[33] 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, p. 073910, Apr. 2011.

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

[35] Y. Shiota, T. Nozaki, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, S. Yuasa, and Y. Suzuki, "Evaluation of write error rate for voltage-driven dynamic magnetization switching in magnetic tunnel junctions with perpendicular magnetization," *Appl. Phys. Express*, vol. 9, p. 013001, Jan. 2016.

[36] D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, pp. 1816–1823, 1990.

[37] 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.

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

[39] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 3rd ed., ser. Wiley series Prob. and Stat. Hoboken, NJ: Wiley, 2017.

[40] 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.

[41] 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.

[42] C. J. García-Cervera, "Numerical micromagnetics: A review," *Bol. Soc. Esp. Mat. Apl.*, vol. 39, pp. 103—135, 2007.

[43] A. Brataas, A. D. Kent, and H. Ohno, "Current-induced torques in magnetic materials," *Nature Mater.*, vol. 11, pp. 372–381, May. 2012.

[44] R. V. Kohn and V. Slastikov, "Effective dynamics for ferromagnetic thin films: a rigorous justification," *Proc. R. Soc. Lond. Ser. A*, vol. 461, no. Jan., pp. 143–154, 2005.

[45] C. J. Garcia-Cervera and W. E, "Effective dynamics in thin ferromagnetic films," *J. Appl. Phys.*, vol. 90, pp. 370–374, 2001.

[46] R. G. Lund, 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.

[47] 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, p. 033901, Jul. 2013.

[48] 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. 184 105–184 105–12, May. 2013.

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

[50] 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.

[51] 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, pp. 1013–1024, Mar. 1995.

[52] 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.

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

[54] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, ser. Springer Series in Synergetics. Berlin: Springer-Verlag, 2004.

[55] P. Glasserman and Y. Wang, "Counterexamples in importance sampling for large deviations probabilities," *Annals Appl. Prob.*, vol. 7, no. 3, pp. 731–746, Aug. 1997.

[56] 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.

[57] 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.

[58] 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, Nov. 2007.