

ON SPIN-INSPIRED REALIZATION OF
QUANTUM AND PROBABILISTIC COMPUTING

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Brian M. Sutton

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Supriyo Datta, Chair

School of Electrical and Computer Engineering

Dr. Gerhard Klimeck

School of Electrical and Computer Engineering

Dr. Yong Chen

School of Electrical and Computer Engineering

Dr. Zhihong Chen

School of Electrical and Computer Engineering

Approved by:

Dr. Michael A. Capano

Head of the Graduate Program

This dissertation is dedicated to Johanna.

ACKNOWLEDGMENTS

I would like to thank my family and the many friends, colleagues, and teachers who have influenced my life and career. They have always encouraged me to have an open and inquisitive approach to learning.

I would like to thank Vinh Diep, Seokmin Hong, Samiran Ganguly, Sherin Sayed, Rafatul Faria, Zeeshan Pervaiz, Orchi Hassan, Shuvro Chowdhury, Anirudh Ghan-tasala, and Jan Kaiser for many engaging group discussions over the years.

I would also like to thank Kerem Camsari for his continual support, guidance, optimism, stimulating perspectives, and friendship.

I would especially like to thank Professor Supriyo Datta for the thought-provoking and insightful discussions that helped shaped this work. He has a remarkable ability to translate complex concepts into an easily understood form. I am grateful to have had the opportunity to work with such an exceptional educator during my time at Purdue. And of course, I would be remiss to not mention the exceeding patience and kindness he demonstrated throughout my degree.

During my professional career, I had the opportunity to pursue graduate studies. This opportunity was in large part due to the support and encouragement of my manager, Jim Gallagher. I want to thank Jim for many years of advice, guidance, and friendship.

I would like to thank my parents for teaching me the importance of hard-work and commitment, characteristics that were invaluable during my academic studies.

Lastly, I would like to thank my wife Johanna for her boundless support, encouragement, and friendship, which helped lead me to where I am today.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	ix
1 INTRODUCTION	1
2 SPIN TORQUE DRIVEN QUANTUM COMPUTING	6
2.1 Quantum Computing with Dots and Donors	6
2.2 Manipulating Quantum Information with Spin Torque	7
2.3 Architecture	8
2.4 Gate Operations	13
2.4.1 Single Qubit Operations	13
2.4.2 Two Qubit Operations	19
2.5 Discussion	33
3 ANNEALING, THE ISING MODEL, AND COMBINATORIAL OPTIMIZA- TION	36
3.1 Solving NP-Hard Problems	36
3.2 Classical Annealing with Stochastic Nanomagnets	37
3.3 Engineering Correlations Through Spin Currents	38
3.3.1 Steady-State Fokker-Planck Description	41
3.3.2 Stochastic Landau-Lifshitz-Gilbert (LLG) Model	42
3.4 Combinatorial Optimization	43
3.5 Considerations for Physical Realization	46
4 AUTONOMOUS PROBABILISTIC COMPUTING	52
4.1 ApC Implementation	56
4.2 Combinatorial Optimization	59

	Page
4.3 Quantum Emulation	60
4.4 Discussion	64
4.5 Methods	67
4.5.1 FPGA p-computing coprocessor	67
4.5.2 FPGA ApC Logical Organization	68
4.5.3 Digital p-bit Implementation Flexibility	71
4.5.4 Benchmarking Eq. (4.5) with stochastic LLG	74
5 LOOKING AHEAD	78
5.1 Quantum and Probabilistic Computing	78
5.2 Realization of Autonomous Probabilistic Computing	79
REFERENCES	81
A CHAPTER 2 SUPPLEMENTARY INFORMATION	98
A.1 Derivations of equations (2.4) and (2.5)	98
A.2 Derivation of Equation (A.4)	100
A.3 Matrices $\tilde{\mathbf{S}}_1$ and $\tilde{\mathbf{S}}_2$ in equations (2.12) and (2.13)	103
A.4 Derivation of (2.21)	105
B STOCHASTIC FERROELECTRIC CAPACITOR	110
B.1 SPICE Model	110
VITA	112

LIST OF TABLES

Table	Page
2.1 Essential Requirements	11
3.1 Options for Physical Realization	49
4.1 Comparison of Sequential and Asynchronous Ising Computers	65

LIST OF FIGURES

Figure	Page
2.1 Quantum Computing with “Spin Torque”	9
2.2 Pseudo-Unitary Evolution	14
2.3 Single and Two Qubit Reflection Matrices	15
2.4 Single Qubit Rotation	17
2.5 Model System for Two-Spin Bell State Preparation	21
2.6 Bell-State Oscillation	25
2.7 Figure of Merit	27
2.8 Two Qubit Rotation	28
2.9 Error Minimization for Practical $\sqrt{\text{SWAP}}$	30
2.10 Controlled-NOT	31
3.1 Response of Stochastic Nanomagnet to Spin Current	40
3.2 Boolean Satisfiability	45
3.3 Annealing Stochastic Nanomagnets for the Traveling Salesman Problem . .	47
3.4 Possibilities for Physical Realization	50
4.1 Autonomous p-Computer (APC)	55
4.2 Max-Cut Combinatorial Optimization	58
4.3 Emulating the Transverse Ising Hamiltonian	62
4.4 Cloud Accessible p -computing Co-processor	69
4.5 Logical Organization and Memory Map	70
4.6 Modular Autonomous Digital p-bit Components	73
4.7 Benchmarking the PPSL Model with sLLG using Euclidean Distance . . .	75
4.8 Benchmarking the PPSL Model with sLLG using Free Energy	76

ABSTRACT

Sutton, Brian M. PhD, Purdue University, December 2019. On Spin-inspired Realization of Quantum and Probabilistic Computing. Major Professor: Supriyo Datta.

The decline of Moore’s law has catalyzed a significant effort to identify beyond-CMOS devices and architectures for the coming decades. A multitude of classical and quantum systems have been proposed to address this challenge, and spintronics has emerged as a promising approach for these post-Moore systems. Many of these architectures are tailored specifically for applications in combinatorial optimization and machine learning. Here we propose the use of spintronics for such applications by exploring two distinct but related computing paradigms. First, the use of spin-currents to manipulate and control quantum information is investigated with demonstrated high-fidelity gate operation. This control is accomplished through repeated entanglement and measurement of a stationary qubit with a flying-spin through spin-torque like effects. Secondly, by transitioning from single-spin quantum bits to larger spin ensembles, we then explore the use of stochastic nanomagnets to realize a probabilistic system that is intrinsically governed by Boltzmann statistics. The nanomagnets explore the search space at rapid speeds and can be used in a wide-range of applications including optimization and quantum emulation by encoding the solution to a given problem as the ground state of the equivalent Boltzmann machine. These applications are demonstrated through hardware emulation using an all-digital autonomous probabilistic circuit.

1. INTRODUCTION

The decline of Moore’s law [1] has generated a significant push to identify beyond-CMOS computing paradigms that can satisfy the commercial demand for ever improving devices. The benefits lost due to the physical limits of device scaling are likely to be overcome through new architectures and computational approaches [2,3], especially those with non-Von Neumann architectures. Already quantum [4,5] and bioinspired architectures [6,7] are opening the door to exciting new possibilities including machine learning [6,8,9], combinatorial optimization [10–18], inference problems [19–21], and general purpose acceleration [22,23]. While general purpose processors can be used in these applications, often hardware tailored to specific applications [22] is leveraged to provide performance gains. As these applications continue to emerge, hardware designed to accelerate their performance [24] are likely to be of significance, especially when provided as a service [25], for example via cloud-based computing.

Amongst the many approaches being pursued, spintronics has emerged as a leading candidate for satisfying this need with demonstrated success in memory applications [26]. The rich array of observed spintronic phenomena [27–36] provide immense promise for new engineering applications beyond those of digital storage. However, after its initial foray into logic applications [37,38], spintronics appears to fall short of the capabilities of CMOS when assessed from an energy-delay perspective [39]. This present limitation of spin in traditional logic applications has motivated the exploration of spintronics in non-traditional, or non-Boolean, applications such as quantum computing [40–49] and more recently probabilistic spin logic [17,20,50].

These two fields, quantum computing and probabilistic computing, in many ways originated with thoughts Feynman gave in his seminal work on simulating physics with computers [51]. In this talk, he explored the use of different computing approaches to efficiently simulate physical systems. Understanding that nature is governed by

quantum phenomena that are inherently unpredictable, he first explored the use of purely probabilistic, though classical, means to simulate the world. Such an approach would lead to a situation in which the classical computer would explicitly disagree with physical reality, effectively recasting Bell’s inequality with additional clarity [52]. Feynman’s conclusion was that the only efficient means to simulate physical reality was with a system governed by its same fundamental constituents: quantum phenomena. Indeed this thought spawned significant study in quantum computing [53–58] that has burgeoned into a vast multi-disciplinary field [4] of which spintronics is a strong contender [40–49]

Given this potential for revolutionary computational power provided via quantum computing, there has been a significant effort to build systems demonstrating this advantage with companies like Google, Intel, IBM, and Microsoft [59] all pursuing the development of commercially viable quantum computers demonstrating “quantum supremacy” [60]. However, there is uncertainty about when scalable quantum computers will be viable [61]. While it is believed by many that experimental quantum supremacy may be demonstrated in the next few years [62], we are likely a decade or more away from a scalable quantum computer that will support large-scale practical use of the numerous quantum algorithms [63] that have been identified.

This prospect for quantum computing, though exciting, has highlighted a computational performance gap between modern computing devices and the scalable quantum computers ahead. It may be possible to help bridge this gap with non-traditional architectures and approaches that overcome the von Neumann bottleneck, disrupting the computational landscape [3]. While quantum approaches may change the asymptotic behavior of various problems, such as factorization, these non-von Neumann approaches seek to improve the constant pre-factor of the classical algorithms for these problems through innovative software and hardware co-design. In this work, we will discuss one such non-von Neumann approach to computing that provides the enticing possibility of bridging the computational gap between the classical systems of today and the quantum computers of tomorrow.

Prior to introducing the necessity of quantum computers to simulate reality, Feynman first suggested a computing approach that uses intrinsically probabilistic devices and highlighted their corresponding limitations. While probabilistic systems are not able to efficiently simulate reality, their practicality for other applications is still relevant. Such probabilistic computers fall naturally between two extremes, classical deterministic computers and the quantum systems as discussed above. Digital computers are built, using spintronics for the analogy, using collections of stable quantum-spins behaving classically as an ensemble. As these stable magnets are scaled down to individual spins [43], they are described using quantum physics [52] and in turn form the “bit” of a quantum computer, a qubit. Quantum computers constructed from single spins are, unsurprisingly, governed by quantum mechanics which is inherently probabilistic. In between these scaling extremes, a computer made from a collection of unstable spins [64] would provide a probabilistic, though classical computer. While it is possible that deterministic, digital, computers can efficiently simulate probabilistic computers [65], i.e. $P \equiv BPP$, in practice, randomness has been very useful to achieve polynomial time solutions to problems [66,67] while searching for an efficient, deterministic counterpart [68]. It is this utility of stochastic operation that helps motivate the exploration of probabilistic devices for beyond-Moore applications.

Algorithms that leverage randomness typically run on digital hardware and obtain entropy using dedicated True Random Number Generators (TRNGs) or perhaps even pseudo-random algorithms [69]. If instead of using RNGs simply as co-processors, the stable bits of the digital hardware are replaced with compact, distributed RNGs, a probabilistic circuit (p-circuit) comprising probabilistic bits (p-bits) is formed. Many of the applications discussed above can leverage such a probabilistic hardware as their implementation inherently involves randomness [70]. Probabilistic Spin Logic (PSL¹) seeks to explore the design and use of probabilistic hardware built-on p-bits as a natural representation of the stochastic elements used in these applications.

¹The use of “spin” in PSL was originally motivated by the strong ties to magnetic-devices. However, it also has a natural connection to spin-glass systems used in a wide-range of problems.

Indeed such stochastic elements are used in several neuromorphic applications forming the neuron for neural networks. Several computational models, such as the Ising model [71] and artificial models for neuromorphic hardware [72], are based on the use of weighted summations where a “neuron” is influence via a weighted sum of “synapse” inputs. The output of a neuron is governed by one of several types of response function such as linear, binary threshold, sigmoidal [73], and stochastic [74, 75] where the output of the neuron has a probabilistic response function

$$p(x_i = 1) = \frac{1}{1 + e^{-b_i + \sum_j x_j w_j}} \quad (1.1)$$

If a natural physical implementation of such a probabilistic bit could be found it would serve as a platform upon which to build dedicated hardware to accelerate applications based on these neurons. As proposed by Behin-Aein et al., magnets [20, 76] are a natural building-block with which to form probabilistic bits, and provide the possibility of continuous, asynchronous operation at GHz speeds, rapidly exploring the configuration space of a problem [17, 77]. When organized into relevant topologies [9, 73], these p-bits form circuits with useful characteristics for post-Moore applications.

In this work, we continue this exploration into applications motivated by the use of spintronics with a specific focus on: (a) extending the established field of spin-based quantum computing to use a non-traditional method of qubit control, manipulation, and readout with spin currents [78], (b) revisiting probabilistic computing with the use of stochastic nanomagnets that are tuned with spin currents [17] to show how the approach can be used to solve similar problems as those being explored in adiabatic quantum computing applications [70], and (c) provide [79] a concrete demonstration of an autonomous probabilistic computer, highlighting its potential to bridge the gap between the limitations of today’s computing devices for such applications and tomorrow’s quantum computers.

The remainder of this report is outlined as follows:

- Chapter 2 will discuss a proposal for a quantum computer based on repeated interaction with ancillary spins as would be realized through carefully controlled spin-currents in a semiconductor
- Chapter 3 will introduce the use of low-barrier stochastic nanomagnets to solve combinatorial optimization problems using the natural behavior of a physical network of coupled devices as an alternative to solving combinatorial optimization problems
- Chapter 4 provides an exploration of autonomous probabilistic computing that eliminates the use of sequencing logic, projecting the possibility of building a chip with petaflips per second with millions of neurons
- Chapter 5 will summarize this work and provide a look ahead, providing some suggested areas of exploration for probabilistic computing

2. SPIN TORQUE DRIVEN QUANTUM COMPUTING

The material in this chapter is based on our Scientific Reports publication “Manipulating quantum information with spin torque” [78] along with some unpublished results.

Over the past several decades quantum computing has emerged as a thriving area of research with multiple physical systems being explored for its realization [4]. Scalability is one of the primary challenges for the realization of these computers and no approach has emerged as the clear answer. To-date only a handful qubits [80] have been successfully manipulated in a coherent fashion to perform traditional gate-based quantum computing (distinct from adiabatic quantum computing [81] that now claims to have over 1000 qubits [82]). One of the limitations to improving the number of qubits is the ability to perform long-range coherent manipulation of qubits in addition to improving their coherence times.

In this chapter, we focus on a leading semiconductor based quantum computer that may provide the answer to scalability. These semiconductor quantum computers do not currently have a convenient mechanism to perform all on-chip single qubit and long-range two qubit interactions. Here, we discuss an alternative method of controlling the qubits that may reduce the required control necessary for implementation: spin-currents.

2.1 Quantum Computing with Dots and Donors

One influential proposal [40, 41] that seeks to satisfy the DiVincenzo criteria [42] is based on the use of donor and nuclear spins of phosphorous ^{31}P atoms in a silicon matrix and much experimental progress has been reported in the last fifteen years towards the realization of structures that could enable proposals of this type [43–49].

In such systems, typically individual qubits, \vec{S} , are selected and rotated using either AC electric or magnetic fields for nuclear magnetic resonance (NMR) and electron spin resonance (ESR) [44], while two qubit operations are realized by activating an effective exchange interaction $J\vec{S}_1 \cdot \vec{S}_2$ between them. These structures require DC magnetic fields and tuned resonant frequencies [83] in order to provide individual addressability [46] of qubits on a chip. This peripheral circuitry represents an obstacle that must be overcome for scalability and a completely on-chip solution.

These systems currently provide a robust method with which to perform single qubit initialization and single-shot readout of the state [43,46], a necessary component of any quantum computing architecture. Furthermore, these nuclear spin qubits provide excellent coherence times at low temperatures. Significant advances have been made in the long-term storage of quantum information in semiconductor systems based around donor and defect spins with experimental results of ^{31}P nuclear spin T_2 times ~ 30 seconds [84], and high-purity silicon donor electrons with T_2 times \sim seconds [85,86]. High-purity silicon also shows promise for spin transport with spin lifetimes of over 500 ns at 60 K reported in undoped Si [87].

The question we seek to explore in the remainder of the chapter is whether or not we can improve on our ability to control the spin qubit while maintaining the excellent properties that have been experimentally demonstrated by these systems.

2.2 Manipulating Quantum Information with Spin Torque

The use of spin torque as a substitute for magnetic fields is now well established for classical operations like the switching of a nanomagnet. Here, our primary objective is to show that “spin torque” like effects can be used to implement quantum processes involving single qubit initialization and rotation as well as two qubit entanglement. Qubit readout using ensemble-measurement can be implemented using the same architecture if a collection of identically initialized and transformed qubits, prepared using replicated physical structures, are available for measurement. Alternatively

the proposed architecture could be used in conjunction with established single shot readout techniques [44], especially for specific applications requiring multi-qubit state tomography or Bell state experiments [49].

In this chapter we will first show that (1) all standard single qubit operations can be effected without any magnetic field through interactions of the form $J\vec{\sigma} \cdot \vec{S}$ with the itinerant or “flying” non-equilibrium spin population $\vec{\sigma}$ while (2) two qubit operations can be implemented through separate interactions of the form $J\vec{\sigma} \cdot \vec{S}_1$ and $J\vec{\sigma} \cdot \vec{S}_2$ with the flying spin population $\vec{\sigma}$. The latter process has been discussed earlier by several authors [88–98] and we draw on this work, but there is a key distinction with the present work as will be explained.

2.3 Architecture

The overall architecture envisioned is shown schematically in Figure 2.1 using four localized spins for illustrative purposes. A complete implementation could include additional qubits as well as multiple versions of the same qubit to allow ensemble readout. All qubits are embedded in a spin-coherent semiconductor channel so that the itinerant or “flying” (f) spins in the conduction band interact with the static (S_i) qubits located at x_i through an interaction of the form

$$\mathcal{H} = J\vec{\sigma} \cdot \vec{S}_i \delta(x - x_i) \quad (2.1)$$

The static qubit could be a charge-neutral nuclear spin like ^{29}Si with J representing the hyperfine interaction, or ^{31}P where the donor level is used to mediate a hyperfine interaction, or it could be the electronic donor spin with J representing the exchange interaction, or perhaps a nanoscale magnet embedded in the semiconductor [99, 100].

Each qubit has gates, R_i and G_i , on either side that can be used to deplete the channel underneath to couple and decouple the qubits from the itinerant spins as desired for specific operations. The spin-coherent channel has additional gates, B_i , that provide isolation and direct the flow of electrons. These gates could be realized

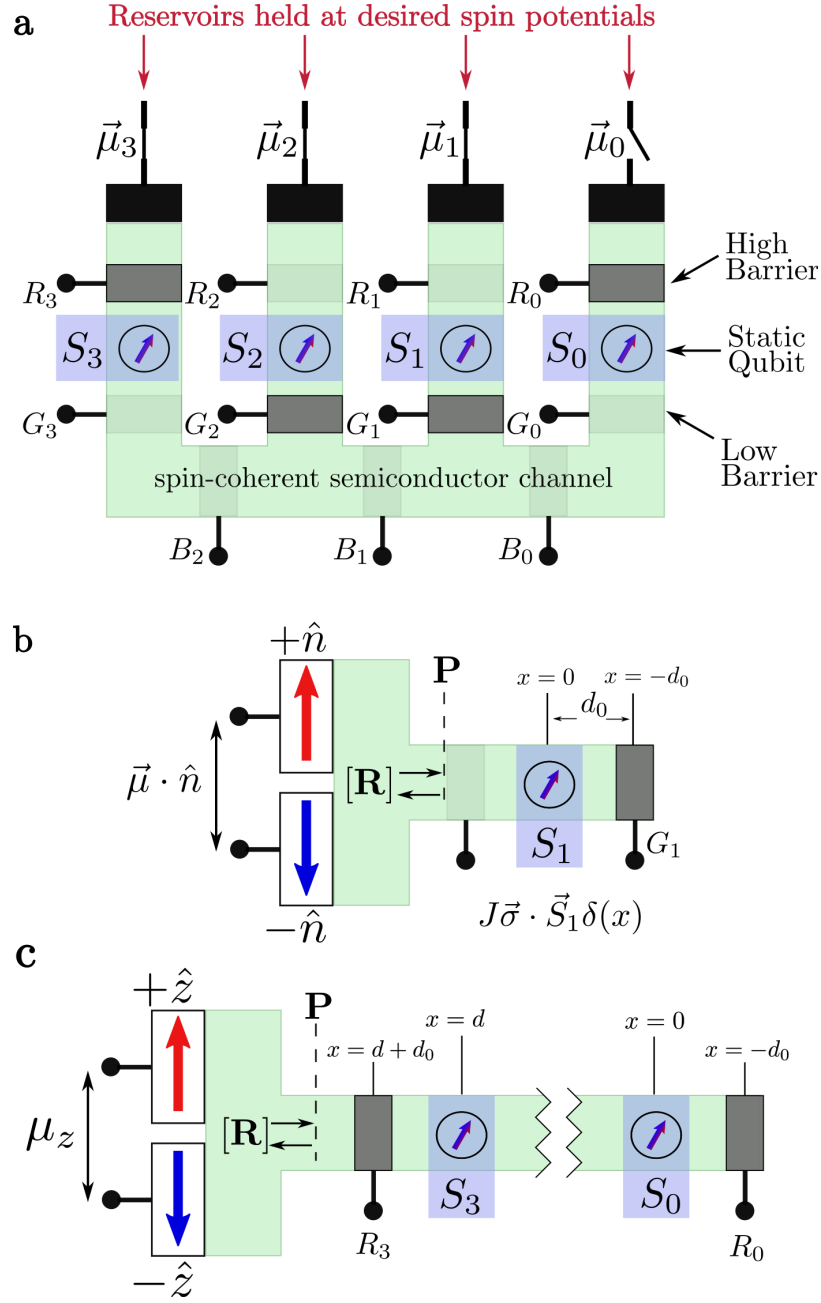


Fig. 2.1. **Quantum computing with “spin torque”**: (a) Schematic showing overall architecture with qubits S_1 and S_2 configured for single qubit operations and qubits S_0 and S_3 for two qubit operations. (b) Equivalent configuration for qubit S_1 redrawn in one dimension. (c) Equivalent configuration for qubits S_0 and S_3 redrawn in one dimension.

using top-gates [84] or through contacts capable of modulating the electrostatics of the channel [46]. A general itinerant electron has a low probability of interacting with a static nuclear spin due to the minuscule size of the nucleus. In order to effect an interaction, the gates, R_i and G_i , are used to produce standing waves for the itinerant electrons. These standing waves should in-turn provide the necessary wavefunction overlap to realize the coupled interaction with a static spin [101].

The semiconducting channel could be realized using silicon [102] or other material supporting coherent spin injection and transport [103, 104]. Itinerant spins can be connected as desired to spin reservoirs held at specific spin potentials [105] along the x, y or z directions. These can be generated using various well-established spintronic phenomena such as magnetic contacts [106, 107], the giant spin Hall effect [32, 108], or spin pumping [34], at both low [109] and room temperatures [110, 111]. Integration of semiconductors with magnetic materials is a viable prospect [36, 112] and has been used in silicon double dot experiments to generate local magnetic fields [113], establishing precedent for the prospect of device fabrication. Table 2.1 summarizes these experimental demonstrations and also outlines the essential requirements necessary for realization of this proposal.

Figure 2.1(a) shows qubits S_1 and S_2 configured for single qubit operations and can be redrawn in one dimension as shown in Figure 2.1(b). Qubits S_0 and S_3 on the other hand are configured for two qubit operations and can be redrawn as shown in Figure 2.1(c). Note that in either case the two reservoirs shown are purely conceptual; in practice these would likely be a single reservoir with a spin potential $\vec{\mu}$ in some direction \hat{n} . With a single contact under steady-state conditions there is no net current flow, however there is a continual exchange of electrons. Electrons are preferentially injected from the contact with spins in direction \hat{n} which interact with the static qubits and are then removed by the same contact. It is this flow of electrons to and from the same contact that drives the scattering phenomena described in the following sections.

Requirement	Discussion
<p>1. Spin-injection into semiconductors</p> <p><i>Threshold:</i> $p > 0$</p> <p><i>Objective:</i> $p = 1.0$</p> <p><i>Precedent:</i></p> <p>$p = 0.1 - 0.3$ in Si at 5 K [114]</p> <p>$p = 0.046$ in n-Si at 300 K [111]</p>	<p>a. Polarization is highly temperature dependent</p> <p>b. Can compensate for low polarization (e.g. $p = 0.01$) through modulation of N via α if barrier height/position is tunable</p> <p>c. $p = 1.0$ is not a fundamental requirement for gate operations</p> <p>d. Heusler alloys may enable higher polarizations [36]</p> <p>e. Two distinct spin-potentials ($\vec{\mu}$) needed for universality</p> <p>f. Polarized current not required for two-qubit operations</p>
<p>2. Coherent spin transport</p> <p><i>Threshold:</i> $\lambda_S \propto d, d_0$</p> <p><i>Objective:</i> $\lambda_S = \infty$</p> <p><i>Precedent:</i></p> <p>$\lambda_S \approx 10\mu\text{m}$ in Si at 85 K [102]</p> <p>$\lambda_S \approx 5\mu\text{m}$ in graphene at room temp [104]</p>	<p>a. Minimum distance governed by proximity of qubits to each other and to neighboring barriers</p> <p>b. Low-temperatures best for long spin-coherence length</p> <p>c. Isotopically enriched semiconductor ideal for coherence</p> <p>d. Non-locality directly related to obtainable λ_S</p> <p>e. Small λ_S may still facilitate nearest-neighbor approaches</p> <p>f. Channel confinement necessary to ensure “1D” operation</p>
<p>3. Exchange/Hyperfine coupling between itinerant and static spins:</p> $\mathcal{H} = J\vec{\sigma} \cdot \vec{S}_i \delta(x - x_i)$ <p><i>Precedent:</i></p> <p>Nuclear polarization with spin-current [101]</p>	<p>a. ^{31}P donors ideal for Si, but interaction with donor electron warrants discussion</p> <p>b. Itinerant spins simultaneously coupled to static qubits may complicate this model</p> <p>c. Quasi-static model used to describe interaction</p> <p>d. Alternate qubits may fit into this model (e.g. quantum dots, nanomagnets)</p>
<p>4. Single-shot initialization and readout compatibility</p> <p><i>Precedent:</i></p> <p>^{31}P donors in Si [43, 44]</p>	<p>a. Spin polarization of 1.0 needed for fiducial state preparation</p> <p>b. Single-shot readout needed for correlation measurements and traditional implementations of quantum algorithms</p> <p>c. Architecture is compatible with demonstrated single-shot initialization and readout mechanisms</p> <p>d. Ensemble-measurement computing is inherently supported</p>

Table 2.1.

Essential Requirements: There are a number of requirements that must be satisfied in order for the proposed architecture to be viable. Shown in the table are a few of the most essential requirements for the proposal along with precedent for their satisfaction.

The density matrix of the incident spins from this reservoir is given by $\rho_f = [\mathbf{I} + \hat{n} \cdot \vec{\sigma}]/2$, where \mathbf{I} is the 2×2 identity matrix and $\vec{\sigma}$ the Pauli spin matrices. The Kronecker product of this flying spin density matrix with the $2^q \times 2^q$ density matrix ρ_S describing the q-qubit system ($q = 1$ for Figure 2.1(b), $q = 2$ for Figure 2.1(c)) gives the initial overall $2^{q+1} \times 2^{q+1}$ density matrix of the system. The initial density matrix gets modified to

$$\rho = [\mathbf{R}][\rho_f \otimes \rho_S][\mathbf{R}^\dagger] \quad (2.2)$$

by the reflection process described by a $2^{q+1} \times 2^{q+1}$ reflection matrix $[\mathbf{R}]$ which is computed taking into account the barrier(s) and the interaction of the itinerant spins with the qubits and depends on the specific structure at hand.

The reflected itinerant spins are returned to the spin reservoir ($+\hat{n}$ or $-\hat{n}$) causing a collapse of the quantum state described by a partial trace of the density matrix over the flying spins represented by Trace_f :

$$\rho_S(n+1) = \text{Trace}_f ([\mathbf{R}][\rho_f \otimes \rho_S(n)][\mathbf{R}^\dagger]) \quad (2.3)$$

Equation (2.3) defines the basic approach used to model the quantum gates discussed in this chapter. It provides a recursive relation expressing the q-qubit density matrix after interacting with $(n+1)$ itinerant spins in terms of the density matrix after interacting with n itinerant spins. A time-independent model for quantum transport is used assuming that the time variation of signals is slow enough to be treated as quasi-static. For example, $\hbar/1 \text{ ns} \approx 1 \text{ } \mu\text{eV}$ which is much smaller than other energy scales of interest.

The use of “flying spins” to manipulate static qubits has been discussed in the past [88–98] and it has been noted that the reflection matrix $[\mathbf{R}]$ in equation (2.2) represents a unitary transformation suitable for quantum operations if the barriers at $x = -d_0$ in Figs. 2.1(b) and 2.1(c) are large enough to reflect the incident electrons completely [97,98], which we employ in this proposal. What is new about the present proposal’s method is the use of sequential interactions with a large number of itinerant electrons, each interaction involving a process of entanglement and reflection, equation

(2.2), followed by a collapse, without post-selection, of the quantum state, equation (2.3), resulting in a deterministic, approximately unitary operation. Every interaction evolves the density matrix according to equation (2.3) which can be used iteratively to determine the final density matrix after interacting with a specified number of electrons.

Note that unlike equation (2.2), the collapse of the density matrix described by equation (2.3) is a non-unitary process and it may seem surprising that an overall operation involving a large number (N) of such non-unitary collapses could still be useful for implementing unitary transformations suitable for quantum computing. However, we will show that with proper choice of parameters the degree of non-unitarity can be made arbitrarily small at the expense of speed.

We seek to show that the non-unitarity that is inevitable with multiple collapses can be held to acceptably low levels so that useful quantum gates can be implemented. This is established first for single qubit operations and then for two qubit operations using the basic approach embodied in equations (2.2) and (2.3). Finally, as a capstone example, we present the implementation of a complete CNOT gate using the proposed architecture, and show that the fidelity under ideal conditions can be made acceptably close to one.

2.4 Gate Operations

2.4.1 Single Qubit Operations

In order to further understand the prescription of Equation 2.3, consider the case of a single nuclear spin repeatedly interacting for a time t with a donor electron governed by the Hamiltonian $H = A\vec{S}_1 \cdot \vec{S}_2$. This can be accomplished with repetitive population and depopulation of the donor state using a SET as in [43].

This system can be modeled using the repetitive measurement process as depicted in Figure 2.2(b). Each load and unload of the donor electron systematically produces some degree of entanglement between the donor and nuclear spin and subsequent col-

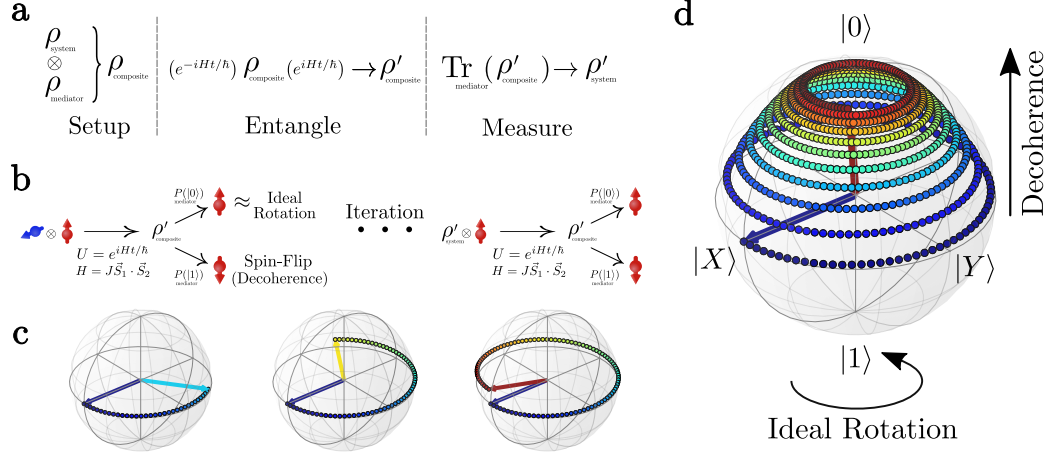


Fig. 2.2. Pseudo-Unitary Evolution: (a) A single iteration of the repeated entanglement and measurement process is depicted. A mediating particle (ρ_{mediator}) and a target quantum system (ρ_{system}) are prepared, both are then evolved to produce some level of entanglement, and finally the ρ_{mediator} is measured. (b) This iterator is applied multiple times to the target system to produce the desired pseudo-unitary operation. Shown in this example are two spin- $\frac{1}{2}$ particles that are entangled through a time-evolved exchange interaction. On each measurement of the mediator (red), the spin will have either flipped with probability $P(|1\rangle)$ or will have remained up with probability $P(|0\rangle)$. If the mediator remained up, the system (blue) will have experience an ideal rotation about the axis specified by the spin-direction of the mediator. (c) If the approximately ideal rotations are plotted on the bloch sphere for various iterations, the pure state evolution can be seen with only a minimal error introduced represented by the off-axis component of the spin vector. (d) The spin-flip processes, coupled with the ideal rotation, produce a mixed state that ultimately converges to the spin-direction of the mediating particle. As shown, an initial state $|X\rangle$ rotates about \hat{z} with each iteration until eventually converging towards $|0\rangle$. The rate of decoherence is related on the interaction strength J with large couplings increasing the rate of decoherence.

lapse of the composite state, see Figure 2.2. This repetitive process produces approximately ideal rotations of the nuclear spin, though with some degree of decoherence, Figure 2.2(d)).

In this example, all of the qubits were stationary, or at least localized, during evolution. Remarkably, even with the repetitive measurement and collapse of the quantum state, an effective rotation can be realized of the nuclear spin state. We can turn this time-based view and extend it to one in which the rotation-mediating electron is now mobile and produced with a spin current.

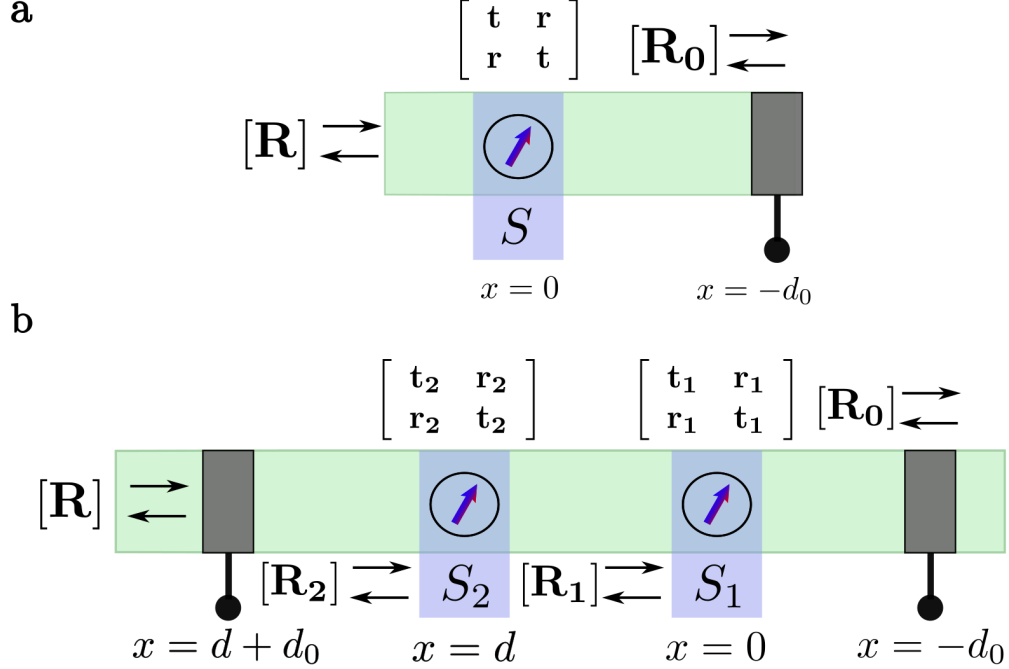


Fig. 2.3. **Single and Two Qubit Reflection Matrices:** Reflection matrix $[R]$ for (a) single qubit operations and (b) two qubit operations.

Figures 2.1(b) and 2.3(a) show the basic configuration for a single qubit operation: in the following discussion we will assume that \hat{n} points along the z-direction, so that the reservoir injects electrons with $+z$ spins and extracts both $\pm z$ spins. Every time an electron is injected it gets entangled with the static spin, while the extraction represents a measurement that collapses the quantum state of the static spin. We

show in Appendix A.1 that after interaction with N electrons, the z-component of the static spin \vec{S} is given by

$$s_z(N) = 1 - (1 - s_z(0))(\cos \alpha)^{2N} \quad (2.4)$$

while the transverse component $s_t \equiv s_x + is_y$ is given by

$$s_t(N) = s_t(0)e^{iN\alpha}(\cos \alpha)^N \quad (2.5)$$

where α represents the effective interaction strength between the flying spin and the static qubit and is given by

$$\begin{aligned} \tan \alpha &= \frac{8\Omega \sin^2 kd_0}{1 - 2\Omega \sin(2kd_0) - 6\Omega(1 - \cos 2kd_0)} \\ &\rightarrow 8\Omega(kd_0)^2, \text{ if } kd_0 \ll 1 \end{aligned} \quad (2.6)$$

where k is the wavenumber of the itinerant electrons, v the corresponding velocity and

$$\Omega \equiv J/\hbar v \quad (2.7)$$

Equations (2.4) and (2.5) describe our numerical results accurately, as evident from Figure 2.4, and provide the basis for single qubit initialization and rotation respectively.

Single qubit initialization: Equation (2.4) tells us that a static spin can be initialized in a state with $s_z \rightarrow 1$, after interaction with a large number of flying spins ($N \rightarrow \infty$). This is similar to the well-known Overhauser effect whereby nuclear spins get polarized through interaction with a spin reservoir driven out of equilibrium [27, 101] or other proposals for state purification using repeated measurements of a coupled quantum system with post-selection [115–117].

Single qubit rotation: Equation (2.5) suggests the possibility of single qubit rotation by an angle $\theta = N\alpha$ around the z-axis through interaction with the itinerant spins. Note, however, that in the process the spin is also attenuated by a factor

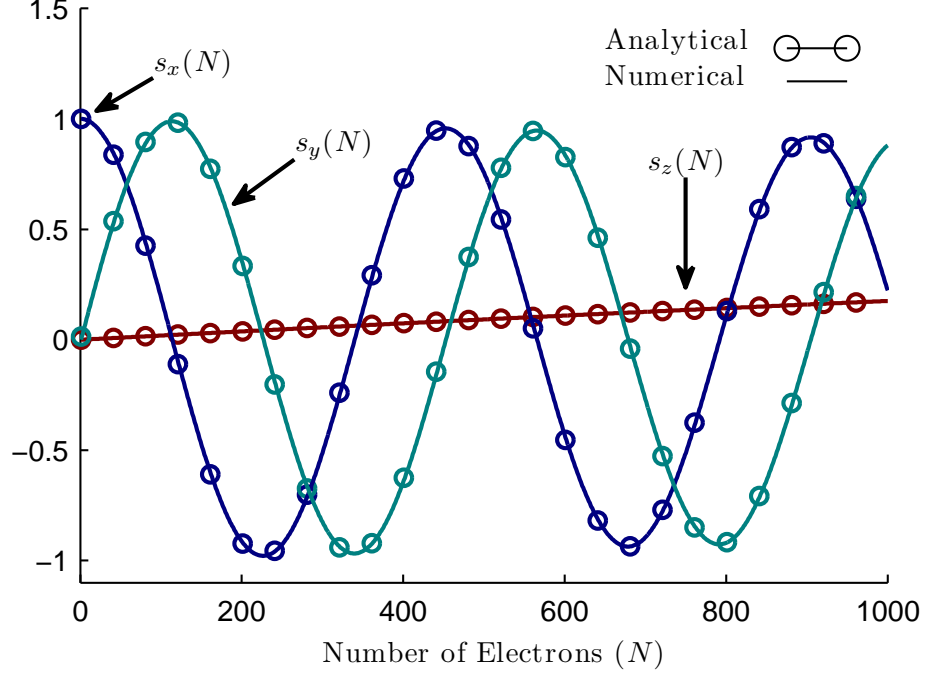


Fig. 2.4. **Single Qubit Rotation:** Evolution of the spin of a single qubit initially pointing along x as it interacts with an increasing number of flying spins, N . The numerical results are described very well by the analytical solutions (2.4) and (2.5) described in the text.

$(\cos \alpha)^N$ which is an undesirable side effect. For a given total rotation $\theta = N\alpha$ we can write the resulting error probability per unit rotation as

$$\frac{P_e}{\theta} = \frac{1}{\theta} \left(1 - \left(\cos \frac{\theta}{N} \right)^N \right) \approx \frac{\theta}{2N}, \text{ if } \frac{\theta}{N} \ll 1 \quad (2.8)$$

which shows that P_e can be made arbitrarily small for a given rotation θ by choosing a large N and hence a small α . The results in Figure 2.4 were obtained with a relatively large value of α with $N \sim 100$ spins for a rotation of $\theta = \pi/2$ in order to make the non-unitary effects apparent with a large error probability. But with $N \sim 10^4$, we have an error probability $P_e \sim 1.23 \times 10^{-4}$.

The use of a large N also gives enhanced control over the process since each electron makes only a small difference to the result. A large N requires a small α ,

and equation (2.6) suggests a convenient mechanism for the control of α , namely by adjusting the effective distance d_0 of the reflective barrier through the barrier voltage.

In this discussion, the polarization of the magnetic contact, \hat{n} , was taken as 100% in the z-direction. In practice, the polarization $P = (N_{\text{up}} - N_{\text{dn}})/(N_{\text{up}} + N_{\text{dn}})$ will be less than 100%. The angle of rotation will then be determined by $N_{\text{up}} - N_{\text{dn}}$, while the non-unitarity will depend on $N_{\text{up}} + N_{\text{dn}}$, making the error probability in equation (2.8) larger by $(1/P)$. For example, a spin polarization of 0.01 would increase the error probability by 10^2 . In order to compensate for this increase in error, we would need $(1/P)$ times more electrons, $N' = P^{-1}N$, which could be accomplished by making α smaller. In silicon, spin polarizations an order of magnitude larger than 0.01 have been obtained at low temperatures [114] with promising progress at room temperature [109].

Single qubit readout: Note that the interaction we just discussed also provides a mechanism for readout if we have multiple replicas of each qubit available so that ensemble measurements can be made, similar to [94]. We could measure the average spin current I_{sz} that flows initially at the terminals

$$(I_{sz})_{N=0} = \frac{dN}{dt} \frac{ds_z}{dN}_{N=0} = (1 - s_z(0)) \frac{dN}{dt} 2 \ln \cos \alpha \quad (2.9)$$

and deduce $s_z(0)$ from it. Similarly the initial spin in the x- and y-directions can be obtained by measuring the spin current that flows when connected to a spin reservoir with an x- and y-component respectively and using equations (2.4)(2.5). Knowing $s_i(0)$, $i = x, y, z$, we can write down the initial density matrix as follows:

$$\rho = \frac{1}{2} \left[\mathbf{I} + s_x(0) \sigma_x + s_y(0) \sigma_y + s_z(0) \sigma_z \right] \quad (2.10)$$

Ensemble measurement techniques provide a mechanism to obtain the expectation value of a given qubit and may be useful for certain computations [118]. However, as the architecture is compatible with single-shot readout methods, replicated physical structures are a non-essential aspect of the proposal.

2.4.2 Two Qubit Operations

Fig. 2.1(c) shows the basic configuration for a two qubit operation which is very similar the configuration for a single qubit operation, Fig. 2.1(b), except that the channel has two embedded qubits instead of one. The overall approach is the same, based on equation (2.3), but the reflection matrix $[\mathbf{R}]$ is 8×8 in size instead of 4×4 , making the algebra less straightforward.

In the two qubit subspace we are seeking to implement a unitary transformation of the form

$$\begin{matrix} & 12 & 1\bar{2} & \bar{1}2 & \bar{1}\bar{2} \\ \begin{matrix} 12 \\ 1\bar{2} \\ \bar{1}2 \\ \bar{1}\bar{2} \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \gamma & -i \sin \gamma & 0 \\ 0 & -i \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad (2.11)$$

which could be viewed as a “rotation” in the $1\bar{2} - \bar{1}2$ space: a rotation of $\gamma = \pi/2$ corresponds to a SWAP while $\gamma = \pi/4$ corresponds to the universal $\sqrt{\text{SWAP}}$ operation that we will use for the CNOT gate.

The approach used is based on the general principle of using itinerant spins as “messengers” that interact with the static spins through separate terms of the form $J\vec{\sigma} \cdot \vec{S}_1$, $J\vec{\sigma} \cdot \vec{S}_2$. Each messenger causes a small rotation, and an overall rotation is achieved through the integrated effect of many messengers.

To see how this works we need the reflection matrix $[\mathbf{R}]$ which is calculated using an extension of the method used in Supplemental Section A.1. The interaction with each of the qubits is described by transmission and reflection matrices given by

$$\mathbf{t}_1 = \left[\mathbf{I}_8 + i\Omega \tilde{\mathbf{S}}_1 \right]^{-1} \quad \text{and} \quad \mathbf{r}_1 = \mathbf{t}_1 - \mathbf{I}_8 \quad (2.12)$$

$$\mathbf{t}_2 = \left[\mathbf{I}_8 + i\Omega \tilde{\mathbf{S}}_2 \right]^{-1} \quad \text{and} \quad \mathbf{r}_2 = \mathbf{t}_2 - \mathbf{I}_8 \quad (2.13)$$

where $\tilde{\mathbf{S}}_1 = \vec{\sigma} \cdot \vec{S}_1$ and $\tilde{\mathbf{S}}_2 = \vec{\sigma} \cdot \vec{S}_2$ are 8×8 matrices (see Supplemental Section A.3 for the explicit form of these matrices) describing the interactions of the itinerant spin

with qubits 1 and 2 respectively. Note that the two qubit structure (Fig. 2.3(b)) has an additional barrier on the left with spin-independent transmission and reflection matrices given by

$$\mathbf{t}_B = \frac{1}{1 + i\Gamma} \mathbf{I}_8 \text{ and } \mathbf{r}_B = \mathbf{t}_B - \mathbf{I}_8 \quad (2.14)$$

where $\Gamma = U/\hbar v$ represents the barrier height normalized to $\hbar v$, assuming a delta function barrier $U\delta(x)$.

The overall reflection matrix $[\mathbf{R}]$ is calculated by repeated cascade of the reflection matrices for the structure in Fig. 2.3(b).

$$\mathbf{R}_1 = \mathbf{r}_1 + e^{i2kd_0} \mathbf{t}_1 [\mathbf{I}_8 - e^{i2kd_0} \mathbf{R}_0 \mathbf{r}_1]^{-1} \mathbf{R}_0 \mathbf{t}_1 \quad (2.15)$$

$$\mathbf{R}_2 = \mathbf{r}_2 + e^{i2kd} \mathbf{t}_2 [\mathbf{I}_8 - e^{i2kd} \mathbf{R}_1 \mathbf{r}_2]^{-1} \mathbf{R}_1 \mathbf{t}_2 \quad (2.16)$$

$$\mathbf{R}_3 = \mathbf{r}_B + e^{i2kd_0} \mathbf{t}_B [\mathbf{I}_8 - e^{i2kd_0} \mathbf{R}_2 \mathbf{r}_B]^{-1} \mathbf{R}_2 \mathbf{t}_B \quad (2.17)$$

It is straightforward to use equations (2.15)(2.16)(2.17) to calculate the reflection matrix $[\mathbf{R}]$ for a given set of parameters Ω , Γ , kd , and kd_0 .

Consider the two impurity spin system shown in Fig. 2.5(a). A single flying spin (σ_f) interacts with the two static spins (S_1, S_0). On either side of the static spins are two barriers, the rightmost barrier is a perfect reflector. The leftmost barrier is an arbitrary height Γ . The static spins are separated from each barrier by a distance d_0 and from each other by a distance of d .

Here we use a scattering model of the system, found by cascading of multiple reflection matrices together. We will approach the problem assuming $kd = \pi$ for ease of computation. With $kd = \pi$, the two static spins are effectively located at the same position. The Hamiltonian for the system is then given by:

$$\mathcal{H} = A\vec{\sigma}_f \cdot \vec{S}_1 \delta(x) + A\vec{\sigma}_f \cdot \vec{S}_0 \delta(x) \quad (2.18)$$

In this model, the two static spins have an effective distance $d = 0$ between them, but are non-interacting. Each static spin only interacts with the flying spin through a hyperfine coupling $A\vec{\sigma}_f \cdot \vec{S}_i$. Given the minuscule size of a nucleus, a typical conduction band-electron will not interact with the nucleus; however, the barriers present in

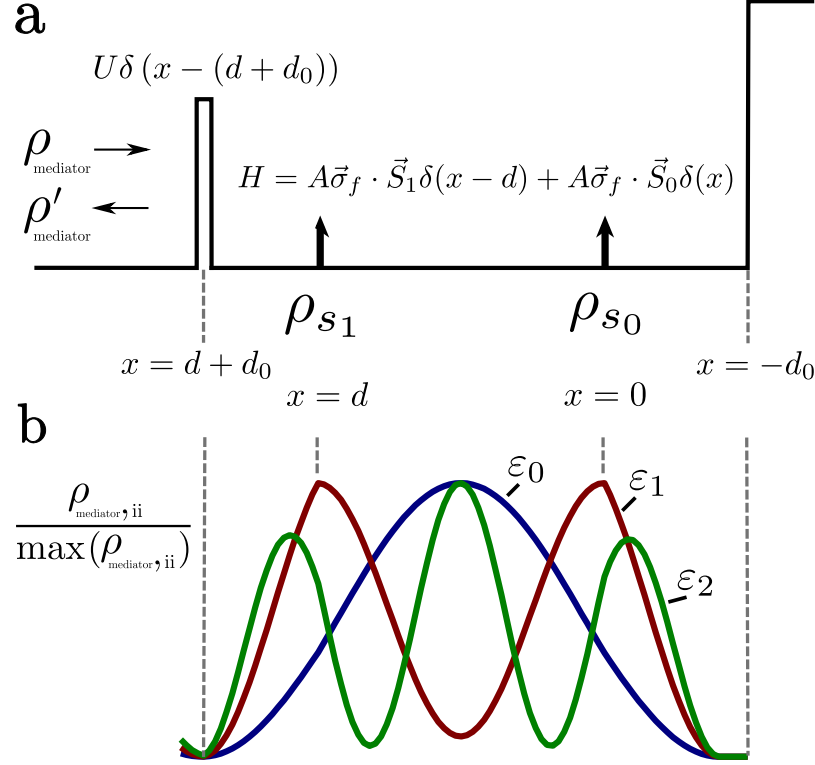


Fig. 2.5. **Model System for Two-Spin Bell State Preparation:** (a) Two stationary spins, ρ_{s1} and ρ_{s0} , are placed before an infinite barrier and after a short delta barrier $U\delta(x - (d + d_0))$. Repetitive interaction with itinerant spins, ρ_{mediator} , according to the interaction Hamiltonian, H , shown in the figure is used to model the dynamics of the two stationary spins. (b) The normalized mediating-spin densities for the three lowest eigenenergies, $\varepsilon_{0,1,2}$, are shown to convey a qualitative perspective of the itinerant spin standing waves. These spin-densities were obtained using numerical modeling of the system with an open boundary on the left.

the system produce standing-waves for the incident spins, producing the necessary wavefunction overlap for the hyperfine interaction. This can be seen in Fig. 2.5(b) which was producing using NEGF modeling of the system with a contact on the left.

We will solve this problem by partitioning the physical model into two regions, the first region is the barrier R_1 and the second is the two-static spin system and barrier R_0 . From the two scattering matrices for these regions, we can solve for the

total system scattering matrix \mathbf{R} . The general solution is found by cascading matrices based on the following expression:

$$R = r + e^{i2kd}t [I_8 - e^{i2kd}R_0r]^{-1} R_0t \quad (2.19)$$

where $t = [I_8 + i\Omega\tilde{S}]^{-1}$, $r = t - I_8$, and we let $\tilde{S} = \tilde{S}_1 + \tilde{S}_2$. Instead of solving the problem in the $2^3 \times 2^3$ subspace, we recognize that the system is block-diagonal. The reflection matrix \mathbf{R} has the form shown in equation (2.20), where the columns have been re-ordered to display the block-diagonal form of the matrix.

$$[\mathbf{R}] : \quad (2.20)$$

$$\begin{array}{c} f12 \quad f1\bar{2} \quad f\bar{1}2 \quad \bar{f}12 \quad f\bar{1}\bar{2} \quad \bar{f}1\bar{2} \quad \bar{f}\bar{1}\bar{2} \quad \bar{f}\bar{1}\bar{2} \\ \left(\begin{array}{cccccccc} f12 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ f1\bar{2} & 0 & a & c & c' & 0 & 0 & 0 \\ f\bar{1}2 & 0 & c & a & c'' & 0 & 0 & 0 \\ \bar{f}12 & 0 & c' & c'' & b & 0 & 0 & 0 \\ f\bar{1}\bar{2} & 0 & 0 & 0 & 0 & b & c' & c' \\ \bar{f}1\bar{2} & 0 & 0 & 0 & 0 & c'' & a & c \\ \bar{f}\bar{1}\bar{2} & 0 & 0 & 0 & 0 & c' & c & a \\ \bar{f}\bar{1}\bar{2} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array}$$

As this matrix is block-diagonal we can focus on the sub-spaces of interest, namely $f1\bar{2}, f\bar{1}2, \bar{f}12$ and $f\bar{1}\bar{2}, \bar{f}1\bar{2}, \bar{f}\bar{1}\bar{2}$. The $f12$ and $\bar{f}\bar{1}\bar{2}$ basis vectors remain unchanged by the scattering operation as spin is conserved. For the remainder of this discussion, we will focus on $f1\bar{2}, f\bar{1}2, \bar{f}12$ as the other sub-space is equivalent with spin-up and spin-down reversed.

The resulting expression for the Bell state wavefunction as a function of the number of electrons is given by (see Appendix A.4):

$$\Psi_0(n) = \frac{1}{2} \left[e^{i\chi n} + \left[\frac{1}{6}(-2e^{i\phi} - 4e^{-i\phi}) \right]^n \right] \Psi_0(0) + \frac{i}{2} \left[e^{i\chi n} + \left[\frac{1}{6}(-2e^{i\phi} - 4e^{-i\phi}) \right]^n \right] \Psi_1(0) \quad (2.21)$$

In order to better understand how this state evolves as a function of the number of electrons n , we can proceed by finding $\Psi_0(n)\Psi_0(n)^*$. Assume that initially the system is in the 0th Bell state such that $\Psi_0(0) = 1$ and $\Psi_1(0) = 0$. Then

$$\Psi_0(n) = \frac{1}{2} \left[e^{i\chi n} + \left[\frac{1}{6}(-2e^{i\phi} - 4e^{-i\phi}) \right]^n \right] = \frac{1}{2} [e^{i\chi n} + \eta^n]$$

Simplifying η :

$$\begin{aligned}\eta &= -\frac{1}{3}(e^{i\phi} + 2e^{-i\phi}) \\ &= -\frac{1}{3}(3\cos\phi - i\sin\phi)\end{aligned}\tag{2.22}$$

Finding $|\eta|^2$ and $\arg(\eta)$:

$$\begin{aligned}|\eta|^2 &= \frac{1}{9}(9\cos^2\phi + \sin^2\phi) \\ &= 1 - \frac{8}{9}\sin^2\phi \\ -\theta &= \arg(\eta) = \arctan\left(-\frac{1}{3}\tan\phi\right)\end{aligned}$$

Therefore

$$\eta = \sqrt{1 - \frac{8}{9}\sin^2\phi} e^{-i\theta} = z e^{-i\theta}\tag{2.23}$$

Now,

$$\begin{aligned}\Psi_0(n)\Psi_0(n)^* &= \frac{1}{4}(e^{i\chi n} + z^n e^{-in\theta})(e^{-i\chi n} + z^n e^{in\theta}) \\ &= \frac{1}{4}[1 + z^n e^{-in(\theta+\chi)} + z^n e^{in(\chi+\theta)} + z^{2n}] \\ &= \frac{1}{4}[1 + z^{2n} + 2z^n \cos(n(\chi + \theta))]\end{aligned}$$

$$\boxed{\Psi_0\Psi_0(n)^* = \frac{1}{4}\left[1 + \left(1 - \frac{8}{9}\sin^2\phi\right)^n + 2\left(1 - \frac{8}{9}\sin^2\phi\right)^{\frac{n}{2}}\cos\left(n\left(\chi + \arctan\left(\frac{1}{3}\tan\phi\right)\right)\right)\right]}\tag{2.24}$$

This can be re-written as

$$\Psi_0\Psi_0(n)^* = \frac{1}{4}[1 + \alpha^n + 2\sqrt{\alpha}^n \cos(n\beta)]\tag{2.25}$$

The α terms represent decay of the wavefunction as a function of the number of electrons as $|\alpha| < 1$. The \cos term describes the oscillation of the Bell state through an angle β for each electron.

$$\beta = \chi + \arctan\left(\frac{1}{3}\tan\phi\right)\tag{2.26}$$

In order to determine the probability of error, we first find

$$\Psi_1 \Psi_1(n)^* = \frac{1}{4} [1 + \alpha^n - 2\sqrt{\alpha^n} \cos(n\beta)] \quad (2.27)$$

For coherent operation, $\Psi_0 \Psi_0^*(n) + \Psi_1 \Psi_1^*(n)$ should equal 1. We define the probability of error, P_E as $1 - (\Psi_0 \Psi_0^*(n) + \Psi_1 \Psi_1^*(n))$:

$$\begin{aligned} P_E &= 1 - \frac{1}{4} [1 + \alpha^n + 2\sqrt{\alpha^n} \cos(n\beta) + 1 + \alpha^n - 2\sqrt{\alpha^n} \cos(n\beta)] \\ &= \frac{1}{2} [1 - \alpha^n] \end{aligned}$$

$$P_E = \frac{1}{2} \left[1 - \left(1 - \frac{8}{9} \sin^2 \phi \right)^n \right] \quad (2.28)$$

Notice that

$$\lim_{n \rightarrow \infty} P_E = \frac{1}{2}$$

as $|\alpha| < 1$. This demonstrates that there is a partial steady-state Bell state preparation.

As shown in Fig. 2.6, the numerical results and analytical result of Eq. (2.24) are in agreement. As the number of electrons increases, the system continues to oscillate between the two Bell states with a non-unitary operation (non-zero probability of failure). Shown in 2.6(b), the density matrix in the standard basis performs a swap between $|01\rangle$ and $|10\rangle$ with a Bell state obtained for $N = 22$.

$\sqrt{\text{SWAP}}$

The direct approach to implementing a two qubit rotation is through an interaction of the form $J \vec{S}_1 \cdot \vec{S}_2$ between the two static spins as, for example, in the Kane architecture [40] or more recently in the quantum dot approaches proposed by Trifunovic et al. [119, 120]. Using the proposed itinerant spin approach to mediate an effective exchange interaction leads to an inherent imperfection in the gate operations. These imperfections arise as a result of the undesirable coefficients c' and c'' of

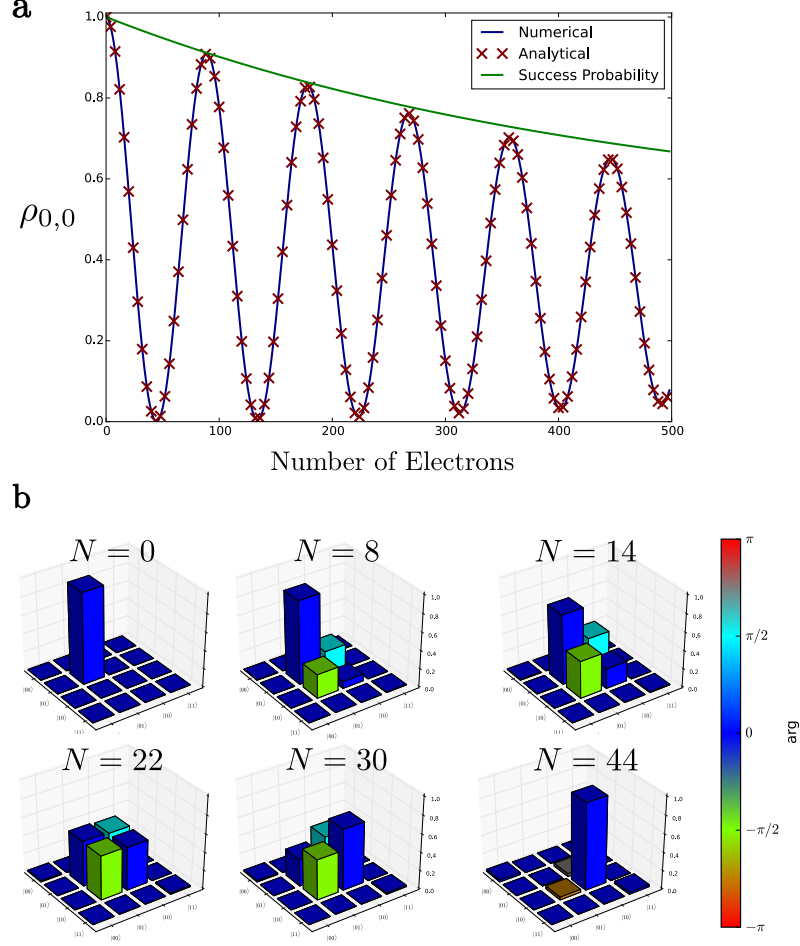


Fig. 2.6. **Bell-State Oscillation:** (a) Numerical and analytical modeling of the system shown in Fig. 2.5 yields oscillations in the Bell-basis as a function of the number of incident electrons. Using $kd_0 = 0.45\pi$, $\Omega = 0.06$, and $\Gamma = 5.5$, the first element of ρ is shown. As can be seen, the numerical and analytical results are in agreement. The success probability, defined as $1 - P_E$, represents the probability the Bell state remains coherent as additional incident spins are provided. The parameters were selected to highlight the oscillations and quickly decreasing success probability. (b) In an alternate depiction, the two-qubit density matrix in the standard basis is shown after N electrons have been injected.

(2.20) due to the interactions $\vec{\sigma} \cdot \vec{S}_1$, $\vec{\sigma} \cdot \vec{S}_2$. This is apparent if we compare the matrix

representation of these operators with $[\mathbf{R}]$ in equation (2.20). On the other hand, the desirable coefficient c arises from product terms of the form

$$[\vec{\sigma} \cdot \vec{\mathbf{S}}_1][\vec{\sigma} \cdot \vec{\mathbf{S}}_2] + [\vec{\sigma} \cdot \vec{\mathbf{S}}_2][\vec{\sigma} \cdot \vec{\mathbf{S}}_1]$$

which are independent of the spin of the itinerant electrons, $\vec{\sigma}$, so that the two qubit operation, unlike the single qubit operations, does not require a spin potential; an ordinary unpolarized reservoir should be fine. The lack of phase symmetry of Eq. (2.20) for up and down flying spins, and hence imperfection in the $\sqrt{\text{SWAP}}$, is somewhat mitigated with the use of un-polarized itinerant spins as, on average, the $|12\rangle$ and $|\bar{1}\bar{2}\rangle$ states will pick up the same overall phase.

Note that if c' , c'' were zero in (2.20), our reflection matrix would nearly provide the transformation we are looking for with term c providing the two qubit rotation γ in equation (2.11). But the terms c' and c'' cause undesirable non-unitary effects leading to an average error probability

$$P_e \sim (|c'|^2 + |c''|^2)/2$$

so that the error probability per unit rotation for the two qubit gate can be estimated from

$$\frac{P_e}{\theta} \sim \frac{|c'|^2 + |c''|^2}{2|c|} \quad (2.29)$$

which is plotted in Fig. 2.7 as function of kd and kd_0 assuming $\Omega \equiv J/\hbar v = 1$, $\Gamma = 5$. Note that with

$$kd = \pi \text{ and } kd_0 \approx \pi/2$$

the error probability is quite small $\sim 10^{-4}$.

The fidelity of the two-qubit $\sqrt{\text{SWAP}}$ gate is strongly dependent on the height of the initial barrier, Γ . To illustrate the importance of adding the additional semi-transparent barrier of strength Γ in front, we have also shown the results without it ($\Gamma = 0$) in dashed lines which show much larger error probability.

Fig. 2.8(a) shows the evolution of the diagonal elements of the density matrix as it interacts with itinerant spins starting from an initial state with $\rho(1\bar{2}, 1\bar{2}) = 1$. Note

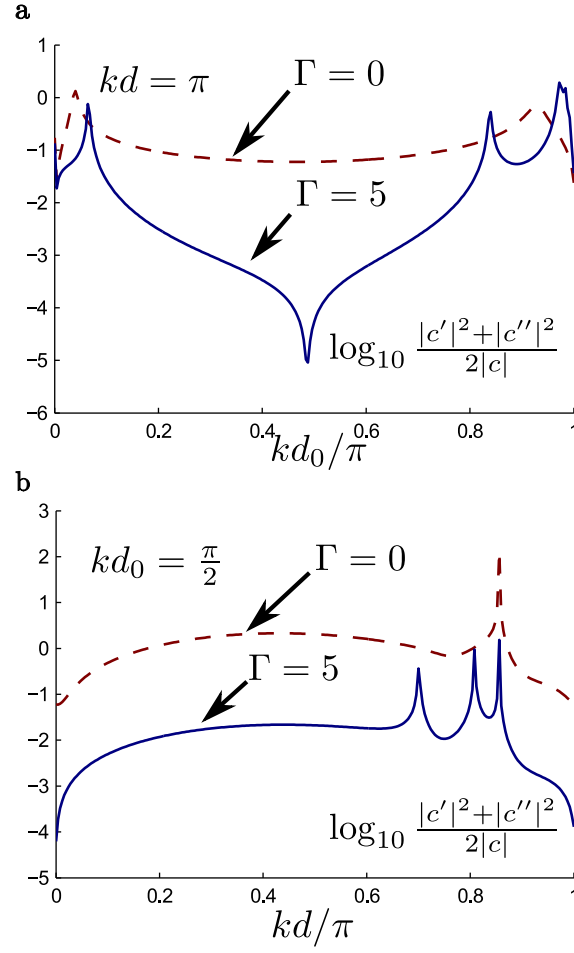


Fig. 2.7. **Figure of merit:** Figure of merit for two qubit operations as a function of (a) kd_0 for $kd = \pi$ and (b) kd for $kd_0 = \pi/2$ with $\Omega = 1$ and $\Gamma = 0$ and $\Gamma = 5$.

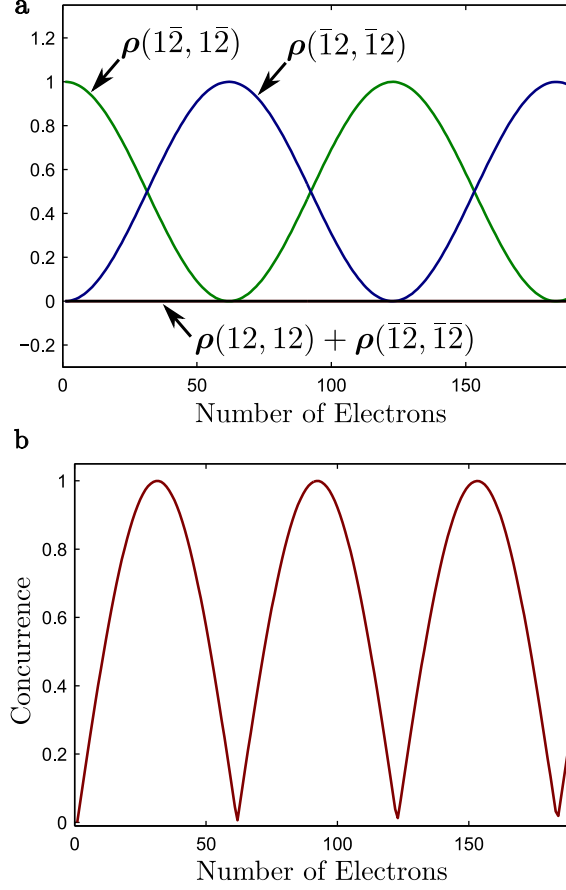


Fig. 2.8. **Two qubit rotation:** (a) Two qubit rotation with $kd = \pi$, $kd_0 = \pi/2$, $\Omega = 0.1$, and $\Gamma = 20$. (b) Two qubit concurrence showing the oscillation of entanglement as a function of the number of incident electrons.

that entangled states with very high concurrence, see equation (10) from [121], are obtained with the proper number of electrons (Fig. 2.8(b)).

For $\Gamma \neq 0$, the enhancement in gate fidelity, being due to multiple reflection, is wavelength dependent, and hence k dependent. At low temperatures the relevant k is the Fermi wavevector k_f corresponding to the Fermi energy which is related to the electron density n_s [122]:

$$k_f = \sqrt{2\pi n_s} \quad (2.30)$$

In general, however, a thermal average over wavevectors is involved and the degree of enhancement from multiple reflections will be averaged accordingly.

It should also be noted that since the reduced error probability with $\Gamma \neq 0$ comes from multiple coherent reflections, this gain in performance can be expected to be more sensitive to processes that cause a loss of spin coherence. Such processes are ignored in our present model.

Using the analytical simplification, that is $kd_0 = \pi$, the error probability and number of electrons as a function of kd_0 , Ω , and P_E for a $\sqrt{\text{SWAP}}$ operation were plotted in Fig. 2.9. Using these results, an optimal operation point can be identified. Using an optimal operation point, and *unpolarized* electrons, quantum process tomography was used to observe a $\sqrt{\text{SWAP}}$ gate operation.

The unpolarized electrons help reduce the overall P_E for the system. This is understood by recognizing that in our analysis, we presumed only z-up electrons were incident. As a result, anytime the system was driven into the $|12\rangle$ state, there was no mechanism whereby a spin could flip. However, with unpolarized electrons, this is now possible as the other 3×3 sub-space we didn't consider during the analysis comes into play. A down flying-spin can drive the $|12\rangle$ state back into the $\{|1\bar{2}\rangle, |\bar{1}2\rangle\}$ states.

CNOT

We end this section with an example of another universal two qubit quantum gate, the CNOT, implemented using the spin potential-based architecture described here. Fig. 2.10(a) shows a CNOT gate in terms of elementary single qubit and two qubit operations [123]. The circuit can be realized using the basic structure shown in Fig. 2.1(a) gated appropriately to obtain two qubits embedded in a spin coherent channel. These qubits can be selectively connected or disconnected from contacts held at specified spin potentials.

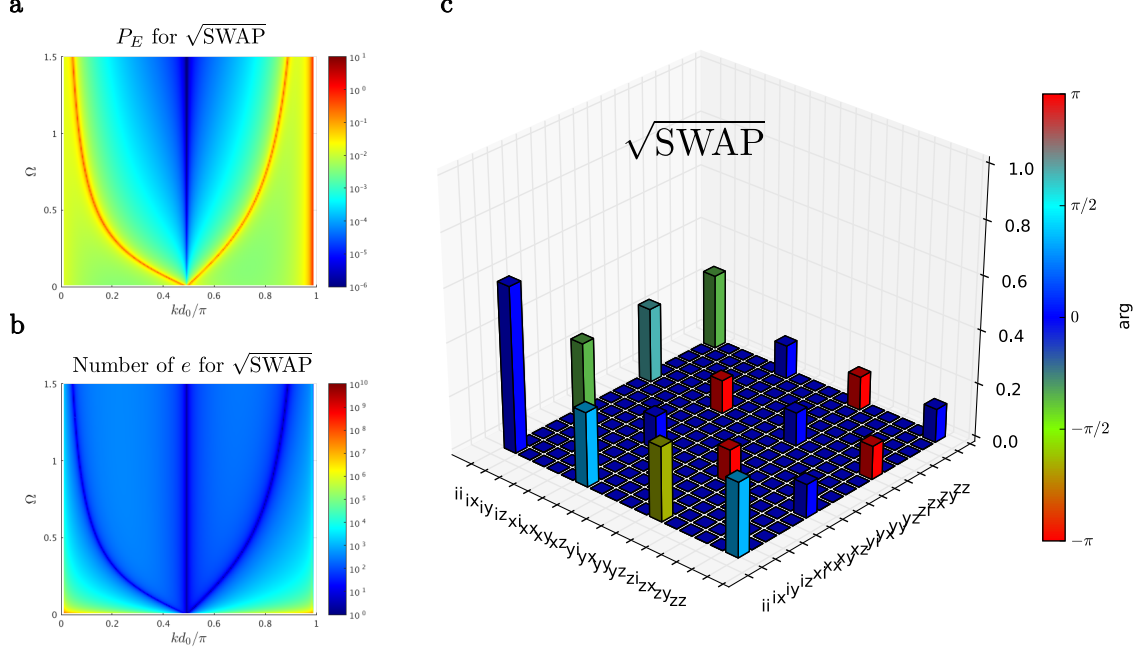


Fig. 2.9. **Error Minimization for Practical $\sqrt{\text{SWAP}}$:** (a) The probability of error for the two-qubit system under a $\sqrt{\text{SWAP}}$ operation is shown as a function of kd_0 and Ω for a fixed $\Gamma = 10$. With increasing Ω , the probability of error decreases. kd_0 values near $\pi/2$ yield lower P_E with specific resonance values yield high P_E . (b) The number of electrons necessary for the gate operation is also shown as a function of Ω and kd_0 . Regions of minimal color gradient provide ideal operating conditions in order to minimize the influence of any single electron on the system, providing increased robustness. (c) Using these results as a guide, with $kd_0 = 0.7\pi$, $\Omega = 1$, and $\Gamma = 5$, state-process tomography for the system was performed for the $\sqrt{\text{SWAP}}$ operation. The operation was performed using un-polarized incident electrons.

The timing diagram in Fig. 2.10(c) shows the sequence of single qubit and two qubit operations needed to implement the required gate:

1. H-gate: This involves $\pi/2$ rotations of $|S_0\rangle$ around the x, z, and x axes in sequence. The barrier gates G_i , R_i and B_i are used to isolate $|S_0\rangle$ which is then

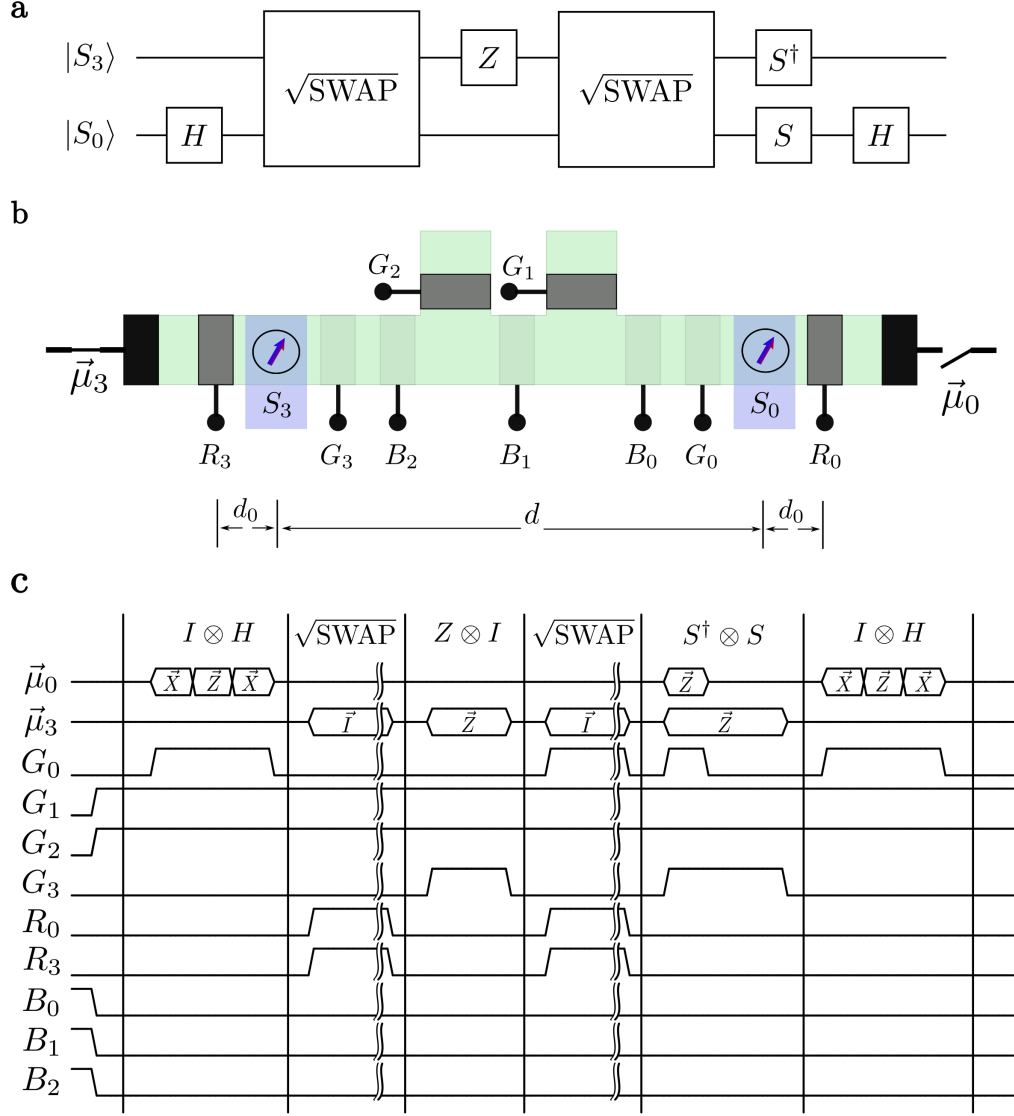


Fig. 2.10. **Controlled-NOT:** (a) Circuit representation for CNOT based on the controlled-Z with Hadamard gates to obtain a controlled-X. (b) Physical picture for single and two qubit operations on $|S_0\rangle$ and $|S_3\rangle$ with electrostatically controlled gates. (c) Waveforms depicting the manipulations necessary for the various nets of Fig. (b) to perform a CNOT operation between $|S_0\rangle$ and $|S_3\rangle$.

rotated by connecting the $\vec{\mu}_0$ terminal to reservoirs with x and z-directed spin potentials, each for the length of time needed to provide the $\pi/2$ rotation.

2. $\sqrt{\text{SWAP}}$ gate: This involves a two qubit rotation on $|S_0\rangle$ and $|S_3\rangle$ of the type discussed in the previous section. This gate is realized with the use of an unpolarized spin potential for the length of time appropriate for a $\pi/4$ “rotation” (equation (2.11)). Note the use of barrier gates R_0 and R_3 to implement the completely reflective barrier and the semitransparent barrier respectively.
3. Z-gate: This rotation of qubit $|S_3\rangle$ around the z-axis is achieved by using gate G_3 for isolation and then connecting to a z-directed spin potential for an appropriate duration.
4. $\sqrt{\text{SWAP}}$ gate: same as step 2.
5. S and S^\dagger gates: This involves a $\pi/2$ rotation of $|S_0\rangle$ and a $3\pi/2$ rotation of $|S_3\rangle$, both around the z-axis.
6. H-gate: same as step 1.

The complete CNOT gate was simulated with single qubit and two qubit operations as described earlier with $\Omega = \pi/16$ and $\Gamma = 10^2$. A 0.5% error in the desired values of kd and kd_0 was assumed:

$$kd = 0.995\pi \text{ and } kd_0 = 0.995\pi/2$$

The fidelity of the gate was estimated using the prescription laid out by Trifunovic et al. [119]. The CNOT gate was simulated with each of the four $(1/\sqrt{2}(|01\rangle \pm |11\rangle)$, and $1/\sqrt{2}(|00\rangle \pm |10\rangle)$ initial states that give rise to Bell states, and the fidelity of the final state ρ' was evaluated by comparing to the ideal Bell state ρ :

$$f(\rho, \rho') = \sqrt{\sqrt{\rho}\rho'\sqrt{\rho}} \quad (2.31)$$

The minimum value of f was 99.8%.

Finally, the fidelity of the complete CNOT implementation of 99.8% is sufficient for fault tolerant quantum computing [119].

2.5 Discussion

The proposed architecture has a number of features that may be advantageous for building a quantum computer in semiconductor based architectures. The itinerant spins generate localized magnetic fields for any target qubit, providing individual qubit selectivity, parallel operation, and qubit isolation. This localized field generation removes the need for AC electric and magnetic fields used for nuclear magnetic resonance (NMR) and electron spin resonance (ESR). Removing the need for these magnetic fields reduces the complexity of system design for qubit manipulation.

External DC magnetic fields can be eliminated if perfect half-metallic contacts can be obtained to produce 100% spin polarized currents suitable for high-fidelity qubit initialization. These ideal spin currents are currently difficult to realize experimentally, and an alternative initialization method is needed in the near-term. Alternatively, this architecture is compatible with existing approaches that leverage an external magnetic field to produce Zeeman splitting such that single qubit initialization and readout can be accomplished with single electron transistors.

Using itinerant spins for gate operations must be compatible with the decoherence times of the qubits. Given the recent advances in long-term storage of quantum information, semiconductor systems are continuing to show promise for the realization of a scalable quantum computer.

While single qubit gates in the architecture require the use of spin polarized currents and hence the integration of magnetic materials, two qubit gates can be realized with traditional contacts that produce un-polarized spin currents. Furthermore, these two qubit gates can be used to obtain non-local entanglement between selective qubits, a limitation of nearest neighbor proposals. The degree of non-locality will be limited by the spin coherent transport length of the itinerant quasi-particles which, depending on the material and the temperature, can range from tens of nanometers to tens of microns. However, coupling to channel contacts along with multiple reflections from barriers, increasing the effective transport length of the itinerant spins,

will limit their range. As long as spin coherence can be maintained over a gate length (see Figure 2.1) between two qubits, it should be possible to entangle them.

The use of a large number of itinerant spins to effect a given qubit operation allows fine tuning and control since a deviation of one electron represents a small error in a process involving, say, 10^4 electrons. We envision controlling the actual number of electrons using gates to connect or disconnect the qubits from the itinerant spins as desired. The use of all-electrical control of qubits is beneficial for producing a scalable architecture using semiconductor based qubits. Other proposals for all-electrical control exist, however, these proposals are largely based on spin-orbit interactions and are likely to be susceptible to charge noise [124].

Many proposals for qubit manipulations require sensitive gate control and the ability to manipulate single electrons. Here, the control for gate operations is based on a large number of electrons drawn from a reservoir which can be controlled accurately. Additionally, the architecture does not require a bound donor electron to perform nuclear spin manipulation which may provide a path for higher temperature operation [125]. As a further example of the reduction in control necessary to implement the architecture, the gate operations do not require precise placement of donors in the lattice.

Nevertheless, there are a number of challenges that are incurred by the architecture. Prominently is the inherent loss of gate fidelity as a result of repeated measurement. This loss of gate fidelity requires error correction even before other sources of decoherence and dephasing are considered. Magnetic material integration into semiconductor processing is another obstacle that must be overcome for experimental realization of high-fidelity initialization, single qubit gates, and ensemble qubit readout. There has been progress towards integration of these materials into fabrication processes as a result of modern magnetic memory technologies, however, this integration is still emerging and does not have mainstream adoption. Additionally, an interaction of the form given by (2.1) was assumed throughout the proposal.

Based on this assumption, a more detailed exploration of the interaction between conduction-band flying electrons and donor-based nuclear spins is warranted.

In summary, we have outlined a quantum computing architecture based on the use of non-equilibrium spin potentials enabled by modern spintronics to perform all basic qubit operations including initialization, arbitrary single qubit rotation, single qubit readout, and two qubit rotation on selected pairs of qubits. A key feature of our architecture is the use of repeated entanglement with itinerant electrons and a subsequent collapse of the quantum state. The latter process is non-unitary, but we have shown that the overall non-unitary component can be kept below tolerable limits with proper design. Finally we presented the implementation of a complete CNOT gate using the proposed spin potential based architecture, and showed that the fidelity under ideal conditions is acceptable for fault tolerant quantum computing. This all-electrical architecture provides a means of qubit control for semiconductor donor systems without the use of magnetic fields while providing qubit selectivity and isolation, and non-local two qubit operation. Future research may include numerical modeling of representative experimental structures and investigation of the interaction between conduction band flying electrons and donor-based nuclear spins. Experimental investigation of the operations described herein would be valuable to assess the validity of the proposal.

3. ANNEALING, THE ISING MODEL, AND COMBINATORIAL OPTIMIZATION

The material in this chapter is based on our publication “Intrinsic optimization with stochastic nanomagnets” [17].

3.1 Solving NP-Hard Problems

The use of Ising computers to solve NP-hard problems has a rich heritage in both theory [126] and practice. These computers seek to solve a wide range of optimization problems by encoding the solution to the problem as the ground-state of an Ising energy expression. Many diverse systems have been proposed to solve NP-hard optimization problems such as those based on simulated annealing [10], DNA [127, 128], quantum annealing [12, 129], Cellular Neural Networks [130–132], CMOS [14], trapped ions [133], electromechanics [134], optics [15, 16, 135–140], and magnets [20, 141, 142]. A common objective of many of the Ising-based approaches is the identification of hardware configurations that can efficiently solve optimization problems of interest.

A popular approach to solving such problems is to map the problem onto an Ising Hamiltonian whose minimum energy represents the solution to the problem. Once mapped, the system is then annealed until it resolves to its ground-state. Commercial companies like D-Wave have developed a business model that centers on the use of quantum processes to perform this annealing [12, 129], harnessing adiabatic evolution. These systems find the ground-state of an Ising energy expression by adiabatically increasing the “spin”-“spin” couplings and biases. Unfortunately, it is easy for the system to escape the ground-state due to thermal noise, requiring operation temperatures on the order of \sim mK.

While quantum annealing provides potential promise, it is worth asking the following question: can a classical system provide similar benefits (such as parallelism) without many of the shortcomings associated with the operation of qubits? In this chapter we introduce once such possibility, namely the use of probabilistic computing elements similar to those proposed by Feynman [51].

3.2 Classical Annealing with Stochastic Nanomagnets

Herein, we demonstrate the possibility of a hardware implementation that does not just mimic the Ising model, but embodies it as a part of its natural physics [20, 141, 142]. It uses a network of N “soft” nanomagnets operating in a stochastic manner [64], each with an energy barrier Δ comparable to $k_B T$ so that they switch between the two Ising states, ± 1 , on time scales $\tau \sim \tau_0 \exp(\Delta/k_B T)$ where $\tau_0 \sim 0.1 - 1$ ns. The natural laws of statistical mechanics guide the network through the 2^N collective states at GHz rates, with an emphasis on low energy states. We show how an optimization problem of interest is solved by engineering the spin-mediated magnet-magnet interactions to encode the problem solution and to simulate annealing without any change in temperature simply by continuously adjusting their overall strength. As proof-of-concept for the potential applications of this natural Ising computer, we present detailed simulation results for standard NP-complete examples, including a 16-city traveling salesman problem. This involves using experimentally benchmarked modules to simulate a suitably designed network of 225 stochastic nanomagnets and letting the hardware itself rapidly identify solutions within the 2^{225} possibilities. It should be possible to integrate such hardware into standard solid state circuits, which will govern the scalability of the solution.

The Ising Hamiltonian for a collection of spins, S_i , which can take on one of two values, ± 1 ,

$$H = - \sum_{i,j} J_{ij} S_i S_j - \sum_i h_i S_i \quad (3.1)$$

was originally developed to describe ferromagnetism where the J_{ij} are positive numbers representing an exchange interaction between neighboring spins S_i and S_j , while h_i represents a local magnetic field for spin S_i . Classically, different spin configurations $\sigma\{S_i\}$ have a probability proportional to $\exp(-H(\sigma)/k_B T)$, T being the temperature, and k_B , the Boltzmann constant. At low temperatures, the system should be in its ground state σ_G , the state with the lowest energy $H(\sigma)$. With $h_i = 0$, and positive J_{ij} , it is easy to see that the ground state is the ferromagnetic configuration σ_F with all spins parallel.

Much of the interest in the Ising Hamiltonian arises from the demonstration of many direct mappings of NP-complete and NP-hard problems to the model [70, 126, 143] such that the desired solution is represented by the spin configuration σ corresponding to the ground state. However, in general this mapping may require a large number of spins, and may require the parameters J_{ij} and h_i to take on a wide range of values, both positive and negative. Finding the ground state of this artificial spin glass is the essence of Ising computing, and broadly speaking it involves abstractly representing an array of spins, their coupling, and thermal noise through software and hardware that attempts to harness the efficiencies of physical equivalence [144]. These representations may take the form of abstract models of the spins, the use of random number generators to produce noise, and logical or digital adders for the weighted summing. If enough layers of abstraction can be eliminated, the underlying hardware will inherently solve a given problem as part of its natural, intrinsic operation and this should be reflected in increased speed and efficiency.

3.3 Engineering Correlations Through Spin Currents

Here we describe a natural hardware for an Ising computer based on the representation of an Ising spin S_k by the magnetization m of a stochastic nanomagnet (SNM), which we believe will compare well with other alternative representations. These SNMs are in the “telegraphic” switching regime [64, 145] requiring the exis-

tence of a small barrier in the magnetic energy ($\Delta \approx k_B T$), that gives a small, but definite preference for a given axis, with two preferred states ± 1 . In the absence of currents, these SNMs continually switch between $+1$ and -1 on the order of nanoseconds, and can be physically realized by a reduction of the magnetic grain volume [146] or by designing weak perpendicular magnetic anisotropy (PMA) magnets [147]. Figure 3.1 shows the response of such a monodomain PMA magnet in the presence of an external spin current in the direction of the magnet's easy axis.

How do we couple the SPMs to implement the Ising Hamiltonian of Equation (3.1)? The usual forms of coupling involve dipolar or exchange interactions that are too limited in range and weightability. Instead, one possibility is an architecture [20] that uses charge currents which can be readily converted locally into spin currents through the spin Hall effect (SHE). These charge currents can be arbitrarily long-range and the total number of cross-couplings is only limited by considerations of routing congestion and delay. The couplings may also be confined to nearest-neighbors, simplifying the hardware design complexity while promoting scalability and retaining universality [143].

The Ising Hamiltonian of Equation (3.1) can be implemented by exposing each SNM m_k to a spin current I_k

$$I_k(m_j) = \frac{2q\alpha}{\hbar} \left(h_k + \sum_j 2J_{kj}m_j \right) \quad (3.2)$$

which has a constant bias determined by h_k together with a term proportional to the magnetization of the j^{th} SNM m_j . The future state of magnet m_i at time $(t + \Delta t)$ is related to the state of the other magnets at time t through the current I_k . This expression is derived analytically in Section 3.3.1 using the Fokker-Planck equation for the system [148].

The spin current I_k can be generated using well-established phenomena and the prospects for physical realization of such a system is discussed further in Section 3.5. The distinguishing feature of the present proposal arises from the intrinsic stochasticity of SNMs and their biasing through the use of weighted spin currents (Figure

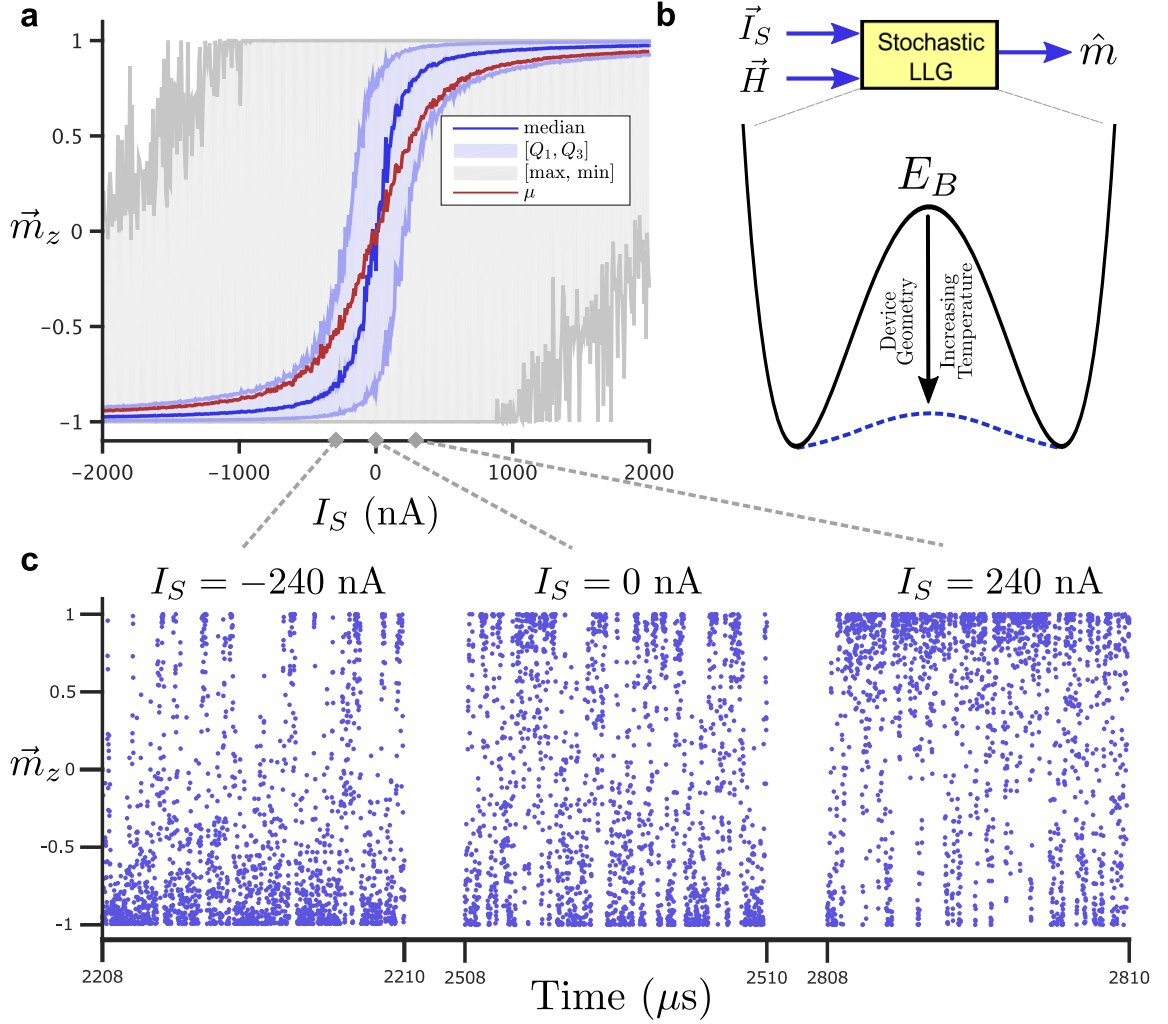


Fig. 3.1. **Response of stochastic nanomagnet to spin current:** (a) The magnetization of a stochastic nanomagnet is shown for varying spin currents. The five number summary of the magnetization \vec{m}_z is shown throughout the simulation. (b) Obtaining stochastic operation for a magnet can be accomplished with a reduction of the energy barrier of the magnet E_B through device geometry or by increasing its temperature. The response of the magnet to thermal noise under these conditions is modeled using a stochastic Landau-Lifshitz-Gilbert (LLG) circuit element based on the input spin current I_S and magnetic field H . (c) Sample time slices are shown at various set points along the sigmoid in order to visualization the magnetization dynamics.

3.1(a)). How the SNMs are interconnected to implement Equation (3.2) can evolve as the field progresses.

Getting a large system to reach its true ground state is non-trivial as it tends to get stuck in local minima [149]. It is common to guide the system towards the ground state through a process of “annealing” [10] which is carried out differently in different hardware implementations. For example, systems based on superconducting flux qubits make use of quantum tunneling, which is referred to as quantum annealing [150], whereas classical CMOS approaches make use of random number generators [151] to produce random transitions out of local minima.

For our system of coupled SNMs, random noise is naturally present and can be easily controlled (Figure 3.1(a)), causing the system of SNMs coupled according to Equation (3.2) to explore the configuration space of the problem on a nanosecond timescale. Annealing could be performed through a controlled lowering of the actual temperature, or equivalently through a controlled increase in the magnitude of the current I_k , even at room temperature. It has been noted that certain annealing schedules can guarantee convergence to the true ground state, but these schedules may be too slow to be used in practice [152]. Here we only present a straightforward annealing process and does not seek out optimal annealing schedules. Consequently, as we show in one of our combinatorial optimization examples, we may find only an approximate solution which, however, may be adequate for many practical problems.

3.3.1 Steady-State Fokker-Planck Description

Our goal is to interconnect magnets such that their equilibrium state is governed by Boltzmann statistics with thermal noise as an inherent characteristic of the system. To see that this is possible, consider a system of N magnets where we want

$$\rho(m_1, \dots, m_N) = \rho_0 e^{-E(m_1, \dots, m_N)/k_B T} \quad (3.3)$$

and

$$E(m_1, \dots, m_N) = \sum_i (A_i m_i^2 + h_i m_i) + \sum_{i,j} J_{ij} m_i m_j \quad (3.4)$$

where m_k represents the z-component of the magnets.

Suppose each magnet is driven by a spin current derived from the others. Using the Fokker-Planck equation [148] for the N -magnet system:

$$\frac{\partial \rho}{\partial \tau} = \frac{\partial}{\partial m_k} \left\{ (1 - m_k^2) \left[(i_k - m_k) \rho + \frac{1}{2\Delta_k} \frac{\partial \rho}{\partial m_k} \right] \right\} \quad (3.5)$$

where $\Delta_k = \mu_0 H_k M_s V / 2k_B T$ and $i_k = I_k / I_0$ with I_0 as the critical switching spin current $I_0 = (2q\alpha/\hbar) 2\Delta_k k_B T$. At equilibrium, $\partial \rho / \partial \tau = 0$ yielding from (3.3) and (3.5):

$$\frac{\partial(\ln \rho)}{\partial m_k} = -2\Delta_k (i_k - m_k) \quad (3.6)$$

$$\frac{\partial(\ln \rho)}{\partial m_k} = -\frac{1}{k_B T} \left(2A_k m_k + h_k + \sum_j (J_{kj} + J_{jk}) m_j \right) \quad (3.7)$$

respectively. Comparing equations (3.6) and (3.7) while assuming symmetric coupling, $J_{kj} \equiv J_{jk}$, for the system we find

$$\Delta_k = -A_k / k_B T \quad (3.8)$$

and arrive at (3.2):

$$i_k = \frac{h_k + \sum_j 2J_{kj} m_j}{\mu_0 H_K M_S V}$$

3.3.2 Stochastic Landau-Lifshitz-Gilbert (LLG) Model

In this section we briefly describe the simulation framework and stochastic LLG model used throughout this chapter. We start with the LLG equation [148] for a monodomain magnet with magnetization m_i in the presence of a spin current ($\vec{I}_s = I_0 \hat{z}$)

$$(1+\alpha^2) \frac{d\hat{m}_i}{dt} = -|\gamma| \hat{m}_i \times \vec{H}_i - \alpha |\gamma| (\hat{m}_i \times \hat{m}_i \times \vec{H}_i) + \frac{1}{qN_i} (\hat{m}_i \times \vec{I}_{Si} \times \hat{m}_i) + \left(\frac{\alpha}{qN_i} (\hat{m}_i \times \vec{I}_{Si}) \right) \quad (3.9)$$

The magnetic thermal noise enters the equation through the effective field of the magnet, $H_i = H_0 + H_n$, as an uncorrelated external magnetic field in three dimensions with the following mean and variance:

$$\langle H_n^{\vec{r}} \rangle = 0 \quad \langle |H_n^{\vec{r}}|^2 \rangle = \frac{2\alpha kT}{|\gamma| M_s \text{Vol.}} \quad (3.10)$$

The numerical model is implemented as an equivalent circuit for SPICE-like simulators and reproduces the equilibrium (Boltzmann) distribution from a Fokker-Planck Equation [148].

A given system of magnets is simulated using a collection of independent, though current-coupled, stochastic LLG models. Delays associated with the communication from one magnet to the next are neglected assuming that the response time of the nanomagnets is much greater than associated wire-delays. Given the response time of experimental stochastic magnets (\sim ms) [146], this is a reasonable approximation. However, this approximation must be revisited if the nanomagnets intrinsic response time is on the same order as any communication delays.

Many options exist for physical realization of the proposed system of stochastic nanomagnets as discussed in Section 3.5. For the simulations in this chapter we simply use Equation (3.2) without assuming any specific hardware to implement it, since it is likely that better alternatives will emerge in the near future, given the rapid pace of discovery in the field of spintronics, see for example [32, 35, 153, 154].

3.4 Combinatorial Optimization

We will focus on two specific examples to demonstrate the ability of such an engineered spin glass to solve problems of interest [155]: an instructive example based on the satisfiability problem (SAT), and a representative example based on the traveling salesman problem (TSP). The first known NP-complete problem is the problem of Boolean satisfiability [156], namely, deciding if some assignment of boolean variables $\{x_i\}$ exists that satisfies a given conjunctive normal form (CNF) expression. Find-

ing the collection of inputs that makes the clauses of the CNF expression true is computationally difficult, but easy to verify.

It is known that any given CNF expression can be mapped to a collection of Ising constraints using the fundamental building blocks of NOT ($m_1 = \bar{m}_2$), AND ($m_1 = m_2 \wedge m_3$), and OR ($m_1 = m_2 \vee m_3$) each subject to the Ising constraints given by [157]:

$$H_{\text{NOT}} = 1 - (-m_1 m_2) \quad (3.11)$$

$$H_{\text{AND}} = 3 - (-m_2 m_3 + 2m_1 m_2 + 2m_1 m_3) - (-2m_1 + m_2 + m_3) \quad (3.12)$$

$$H_{\text{OR}} = 3 - (-m_2 m_3 + 2m_1 m_2 + 2m_1 m_3) - (2m_1 - m_2 - m_3) \quad (3.13)$$

Using these building blocks, a network capable of finding the truth table for XOR ($m_1 = (m_2 \vee m_3) \wedge \overline{(m_2 \wedge m_3)}$) was prepared (Figure 3.2). For simplicity, the solution uses a naive method to construct the network and leverages the use of ancillary spins to represent $(m_2 \vee m_3)$ and $(m_2 \wedge m_3)$ respectively (note that four spins could have been used [158]). The array of spins from Figure 3.2(b) are connected as specified by (3.11),(3.12),(3.13), driven by a reference current I_0 . As the magnets explore the configuration space, their outputs are digitized and used to compute the overall energy of the system (Figure 3.2(c)). The regions of zero energy correspond to solutions of the problem. The digitized outputs are aggregated to determine their probability of occurrence. By looking at the first three bits of the most probable outputs, the solution to the problem can be directly found (Figures 3.2(d) and 3.2(e)). While this problem helps convey the essence of the approach, a more demonstrative application is worth considering.

The decision form of the TSP is NP-complete, that is, for a collection of N cities, does there exist a closed path for which each city is visited exactly once that has a tour length less than some value d . Finding tours that satisfy this problem is computationally challenging and also of great practical interest. There are well-known mappings that translate the TSP to the Ising model [70, 159]. Here we adopt the following:

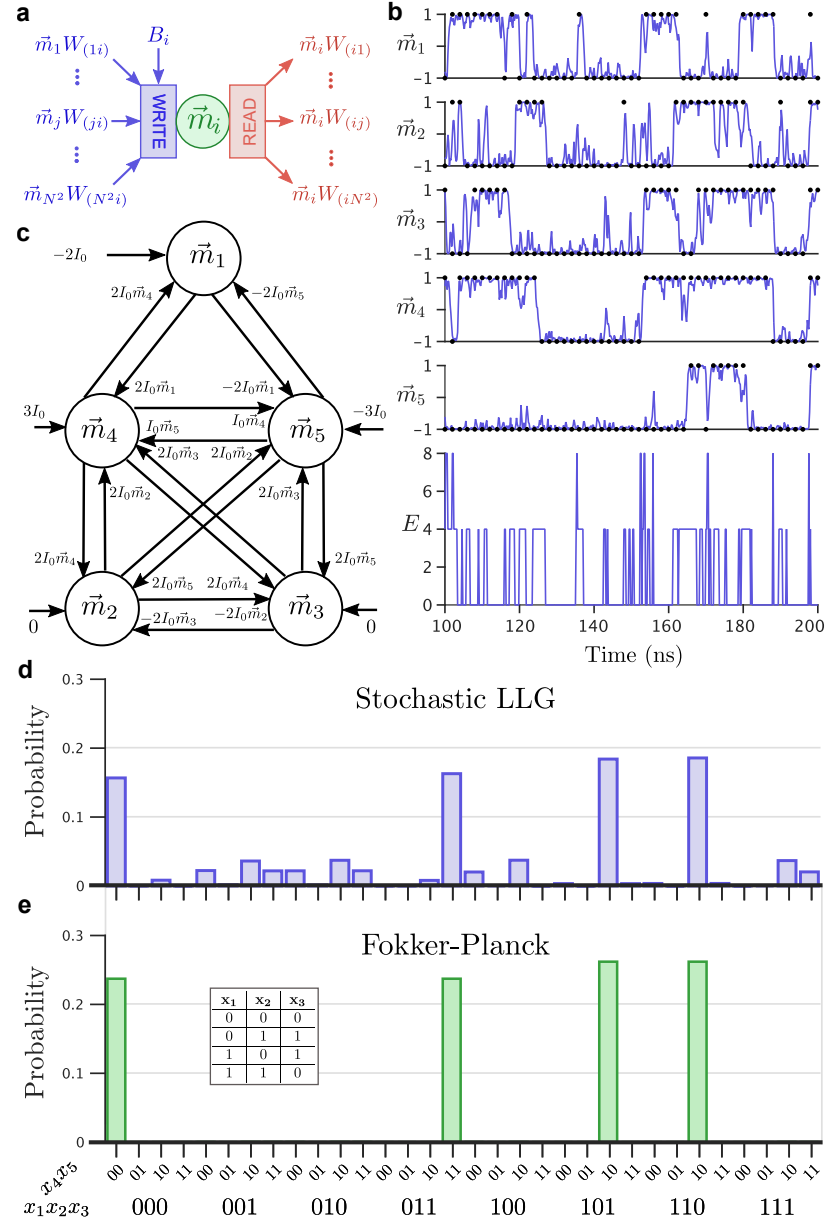


Fig. 3.2. **Boolean satisfiability:** (a) Each magnet \vec{m}_i is influenced by problem-specific bias, B_i , and weighted-coupling, $W_{(ji)}$, to magnet \vec{m}_j . (b) The truth table of the XOR operation is found by mapping the problem on an Ising expression. (c) The magnetization response of the magnets during a time slice is digitized in order to compute the energy of the system over time (d) Each digitized magnetization represents the logical bits x_i . (e) Steady-State Fokker-Planck equation analytical solution.

$$H = \sum_{v=1}^N \left(1 - \sum_{j=1}^N x_{v,j} \right)^2 + \sum_{j=1}^N \left(1 - \sum_{v=1}^N x_{v,j} \right)^2 + \lambda \sum_{uvj} W_{(uv)} x_{u,j} x_{v,j+1} \quad (3.14)$$

where $x_{i,j}$ is a Boolean variable that is TRUE when city i is stop number j and FALSE otherwise, and $W_{(uv)}$ are directed weights based on the distance between cities u and v . This Hamiltonian is mapped to a spin system by replacing each x_{ij} with $1/2(m_{ij} + 1)$ and weights $W_{(uv)}$ with $i_{(uv)}$ given by (3.2).

If the interconnections between each city are symmetric, then a Boltzmann machine [74] with each of the $2^{N \times N}$ states associated with an effective energy H is realized, and the probability of the system visiting a particular state is proportional to $\exp(-H/k_B T)$. In order to find low-energy, optimized states, direct annealing of the glass can be performed. Using the ulysses16 reference dataset [160], annealing of a problem specific magnetic array through control of the effective temperature was performed (Figure 3.3). Two specific traits of interest arise, namely the energy decays in a sigmoidal relationship with the $\ln T$, and the specific heat of the system, $C(T) = (\langle E(T)^2 \rangle - \langle E(T) \rangle^2)/k_B T^2$, shows a defined peak about a critical temperature. At high temperatures, the system is disordered and corresponds to high energy states (Figure 3.3(c)). As the temperature is reduced, the system continues to explore the energy landscape on a nanosecond timescale while gradually converging to a low-energy solution. For the given annealing profile and simulation duration, a low-energy, though not ideal, solution is found to the problem, highlighting the heuristic nature of the optimization [159]. Note that in principle these simulation results could be obtained directly from actual hardware. For example, Figures 3.2(d) and 3.3(d) could be obtained by continuously monitoring the states of the individual SNMs using spin valves.

3.5 Considerations for Physical Realization

Physical realization of these engineered spin glasses requires the integration of multiple functional elements as highlighted in Figure 3.2(a). The magnetization of

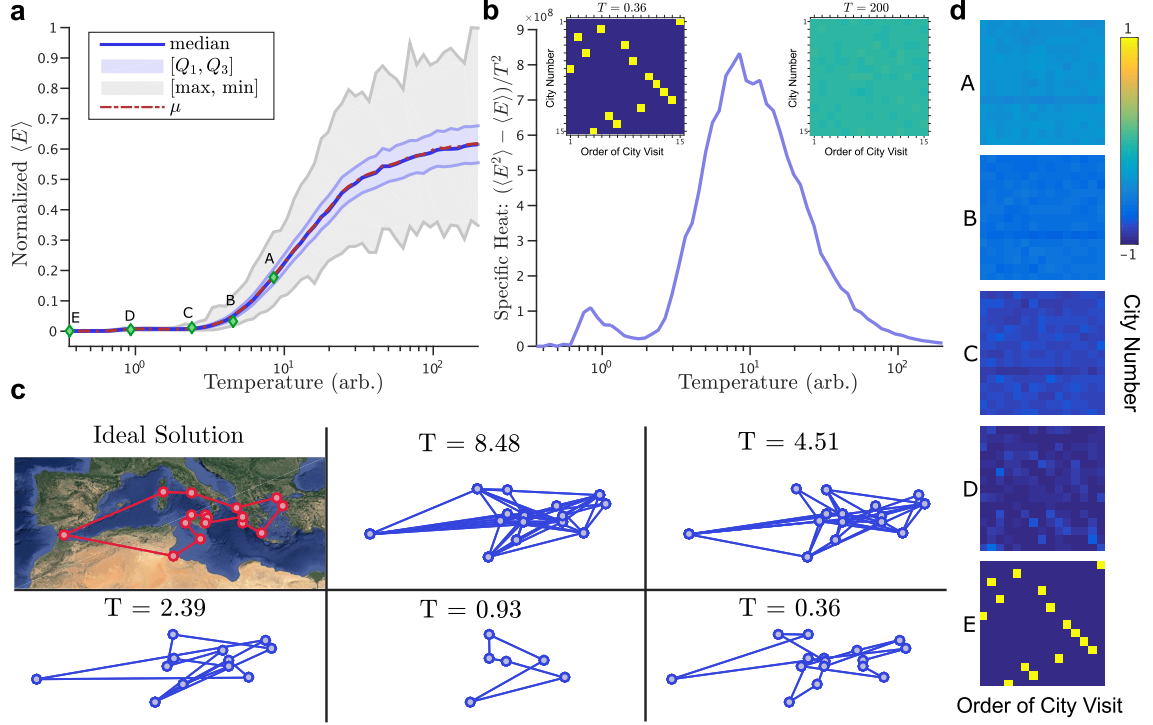


Fig. 3.3. Annealing stochastic nanomagnets for heuristic optimization of the traveling salesman problem: (a) An $N = 16$ city traveling salesman problem based on the ulysses16 data set [160] was simulated using an array of $(N - 1)^2 = 225$ stochastic magnets, assuming a fixed starting city. Each magnet represents if city i was stop j using $m_z = +1$ or was skipped $m_z = -1$ (insets). The magnets are prepared in a random initial configuration and gradually annealed until eventually frozen in a low-energy configuration. The normalized average energy of the system at each temperature is shown as the system is gradually annealed. (b) The specific heat of the array versus temperature is shown along with insets of the array configuration at early and late temperatures. (c) The state of the array is shown as a TSP graph at various temperatures, shown as green diamonds in (a), during the annealing process with the ideal configuration shown on the top left. (d) Average magnetization shown at the temperatures of (c).

each magnet m_i is first sensed with a read unit. The signal produced by this read unit is then propagated to all of the magnets with couplings dependent on the read magnetization m_i . Each of these connections is independently weighted with weights

$W_{(ij)}$ and provided as input, along with an on-site bias B_i to write units. The write units in turn influence and control the state of magnet m_j .

There are a number of design options available for each functional unit as shown in Table 3.1 and Figure 3.4. Write-control of the magnets can be affected through a number of means including the spin Hall effect (SHE) [32] or perhaps through voltage control [154]. The use of the SHE effect provides a convenient mechanism with which to sum several, independently weighted, input currents. Readout of the magnetization can be accomplished using well-established tunnel junctions [161] which have been demonstrated for stochastic nanomagnets [146]. Alternatively, readout could perhaps be accomplished using the inverse SHE [32]. Assuming the use of a SHE material and tunnel junction stack, care must be given to accomodate the simultaneous use of write and read currents. One approach would be to introduce the use of a time-multiplexed scheme that disassociates the write and read operations [162]. Alternatively, structures that provide write and read isolation may be used [163].

The ability to write and read the magnetization is of fundamental importance, however, once read, the likely weak signal must be amplified to satisfy the fanout requirements of the network. This transistor-like gain can be realized using all-spin based approaches [20,163] or perhaps with the use of a hybrid-CMOS design [164,167]. These proposed approaches may introduce power dissipation challenges during the read operation, e.g. the short-circuit current produced with the use of amplifying inverters. Power dissipation considerations must be carefully evaluated to assess the viability of scaling the proposed system.

The output from the amplification stage can be selectively weighted so that a wide range of problems based on (3.1) can be encoded onto the network. The weighting of inputs can be based on an approach using re-programmable floating-gate voltages [165] that would enable the use of analog weights for the circuit. While floating-gate regulation would enable convenient re-programmability, the design would be complicated with the requirement for peripheral drivers to control the floating-gate array. Others proposals have suggested the use of memristors [64,166,167] or other

Function	Technique
Writing	Spin-Orbit Torque [32, 35]
	Voltage Control [154]
Reading	Spin-Valves/Tunnel Junctions [161]
	Inverse Spin-Hall Effect [32]
Amplification	Spin-Switches [163]
	CMOS [164]
Weighting	Floating-gate Regulators [165]
	Memristive Elements [64, 166, 167]
	Digital Logic [14]
	Fixed Voltages
Routing	Tailored Topologies [168]
	FPGA-Like Interconnect [169]

Table 3.1.

Options for Physical Realization: Many options exist for physical realization of the proposed system of stochastic nanomagnets. These magnets must be written, read, possibly amplified, weighted, and routed for the network to form a Boltzmann machine. The design options shown in this table reflect various approaches that can be used to perform each of these functions.

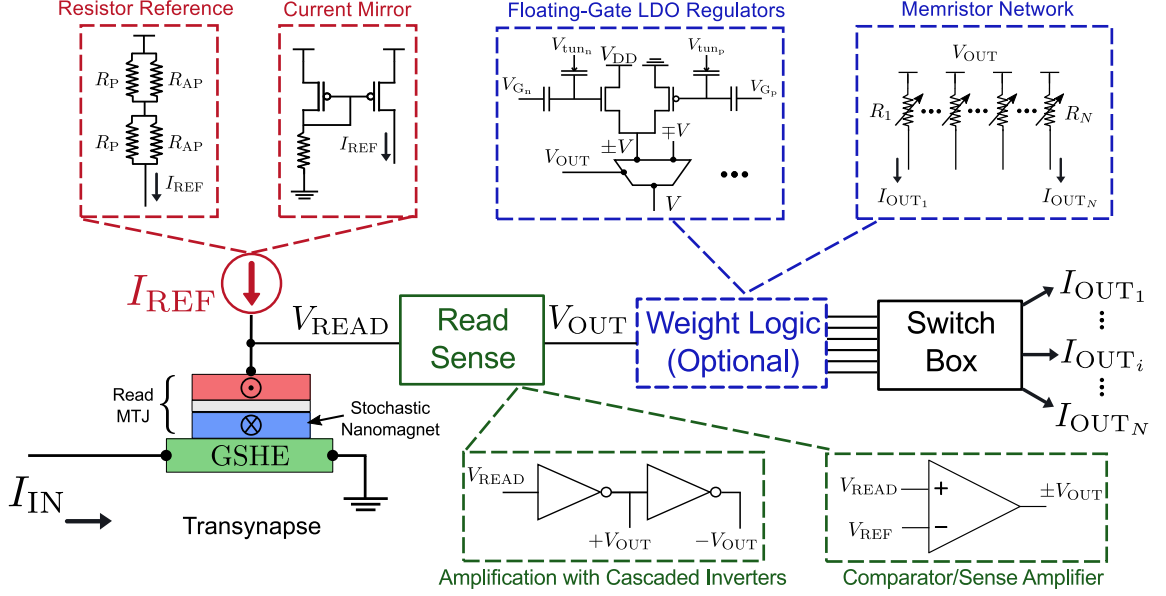


Fig. 3.4. **Possibilities for Physical Realization:** Numerous possibilities exist for physical realization of a nanomagnet based probabilistic architecture. Shown in the figure are a number of functional elements that must be integrated to form an engineered spin glass. Shown in the dotted boxes are examples (not exhaustive) of the types of design elements that could implement the functions.

programmable elements in a cross-bar like configuration [162,170], though with constrained fanout. Note that one weighting scheme that still retains the ability to encode NP-hard problems onto the network is with the use of $\{-1, 0, 1\}$ weights [126]. Using this simple approach removes the necessity for tunable weights and instead relegates the problem to one of routing, connectivity, and area.

All of the simulations used in this chapter assume a fully connected network of magnets in which each magnet talks to all other magnets. For small networks this is OK, however, such an assumption is invalid for large networks as the number of routes grows as N^N . Instead, different topologies [168] and routing considerations must be made to account for congestion and long-distance communication. By limiting the connections to local-neighbors [14,132], the network may still be used to perform NP-hard optimization while also simplifying routing complexity. One design possibility is

to leverage the lessons learned from the advances in the design of Field-Programmable Gate Array (FPGA) interconnects [169]. FPGAs are designed with routing topologies that facilitate both short and long-range interconnections while also providing re-programmability.

The fidelity of the programmed weights and number of high-fanout signals needed for robust solutions may impose challenges on the selected weighting and routing schemes. Additionally, the propagation delay of these high-fanout signals must be balanced with the response time of the magnets in order for the system to be governed by (3.1). While flexibility in the allowed weights and number of couplings is convenient for encoding problems onto the model [70], it is important to note that discrete nearest-neighbor couplings still retain NP-hardness [126] and may greatly simplify the hardware design, improving scalability at the expense of increased encoding complexity and area.

The main point of this chapter is the remarkable high-speed search through Fock space enabled by the intrinsic physics of a network of stochastic nanomagnets interacting via spin-mediated interactions. We hope this work fosters an interest in the physical realization and exploration of stochastic nanomagnets as a viable Ising computer.

4. AUTONOMOUS PROBABILISTIC COMPUTING

The material in this chapter is based on our pre-print “Autonomous Probabilistic Coprocessing with Petaflips per Second” [79].

Stochastic neural networks (SNN) are widely used for machine learning, inference and many other emerging applications [9]. As discussed earlier, a common version of such algorithms is based on the concept of a binary stochastic neuron (BSN) [74, 75] which fluctuates between -1 and +1 with probabilities that can be controlled through an input, I_i , constructed from the outputs of other BSNs, m_j . The synaptic function, $I_i(\{m\})$, can have many different forms depending on the desired functionality, but we will restrict this discussion to linear functions defined by a set of weights W_{ij} such that

$$I_i(t + \tau_S) = \beta \sum_j W_{ij} m_j(t) \quad (4.1)$$

where β is a constant and τ_S is the ‘synapse time’, that is the time it takes to recompute the inputs $\{I\}$ every time the outputs $\{m\}$ change. In software implementations, each BSN is updated repeatedly according to

$$m_i(t + \tau_N) = \text{sgn} [\tanh (I_i(t)) - r_{[-1,+1]}] \quad (4.2)$$

where $r_{[a,b]}$ represents a random number in the range $[a, b]$, and τ_N is the ‘neuron’ time, that is the time it takes for a neuron to provide stochastic output m_i with the correct statistics dictated by a new input I_i .

It is well-known [171] that to ensure fidelity of operation it is important to avoid *simultaneous* updates of two BSNs that are causally connected through a non-zero W_{ij} . The standard approach is to update each BSN sequentially according to Eq. (4.2), recomputing the input from Eq. (4.1) after each update, a procedure known as Gibbs sampling [172]. By contrast, the objective of this chapter is to explore the feasibility

of ultrafast operation through an autonomous architecture whereby each BSN continually fluctuates between -1 and +1 with probabilities that are controlled by the input I_i . We refer to this autonomous BSN as a *p-bit* to highlight its role as the key element of an autonomous p-computer (ApC), similar to the role of a qubit in a quantum computer.

As a quantitative measure of an ApC's speed of operation we use the number of *flips* per second (f), a *flip* being defined as a p-bit update *attempt*. For purely *sequential* updating, the number of flips per second is $\sim 1/(\tau_S + \tau_N)$. However, updating need not be purely sequential since unconnected BSNs can be simultaneously updated without loss of fidelity. If a number (N_p) out of the total number (N) of BSNs can be updated in parallel, then the number of flips per second will be much larger $\sim N_p/(\tau_S + \tau_N)$. Note, however, that in order to achieve this enhanced flip rate, the number of neurons that are simultaneously updated, N_p , have to be deliberately selected using a digital sequencer, so that the clock period, τ_{clock} , limits the maximum number of flips per second:

$$f \leq \frac{N_p}{\tau_{\text{clock}}} \quad (\text{Sequenced mode}) \quad (4.3a)$$

This clock speed will be limited by τ_S , τ_N , and the overhead associated with clock distribution [173].

The objective of this chapter is to present a framework for *clockless* operation [171, 174] whose speed is limited only by the neuron and synapse speeds

$$f \leq \frac{N}{\tau_N} = \frac{sN}{\tau_S} \quad (\text{Autonomous mode}) \quad (4.3b)$$

where $s \equiv \tau_S/\tau_N$. We will show that this autonomous mode provides high fidelity results without supervision keeping the fraction of detrimental simultaneous updates down to an acceptably low. Detrimental updates are managed simply by choosing a small s so that τ_N is much longer than τ_S , without using a digital sequencer to enforce a deliberate update order. As a result, f is not limited by τ_{clock} and can continue to operate faster as the synapse time τ_S is lowered. For example if we have

nearest neighbor connections with weights of $\{-1, 0, +1\}$, then the synapse can be implemented with short wires which respond in times less than 10 to 100 ps [175], much shorter than typical clock periods, reminiscent of the “intrinsic parallelism” observed in dynamical systems [176].

Eqs. (4.3a) and (4.3b) suggest that an autonomous design will allow faster operation (that is, more flips per second) if

$$\tau_S < \frac{sN}{N_p} \tau_{\text{clock}} \quad (4.4)$$

The factor N_p/N represents the fraction of neurons that can be updated simultaneously, which depends on the fan-in and the nature of interconnections and problem topology. For example, with nearest neighbor connections on a 2D square lattice as in Fig. 4.1(b), half the nodes can be updated simultaneously so that $N_p/N = 1/2$. The factor s also depends on the interconnections, but it additionally depends on the nature of the problem and the degree of solution fidelity needed, as we will show in this chapter.

This chapter demonstrates the feasibility of an ApC that performs the weight logic and p-bit functions defined by Eqs. (4.1) and (4.2) without the aid of sequencers as portrayed in Figs. 4.1(a) and 4.1(b). Our work is motivated by the compact, fast, energy efficient hardware that are currently being developed for the implementation of these functions [177, 178] as shown in Fig. 4.1(c), which we emulate using existing CMOS devices on an easily reconfigurable, cloud accessible digital FPGA platform, Fig. 4.1(d). The detailed methodology for constructing this p-computing coprocessor is described in the Methods section of this chapter. In the following sections we will describe the implementation approach and discuss results for two distinct applications, one involving combinatorial optimization and one involving emulated quantum annealing. These applications will help show that accurate results can be obtained with a sequencerless probabilistic computer of the type envisioned by Feynman [51], implemented using modern devices to enable operation at ultrafast rates unconstrained by the available clock speed.

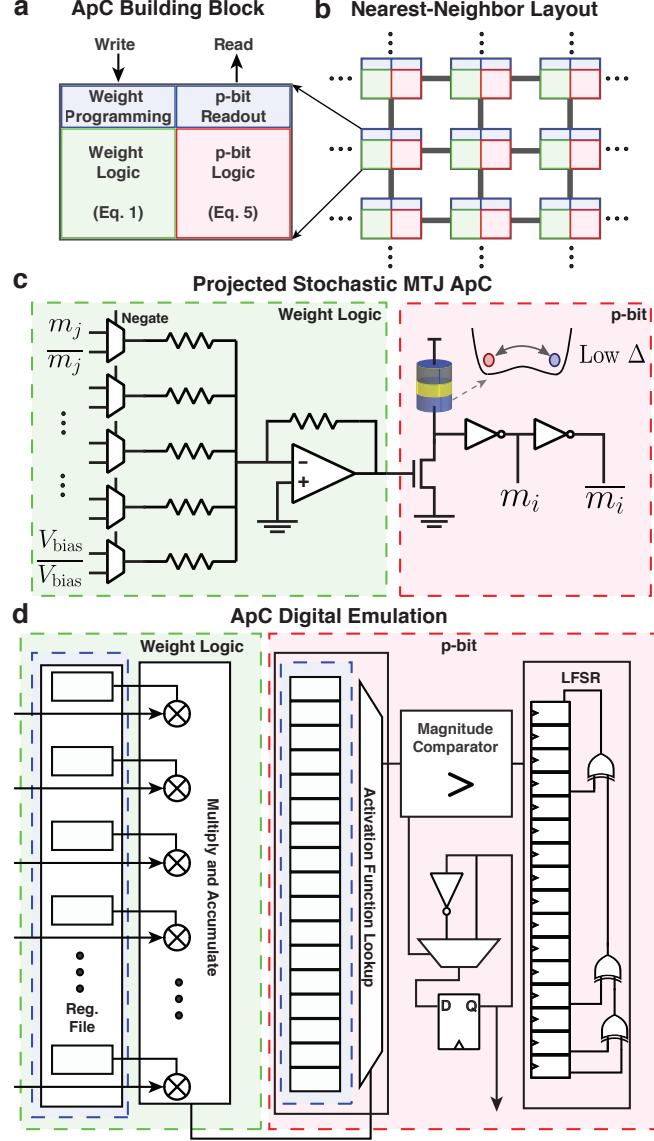


Fig. 4.1. **Autonomous p-Computer (ApC):** (a) A weighted p-bit building block is used to construct an ApC that comprises four components supporting weight logic, p-bit logic, weight programming, and p-bit readout. (b) The individual building blocks are interconnected to construct an ApC with a desired topology. A nearest-neighbor coupling layout is depicted. (c) A projected MRAM based ApC using nanomagnetic devices for the p-bit with resistor-based weight logic can be used to form a compact, efficient building block. (d) Using all-digital technology, an FPGA was used to construct an ApC that emulates the MRAM based design. An example composition of a p-bit with a linear, register based weight logic, a lookup table based activation function, a linear-feedback shift register based pseudo random number generator, and a pseudo-asynchronous attempt logic is shown.

4.1 ApC Implementation

The building block for our ApC has four components as shown in Fig. 4.1(a):

- *weight logic* to implement Eq. (4.1),
- *p-bit* to implement Eq. (4.5),
- *write unit* to program the weights W_{ij} and β
- *read unit* to access the individual p-bit outputs

Fig. 4.1(b) shows how multiple building blocks can be interconnected to form a p-computer. The tiling shown is based on nearest-neighbor connections, but the connections need not be limited to nearest-neighbor. We have also implemented all-to-all networks using the digital emulator shown in Fig. 4.1(d), as discussed in the Methods section.

Superficially Fig. 4.1 looks like other existing neural network architectures, like the one used for TrueNorth [179]. However, to our knowledge, earlier implementations have used time-multiplexing [179, 180] to share the same resource among different neurons and synapses, while our objective is to eliminate sequencing and time-multiplexing altogether so that we are not constrained by available clock speeds. TrueNorth for example uses 4096 neuronsynaptic cores, each core having dedicated neuron and synaptic memory forming 256 logical neurons that are time multiplexed sequentially to implement $4096 \times 256 \approx 1\text{M}$ logical neurons [179]. For our sequencerless operation we would need 1M distinct building blocks for the same number of neurons. This would be impractical if we were relying on fully digital implementations, but the compact hardware implementations currently being developed makes such a design feasible. As an example, stochastic behavior of nanomagnets has recently attracted attention in the context of novel computing paradigms, and they show promise in probabilistic and neuromorphic applications [7, 20, 181–185]. For example, an MRAM-based p-bit requires only 3 transistors and a magnetic tunnel junction [186], while its digital emulation requires significantly more transistors. With such compact hardware, it is feasible to have one building block for every p-bit in

order to support sequencerless, autonomous operation that is not limited by clock speeds.

Since digital platforms are inherently synchronous, we mimic autonomous operation by replacing Eq. (4.2) with a new hardware-inspired model, Eq. (4.5) (PPSL), that we benchmarked against established state-of-the-art physical models as described in the Methods section. These equations are based on SPICE simulations of Boltzmann networks where the update order of p-bits is irrelevant given symmetric coupling between connected p-bits. However, for certain networks such as those with directed connections, the update ordering of p-bits may be important and other hardware models more appropriate for these systems are likely required. These models are not discussed herein, but the overall FPGA architecture was designed to support the exploration of different hardware models and network topologies, hence they can be included here with minor effort.

At each time step, all p-bits are free to flip and they do so with a probability $\sim s$ that is controlled by the input I_i having a zero-input value $s(I_i = 0) = s_0 \ll 1$.

$$m_i(n+1) = m_i(n) \times \text{sgn}[e^{-s} - r_{[0,1]}] \quad (4.5a)$$

$$s = s_0 e^{-m_i(n)I_i(n)} \quad (4.5b)$$

In each time step the p-bit flips with a probability $\sim s$, so that the average time taken for it to respond is $1/s$. Since time steps are measured in units of τ_S , we have $\tau_N = (1/s) \times \tau_S$ as stated earlier. Unlike Eq. (4.2), Eq. (4.5) can be used to update all p-bits in parallel without explicitly worrying about simultaneous updates. With small values of s_0 , the fraction of simultaneous updates is sufficiently small such that Eq. (4.5) in an unsequenced mode gives results equivalent to those obtained from Eq. (4.2) with careful sequencing.

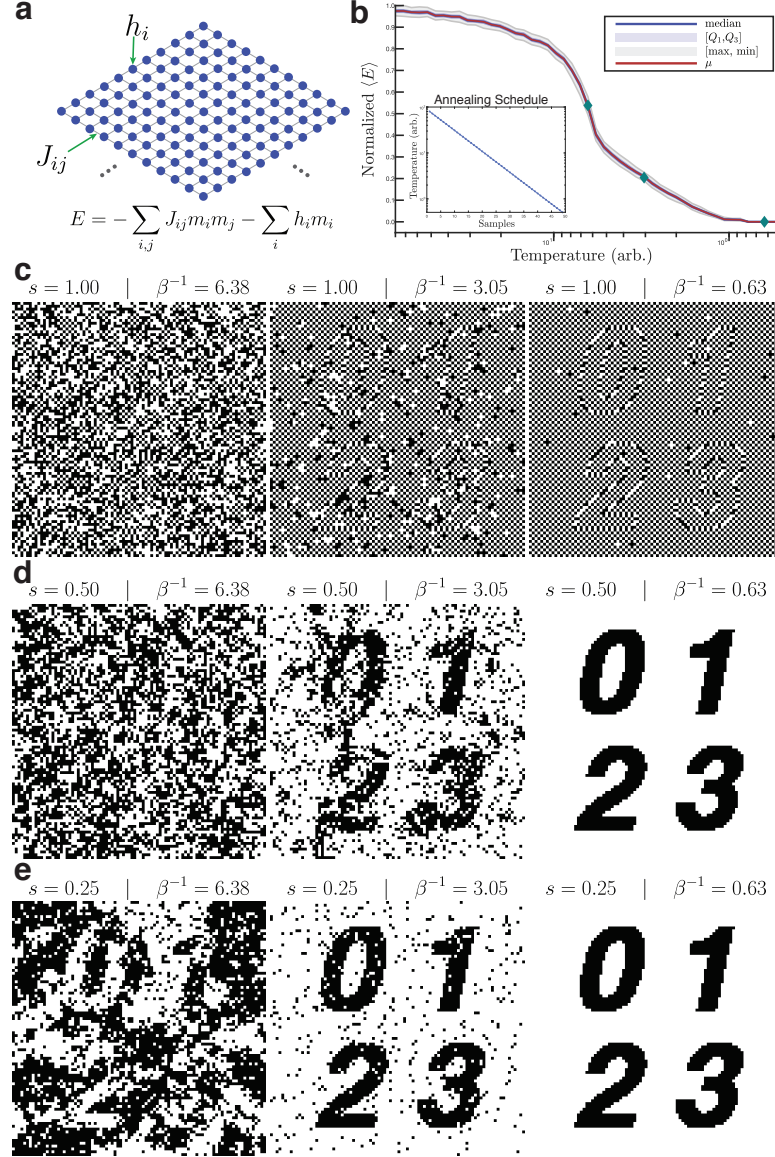


Fig. 4.2. Max-Cut Combinatorial Optimization: (a) An 8K spin (90x90) nearest-neighbor Ising lattice of p-bits supporting $\{\pm 1, 0\}$ weights can be used to solve a Max-Cut combinatorial optimization problem. Here a black and white image is used to generate magnetic domains corresponding to each character. The problem is then solved leveraging an FPGA ApC co-processor (see Methods). (b) Online annealing is used to transition the network from a high to low temperature through a reduction of $T(n+1) = 0.9 T(n)$ where $T = \beta^{-1}$ (inset). During this process the network converges to a low energy state solution to the problem of interest. (c) The network features begin to emerge as lower temperatures are reached. The probability of an individual p-bit flipping is controlled to explore how simultaneous updates effect the network or $s = 1$ (c), $s = 1/2$ (d), and $s = 1/4$ (e). If s is too large, there is no convergence to the ideal solution. As s is lowered, the network is more effective at finding low energy solutions, even at higher temperatures.

4.2 Combinatorial Optimization

Ising computers designed to solve problems in combinatorial optimization continue to emerge using a wide-range of underlying technologies. Solvers for such problems have been explored using quantum effects, optical approaches, digital logic, and magnetic technologies [5, 14–17, 187–191]. In general, these systems map a given optimization problem onto a hardware whose operation is guided by a cost function [70]. A common architecture used for such problems is the nearest-neighbor Ising model as shown in Fig. 4.2(a).

In the Ising model, each p-bit is connected to its neighbor through a coupling matrix, J_{ij} , and is influenced by an on-site bias h_i . Note that this is trivially mapped into W_{ij} as in Eq. (4.1). By mapping a problem of interest to this system, an Ising computer will intrinsically search for the lowest energy solution to the problem.

Typical implementations of these systems leverage some form of simulated annealing in hardware to guide the system into a low energy solution, using careful control and sequencing of spin-flip updates to avoid non-ideal spin updates [14, 192, 193]. In the context of a nearest-neighbor topology, the network can be split into two groups in a checkerboard-like pattern such that all spins in a given group can be updated in parallel [194]. This results in the ability to update half of all spins within a given clock period. According to Eq. (4.3a) this results in

$$f = \frac{N}{2\tau_{clock}} \quad (4.6)$$

An ApC can be used to solve the same class of problems, but to do so a value of s must be found that produces a solution with the desired fidelity. By identifying this limit for s , an upper bound is placed on the synapse speed that must be obtained to achieve a competitive f from Eq. (4.4).

Shown in Fig. 4.2 is a Max-Cut problem for which a black and white image was used to encode magnetic domains for the p-bits. The network is initialized to run at a high effective-temperature (low β) resulting in an effectively uncoupled network. The temperature is then gradually reduced according to an annealing scheduling

until the network crystallizes at a low energy, in this case ideal, solution as shown in Fig. 4.2(b). Shown in Figs. 4.2(c), 4.2(d), 4.2(e), different values of s are used to convey how simultaneous updating affects the ability of the network to converge to the solution. As s is decreased from 1 to $1/2$ and $1/4$, the smaller values of s result in a more effective convergence. While these results are heuristic in nature given their visual display, a quantitative energy based analysis conveys the same relationship as discussed in the following section with a demonstration of simulated quantum annealing.

For this problem, a value of $s = 1/4$ resulted in effective convergence to the ideal solution and well-behaved network operation. Given this result, the ApC has

$$f = \frac{N}{4\tau_S} \quad (4.7)$$

Comparing (4.6) and (4.7), as long as the synapse is twice as fast as the best τ_{clock} that could be obtained from a synchronous implementation, the network will perform the same number of flips per second without needing a sequencer.

4.3 Quantum Emulation

Simulating the behavior of quantum systems using classical models has long attracted interest. Since the seminal work of Suzuki [195], it has been known that thermodynamic features of quantum systems that avoid the “sign problem” [196] can be efficiently simulated by classical computers. This allows Quantum Annealing (QA) algorithms designed for quantum systems to be simulated on classical computers, an approach that is called Simulated Quantum Annealing (SQA). Computationally, SQA uses a finite number of “replicas” where each replica is sized to match the size of the original quantum system. This replication enables a mapping of the quantum system to a classical collection of p-bits. The number of replicas that are needed depends on the temperature of the quantum system [197,198] and the desired accuracy to emulate the quantum system. SQA algorithms are typically run on software [197–199] and on sequencer-based hardware designs [187,200]. Here, we demonstrate how quantum

systems can be emulated using our ApC by solving a model quantum system, the Transverse Ising Hamiltonian, and establish how ApC exactly reproduces the thermodynamics of a many-body quantum system at finite temperatures and magnetic fields.

Recently, it was theoretically argued [201] that a network of p-bits described by Eq. (4.1) and interconnected according to Eq. (4.2) can be used to emulate a quantum system with a finite number of classical replicas, where a p-bit is represented in hardware by the stochastic MTJ-based implementation of Fig. 4.1(c).

We start from the nearest neighbor Transverse Ising Hamiltonian in 1D [202]:

$$\mathcal{H}_Q = - \left(\sum_i^M J_{i,i+1} \sigma_i^z \sigma_{i+1}^z + \Gamma_x \sum_i^M \sigma_i^x + \Gamma_z \sum_i^M \sigma_i^z \right) \quad (4.8)$$

where $J_{i,i+1}$ represents the interaction between neighboring spins, Γ_x and Γ_z are local magnetic fields in the \hat{x} and \hat{z} directions. After a Suzuki-Trotter mapping, the 2D classical Hamiltonian that approximates this system with n replicas is given as:

$$\mathcal{H}_C = - \left(\sum_{k=1}^n \sum_{i=1}^M (J_{\parallel})_{i,i+1} m_{i,k} m_{i+1,k} + \gamma_z m_{i,k} + J_{\perp} m_{i,k} m_{i,k+1} \right) \quad (4.9)$$

where $(J_{\parallel})_{i,j} = J_{i,j}/n$, $\gamma_z = \Gamma_z/n$ and the vertical coupling term is $J_{\perp} = -1/(2\beta) \log \tanh(\beta \Gamma_x/n)$ and $m_{i,j} \in \{-1, +1\}$.

Eq. (4.9) can be used to find each diagonal element of the quantum density matrix and therefore, all diagonal operators, and their correlations can be calculated from it. The corresponding interaction matrix (J_{ij}) to perform a p-bit simulation can be calculated from the mapped classical Hamiltonian, such that $I_i = -\partial \mathcal{H}_C / \partial m_i$ to yield the weight coefficients. Note that the Suzuki-Trotter decomposition adds another dimension to the classical system, therefore a 1D linear chain for a quantum system is emulated by a 2D classical system.

In Fig. 4.3, we simulate a ferromagnetic linear chain ($J_{i,j} = +2$) using $M = 8$ spins with periodic boundary conditions at different transverse magnetic fields and we compute the average magnetization, $\langle m_z \rangle$. We include a symmetry breaking field in the $+z$ direction ($\Gamma_z = +1$) to obtain a net magnetization at vanishing transverse fields

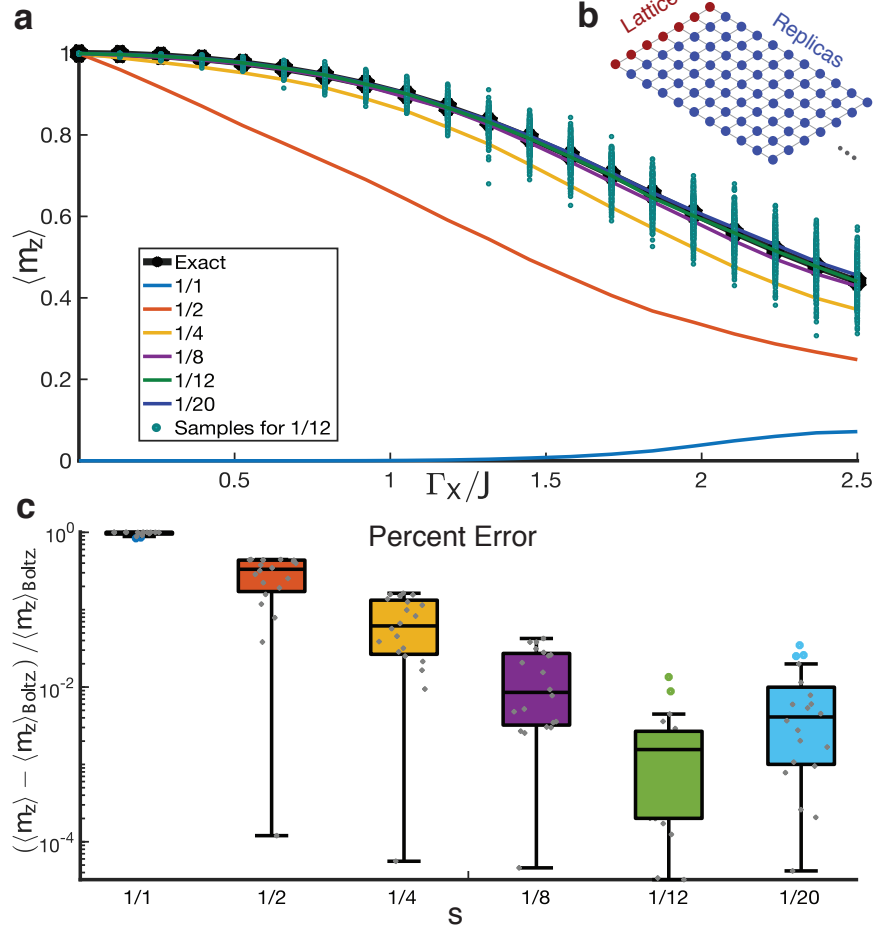


Fig. 4.3. **Emulating the Transverse Ising Hamiltonian:** (a) A 1D ferromagnetic linear chain ($J_{ij}=+2$) with $M = 8$ spins described by the quantum transversing Ising Hamiltonian is solved as a function of the transverse magnetic field Γ_x at an inverse temperature of $\beta = 20$ using a network of 8×250 p-bits with periodic boundary conditions. The 250 p-bits serve as replicas (b) generated using a Suzuki-Trotter decomposition. The exact quantum Boltzmann solution is compared against the average z magnetization for different values of s . As s is decreased, the system converges to the Boltzmann solution for a value of $s = 1/12$. As s decreases beyond $s = 1/12$, the system begins to diverge from the solution for larger Γ_x/J due to the chosen precision of the digital implementation. For each Γ_x and s , 200 samples from a free-running FPGA implementation were collected spaced $\approx 30,000$ synapse delays apart. (c) Boxplots are shown of the difference in computed mean at each Γ_x/J from the ideal Boltzmann solution, for each value of s .

($\Gamma_x = 0$). We obtain the exact density matrix by directly diagonalizing the quantum Hamiltonian (Eq. (4.9)): $\rho = \exp(-\beta\mathcal{H}_Q)$, where we chose an inverse temperature of $\beta = 20$. Once ρ is known, we compute the average magnetization by tracing it with the magnetization operator S^z , $\langle m \rangle = \text{tr}[S^z \rho] / \text{tr}[\rho]$. The exact solution is shown as a black solid line in Fig. 4.3(a).

The corresponding average magnetization is obtained by running the classical system that is described by Eq. (4.9) with $n = 250$ replicas ($250 \times 8 = 2000$ spins) using Eq. (4.5) in the FPGA emulator. Fig. 4.3(a) shows different s values that are used to obtain the exact result. Clearly, choosing s too high ($s = 1/1$) fails, but gradually decreasing s allows the result to approach the exact result. We observe that $s = 1/12$ seems to be an optimal choice, as decreasing s further does not yield more improvement due to the chosen precision in the digital implementation. Specifically, as s continues to reduce, the chosen precision of the arithmetic logic and look-up tables in the FPGA emulator design limits the accuracy of the solution (please see Methods). Fig. 4.3(b) quantitatively shows a boxplot of the error incurred at each s value across 200 trials. In the the Suzuki-Trotter decomposition, increasing Γ_X/J systematically increases the error but a reasonably close agreement is observed for $s \leq 1/8$.

Typically, SQA algorithms initialize the system at a high magnetic field (Γ_X) and slowly remove it to keep the system in its ground state and guide it to the desired ground state of the Ising Hamiltonian. In Fig. 4.3, however, we have not changed the magnetic field as a function of time, but rather sampled from 200 ensembles, each separated by $\sim 30,000$ synapse delays, to obtain the system statistics. This means that not only was the system guided to the expected ground state ($\Gamma_X \rightarrow 0, \langle m \rangle \rightarrow 1$), but it also followed the correct average magnetization at high magnetic fields. As such, this could be viewed as an example of sampling a probability distribution rather than finding the ground state of the system [74], an important problem space where an ApC could be useful.

It is important to note that to solve optimization problems using SQA, an exact mapping of the quantum system may not be optimal and a finite number of replicas that only approximate the thermodynamics of the quantum system could lead to more efficient results. In the present context, optimizing the number of replicas for a given system can be arranged since our system is *not* a natural quantum system nor is it trying to faithfully reproduce the behavior of such a system. Therefore, optimizing the number of physical replicas or finding the replica with the lowest energy becomes possible in our engineered system [198].

4.4 Discussion

These examples demonstrate the feasibility of an ApC governed by Eq. (4.5) to follow Boltzmann statistics without the need for a controlling sequencer. But in order to make use of ultrafast flip rates, f , enabled by short τ_S times, it is important that the building blocks be energy efficient to ensure that power levels are acceptable:

$$P = f \varepsilon \quad (4.10)$$

where P is the power budget and ε is the energy required per flip.

In Table 4.1, we compare representative Ising computers based on sequenced designs [14, 192] to the autonomous designs implemented in and projected by this work. The last row of the table shows ε for both the sequenced and autonomous designs. As shown in the table, the FPGA based 8K-spin ApC achieves an energy per flip that is similar to the Janus II sequenced FPGA implementation. However, this comparison applies some artificial constraints on the Janus II design: namely that all spins must be simultaneously resident in the device at one time. This is a requirement for unlocked autonomous designs that for the sake of comparison we applied to the sequenced design. A clocked, sequenced design can pause the system and leverage external memory for storage of neuron state, providing a large pool of logical neurons, though limited by available memory bandwidth. Additionally, the ability to

	Sequenced		Autonomous (This Work)		
	Hitachi ^a	Janus II ^b	2K QA ^c	8K Ising ^c	Projected ^d
Technology	CMOS (SRAM)	CMOS (FPGA)	CMOS (FPGA)	CMOS (FPGA)	CMOS + MTJ
Total Power (W)	0.05	25	55	32	19.25
Number of Neurons (N)	20K (80×256)	2K	2K (8×250)	8.1K (90×90)	1M
$s \equiv \tau_S/\tau_N$	1^\dagger	1^\dagger	1/12	1/4	1/10
Synapse Delay (τ_S) (ps)	10,000	4,000	8,000	8,000	10
Neuron Delay (τ_N) (ps)	10,000	4,000	96,000	32,000	100
Flips per Second (f)	1×10^{12}	2.5×10^{11}	2.08×10^{10}	2.5×10^{11}	1×10^{16}
Spin Update Time ($1/f$) (ps/flip)	1	4	48	4	0.0001
Energy per Flip (ϵ) (nJ/flip)	5×10^{-5}	0.1	2.64	0.13	1.93×10^{-6}

^a Calculations based on information extracted from [14]. Design used sequential implementation, therefore we assume (\dagger) $s = 1$ and an optimal updating scheme such that half of all neurons are updated every step such that $f = (0.5)N/\tau_S$. Synapse and neuron delays assumed to operate at interaction frequency of 100 MHz.

^b Calculations based on information extracted from [192]. Janus II supports a much larger number of neurons using external memory, multiple FPGAs, and data shuttling between devices. Here we limit the comparison to only a single spin processor assuming all spins are co-located on one FPGA, as would be required for an autonomous design. As with (a), we assume (\dagger) $s = 1$. With all spins on one chip, an optimal updating scheme updates half of all neurons every step for $f = (0.5)N/\tau_S$. Based on this, the spin update time calculated here is 4 instead of 2 as in [192]. Design assumed to operate at operating frequency of 250 MHz.

^c Power consumption measured from maximum power draw during computation. 2K QA topology based on 16-bit precision operations and 8K Ising is based on 2-bit precision operations. Selected s values based on Figs. 4.2 and 4.3.

^d Assuming chip density of 1M based on memory density available from commercial ST-MRAM [203]. With nearest-neighbor connections, a synapse delay of 10 ps assumed [175]. Assumed neuron delay of ~ 100 ps and a steady-state neuron power consumption $10 \mu\text{W}$ [177]. Overhead from nearest-neighbor resistive cross-bar and external communication logic estimated as ~ 9.25 W using 5 memristors and 1 op-amp per neuron [204].

Table 4.1.

Comparison of Sequential and Asynchronous Ising Computers:

Ising computers proposed to-date have used sequential updating mechanisms based on CMOS technology. These designs achieve spin update times of $\sim 1 - 4$ ps/flip. The 20K SRAM chip from [14] achieves an energy of 50 fJ per flip. The autonomous CMOS implementations demonstrated in this work obtain a similar spin update time as sequenced implementations and comparative energy per flip as sequential FPGA implementations. Using CMOS and MTJ technologies as proposed herein, a highly efficient design with petaflips per second with 2 fJ per flip is projected.

pause the network enables time-multiplexing and can harness the ability to re-use logic resources.

The FPGA results naturally consume more power and have reduced density [205]. The 8K spin result of Table 4.1 has ~ 18 W of static power dissipation due to the periphery included in the FPGA design, not all of which is used. Using the FPGA to ASIC power ratio [205] of 14, a naive migration of the design to an ASIC would result in an ε of $\sim 4.0 \times 10^{-3}$. By translating these approaches to a tailored ASIC implementation, the energy efficiency per flip increases and the ability to increase the density of resident neurons within a device increases substantially. Extending this further, it should be feasible to obtain designs with ~ 10 *petaflips per second* with a power budget of ~ 10 W, but we need devices with $\varepsilon \sim 1$ fJ that also support a density of 1M devices. We note that the corresponding flip time (100 attoseconds per flip) is orders of magnitude shorter than what is achieved in highly parallelized GPU implementations (33 picoseconds per flip) that implement Monte Carlo simulations of Ising models [206].

The CMOS based SRAM design of Hitachi [14] has an estimated energy per flip of $\varepsilon \sim 50$ fJ, though the neuron density limits the ability of the approach to obtain an f of petaflips per second. Using a hybrid CMOS and MTJ design as would be encountered in modern commercial MRAMs [203], it should be feasible to obtain ~ 10 petaflips per sec as projected in the last column of Table 4.1 [177]. Modern MRAMs can achieve Gb densities; however, we limit the projection to a 1 Mb density based on a target power consumption of ~ 20 W for the neurons and synapses in the design.

It should be noted that the 20 W power target was arbitrarily chosen. In principle the autonomous designs can leverage clocking to choose when global p-bit updates should occur, much like the FPGA emulator discussed in the methods section. While this approach can save power, it limits the utility of the MTJ based approach by constraining the flips per second (f) to the same limits of Eq. (4.3a) due to the clocking scheme. Even with this limit in place, as the connectivity between neurons increases beyond nearest-neighbor, the sequencing logic becomes more complex while

the ApC only requires a balance of s to ensure proper convergence. In the case of all-to-all connectivity, the sequencing logic reduces in complexity and technically can be implemented with a single time-multiplex weighted p-bit. In this situation, the benefits of an ApC begin to degrade, except for the elimination of memory bottlenecks with the use of distributed weights and the avoidance of time-multiplexing. A 500 node all-to-all network was implemented using the FPGA emulator, and the resulting s was directly proportional to the number of neurons, affirming the reduced benefit.

Based on these results, the removal of sequencers, ability to run at speeds limited only by synapse delays, and ability scale to millions of neurons, all within an accessible power budget, makes an ApC a compelling alternative to clocked, sequential designs for SNN coprocessing.

4.5 Methods

4.5.1 FPGA p-computing coprocessor

An all-digital framework based on Eqs. (4.5a) and (4.5b) was developed, Fig. 4.4, to facilitate architectural exploration of an ApC, study various trade-offs in p-bit, weight logic, and topology design, and to accelerate the combinatorial optimization and sampling problems used in this work. The digital framework leverages reconfigurable computing devices to support rapid exploration of different designs. A Xilinx Virtex Ultrascale+ xcvu9p-flgb2104-2-i provided via Amazon Web Services F1 cloud-accessible EC2 compute instances was used for the problems of Figs. 4.2 and 4.3. While a Xilinx FPGA was used in this work, the design is hardware agnostic and another device, e.g. an Intel Stratix FPGA, can be used.

As shown in Fig. 4.1(b) and Fig. 4.4(b), an ApC comprises multiple weighted p-bits arranged in various topologies, each supporting programmable problem instances. There are many options for the implementation of programmable control, weight logic, p-bits, and p-bit readout in a digital platform. The digital ApC of Fig. 4.4 comprises a modular weighted p-bit, Fig. 4.4(c), that can be organized into various topologies,

Fig. 4.4(b), supporting programmed problem instances. An example weighted p-bit implementation is shown in Fig. 4.1(d) leveraging a memory-mapped weight-logic register bank supporting linear weight coupling. The output of the weight logic block is provided to a programmable, activation function look-up table that is used in conjunction with a Linear Feedback Shift Register (LFSR) based pseudo-random function to implement Eqs. (4.5a) and (4.5b). Multiple options were developed and explored for these building block elements, see Fig. 4.6 beyond what is shown in Fig. 4.1(d).

Interaction with the framework is provided through MATLAB MEX programs in a client-server command based model, Fig. 4.4(a). Clients issue commands to select which of the pre-built topologies to program into the cloud FPGA instance, the current pseudo-temperature for the network, problem specific weights, and options to pause or resume the network. Commands are also provided to support random sampling from the network for readout operation. Online annealing is directly supported through global update operations of the activation function look-up. All weights are dynamically programmable through memory-mapped operations as discussed in the following section. The server interfaces with a PCIe attached FPGA, interacting with the programmed design as commanded. Other clients such as Octave and Python are readily supported through a C++ abstraction layer leveraging networking and serialization libraries.

4.5.2 FPGA ApC Logical Organization

Fig. 4.5 depicts the logical organization of the ApC used in this work. All interaction with the FPGA is performed using a PCIe Gen3 x16 interface supporting direct memory access (DMA) transactions. Programmable control of the design is accomplished using an AXI lite 32-bit slave interface and memory map decoder. The address space for the ApC is divided into a few regions: global control and informa-

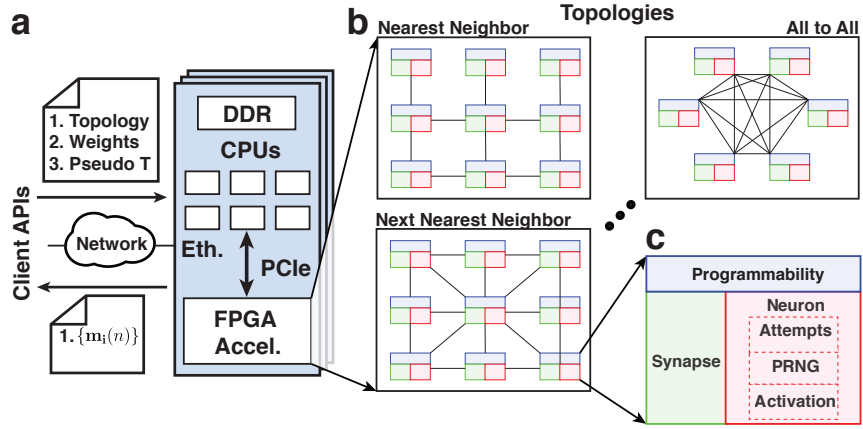


Fig. 4.4. **Cloud Accessible p -computing Co-processor:** (a) Client applications are used to specify a desired network topology, problem specific weights, and current pseudo-temperature through a network accessible pool of dedicated servers. The servers provide general purpose processing, network connectivity, and PCIe accessible FPGA accelerators. After loading the desired topology into the FPGA, the system operates and provides a sampling interface for the current state of the network p -bits. (b) The architecture supports multiple topologies depending on the problem of interest ranging from nearest-neighbor connectivity to all-to-all connectivity. (c) Each of the modular p -bits in the network comprises a bus accessible programmable interface, synapse block, and neuron. The neuron is further modularized to support different approaches for flip-attempt logic, pseudo random number generation, and activation function implementations.

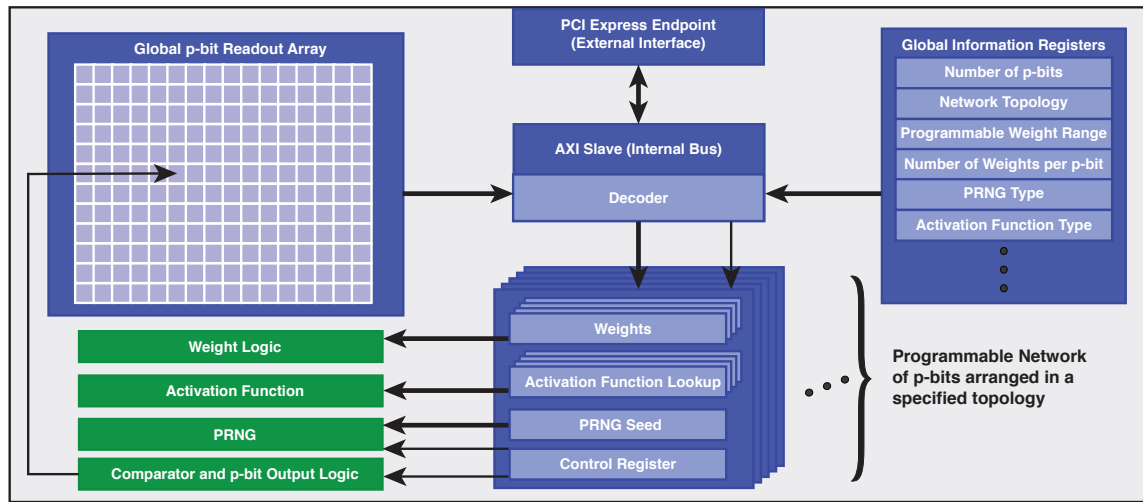


Fig. 4.5. **Logical Organization and Memory Map:** The FPGA based ApC coprocessor is accessible to a host processor from a dedicated PCIe endpoint within the FPGA. This endpoint provides logical access to the internal memory map of the ApC. The ApC contains global information and control registers, a global p-bit array facilitating readout, and finally individual control of each weighted p-bit through a localized, dedicated address space.

tion registers, p-bit array for readout, and individual control for each weighted p-bit in the design.

The global address space provides information on the ApC pertinent for client interaction with the system. This includes information such as the total number of p-bits in the design, the p-bit connectivity graph topology, weight precision, and other useful runtime information such as the number of elapsed synapse delays between client sampling requests. Additionally, global control functions include the ability to pause and resume the network so that a client can access the global p-bit readout array. This readout array holds the current output of all p-bits sampled on each system clock cycle. Reads from this interface are performed when the network is paused to ensure atomic readout of all p-bits given the limited bandwidth of the readout interface.

Each p-bit in the system has a local memory space supporting programmable control of its function. Each p-bit has localized programmable weights enabled through registers or internal memory (RAM), a programmable activation function lookup table supporting direct look-up or interpolation, support for seeding the chosen pseudo random number generation (PRNG) function, and finally a set of control registers for the p-bit. The output of the p-bit programmable elements directly interface with the weight logic, activation function, PRNG, and comparator operation of the weighted p-bits. Online annealing is accomplished using a global bus broadcast when programming the activation function lookup tables, so that all p-bits are updated simultaneously. Alternatively, each p-bit's activation lookup table can be independently controlled, allowing the exploration of non-uniform bias, local temperature effects, and other non-idealities, facilitating future opportunities for exploration.

4.5.3 Digital p-bit Implementation Flexibility

A modular digital p-bit was designed to support different options for the p-bit building block elements. Each p-bit is logically partitioned into a unit for pseudo-

randomness or “entropy”, a block for computing the activation function, and a portion that uses the results of a comparison between the activation function and PRNG to determine if a p-bit update attempt should occur. There are various ways to construct these elements using digital logic. In this design, a few select implementations were explored as shown in Fig. 4.6.

Fig 4.6(a) and 4.6(b) show two methods for performing autonomous updates of each p-bit. Shown in Fig. 4.6(a) is the logic corresponding to Eq. (4.5a) where the comparator provides $\text{sgn}(e^{-s} - r_{[0,1]})$. This approach emulates autonomy while preserving fully synchronous operation of the digital design. This p-bit update logic was used for the problems in this work. The activation look-up is programmed with e^{-s} .

Alternative approaches were explored and implemented for the p-bit updates including the use of free-running ring oscillators, Fig. 4.6(b), to mimic the naturally stochastic update frequency of an MTJ based p-bit as described by (4.2). In this implementation, each p-bit has a dedicated free running ring oscillator that generates asynchronous “attempt” edges that are synchronized into a system clock domain. Each asynchronous edge determines when the p-bit should attempt to update based on the current output from a comparator according to Eq. (4.2), in which case the activation look-up is programmed with \tanh .

While the attempt logic is used to determine *when* an update attempt should be made, the logic relies on the output of a comparator to determine *if* a flip should occur. The comparator computes the sign of the difference between the output of a PRNG and the output of the programmed activation function. Shown in the second row of Fig. 4.6 are two PRNG implementations. A LFSR is a PRNG that provides pseudo randomness in a compact design at the expense of output quality, Fig. 4.6(c). While the LFSR may be sufficient for many problems, a higher-quality PRNG was implemented that requires minimal FPGA resources, but significantly improves the PRNG quality [207].

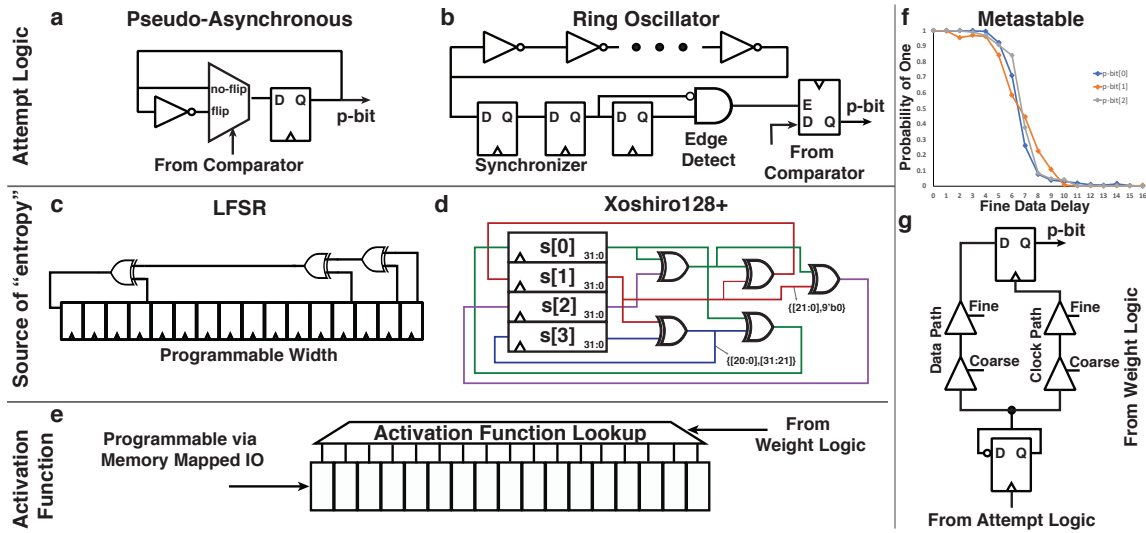


Fig. 4.6. **Modular Autonomous Digital p-bit Components:** (a) Using output from a comparator, the state of a given p-bit is probabilistically flipped on any given system clock edge. (b) Alternatively, a free-running ring oscillator can be used to generated edges centered around a oscillator characteristic frequency, mimicking the attempt rate of a stochastic nanomagnet. These edges control the enable of the p-bit which is then updated based on the output of a comparator. (c) Pseudo-randomness is used to provide the rand() function. A linear feedback shift register is an area efficient means to generate pseudo random values, however the quality is limited. (d) For improved quality of pseudo random output at the expense of more area, a Xoshiro128+ [207] generator can be used. (e) The p-bit activation function can be implemented using a straight-forward look-up table or through an interpolated look-up table output. (g) By combining the source of “entropy” with the activation function, a tunable delay chain can be used to leverage controlled metastability at the input of a flip-flop to produce a probabilistic output. Outputs from three independent p-bits are shown overlaid in (f).

The second input to the comparator is from an activation function output, shown on the third row of Fig. 4.6(e). As implemented, a straight-forward look-up table was sufficient for the designs in this chapter. However, an improved interpolation based activation function [208] would improve the accuracy of the lookup results and may be necessary for certain problem classes.

Finally, an additional building block was created to leverage *physical* randomness and a built-in sigmoidal response from within a digital design as shown in Fig. 4.6(g). Leveraging a delay based building block [209] and flip-flop metastability, “true” entropy was used to construct a sigmoidal response as depicted in Fig. 4.6(f). However, over continued operation within the device, temperature and other variations caused the sigmoidal curves to drift, resulting in non-uniform bias and operation. As a result, this building block is not currently being used in the design; however, it does provide insight into non-idealities that may be encountered in a chip design.

4.5.4 Benchmarking Eq. (4.5) with stochastic LLG

A coupled stochastic Landau-Lifshitz-Gilbert (sLLG) equation is solved and benchmarked against the autonomous p -bit model of (4.5a) and (4.5b) (PPSL). Magnetization dynamics of a circular stochastic nanomagnet are captured by solving the sLLG equation in the macrospin assumption within a modular [153] SPICE framework ,

$$(1 + \alpha^2) \frac{d\hat{m}}{dt} = -|\gamma|\hat{m} \times \vec{H} - \alpha|\gamma|(\hat{m} \times \hat{m} \times \vec{H}) + \frac{1}{qN_s}(\hat{m} \times \vec{I}_S \times \hat{m}) + \left(\frac{\alpha}{qN_s}(\hat{m} \times \vec{I}_S) \right) \quad (4.11a)$$

where α is the damping coefficient, γ is the electron gyromagnetic ratio, $N_s = M_s \text{Vol.} / \mu_B$ is the total number of Bohr magnetons in the magnet, M_s is the saturation magnetization, $\vec{H} = \vec{H}_d + \vec{H}_n$ is the effective field including the out-of-plane (\hat{x} directed) demagnetization field $\vec{H}_d = -4\pi M_s m_x \hat{x}$, as well as the thermally fluctuating magnetic field due to the three dimensional uncorrelated thermal noise H_n with zero mean $\langle H_n \rangle = 0$ and standard deviation $\langle H_n^2 \rangle = 2\alpha kT / |\gamma| M_s \text{Vol.}$ along each direction, \vec{I}_S is the applied spin current to the nanomagnet.

Individual p-bits are coupled according to:

$$I_{s,i}^z(t + \Delta_t) = \beta I_{s0} \sum_j W_{ij} \text{sgn}(m_j^z(t)) \quad (4.12)$$

where, I_{s0} is the tanh fitting parameter of the sigmoidal response ($\text{sgn}(m_z)$ versus spin current I_s^z along z -direction). In the benchmark, a circular disk magnet with a vanishing anisotropy (H_K) is used with the parameters: diameter $D = 150$ nm and thickness $t = 2$ nm, $\alpha = 0.01$, $M_s = 1100$ emu/cc, $H_K = 1$ Oe resulting in an autocorrelation time of $\tau_{corr} = 1.372$ ns and $I_{s0} = 1$ mA. A fitting parameter of 1.4 is used in the PPSL model for τ_N , i.e. $\tau_N = 1.4\tau_{corr}$.

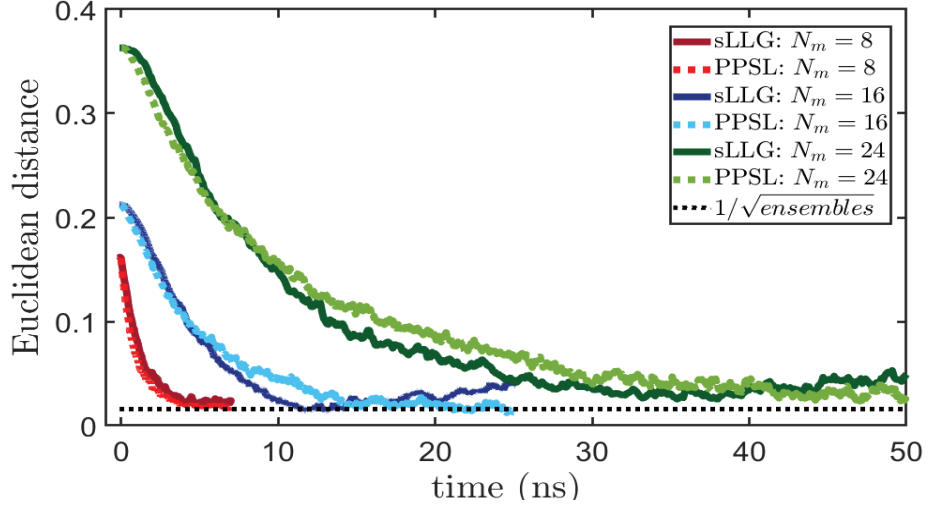


Fig. 4.7. **Benchmarking the PPSL Model with sLLG using Euclidean distance:** Using a random Sherrington-Kirkpatrick spin glass instance for different network sizes, N , the PPSL model is benchmarked against sLLG as a function of time. Each point on the graph represents the Euclidean distance from the ideal Boltzmann distribution and the ensemble solution obtained from PPSL and sLLG. The steady state error will depend on the number of ensembles as shown by the black dotted line.

The simulated network is a Sherrington-Kirkpatrick [210] spin glass with a random coupling matrix and random bias between -1 and +1. The benchmarking of the

proposed PPSL model with the coupled sLLG network, analogous to the probabilistic circuit proposed in [50], is accomplished by comparing two different quantities: (1) Euclidean distance and (2) Free energy.

Euclidean distance is defined by:

$$ED = \sqrt{\sum_{i=1}^{2^N} (P_i - P_{i,\text{Boltzmann}})^2} \quad (4.13)$$

where P_i is the probability of occurrence of the i -th configuration computed out of 4000 ensembles at each time step of the simulation. $P_{i,\text{Boltzmann}}$ is computed from the joint probability distribution obtained from a Boltzmann law.

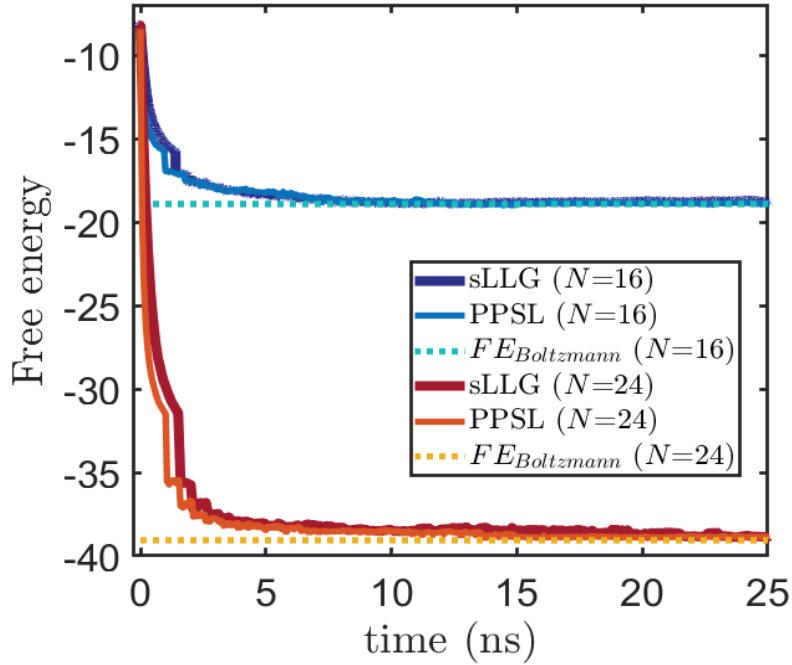


Fig. 4.8. **Benchmarking the PPSL Model with sLLG using Free Energy:** The free energy calculated for the random Sherrington-Kirkpatrick spin glass instance of Fig. 4.7 from the PPSL model is benchmarked against sLLG as a function of time for network sizes $N = 16$ and $N = 24$, showing convergence to the free energy obtained from Boltzmann law.

The second benchmark approach is based on a comparison of the free energy of the system with what is expected from the principles of statistical mechanics. Free energy is defined by [211] the partition function Z :

$$FE = \frac{\ln(Z)}{-\beta} \quad (4.14)$$

where, β is the pseudo-inverse temperature. Partition function Z is given by:

$$Z = \sum_k \exp(-\beta E_k) \quad (4.15)$$

where k represents different configurations of the network. Energy of a specific configuration is defined by:

$$E_k = -0.5 \sum_{\substack{i,j \\ i \neq j}} W_{ij} m_i m_j - h_i m_i - h_j m_j \quad (4.16)$$

When numerically calculating free energy from the sLLG data, the following steps have been applied (similar to the importance sampling method described in [212]):

1. The probability of different configurations, P_i , are calculated out of 4000 ensembles for each time step
2. For each P_i larger than a certain threshold value P_{th} , the partition function $Z_i = \exp(-I_0 E_i)/P_i$ is calculated, so that outliers are excluded
3. For each Z_i , the free energy $FE_i = -\ln(Z_i)/I_0$ is calculated.
4. Finally the mean of all FE_i is computed.

The above method is suitable for small examples, but may not scale to large examples due to the difficulty in empirically calculating different probabilities P_i as the network size grows. The striking agreement between the sLLG model and the behavioral model given by Eq. (4.5) shown in Figs. 4.7 and ref4.14 establishes the validity of Eq. 4.5 as a suitable model for the projected autonomous, stochastic MTJ-based computer.

5. LOOKING AHEAD

In this work, we explored areas related to reducing the control complexity for semiconductor spin-based quantum systems, Chapter 2, discussed in Chapter 3 how the intrinsic nature of stochastic nanomagnets forms a natural Boltzmann machine for use in combinatorial optimization, and finally introduced a framework to emulate autonomous probabilistic computers in Chapter 4 and highlighted the potential speed-up enabled by a truly clockless, scaled ApC. As we look ahead towards additional areas of exploration, a number of opportunities stand-apart in the near-term. In the following sections we will highlight some of these areas, extending on the work presented herein.

5.1 Quantum and Probabilistic Computing

With the promise of quantum computing, numerous endeavors are underway to build a viable quantum computer with spin-based quantum computing highlighted in Chapter 2 amongst them. The continued prevalence of systems that require sophisticated and complex control of qubits, e.g. through microwave control, continues to imply scaling will remain an elusive challenge. It is my hope that methods to reduce the difficulty of qubit control, for example spin-torque based manipulation of a semiconductor spin, will help enable continued scalability of these systems.

Throughout this report, we have continually alluded to the seminal work of Feynman [51] who initially explored the relationship between computing with classical probabilities and those of quantum. He conveyed a clear distinction between the two approaches leveraging an example of Bell's inequality in the process [52]. However, while this fundamental difference is clearly present in nature, many advances in quan-

tum computing translate directly to probabilistic systems either through algorithmic advances or in some cases efficient emulation of such systems, see 4.3.

Indeed, continued research into the area of quantum computing, especially for those systems based on Ising like connectivities naturally aids the probabilistic computers discussed in this work. There are a number of open areas to explore such as finding embeddings of problems onto a given hardware topology given finite connectivity graphs and weight values [213–216]. While it is straightforward, based on the work of Lucas [70], to identify a mapping from a given problem to a general connectivity graph, in practice the physical topology of a hardware places limitations on the available connections and weights. As additional embedding approaches are identified, the embeddings can be applied to both adiabatic quantum systems as well as their probabilistic counterparts.

5.2 Realization of Autonomous Probabilistic Computing

Beyond the influences of quantum computing, the probabilistic computers discussed in this work, Chapters 3 and 4 and in several other sources help motivate continued investigation into beyond-Moore, non-Von Neumann architectures. Within the field of probabilistic computing, there are numerous areas ripe for additional exploration.

While an efficient magnet based p-bit has been identified as a viable building-block [186], other materials and approaches may enable an alternate approach. For example, nanoscale ferroelectric materials [217] could in theory be used to realize a probabilistic element that conveniently couples to well-established CMOS technology [218, 219]. By leveraging a stochastic ferroelectric material, it should be possible to amplify the noise at the input of a MOSFET or inverter, providing a source of amplified entropy. How this can be engineered into a competitive p-bit remains to be seen, but to that end a time-domain stochastic SPICE model based on [217, 220] is provided in Appendix B to facilitate such exploration.

In addition to exploring alternatives for the p-bit building block, and the continued exploration to find an efficient, compact synaptic implementation, see for example Table 3.1., an important area to explore is the scaled physical realization of the autonomous probabilistic computer highlighted in Table 4.1. Combining the predictive emulation of Chapter 4 and the recent experimental results demonstrating viable stochastic nanomagnets coupled in a clockless design [178], realization of a scaled heterogeneous autonomous probabilistic computer should be possible in the coming years. Such a device could provide a concrete performance improvement over existing clock-based Ising computers used in optimization. Indeed the discussion of 4.3 highlights the pragmatic means with which such a system could effectively emulate quantum, stoquastic systems.

REFERENCES

- [1] M. M. Waldrop, “The chips are down for Moore’s law,” *Nature News*, vol. 530, no. 7589, p. 144, Feb. 2016.
- [2] T. N. Theis and H. S. P. Wong, “The End of Moore’s Law: A New Beginning for Information Technology,” *Computing in Science Engineering*, vol. 19, no. 2, pp. 41–50, Mar. 2017.
- [3] T. M. Conte, E. P. DeBenedictis, P. A. Gargini, and E. Track, “Rebooting Computing: The Road Ahead,” *Computer*, vol. 50, no. 1, pp. 20–29, Jan. 2017.
- [4] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O’Brien, “Quantum computers,” *Nature*, vol. 464, no. 7285, pp. 45–53, Mar. 2010.
- [5] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, “Evidence for quantum annealing with more than one hundred qubits,” *Nature Physics*, vol. 10, no. 3, pp. 218–224, Mar. 2014.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [7] J. Grollier, D. Querlioz, and M. D. Stiles, “Spintronic Nanodevices for Bioinspired Computing,” *Proceedings of the IEEE*, vol. 104, no. 10, pp. 2024–2039, Oct. 2016.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [9] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, “A Survey of Neuromorphic Computing and Neural Networks in Hardware,” *arXiv:1705.06963 [cs]*, May 2017.
- [10] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [11] J. J. Hopfield and D. W. Tank, ““Neural” computation of decisions in optimization problems,” *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, Jul. 1985.
- [12] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, “Quantum annealing with manufactured spins,” *Nature*, vol. 473, no. 7346, pp. 194–198, May 2011.

- [13] F. L. Traversa, C. Ramella, F. Bonani, and M. D. Ventra, “Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states,” *Science Advances*, vol. 1, no. 6, p. e1500031, Jul. 2015.
- [14] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, and H. Mizuno, “A 20k-Spin Ising Chip to Solve Combinatorial Optimization Problems With CMOS Annealing,” *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 303–309, Jan. 2016.
- [15] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-i. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, “A coherent Ising machine for 2000-node optimization problems,” *Science*, vol. 354, no. 6312, pp. 603–606, Nov. 2016.
- [16] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, “A fully-programmable 100-spin coherent Ising machine with all-to-all connections,” *Science*, p. aah5178, Oct. 2016.
- [17] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, “Intrinsic optimization using stochastic nanomagnets,” *Scientific Reports*, vol. 7, p. 44370, Mar. 2017.
- [18] Y. Shim, A. Jaiswal, and K. Roy, “Ising computation based combinatorial optimization using spin-Hall effect (SHE) induced stochastic magnetization reversal,” *Journal of Applied Physics*, vol. 121, no. 19, p. 193902, May 2017.
- [19] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, “Taking the Human Out of the Loop: A Review of Bayesian Optimization,” *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.
- [20] B. Behin-Aein, V. Diep, and S. Datta, “A building block for hardware belief networks,” *Scientific Reports*, vol. 6, p. 29893, Jul. 2016.
- [21] R. Faria, K. Y. Camsari, and S. Datta, “Implementing Bayesian networks with embedded stochastic MRAM,” *AIP Advances*, vol. 8, no. 4, p. 045101, Apr. 2018.
- [22] S. Che, J. Li, J. W. Sheaffer, K. Skadron, and J. Lach, “Accelerating Compute-Intensive Applications with GPUs and FPGAs,” in *2008 Symposium on Application Specific Processors*, Jun. 2008, pp. 101–107.
- [23] S. Mittal and J. S. Vetter, “A Survey of CPU-GPU Heterogeneous Computing Techniques,” *ACM Comput. Surv.*, vol. 47, no. 4, pp. 69:1–69:35, Jul. 2015.
- [24] J. Teich, “Hardware/Software Codesign: The Past, the Present, and Predicting the Future,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1411–1430, May 2012.
- [25] D. Castelvecchi, “IBM’s quantum cloud computer goes commercial,” *Nature News*, vol. 543, no. 7644, p. 159, Mar. 2017.

- [26] E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S. Wolf, A. W. Ghosh, J. Lu, S. J. Poon, M. Stan, W. Butler, S. Gupta, C. K. A. Mewes, T. Mewes, and P. Visscher, "Advances and Future Prospects of Spin-Transfer Torque Random Access Memory," *IEEE Transactions on Magnetism*, vol. 46, no. 6, pp. 1873–1878, Jun. 2010.
- [27] A. W. Overhauser, "Polarization of Nuclei in Metals," *Physical Review*, vol. 92, no. 2, pp. 411–415, Oct. 1953.
- [28] M. N. Baibich, J. M. Broto, A. Fert, F. N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Physical Review Letters*, vol. 61, no. 21, pp. 2472–2475, Nov. 1988.
- [29] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, "Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches," *Physical Review B*, vol. 63, no. 5, p. 054416, Jan. 2001.
- [30] J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, and D. C. Ralph, "Current-Driven Magnetization Reversal and Spin-Wave Excitations in Co/Cu/Co Pillars," *Physical Review Letters*, vol. 84, no. 14, pp. 3149–3152, Apr. 2000.
- [31] I. Mihalai Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, "Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer," *Nature Materials*, Jan. 2010.
- [32] L. Liu, C.-F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin-Torque Switching with the Giant Spin Hall Effect of Tantalum," *Science*, vol. 336, no. 6081, pp. 555–558, Apr. 2012.
- [33] Y. L. Chen, J. G. Analytis, J.-H. Chu, Z. K. Liu, S.-K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen, "Experimental Realization of a Three-Dimensional Topological Insulator, Bi₂Te₃," *Science*, vol. 325, no. 5937, pp. 178–181, Jul. 2009.
- [34] B. Kaestner, C. Leicht, V. Kashcheyevs, K. Pierz, U. Siegner, and H. W. Schumacher, "Single-parameter quantized charge pumping in high magnetic fields," *Applied Physics Letters*, vol. 94, no. 1, p. 012106, Jan. 2009.
- [35] J. C. R. Sánchez, L. Vila, G. Desfonds, S. Gambarelli, J. P. Attané, J. M. De Teresa, C. Magén, and A. Fert, "Spin-to-charge conversion using Rashba coupling at the interface between non-magnetic materials," *Nature Communications*, vol. 4, p. 2944, Dec. 2013.
- [36] R. Farshchi and M. Ramsteiner, "Spin injection from Heusler alloys into semiconductors: A materials perspective," *Journal of Applied Physics*, vol. 113, no. 19, p. 191101, May 2013.
- [37] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotechnology*, vol. 5, no. 4, pp. 266–270, Apr. 2010.
- [38] B. Behin-Aein, J.-P. Wang, and R. Wiesendanger, "Computing with spins and magnets," *MRS Bulletin*, vol. 39, no. 08, pp. 696–702, 2014.

- [39] S. Ganguly, K. Y. Camsari, and S. Datta, "Evaluating Spintronic Devices Using the Modular Approach," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 2, pp. 51–60, Dec. 2016.
- [40] B. E. Kane, "A silicon-based nuclear spin quantum computer," *Nature*, vol. 393, no. 6681, pp. 133–137, May 1998.
- [41] A. J. Skinner, M. E. Davenport, and B. E. Kane, "Hydrogenic Spin Quantum Computing in Silicon: A Digital Approach," *Physical Review Letters*, vol. 90, no. 8, p. 087901, Feb. 2003.
- [42] D. P. DiVincenzo, "The Physical Implementation of Quantum Computation," *arXiv:quant-ph/0002077*, Feb. 2000.
- [43] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. L. Morton, D. N. Jamieson, A. S. Dzurak, and A. Morello, "A single-atom electron spin qubit in silicon," *Nature*, vol. 489, no. 7417, pp. 541–545, Sep. 2012.
- [44] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. L. Morton, F. A. Zwanenburg, D. N. Jamieson, A. S. Dzurak, and A. Morello, "High-fidelity readout and control of a nuclear spin qubit in silicon," *Nature*, vol. 496, no. 7445, pp. 334–338, Apr. 2013.
- [45] F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, "Silicon quantum electronics," *Reviews of Modern Physics*, vol. 85, no. 3, pp. 961–1019, Jul. 2013.
- [46] H. Büch, S. Mahapatra, R. Rahman, A. Morello, and M. Y. Simmons, "Spin readout and addressability of phosphorus-donor clusters in silicon," *Nature Communications*, vol. 4, Jun. 2013.
- [47] A. Morello, "Silicon quantum dots: Fine-tuning to maturity," *Nanotechnology*, vol. 26, no. 50, p. 502501, 2015.
- [48] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, "A two-qubit logic gate in silicon," *Nature*, vol. 526, no. 7573, pp. 410–414, Oct. 2015.
- [49] J. P. Dehollain, S. Simmons, J. T. Muhonen, R. Kalra, A. Laucht, F. Hudson, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, "Bell's inequality violation with spins in silicon," *Nature Nanotechnology*, vol. 11, no. 3, pp. 242–246, Mar. 2016.
- [50] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p -Bits for Invertible Logic," *Physical Review X*, vol. 7, no. 3, Jul. 2017.
- [51] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6-7, pp. 467–488, Jun. 1982.
- [52] J. S. Bell, *On the Einstein Podolsky Rosen Paradox*, 1964.

- [53] D. Deutsch, “Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer,” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 400, no. 1818, pp. 97–117, Jul. 1985.
- [54] D. Deutsch, “Quantum Computational Networks,” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 425, no. 1868, pp. 73–90, Sep. 1989.
- [55] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, “Elementary gates for quantum computation,” *Physical Review A*, vol. 52, no. 5, pp. 3457–3467, Nov. 1995.
- [56] P. W. Shor, “Algorithms for quantum computation: Discrete logarithms and factoring,” in *35th Annual Symposium on Foundations of Computer Science, 1994 Proceedings*, Nov. 1994, pp. 124–134.
- [57] D. R. Simon, “On the power of quantum computation,” in *In Proc. 35th IEEE Symp. on Foundations of Computer Science*, 1994, pp. 116–123.
- [58] L. K. Grover, “Quantum Mechanics Helps in Searching for a Needle in a Haystack,” *Physical Review Letters*, vol. 79, no. 2, pp. 325–328, Jul. 1997.
- [59] L. Gomes, “Quantum computing: Both here and not here,” *IEEE Spectrum*, vol. 55, no. 4, pp. 42–47, Apr. 2018.
- [60] J. Preskill, “Quantum computing and the entanglement frontier,” *arXiv:1203.5813 [cond-mat, physics:quant-ph]*, Mar. 2012.
- [61] Committee on Technical Assessment of the Feasibility and Implications of Quantum Computing, Computer Science and Telecommunications Board, Intelligence Community Studies Board, Division on Engineering and Physical Sciences, and National Academies of Sciences, Engineering, and Medicine, *Quantum Computing: Progress and Prospects*, E. Grumbling and M. Horowitz, Eds. Washington, D.C.: National Academies Press, Mar. 2019.
- [62] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019.
- [63] A. Montanaro, “Quantum algorithms: An overview,” *npj Quantum Information*, vol. 2, no. 1, pp. 1–8, Jan. 2016.

- [64] N. Locatelli, V. Cros, and J. Grollier, “Spin-torque building blocks,” *Nature Materials*, vol. 13, no. 1, pp. 11–20, Jan. 2014.
- [65] R. Impagliazzo and A. Wigderson, “ $P = BPP$ if E Requires Exponential Circuits: Derandomizing the XOR Lemma,” in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, ser. STOC ’97. New York, NY, USA: ACM, 1997, pp. 220–229.
- [66] G. L. Miller, “Riemann’s Hypothesis and Tests for Primality,” in *Proceedings of the Seventh Annual ACM Symposium on Theory of Computing*, ser. STOC ’75. New York, NY, USA: ACM, 1975, pp. 234–239.
- [67] M. O. Rabin, “Probabilistic algorithm for testing primality,” *Journal of Number Theory*, vol. 12, no. 1, pp. 128–138, Feb. 1980.
- [68] M. Agrawal, N. Kayal, and N. Saxena, “PRIMES is in P,” *Annals of mathematics*, pp. 781–793, 2004.
- [69] B. E. S. Akgul, L. N. Chakrapani, P. Korkmaz, and K. V. Palem, “Probabilistic CMOS Technology: A Survey and Future Directions,” in *2006 IFIP International Conference on Very Large Scale Integration*, Oct. 2006, pp. 1–6.
- [70] A. Lucas, “Ising formulations of many NP problems,” *Interdisciplinary Physics*, vol. 2, p. 5, 2014.
- [71] S. G. BRUSH, “History of the Lenz-Ising Model,” *Reviews of Modern Physics*, vol. 39, no. 4, pp. 883–893, Oct. 1967.
- [72] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943.
- [73] A. K. Jain, J. Mao, and K. M. Mohiuddin, “Artificial neural networks: A tutorial,” *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996.
- [74] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, “A Learning Algorithm for Boltzmann Machines,” *Cognitive Science*, vol. 9, no. 1, pp. 147–169, Jan. 1985.
- [75] R. M. Neal, “Connectionist learning of belief networks,” *Artificial Intelligence*, vol. 56, no. 1, pp. 71–113, Jul. 1992.
- [76] P. Debashis, R. Faria, K. Y. Camsari, J. Appenzeller, S. Datta, and Z. Chen, “Experimental demonstration of nanomagnet networks as hardware for Ising computing,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2016, pp. 34.3.1–34.3.4.
- [77] R. Faria, K. Y. Camsari, and S. Datta, “Low-Barrier Nanomagnets as p-Bits for Spin Logic,” *IEEE Magnetism Letters*, vol. 8, pp. 1–5, 2017.
- [78] B. Sutton and S. Datta, “Manipulating quantum information with spin torque,” *Scientific Reports*, vol. 5, p. 17912, Dec. 2015.
- [79] B. Sutton, R. Faria, L. A. Ghantasala, K. Y. Camsari, and S. Datta, “Autonomous Probabilistic Coprocessing with Petaflips per Second,” *arXiv:1907.09664 [cond-mat]*, Jul. 2019.

- [80] T. Monz, D. Nigg, E. A. Martinez, M. F. Brandl, P. Schindler, R. Rines, S. X. Wang, I. L. Chuang, and R. Blatt, “Realization of a scalable Shor algorithm,” *Science*, vol. 351, no. 6277, pp. 1068–1070, Mar. 2016.
- [81] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, “A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem,” *Science*, vol. 292, no. 5516, pp. 472–475, Apr. 2001.
- [82] J. King, S. Yarkoni, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, “Benchmarking a quantum annealing processor with the time-to-target metric,” *arXiv:1508.05087 [quant-ph]*, Aug. 2015.
- [83] Y. Wang, C.-Y. Chen, G. Klimeck, M. Y. Simmons, and R. Rahman, “Characterizing Si:P quantum dot qubits with spin resonance techniques,” *Scientific Reports*, vol. 6, Aug. 2016.
- [84] J. T. Muhonen, J. P. Dehollain, A. Laucht, F. E. Hudson, R. Kalra, T. Sekiguchi, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, “Storing quantum information for 30 seconds in a nanoelectronic device,” *Nature Nanotechnology*, vol. advance online publication, Oct. 2014.
- [85] A. M. Tyryshkin, S. Tojo, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, T. Schenkel, M. L. W. Thewalt, K. M. Itoh, and S. A. Lyon, “Electron spin coherence exceeding seconds in high-purity silicon,” *Nature Materials*, vol. 11, no. 2, pp. 143–147, Feb. 2012.
- [86] M. Steger, K. Saeedi, M. L. W. Thewalt, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, and H.-J. Pohl, “Quantum Information Storage for over 180 s Using Donor Spins in a ^{28}Si “Semiconductor Vacuum”,” *Science*, vol. 336, no. 6086, pp. 1280–1283, Aug. 2012.
- [87] B. Huang, D. J. Monsma, and I. Appelbaum, “Coherent Spin Transport through a 350 Micron Thick Silicon Wafer,” *Physical Review Letters*, vol. 99, no. 17, p. 177209, Oct. 2007.
- [88] A. T. Costa, S. Bose, and Y. Omar, “Entanglement of Two Impurities through Electron Scattering,” *Physical Review Letters*, vol. 96, no. 23, p. 230501, Jun. 2006.
- [89] F. Ciccarello, G. M. Palma, M. Zarcone, Y. Omar, and V. R. Vieira, “Electron Fabry–Perot interferometer with two entangled magnetic impurities,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 28, p. 7993, Jul. 2007.
- [90] K. Yuasa and H. Nakazato, “Resonant scattering can enhance the degree of entanglement,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 2, p. 297, Jan. 2007.
- [91] F. Ciccarello, M. Paternostro, M. S. Kim, and G. M. Palma, “Extraction of Singlet States from Noninteracting High-Dimensional Spins,” *Physical Review Letters*, vol. 100, no. 15, p. 150501, Apr. 2008.
- [92] F. Ciccarello, M. Paternostro, G. M. Palma, and M. Zarcone, “Reducing quantum control for spin–spin entanglement distribution,” *New Journal of Physics*, vol. 11, no. 11, p. 113053, Nov. 2009.

- [93] K. Yuasa, “Extraction of an entanglement by repetition of the resonant transmission of an ancilla qubit,” *Journal of Physics A: Mathematical and Theoretical*, vol. 43, no. 9, p. 095304, Feb. 2010.
- [94] A. De Pasquale, K. Yuasa, and H. Nakazato, “State tomography of a qubit through scattering of a probe qubit,” *Physical Review A*, vol. 80, no. 5, p. 052111, Nov. 2009.
- [95] K. Yuasa, D. Burgarth, V. Giovannetti, and H. Nakazato, “Efficient generation of a maximally entangled state by repeated on- and off-resonant scattering of ancilla qubits,” *New Journal of Physics*, vol. 11, no. 12, p. 123027, Dec. 2009.
- [96] G. Cordourier-Maruri, F. Ciccarello, Y. Omar, M. Zarcone, R. de Coss, and S. Bose, “Implementing quantum gates through scattering between a static and a flying qubit,” *Physical Review A*, vol. 82, no. 5, p. 052313, Nov. 2010.
- [97] F. Ciccarello, D. E. Browne, L. C. Kwek, H. Schomerus, M. Zarcone, and S. Bose, “Quasideterministic realization of a universal quantum gate in a single scattering process,” *Physical Review A*, vol. 85, no. 5, p. 050305, May 2012.
- [98] A. D. Pasquale, F. Ciccarello, K. Yuasa, and V. Giovannetti, “Selective writing and read-out of a register of static qubits,” *New Journal of Physics*, vol. 15, no. 4, p. 043012, Apr. 2013.
- [99] J. Tejada, E. M. Chudnovsky, E. Del Barco, J. M. Hernandez, and T. P. Spiller, “Magnetic qubits as hardware for quantum computers,” *Nanotechnology*, vol. 12, no. 2, p. 181, 2001.
- [100] A. A. Khajetoorians, B. Baxevanis, C. Hübner, T. Schlenk, S. Krause, T. O. Wehling, S. Lounis, A. Lichtenstein, D. Pfannkuche, J. Wiebe, and R. Wiesendanger, “Current-Driven Spin Dynamics of Artificially Constructed Quantum Magnets,” *Science*, vol. 339, no. 6115, pp. 55–59, Apr. 2013.
- [101] C. J. Trowbridge, B. M. Norman, Y. K. Kato, D. D. Awschalom, and V. Sih, “Dynamic nuclear polarization from current-induced electron spin polarization,” *Physical Review B*, vol. 90, no. 8, p. 085122, Aug. 2014.
- [102] I. Appelbaum, B. Huang, and D. J. Monsma, “Electronic measurement and control of spin transport in silicon,” *Nature*, vol. 447, no. 7142, pp. 295–298, May 2007.
- [103] H. C. Koo, J. H. Kwon, J. Eom, J. Chang, S. H. Han, and M. Johnson, “Control of Spin Precession in a Spin-Injected Field Effect Transistor,” *Science*, vol. 325, no. 5947, pp. 1515–1518, Sep. 2009.
- [104] Y. Gao, Y. Kubo, C.-C. Lin, Z. Chen, and J. Appenzeller, “Optimized spin relaxation length in few layer graphene at room temperature,” in *Electron Devices Meeting (IEDM), 2012 IEEE International*, Dec. 2012, pp. 4.4.1–4.4.4.
- [105] A. Brataas, G. E. W. Bauer, and P. J. Kelly, “Non-collinear magnetoelectronics,” *Physics Reports*, vol. 427, no. 4, pp. 157–255, Apr. 2006.
- [106] R. Fiederling, M. Keim, G. Reuscher, W. Ossau, G. Schmidt, A. Waag, and L. W. Molenkamp, “Injection and detection of a spin-polarized current in a light-emitting diode,” *Nature*, vol. 402, no. 6763, pp. 787–790, Dec. 1999.

- [107] G. Schmidt, “Concepts for spin injection into semiconductors—a review,” *Journal of Physics D: Applied Physics*, vol. 38, no. 7, p. R107, Apr. 2005.
- [108] T. Seki, Y. Hasegawa, S. Mitani, S. Takahashi, H. Imamura, S. Maekawa, J. Nitta, and K. Takanashi, “Giant spin Hall effect in perpendicularly spin-polarized FePt/Au devices,” *Nature Materials*, vol. 7, no. 2, pp. 125–129, Feb. 2008.
- [109] R. Jansen, “Silicon spintronics,” *Nature Materials*, vol. 11, no. 5, pp. 400–408, May 2012.
- [110] X. Jiang, R. Wang, R. M. Shelby, R. M. Macfarlane, S. R. Bank, J. S. Harris, and S. S. P. Parkin, “Highly Spin-Polarized Room-Temperature Tunnel Injector for Semiconductor Spintronics using MgO(100),” *Physical Review Letters*, vol. 94, no. 5, p. 056601, Feb. 2005.
- [111] S. P. Dash, S. Sharma, R. S. Patel, M. P. de Jong, and R. Jansen, “Electrical creation of spin polarization in silicon at room temperature,” *Nature*, vol. 462, no. 7272, pp. 491–494, Nov. 2009.
- [112] O. M. J. van ’t Erve, A. L. Friedman, E. Cobas, C. H. Li, J. T. Robinson, and B. T. Jonker, “Low-resistance spin injection into silicon using graphene tunnel barriers,” *Nature Nanotechnology*, vol. 7, no. 11, pp. 737–742, Nov. 2012.
- [113] X. Wu, D. R. Ward, J. R. Prance, D. Kim, J. K. Gamble, R. T. Mohr, Z. Shi, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, and M. A. Eriksson, “Two-axis control of a singlet–triplet qubit with an integrated micromagnet,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 33, pp. 11 938–11 942, Aug. 2014.
- [114] B. T. Jonker, G. Kioseoglou, A. T. Hanbicki, C. H. Li, and P. E. Thompson, “Electrical spin-injection into silicon from a ferromagnetic metal/tunnel barrier contact,” *Nature Physics*, vol. 3, no. 8, pp. 542–546, Aug. 2007.
- [115] H. Nakazato, T. Takazawa, and K. Yuasa, “Purification through Zeno-Like Measurements,” *Physical Review Letters*, vol. 90, no. 6, p. 060401, Feb. 2003.
- [116] H. Nakazato, M. Unoki, and K. Yuasa, “Preparation and entanglement purification of qubits through Zeno-like measurements,” *Physical Review A*, vol. 70, no. 1, p. 012303, Jul. 2004.
- [117] B. Militello, K. Yuasa, H. Nakazato, and A. Messina, “Influence of dissipation on the extraction of quantum states via repeated measurements,” *Physical Review A*, vol. 76, no. 4, p. 042110, Oct. 2007.
- [118] P. O. Boykin, T. Mor, V. Roychowdhury, and F. Vatan, “Algorithms on ensemble quantum computers,” *Natural Computing*, vol. 9, no. 2, pp. 329–345, May 2009.
- [119] L. Trifunovic, O. Dial, M. Trif, J. R. Wootton, R. Abebe, A. Yacoby, and D. Loss, “Long-Distance Spin-Spin Coupling via Floating Gates,” *Physical Review X*, vol. 2, no. 1, p. 011006, Jan. 2012.
- [120] L. Trifunovic, F. L. Pedrocchi, and D. Loss, “Long-Distance Entanglement of Spin Qubits via Ferromagnet,” *Physical Review X*, vol. 3, no. 4, p. 041023, Dec. 2013.

- [121] W. K. Wootters, "Entanglement of Formation of an Arbitrary State of Two Qubits," *Physical Review Letters*, vol. 80, no. 10, pp. 2245–2248, Mar. 1998.
- [122] Datta, Supriyo, *Electronic Transport in Mesoscopic Systems*. Cambridge university press, 1997.
- [123] D. Loss and D. P. DiVincenzo, "Quantum computation with quantum dots," *Physical Review A*, vol. 57, no. 1, pp. 120–126, Jan. 1998.
- [124] S. Nadj-Perge, S. M. Frolov, E. P. a. M. Bakkers, and L. P. Kouwenhoven, "Spin-orbit qubit in a semiconductor nanowire," *Nature*, vol. 468, no. 7327, pp. 1084–1087, Dec. 2010.
- [125] K. Saeedi, S. Simmons, J. Z. Salvail, P. Dluhy, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, J. J. L. Morton, and M. L. W. Thewalt, "Room-Temperature Quantum Bit Storage Exceeding 39 Minutes Using Ionized Donors in Silicon-28," *Science*, vol. 342, no. 6160, pp. 830–833, Nov. 2013.
- [126] F. Barahona, "On the computational complexity of Ising spin glass models," *Journal of Physics A Mathematical General*, vol. 15, pp. 3241–3253, Oct. 1982.
- [127] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, Nov. 1994.
- [128] Q. Ouyang, P. D. Kaplan, S. Liu, and A. Libchaber, "DNA Solution of the Maximal Clique Problem," *Science*, vol. 278, no. 5337, pp. 446–449, Oct. 1997.
- [129] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, "Finding low-energy conformations of lattice protein models by quantum annealing," *Scientific Reports*, vol. 2, Aug. 2012.
- [130] L. O. Chua and L. Yang, "Cellular neural networks: Applications," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1273–1290, Oct. 1988.
- [131] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 147–156, Mar. 1993.
- [132] M. Ercsey-Ravasz, T. Roska, and Z. Néda, "Cellular Neural Networks for NP-hard Optimization," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 2:1–2:7, Jan. 2009.
- [133] K. Kim, M.-S. Chang, S. Korenblit, R. Islam, E. E. Edwards, J. K. Freericks, G.-D. Lin, L.-M. Duan, and C. Monroe, "Quantum simulation of frustrated Ising spins with trapped ions," *Nature*, vol. 465, no. 7298, pp. 590–593, Jun. 2010.
- [134] I. Mahboob, H. Okamoto, and H. Yamaguchi, "An electromechanical Ising Hamiltonian," *Science Advances*, vol. 2, no. 6, p. e1600236, Jun. 2016.
- [135] N. T. Shaked, S. Messika, S. Dolev, and J. Rosen, "Optical solution for bounded NP-complete problems," *Applied Optics*, vol. 46, no. 5, pp. 711–724, Feb. 2007.
- [136] M. Oltean, "Solving the Hamiltonian path problem with a light-based computer," *Natural Computing*, vol. 7, no. 1, pp. 57–70, Mar. 2008.

- [137] S. Utsunomiya, K. Takata, and Y. Yamamoto, “Mapping of Ising models onto injection-locked laser systems,” *Optics Express*, vol. 19, no. 19, pp. 18 091–18 108, Sep. 2011.
- [138] K. Wu, J. García de Abajo, C. Soci, P. Ping Shum, and N. I. Zheludev, “An optical fiber network oracle for NP-complete problems,” *Light: Science & Applications*, vol. 3, no. 2, p. e147, Feb. 2014.
- [139] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, “Coherent Ising machine based on degenerate optical parametric oscillators,” *Physical Review A*, vol. 88, no. 6, p. 063853, Dec. 2013.
- [140] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, “Network of time-multiplexed optical parametric oscillators as a coherent Ising machine,” *Nature Photonics*, vol. 8, no. 12, pp. 937–942, Dec. 2014.
- [141] S. Bhanja, D. K. Karunaratne, R. Panchumorthy, S. Rajaram, and S. Sarkar, “Non-Boolean computing with nanomagnets for computer vision applications,” *Nature Nanotechnology*, vol. 11, no. 2, pp. 177–183, Oct. 2015.
- [142] U. B. Arnalds, J. Chico, H. Stopfel, V. Kapaklis, O. Bärenbold, M. A. Verschuuren, U. Wolff, V. Neu, A. Bergman, and B. Hjörvarsson, “A new look on the two-dimensional Ising model: Thermal artificial spins,” *New Journal of Physics*, vol. 18, no. 2, p. 023008, Jan. 2016.
- [143] G. De las Cuevas and T. S. Cubitt, “Simple universal models capture all classical spin physics,” *Science*, vol. 351, no. 6278, pp. 1180–1183, Mar. 2016.
- [144] S. Khasanvis, M. Li, M. Rahman, M. Salehi-Fashami, A. K. Biswas, J. Atulasimha, S. Bandyopadhyay, and C. A. Moritz, “Physically equivalent magneto-electric nanoarchitecture for probabilistic reasoning,” in *Proceedings of the 2015 IEEE/ACM International Symposium on Nanoscale Architectures*, Jul. 2015, pp. 25–26.
- [145] M. Bapna, S. K. Piotrowski, S. D. Oberdick, M. Li, C.-L. Chien, and S. A. Majetich, “Magnetostatic effects on switching in small magnetic tunnel junctions,” *Applied Physics Letters*, vol. 108, no. 2, p. 022406, Jan. 2016.
- [146] N. Locatelli, A. Mizrahi, A. Accioly, R. Matsumoto, A. Fukushima, H. Kubota, S. Yuasa, V. Cros, L. G. Pereira, D. Querlioz, J.-V. Kim, and J. Grollier, “Noise-Enhanced Synchronization of Stochastic Magnetic Oscillators,” *Physical Review Applied*, vol. 2, no. 3, Sep. 2014.
- [147] R. P. Cowburn, D. K. Koltsov, A. O. Adeyeye, M. E. Welland, and D. M. Tricker, “Single-domain circular nanomagnets,” *Physical Review Letters*, vol. 83, no. 5, p. 1042, 1999.
- [148] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Rippard, S. E. Russek, and R. Heindl, “Switching Distributions for Perpendicular Spin-Torque Devices Within the Macrospin Approximation,” *IEEE Transactions on Magnetism*, vol. 48, no. 12, pp. 4684–4700, Dec. 2012.
- [149] S. Aaronson, “Guest Column: NP-complete Problems and Physical Reality,” *SIGACT News*, vol. 36, no. 1, pp. 30–52, Mar. 2005.

- [150] T. Kadowaki and H. Nishimori, “Quantum annealing in the transverse Ising model,” *Physical Review E*, vol. 58, no. 5, p. 5355, 1998.
- [151] S. Cheemalavagu, P. Korkmaz, K. V. Palem, B. E. S. Akgul, and L. N. Chakrapani, “A probabilistic CMOS switch and its realization by exploiting noise,” in *The Proceedings of the IFIP International*, 2005.
- [152] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
- [153] K. Y. Camsari, S. Ganguly, and S. Datta, “Modular Approach to Spintronics,” *Scientific Reports*, vol. 5, p. 10571, Jun. 2015.
- [154] J. T. Heron, J. L. Bosse, Q. He, Y. Gao, M. Trassin, L. Ye, J. D. Clarkson, C. Wang, J. Liu, S. Salahuddin, D. C. Ralph, D. G. Schlom, J. Íñiguez, B. D. Huey, and R. Ramesh, “Deterministic switching of ferromagnetism at room temperature using an electric field,” *Nature*, vol. 516, no. 7531, pp. 370–373, Dec. 2014.
- [155] R. M. Karp, “Reducibility among Combinatorial Problems,” in *Complexity of Computer Computations*, ser. The IBM Research Symposia Series, R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, Eds. Springer US, 1972, pp. 85–103.
- [156] S. A. Cook, “The Complexity of Theorem-proving Procedures,” in *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, ser. STOC ’71. New York, NY, USA: ACM, 1971, pp. 151–158.
- [157] Z. Bian, F. Chudak, W. G. Macready, and G. Rose, “The Ising model: Teaching an old problem new tricks,” *D-Wave Systems*, vol. 2, 2010.
- [158] J. D. Biamonte, “Nonperturbative k-body to two-body commuting conversion Hamiltonians and embedding problem instances into Ising spins,” *Physical Review A*, vol. 77, no. 5, p. 052331, May 2008.
- [159] J. J. Schneider and S. Kirkpatrick, *Stochastic Optimization (Scientific Computation)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [160] G. Reinelt, “TSPLIB— A Traveling Salesman Problem Library,” *ORSA Journal on Computing*, vol. 3, no. 4, p. 376.
- [161] S. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, and M. Samant, “Magnetically engineered spintronic sensors and memory,” *Proceedings of the IEEE*, vol. 91, no. 5, pp. 661–680, May 2003.
- [162] A. Sengupta, M. Parsa, B. Han, and K. Roy, “Probabilistic Deep Spiking Neural Systems Enabled by Magnetic Tunnel Junction,” *IEEE Transactions on Electron Devices*, vol. 63, no. 7, pp. 2963–2970, Jul. 2016.
- [163] S. Datta, S. Salahuddin, and B. Behin-Aein, “Non-volatile spin switch for Boolean and non-Boolean logic,” *Applied Physics Letters*, vol. 101, no. 25, p. 252411, Dec. 2012.

- [164] A. Tangel and K. Choi, ““The CMOS Inverter” as a Comparator in ADC Designs,” *Analog Integrated Circuits and Signal Processing*, vol. 39, no. 2, pp. 147–155, May 2004.
- [165] V. Q. Diep, B. Sutton, B. Behin-Aein, and S. Datta, “Spin switches for compact implementation of neuron and synapse,” *Applied Physics Letters*, vol. 104, no. 22, p. 222405, Jun. 2014.
- [166] J. J. Yang, D. B. Strukov, and D. R. Stewart, “Memristive devices for computing,” *Nature Nanotechnology*, vol. 8, no. 1, pp. 13–24, Jan. 2013.
- [167] A. Sengupta, S. H. Choday, Y. Kim, and K. Roy, “Spin orbit torque based electronic neuron,” *Applied Physics Letters*, vol. 106, no. 14, p. 143701, Apr. 2015.
- [168] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, and J. Whittaker, “Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor,” *IEEE Transactions on Applied Superconductivity*, vol. 24, no. 4, pp. 1–10, Aug. 2014.
- [169] G. Lemieux and D. Lewis, *Design of Interconnection Networks for Programmable Logic*. Boston, MA: Springer US, 2004.
- [170] A. Sengupta, Y. Shim, and K. Roy, “Proposal for an All-Spin Artificial Neural Network: Emulating Neural and Synaptic Functionalities Through Domain Wall Motion in Ferromagnets,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. PP, no. 99, pp. 1–9, 2016.
- [171] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York, NY, USA: John Wiley & Sons, Inc., 1989.
- [172] S. S. Haykin and S. S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. New York: Prentice Hall, 2009, oCLC: ocn237325326.
- [173] B. H. Calhoun, Y. Cao, X. Li, K. Mai, L. T. Pileggi, R. A. Rutenbar, and K. L. Shepard, “Digital Circuit Design Challenges and Opportunities in the Era of Nanoscale CMOS,” *Proceedings of the IEEE*, vol. 96, no. 2, pp. 343–365, Feb. 2008.
- [174] T. P. Hayes and A. Sinclair, “A general lower bound for mixing of single-site dynamics on graphs,” *The Annals of Applied Probability*, vol. 17, no. 3, pp. 931–952, Jun. 2007.
- [175] Peng Gu, Boxun Li, Tianqi Tang, S. Yu, Yu Cao, Y. Wang, and H. Yang, “Technological exploration of RRAM crossbar array for matrix-vector multiplication,” in *The 20th Asia and South Pacific Design Automation Conference*, Jan. 2015, pp. 106–111.
- [176] M. Di Ventra and F. L. Traversa, “Perspective: Memcomputing: Leveraging memory and physics to compute efficiently,” *Journal of Applied Physics*, vol. 123, no. 18, p. 180901, May 2018.

- [177] O. Hassan, R. Faria, K. Y. Camsari, J. Z. Sun, and S. Datta, "Low-Barrier Magnet Design for Efficient Hardware Binary Stochastic Neurons," *IEEE Magnetics Letters*, vol. 10, pp. 1–5, 2019.
- [178] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions," *Nature*, vol. 573, no. 7774, pp. 390–393, Sep. 2019.
- [179] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.
- [180] M. Davies, N. Srinivasa, T. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, Jan. 2018.
- [181] C. M. Liyanagedera, A. Sengupta, A. Jaiswal, and K. Roy, "Stochastic Spiking Neural Networks Enabled by Magnetic Tunnel Junctions: From Nontelegraphic to Telegraphic Switching Regimes," *Physical Review Applied*, vol. 8, no. 6, p. 064017, Dec. 2017.
- [182] Y. Lv and J. Wang, "A single magnetic-tunnel-junction stochastic computing unit," in *2017 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 36.2.1–36.2.4.
- [183] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, "Neural-like computing with populations of superparamagnetic basis functions," *Nature Communications*, vol. 9, no. 1, p. 1533, Apr. 2018.
- [184] K. Y. Camsari, B. M. Sutton, and S. Datta, "P-bits for probabilistic spin logic," *Applied Physics Reviews*, vol. 6, no. 1, p. 011305, Mar. 2019.
- [185] R. Zand, K. Y. Camsari, S. Datta, and R. F. Demara, "Composable Probabilistic Inference Networks Using MRAM-based Stochastic Neurons," *J. Emerg. Technol. Comput. Syst.*, vol. 15, no. 2, pp. 17:1–17:22, Mar. 2019.
- [186] K. Y. Camsari, S. Salahuddin, and S. Datta, "Implementing p-bits with Embedded MTJ," *IEEE Electron Device Letters*, vol. PP, no. 99, pp. 1–1, 2017.
- [187] T. Okuyama, M. Hayashi, and M. Yamaoka, "An Ising Computer Based on Simulated Quantum Annealing by Path Integral Monte Carlo Method," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, Nov. 2017, pp. 1–6.
- [188] R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasa-hara, S. Utsunomiya, S. Kako, K.-i. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, D. Englund, E. Rieffel, H. Takesue, and Y. Yamamoto, "Experimental investigation of performance differences between coherent Ising machines and a quantum annealer," *Science Advances*, vol. 5, no. 5, p. eaau0823, May 2019.

- [189] T. Wang and J. Roychowdhury, “Oscillator-based Ising Machine,” *arXiv:1709.08102 [physics]*, Sep. 2017.
- [190] T. Wang and J. Roychowdhury, “OIM: Oscillator-Based Ising Machines for Solving Combinatorial Optimisation Problems,” in *Unconventional Computation and Natural Computation*, ser. Lecture Notes in Computer Science, I. McQuillan and S. Seki, Eds. Springer International Publishing, 2019, pp. 232–256.
- [191] A. Raychowdhury, A. Parihar, G. H. Smith, V. Narayanan, G. Csaba, M. Jerry, W. Porod, and S. Datta, “Computing With Networks of Oscillatory Dynamical Systems,” *Proceedings of the IEEE*, vol. 107, no. 1, pp. 73–89, Jan. 2019.
- [192] M. Baity-Jesi, R. A. Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, D. Iñiguez, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, M. Pivanti, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccion, and D. Yllanes, “Janus II: A new generation application-driven computer for spin-system simulations,” *Computer Physics Communications*, vol. 185, no. 2, pp. 550–559, Feb. 2014.
- [193] C. Yoshimura, M. Hayashi, T. Okuyama, and M. Yamaoka, “FPGA-based Annealing Processor for Ising Model,” in *2016 Fourth International Symposium on Computing and Networking (CANDAR)*, Nov. 2016, pp. 436–442.
- [194] F. Ortega-Zamorano, M. A. Montemurro, S. A. Cannas, J. M. Jerez, and L. Franco, “FPGA Hardware Acceleration of Monte Carlo Simulations for the Ising Model,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 9, pp. 2618–2627, Sep. 2016.
- [195] M. Suzuki, “Relationship between d-Dimensional Quantal Spin Systems and (d+1)-Dimensional Ising Systems Equivalence, Critical Exponents and Systematic Approximants of the Partition Function and Spin Correlations,” *Progress of Theoretical Physics*, vol. 56, no. 5, pp. 1454–1469, Nov. 1976.
- [196] M. Troyer and U.-J. Wiese, “Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations,” *Physical Review Letters*, vol. 94, no. 17, p. 170201, May 2005.
- [197] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, “Theory of Quantum Annealing of an Ising Spin Glass,” *Science*, vol. 295, no. 5564, pp. 2427–2430, Mar. 2002.
- [198] B. Heim, T. F. Rønnow, S. V. Isakov, and M. Troyer, “Quantum versus classical annealing of Ising spin glasses,” *Science*, vol. 348, no. 6231, pp. 215–217, Apr. 2015.
- [199] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, “What is the Computational Value of Finite-Range Tunneling?” *Physical Review X*, vol. 6, no. 3, p. 031015, Aug. 2016.
- [200] H. M. Waidyasooriya, M. Hariyama, M. J. Miyama, and M. Ohzeki, “OpenCL-based design of an FPGA accelerator for quantum annealing simulation,” *The Journal of Supercomputing*, Feb. 2019.

- [201] K. Y. Camsari, S. Chowdhury, and S. Datta, “Scalable Emulation of Sign-Problem-Free Hamiltonians with Room-Temperature p -bits,” *Physical Review Applied*, vol. 12, no. 3, p. 034061, Sep. 2019.
- [202] P. Pfeuty, “The one-dimensional Ising model with a transverse field,” *Annals of Physics*, vol. 57, no. 1, pp. 79–90, Mar. 1970.
- [203] N. D. Rizzo, D. Houssameddine, J. Janesky, R. Whig, F. B. Mancoff, M. L. Schneider, M. DeHerrera, J. J. Sun, K. Nagel, S. Deshpande, H. Chia, S. M. Alam, T. Andre, S. Aggarwal, and J. M. Slaughter, “A Fully Functional 64 Mb DDR3 ST-MRAM Built on 90 nm CMOS Technology,” *IEEE Transactions on Magnetics*, vol. 49, no. 7, pp. 4441–4446, Jul. 2013.
- [204] B. Li, Y. Shan, M. Hu, Y. Wang, Y. Chen, and H. Yang, “Memristor-based approximated computation,” in *International Symposium on Low Power Electronics and Design (ISLPED)*. Beijing, China: IEEE, Sep. 2013, pp. 242–247.
- [205] I. Kuon and J. Rose, “Measuring the Gap Between FPGAs and ASICs,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 2, pp. 203–215, Feb. 2007.
- [206] Y. Fang, S. Feng, K.-M. Tam, Z. Yun, J. Moreno, J. Ramanujam, and M. Jarrell, “Parallel tempering simulation of the three-dimensional Edwards–Anderson model with compact asynchronous multispin coding on GPU,” *Computer Physics Communications*, vol. 185, no. 10, pp. 2467–2478, Oct. 2014.
- [207] D. Blackman and S. Vigna, “Scrambled Linear Pseudorandom Number Generators,” *arXiv:1805.01407 [cs]*, May 2018.
- [208] F. Ortega-Zamorano, J. M. Jerez, G. Juárez, J. O. Pérez, and L. Franco, “High precision FPGA implementation of neural network activation functions,” in *2014 IEEE Symposium on Intelligent Embedded Systems (IES)*, Dec. 2014, pp. 55–60.
- [209] M. Majzoobi, F. Koushanfar, and S. Devadas, “FPGA PUF using programmable delay lines,” in *2010 IEEE International Workshop on Information Forensics and Security*, Dec. 2010, pp. 1–6.
- [210] D. Sherrington and S. Kirkpatrick, “Solvable Model of a Spin-Glass,” *Physical Review Letters*, vol. 35, no. 26, pp. 1792–1796, Dec. 1975.
- [211] L. E. Reichl, “A Modern Course in Statistical Physics, 2nd Edition,” *American Journal of Physics*, vol. 67, no. 12, pp. 1285–1287, Nov. 1999.
- [212] Q. Liu, A. Ihler, J. Peng, and J. Fisher, “Estimating the partition function by discriminative sampling,” in *Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI 2015*. AUAI Press, 2015, pp. 514–522.
- [213] V. Dumoulin, I. J. Goodfellow, A. C. Courville, and Y. Bengio, “On the Challenges of Physical Implementations of RBMs,” in *AAAI*, vol. 2014, 2014, pp. 1199–1205.
- [214] V. Choi, “Minor-embedding in adiabatic quantum computation: I. The parameter setting problem,” *Quantum Information Processing*, vol. 7, no. 5, pp. 193–209, Oct. 2008.

- [215] V. Choi, “Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design,” *Quantum Information Processing*, vol. 10, no. 3, pp. 343–353, Jun. 2011.
- [216] W. Vinci, T. Albash, G. Paz-Silva, I. Hen, and D. A. Lidar, “Quantum annealing correction with minor embedding,” *Physical Review A*, vol. 92, no. 4, Oct. 2015.
- [217] S. R. Etesami, A. Sukhov, and J. Berakdar, “Kinetics of nanosize ferroelectrics,” *Physical Review B*, vol. 94, no. 17, Nov. 2016.
- [218] S. Salahuddin and S. Datta, “Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices,” *Nano Letters*, vol. 8, no. 2, pp. 405–410, Feb. 2008.
- [219] A. I. Khan, K. Chatterjee, B. Wang, S. Drapcho, L. You, C. Serrao, S. R. Bakaul, R. Ramesh, and S. Salahuddin, “Negative capacitance in a ferroelectric capacitor,” *Nature Materials*, vol. 14, no. 2, pp. 182–186, Dec. 2014.
- [220] A. Aziz, S. Ghosh, S. Datta, and S. K. Gupta, “Physics-Based Circuit-Compatible SPICE Model for Ferroelectric Transistors,” *IEEE Electron Device Letters*, vol. 37, no. 6, pp. 805–808, Jun. 2016.

A. CHAPTER 2 SUPPLEMENTARY INFORMATION

A.1 Derivations of equations (2.4) and (2.5)

Since the discussion of single qubit operations using non-equilibrium spins is all based on equations (2.4) and (2.5) let us now describe how these are obtained. Our starting point is equation (2.3) which relates the qubit density matrix after it has interacted with $n + 1$ spins to the earlier density matrix after it had interacted with n spins.

The density matrix for the qubit can be written as

$$\boldsymbol{\rho}_s(n) = \frac{1}{2} \begin{bmatrix} 1 + s_z(n) & s_t^*(n) \\ s_t(n) & 1 - s_z(n) \end{bmatrix} \quad (\text{A.1})$$

while the itinerant z-directed spins are describe by a density matrix $\boldsymbol{\rho}_f = \begin{bmatrix} 1 & 0 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \end{bmatrix}$. We can write from equation (2)

$$\boldsymbol{\rho} = \frac{1}{2} \mathbf{R} \begin{bmatrix} 1 + s_z(n) & s_t^*(n) & 0 & 0 \\ s_t(n) & 1 - s_z(n) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{R}^\dagger \quad (\text{A.2})$$

To find the reflection matrix $[\mathbf{R}]$ we note that it is related to the reflection matrix $[\mathbf{R}_0]$ for the barrier and the reflection and transmission matrices $[\mathbf{r}]$, $[\mathbf{t}]$ describing the interaction of the itinerant spins with the static qubit (Figure 3(a))

$$\mathbf{R} = \mathbf{r} + e^{i2kd_0} \mathbf{t} [\mathbf{I}_4 - e^{i2kd_0} \mathbf{R}_0 \mathbf{r}]^{-1} \mathbf{R}_0 \mathbf{t} \quad (\text{A.3})$$

Here \mathbf{I}_4 is the 4×4 identity matrix and the matrices $[\mathbf{r}]$ and $[\mathbf{t}]$ can be written as (see Supplementary Section A.2)

$$\mathbf{t} = \left[\mathbf{I}_4 + i\Omega \tilde{\mathbf{S}} \right]^{-1} \quad \text{and} \quad \mathbf{r} = \mathbf{t} - \mathbf{I}_4 \quad (\text{A.4})$$

$\tilde{\mathbf{S}}$ being the 4×4 matrix, $\vec{\sigma}_f \cdot \tilde{\mathbf{S}}$, describing the basic itinerant $\text{spin}(f)$ -qubit(s) interaction in equation (1):

$$\tilde{\mathbf{S}} = \begin{matrix} & fs & f\bar{s} & \bar{f}s & \bar{f}\bar{s} \\ \begin{matrix} fs \\ f\bar{s} \\ \bar{f}s \\ \bar{f}\bar{s} \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad (\text{A.5})$$

Assuming the barrier reflection coefficient to be given by $\mathbf{R}_0 = -[\mathbf{I}_4]$ and using equations (A.4) and (A.5), we obtain the desired reflection matrix $[\mathbf{R}]$ from equation (A.3) (see Supplementary Section A.1 for details)

$$\mathbf{R} = \begin{matrix} & fs & f\bar{s} & \bar{f}s & \bar{f}\bar{s} \\ \begin{matrix} fs \\ f\bar{s} \\ \bar{f}s \\ \bar{f}\bar{s} \end{matrix} & \begin{pmatrix} e^{-i\alpha} & 0 & 0 & 0 \\ 0 & \cos \alpha & -i \sin \alpha & 0 \\ 0 & -i \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & e^{-i\alpha} \end{pmatrix} \end{matrix} \quad (\text{A.6})$$

where α is given by equation (2.6) as stated earlier.

Using $[\mathbf{R}]$ in equation (2.3) we have for $\boldsymbol{\rho}$ (dropping the argument n for clarity)

$$\frac{1}{2} \begin{bmatrix} 1 + s_z & s_t^* \cos \alpha e^{-i\alpha} & iS_t^* \sin \alpha e^{-i\alpha} & 0 \\ s_t \cos \alpha e^{i\alpha} & (1 - s_z) \cos^2 \alpha & i(1 - s_z) \cos \alpha \sin \alpha & 0 \\ -iS_t \sin \alpha e^{i\alpha} & -i(1 - s_z) \cos \alpha \sin \alpha & (1 - s_z) \sin^2 \alpha & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which upon collapsing, as indicated in equation (2.3), gives the static qubit density matrix after interacting with $(n + 1)$ spins

$$\boldsymbol{\rho}_s(n+1) = \frac{1}{2} \begin{bmatrix} (1 + s_z(n)) + (1 - s_z(n)) \sin^2 \alpha & s_t^*(n) \cos \alpha e^{-i\alpha} \\ s_t(n) \cos \alpha e^{i\alpha} & (1 - s_z(n)) \cos^2 \alpha \end{bmatrix}$$

which by definition must equal

$$\boldsymbol{\rho}_s(n+1) = \frac{1}{2} \begin{bmatrix} 1 + s_z(n+1) & s_t^*(n+1) \\ s_t(n+1) & 1 - s_z(n+1) \end{bmatrix}$$

so that we can write

$$s_z(n+1) = s_z(n) + (1 - s_z(n)) \sin^2 \alpha \quad (\text{A.7})$$

$$s_t(n+1) = s_t(n) \cos \alpha e^{i\alpha} \quad (\text{A.8})$$

It is straightforward to check that the solutions stated earlier in (2.4) and (2.5) satisfy the recurrence relations in equations (A.7) and (A.8).

A.2 Derivation of Equation (A.4)

$\tilde{\mathbf{S}}$ can be written in the standard basis as

$$\tilde{\mathbf{S}} = \vec{\sigma}_f \cdot \vec{\mathbf{S}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.9})$$

Under basis transformation \mathbf{U} to the $|fs\rangle$, $(|f\bar{s}\rangle + |\bar{f}s\rangle)/\sqrt{2}$, $(|f\bar{s}\rangle - |\bar{f}s\rangle)/\sqrt{2}$, $|\bar{f}\bar{s}\rangle$ basis, this matrix becomes diagonal.

$$\mathbf{U}^\dagger \tilde{\mathbf{S}} \mathbf{U} = \begin{matrix} & fs & \frac{f\bar{s} + \bar{f}s}{\sqrt{2}} & \frac{f\bar{s} - \bar{f}s}{\sqrt{2}} & \bar{f}\bar{s} \\ \begin{matrix} fs \\ \frac{f\bar{s} + \bar{f}s}{\sqrt{2}} \\ \frac{f\bar{s} - \bar{f}s}{\sqrt{2}} \\ \bar{f}\bar{s} \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad (\text{A.10})$$

As a result the scattering problem can be treated as four independent one-dimensional channels with transmission matrix $\mathbf{U}^\dagger \mathbf{t} \mathbf{U}$ given by

$$\mathbf{U}^\dagger \mathbf{t} \mathbf{U} = \begin{bmatrix} (1 + i\Omega)^{-1} & 0 & 0 & 0 \\ 0 & (1 + i\Omega)^{-1} & 0 & 0 \\ 0 & 0 & (1 - 3i\Omega)^{-1} & 0 \\ 0 & 0 & 0 & (1 + i\Omega)^{-1} \end{bmatrix}$$

$$\mathbf{U}^\dagger \mathbf{t} \mathbf{U} = [\mathbf{I}_4 + i\Omega(\mathbf{U}^\dagger \tilde{\mathbf{S}} \mathbf{U})]^{-1}$$

Which yields equation (A.4)

$$\mathbf{t} = [\mathbf{I}_4 + i\Omega\tilde{\mathbf{S}}]^{-1} \quad (\text{A.11})$$

We can write from equation (A.3) with $\mathbf{R}_0 = -\mathbf{I}_4$

$$\mathbf{R} = \mathbf{r} - e^{i2kd_0}\mathbf{t} [\mathbf{I}_4 + e^{i2kd_0}\mathbf{r}]^{-1}\mathbf{t} \quad (\text{A.12})$$

where

$$\mathbf{t} = [\mathbf{I}_4 + i\Omega\tilde{\mathbf{S}}]^{-1} \text{ and } \mathbf{r} = \mathbf{t} - \mathbf{I}_4 \quad (\text{A.13})$$

We can diagonalize \mathbf{r} and \mathbf{t} as well as \mathbf{R} by diagonalizing $\tilde{\mathbf{S}}$:

$$\tilde{\mathbf{S}} = \vec{\sigma} \cdot \vec{\mathbf{S}} = \begin{matrix} & fs & \frac{f\bar{s}+\bar{f}s}{\sqrt{2}} & \frac{f\bar{s}-\bar{f}s}{\sqrt{2}} & \bar{f}\bar{s} \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.14})$$

so that

$$\mathbf{R} = \vec{\sigma} \cdot \vec{\mathbf{S}} = \begin{matrix} & fs & \frac{f\bar{s}+\bar{f}s}{\sqrt{2}} & \frac{f\bar{s}-\bar{f}s}{\sqrt{2}} & \bar{f}\bar{s} \end{matrix} \begin{pmatrix} R_T & 0 & 0 & 0 \\ 0 & R_T & 0 & 0 \\ 0 & 0 & R_S & 0 \\ 0 & 0 & 0 & R_T \end{pmatrix} \quad (\text{A.15})$$

where

$$R_m = r_m - \frac{t_m^2 e^{2ikd_0}}{1 + r_m e^{2ikd_0}}, \quad m = T, S \quad (\text{A.16})$$

$$t_T = 1 + r_T = \frac{1}{1 + i\Omega} \quad (\text{A.17})$$

$$t_S = 1 + r_S = \frac{1}{1 - 3i\Omega} \quad (\text{A.18})$$

Using equations (A.16), (A.18), and (A.17) we can write

$$R_m = -e^{2i\theta_m} \quad (\text{A.19})$$

where

$$\tan \theta_T = \frac{\sin kd_0}{\cos kd_0 + 2\Omega \sin kd_0} \quad (\text{A.20})$$

$$\tan \theta_S = \frac{\sin kd_0}{\cos kd_0 - 6\Omega \sin kd_0} \quad (\text{A.21})$$

Consider first the triplet coefficient R_T with $t_T = 1 + r_T$ and $r_T = (-i\Omega)/(1 + i\Omega)$ and let $P = e^{ikd_0}$. We can then write, dropping the sub-scripts for convenience

$$\begin{aligned} R &= r - \frac{(1+r)^2 P^2}{1+rP^2} \\ &= \frac{r + (r^2 - (1+r)^2)P^2}{1+rP^2} = \frac{r - (2r+1)P^2}{1+rP^2} \frac{1+i\Omega}{1+i\Omega} \\ &= \frac{-i\Omega - (1-i\Omega)P^2}{1+i\Omega - i\Omega P^2} = -\frac{i\Omega P^* + (1-i\Omega)P}{(1+i\Omega)P^* - i\Omega P} \\ &= -\frac{Xe^{i\theta}}{Xe^{-i\theta}} \end{aligned}$$

where

$$\begin{aligned} Xe^{i\theta} &= i\Omega P^* + (1-i\Omega)P \\ &= i\Omega(\cos kd_0 - i \sin kd_0) + (1-i\Omega)(\cos kd_0 + i \sin kd_0) \\ &= (\cos kd_0 + 2\Omega \sin kd_0) + i \sin kd_0 \end{aligned}$$

which allows us to write

$$\tan \theta_T = \frac{\sin kd_0}{\cos kd_0 + 2\Omega \sin kd_0} \quad (\text{A.22})$$

Similarly, $\tan \theta_S$ can be found by replacing $\Omega \rightarrow -3\Omega$

$$\tan \theta_S = \frac{\sin kd_0}{\cos kd_0 - 6\Omega \sin kd_0} \quad (\text{A.23})$$

corresponding to equations (A.20) and (A.21) respectively. We can then write the reflection matrix as

$$\mathbf{R} = \begin{matrix} & \begin{matrix} fs & \frac{f\bar{s}+\bar{f}s}{\sqrt{2}} & \frac{f\bar{s}-\bar{f}s}{\sqrt{2}} & \bar{f}\bar{s} \end{matrix} \\ \begin{matrix} fs \\ \frac{f\bar{s}+\bar{f}s}{\sqrt{2}} \\ \frac{f\bar{s}-\bar{f}s}{\sqrt{2}} \\ \bar{f}\bar{s} \end{matrix} & \begin{pmatrix} e^{-i\alpha} & 0 & 0 & 0 \\ 0 & e^{-i\alpha} & 0 & 0 \\ 0 & 0 & e^{+i\alpha} & 0 \\ 0 & 0 & 0 & e^{+i\alpha} \end{pmatrix} \end{matrix} \quad (\text{A.24})$$

where $\alpha \equiv \theta_S - \theta_T$ and we have dropped the overall phase factor of $e^{-i(\theta_T + \theta_S)}$. Transforming back to the original basis we have:

$$\mathbf{R} = \begin{matrix} & f s & f \bar{s} & \bar{f} s & \bar{f} \bar{s} \\ \begin{matrix} f s \\ f \bar{s} \\ \bar{f} s \\ \bar{f} \bar{s} \end{matrix} & \begin{pmatrix} e^{-i\alpha} & 0 & 0 & 0 \\ 0 & \cos \alpha & -i \sin \alpha & 0 \\ 0 & -i \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & e^{-i\alpha} \end{pmatrix} \end{matrix} \quad (\text{A.25})$$

Finally, making use of equations (A.20) and (A.21) we can write

$$\begin{aligned}\tan \alpha &= \tan(\theta_S - \theta_T) = \frac{\tan \theta_S - \tan \theta_T}{1 + \tan \theta_S \tan \theta_T} \\ &= \frac{\frac{\sin kd_0}{\cos kd_0 - 6\Omega \sin kd_0} - \frac{\sin kd_0}{\cos kd_0 + 2\Omega \sin kd_0}}{1 + \frac{\sin^2 kd_0}{(\cos kd_0 - 6\Omega \sin kd_0)(\cos kd_0 + 2\Omega \sin kd_0)}} \\ &= \frac{(\cos kd_0 + 2\Omega \sin kd_0) \sin kd_0 - (\cos kd_0 - 6\Omega \sin kd_0) \sin kd_0}{\cos^2 kd_0 - 4\Omega \cos kd_0 \sin kd_0 - 12\Omega \sin^2 kd_0 + \sin^2 kd_0} \\ &= \frac{8\Omega \sin^2 kd_0}{1 - 4\Omega \cos kd_0 \sin kd_0 - 12\Omega \sin^2 kd_0}\end{aligned}$$

which yields the result of (2.6)

$$\tan \alpha = \frac{8\Omega \sin^2 kd_0}{1 - 2\Omega \sin(2kd_0) - 6\Omega(1 - \cos 2kd_0)} \quad (\text{A.26})$$

A.3 Matrices $\tilde{\mathbf{S}}_1$ and $\tilde{\mathbf{S}}_2$ in equations (2.12) and (2.13)

$$\tilde{\mathbf{S}}_1 = \tilde{\sigma} \cdot \tilde{\mathbf{S}}_1 \rightarrow$$

	$f12$	$f\bar{1}\bar{2}$	$f\bar{1}\bar{2}$	$f\bar{1}\bar{2}$	$\bar{f}12$	$\bar{f}\bar{1}\bar{2}$	$\bar{f}\bar{1}\bar{2}$	$\bar{f}\bar{1}\bar{2}$
$f12$	1	0	0	0	0	0	0	0
$f\bar{1}\bar{2}$	0	1	0	0	0	0	0	0
$f\bar{1}\bar{2}$	0	0	-1	0	2	0	0	0
$f\bar{1}\bar{2}$	0	0	0	-1	0	2	0	0
$\bar{f}12$	0	0	2	0	-1	0	0	0
$\bar{f}\bar{1}\bar{2}$	0	0	0	2	0	-1	0	0
$\bar{f}\bar{1}\bar{2}$	0	0	0	0	0	0	1	0
$\bar{f}\bar{1}\bar{2}$	0	0	0	0	0	0	0	1

A.4 Derivation of (2.21)

In the 3×3 sub-space we find

$$\begin{aligned}
 \tilde{S}_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 2 \\ 0 & 2 & 1 \end{pmatrix} = \vec{\sigma}_f \cdot \vec{S}_1 \\
 \tilde{S}_2 &= \begin{pmatrix} -1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{pmatrix} = \vec{\sigma}_f \cdot \vec{S}_2 \\
 \tilde{S} &= 2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} = \tilde{S}_1 + \tilde{S}_2
 \end{aligned} \tag{A.27}$$

In order to simplify the algebra, we first diagonalize \tilde{S} in our 3 dimensional subspace. This yields

$$\tilde{S} = UJU^{-1} = \begin{pmatrix} -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} -4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} \tag{A.28}$$

With basis vectors

$$\begin{aligned}
 |\Phi_0\rangle &= \frac{1}{\sqrt{6}} (-|f1\bar{2}\rangle - |f\bar{1}2\rangle + 2|\bar{f}12\rangle) \\
 |\Phi_1\rangle &= \frac{1}{\sqrt{2}} (-|f1\bar{2}\rangle + |f\bar{1}2\rangle) \\
 |\Phi_2\rangle &= \frac{1}{\sqrt{3}} (|f1\bar{2}\rangle + |f\bar{1}2\rangle + |\bar{f}12\rangle)
 \end{aligned}$$

Now consider the impurity system with the infinite barrier reflective R_0 . Let $R_0 = -I_3$ and $e^{i2kd_0} = \delta$.

$$[I_3 - e^{i2kd_0} R_0 r']^{-1} = [I_3 + \delta r']^{-1} \tag{A.29}$$

The only term missing in the expression is r' . We find r' by first finding t and r in the standard basis and then perform a basis transformation to the diagonal basis $\{\Phi\}$.

$$\begin{aligned}
t &= [I + i\Omega\tilde{S}]^{-1} \\
U^{-1}tU &= U^{-1}[I + i\Omega\tilde{S}]^{-1}U \\
t' &= [I + i\Omega U^{-1}\tilde{S}U]^{-1} \\
t' &= \begin{pmatrix} (1 - 4i\Omega)^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1 + 2i\Omega)^{-1} \end{pmatrix} = \begin{pmatrix} \alpha' & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \beta' \end{pmatrix} \quad (\text{A.30})
\end{aligned}$$

We can now easily write $r' = t' - I$.

$$r' = \begin{pmatrix} \alpha' - 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \beta' - 1 \end{pmatrix} \quad (\text{A.31})$$

Equation (A.29) can be written as

$$[I_3 + \delta r']^{-1} = \begin{pmatrix} (1 + \delta(\alpha' - 1))^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1 + \delta(\beta' - 1))^{-1} \end{pmatrix}$$

Let $\eta = [1 + \delta(\alpha' - 1)]^{-1}$ and $\mu = [1 + \delta(\beta' - 1)]^{-1}$. We can now write the complete reflection matrix for first region

$$\begin{aligned}
R'_{\text{region } 1} &= r' - \delta t' [I_3 + \delta r']^{-1} t' \\
R'_{\text{region } 1} &= \begin{pmatrix} \alpha' - 1 - \delta\alpha'\eta\alpha' & 0 & 0 \\ 0 & -\delta & 0 \\ 0 & 0 & \beta' - 1 - \delta\beta'\mu\beta' \end{pmatrix} \quad (\text{A.32})
\end{aligned}$$

We now find the complete reflection matrix for the system by cascading this matrix with the first arbitrary height barrier. The barrier is a diagonal reflection matrix with $t_B = 1/(1 + i\Gamma) = \xi$ and $r = t_B - I$. Therefore

$$\mathbf{R}' = r'_b + \delta t'_B [I_3 - \delta R'_{\text{region } 1} r'_B]^{-1} R'_{\text{region } 1} t'_B \quad (\text{A.33})$$

We can write $[I_3 - \delta R'_{\text{region } 1} r'_B]^{-1}$ as

$$[I_3 - \delta R'_{\text{region } 1} r'_B]^{-1} = \begin{pmatrix} 1 - \delta(\alpha' - 1 - \delta(\alpha')^2 \eta)(\xi - 1) & 0 & 0 \\ 0 & 1 - \delta^2(1 - \xi) & 0 \\ 0 & 0 & 1 - \delta(\beta' - 1 - \delta(\beta')^2 \mu)(\xi - 1) \end{pmatrix}$$

Letting $\eta_2 = (1 - \delta(\alpha' - 1 - \delta(\alpha')^2 \eta)(\xi - 1))^{-1}$, $\mu_2 = (1 - \delta(\beta' - 1 - \delta(\beta')^2 \mu)(\xi - 1))^{-1}$,

and $\kappa_2 = (1 - \delta^2(1 - \xi))^{-1}$ we find

$$\mathbf{R}' = \begin{pmatrix} (\xi - 1) + \delta \xi^2 \eta_2 (\alpha' - 1 - \delta(\alpha')^2 \eta) & 0 & 0 \\ 0 & (\xi - 1) - \delta^2 \xi^2 \kappa_2 & 0 \\ 0 & 0 & (\xi - 1) + \delta \xi^2 \mu_2 (\beta' - 1 - \delta(\beta')^2 \mu) \end{pmatrix}$$

We can now simplify these expressions to find a solution of the form

$$\mathbf{R}' = \begin{pmatrix} -e^{i2\theta_a} & 0 & 0 \\ 0 & e^{i2\theta_b} & 0 \\ 0 & 0 & -e^{i2\theta_c} \end{pmatrix} = e^{i(\theta_a + \theta_c)} \begin{pmatrix} -e^{i\phi} & 0 & 0 \\ 0 & e^{i\chi} & 0 \\ 0 & 0 & -e^{-i\phi} \end{pmatrix}$$

where $\phi = \theta_a - \theta_c$ and $\chi = 2\theta_b - (\theta_a + \theta_c)$. With some algebra, we find:

$$\tan \theta_a = \frac{\sin(2kd_0) - 8\Omega \sin^2(kd_0)}{-16\Omega\Gamma \sin^2(kd_0) + \cos(2kd_0) + 2(\Gamma - 2\Omega) \sin(2kd_0)} \quad (\text{A.34})$$

$$\tan \theta_b = -\frac{2\Gamma \sin(2kd_0) + \cos(2kd_0)}{\sin(2kd_0)} \quad (\text{A.35})$$

$$\tan \theta_c = \frac{\sin(2kd_0) + 4\Omega \sin^2(kd_0)}{8\Omega\Gamma \sin^2(kd_0) + \cos(2kd_0) + 2(\Gamma + \Omega) \sin(2kd_0)} \quad (\text{A.36})$$

We are now in a position to transform back to the standard basis. Doing so we find

$$\mathbf{R} = -e^{i(\theta_a + \theta_c)} \frac{1}{6} \begin{pmatrix} e^{i\phi} - 3e^{i\chi} + 2e^{-i\phi} & e^{i\phi} + 3e^{i\chi} + 2e^{-i\phi} & -2e^{i\phi} + 2e^{-i\phi} \\ e^{i\phi} + 3e^{i\chi} + 2e^{-i\phi} & e^{i\phi} - 3e^{i\chi} + 2e^{-i\phi} & -2e^{i\phi} + 2e^{-i\phi} \\ -2e^{i\phi} + 2e^{-i\phi} & -2e^{i\phi} + 2e^{-i\phi} & 4e^{i\phi} + 2e^{-i\phi} \end{pmatrix} \quad (\text{A.37})$$

We can neglect the phase factor in-front without loss of generality. We immediately recognize that this matches the form of (2.20) where $c' \equiv c''$.

As we are interesting in two-qubit entanglement, it is helpful to transform to the Bell-state basis $\{1/\sqrt{2}(i1\bar{2} - \bar{1}2), 1/\sqrt{2}(i1\bar{2} + \bar{1}2)\}$ to directly observe rotations between the Bell states.

$$\begin{aligned} \mathbf{R}'' = U_B^\dagger \mathbf{R} U_B &= \begin{pmatrix} -i/\sqrt{2} & -1/\sqrt{2} & 0 \\ -i/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & c & d \\ c & a & d \\ d & d & b \end{pmatrix} \begin{pmatrix} i/\sqrt{2} & i/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ \mathbf{R}'' &= \begin{pmatrix} a & -ic & -\frac{d}{\sqrt{2}}(i+1) \\ ic & a & -\frac{d}{\sqrt{2}}(i-1) \\ \frac{d}{\sqrt{2}}(i-1) & \frac{d}{\sqrt{2}}(i+1) & b \end{pmatrix} \end{aligned} \quad (\text{A.38})$$

We are interested in understanding how this matrix manipulates a static two-spin wavefunction $|\Psi\rangle = \{\Psi_0, \Psi_1, \Psi_2\}$. Assume for the sake of analysis that each flying electron is provided in the up-state. If this is true, the wavefunction has only two components involved in the operation, Ψ_0 and Ψ_1 with Ψ_2 zero ($\bar{f}12$ has no component). Therefore $\mathbf{R}''|\Psi\rangle$ yields the following three recurrence relations:

$$|\Psi(n+1)\rangle = \begin{cases} a\Psi_1(n) - ic\Psi_2(n) \\ ic\Psi_1(n) - a\Psi_2(n) \\ \frac{d}{\sqrt{2}}(i-1)\Psi_1(n) + \frac{d}{\sqrt{2}}(i+1)\Psi_2(n) \end{cases} \quad (\text{A.39})$$

Recognizing that the eigenvalues of equation (A.38) are $a - c$ and $a + c$ we can find the solution to the recurrence equation:

$$\Psi_1(n) = \frac{1}{2} [(a - c)^n + (a + c)^n] \Psi_1(0) + \frac{i}{2} [(a - c)^n - (a + c)^n] \Psi_2(0) \quad (\text{A.40})$$

$$\Psi_2(n) = -\frac{i}{2} [(a - c)^n - (a + c)^n] \Psi_1(0) + \frac{1}{2} [(a - c)^n + (a + c)^n] \Psi_2(0) \quad (\text{A.41})$$

Notice that, neglecting the overall global phase factor $e^{i(\theta_a+\theta_c)}$

$$\begin{aligned} a - c &= -\frac{1}{6} [e^{i\phi} - 3e^{i\chi} + 2e^{-i\phi} - e^{i\phi} - 3e^{i\chi} - 2e^{-i\phi}] \\ a - c &= e^{i\chi} \end{aligned} \tag{A.42}$$

$$a + c = -\frac{1}{6} [2e^{i\phi} + 4e^{-i\phi}] \tag{A.43}$$

We can plug these in to find

$$\Psi_0(n) = \frac{1}{2} \left[e^{i\chi n} + \left[\frac{1}{6} (-2e^{i\phi} - 4e^{-i\phi}) \right]^n \right] \Psi_0(0) + \frac{i}{2} \left[e^{i\chi n} + \left[\frac{1}{6} (-2e^{i\phi} - 4e^{-i\phi}) \right]^n \right] \Psi_1(0) \tag{A.44}$$

B. STOCHASTIC FERROELECTRIC CAPACITOR

B.1 SPICE Model

```

1  ** 1D Time-Dependent Landau-Khalatnikov solver for ferroelectric
    capacitors
2  ** Adapted from Aziz, VOL. 32, NO. 6, IEEE (2016)
3  ** Includes support for thermal noise
4  ** Brian M. Sutton
5  ** bmsutton@purdue.edu
6
7  **Parameters:
8  **alpha, beta, and gamma are the Landau coefficients
9  **rho is the resistivity of the material
10 **A_FE is the area of the ferroelectric capacitor
11 **T_FE is the thickness of the ferroelectric capacitor
12 **TEMP is the temperature of the material
13 **nS is the noise suppression factor
14
15 .subckt FEC V_FE_P V_FE_N alpha='-11.57e7' beta='2.1e8' gamma='25e-6'
    rho='25e-6' A_FE='1e-9*1e-9' T_FE='1e-9' TEMP='200' nS='1'
16
17 *SPICE noise suppression factor
18 .param noiseSupress='nS'
19
20 *Permittivity of free space
21 .param e0='8.85418782e-12'
22
23 *Boltzmann Constant
24 .param kB='1.38064852e-23'
25
26 *Thermal Energy

```

```

27 .param kT='kB*TEMP'
28
29 *****NOISE MODEL*****
30 *Polarization modeled using a capacitor and current source
31 C_P P 0 c='rho'
32 G_P 0 P cur='-alpha*v(P) - beta*v(P)^3 -gamma*v(P)^5 + V(V_FE_P,V_FE_N)/
      T_FE'
33
34 *Noise modeled with a current source out of P
35 G_PN 0 P noise='4*kT*rho/(A_FE*T_FE)*(1/noiseSupress^2)'
36
37 *Voltage drop across the ferroelectric capacitor
38 E_VD V_DEL_FE 0 vol='V(V_FE_P,V_FE_N)'
39
40 *Charge conserved capacitor model of ferroelectric
41 C_FE V_FE_P V_FE_N Q='V(P)*A_FE'
42
43 *Linear capacitance
44 C_0 V_FE_P V_FE_N C='e0*A_FE/T_FE'
45 *****
46
47 .ends

```


VITA

Brian Sutton is a Senior Manager in the Embedded Systems Engineering group at Mercury Systems. He has over 13 years of experience in embedded systems with an emphasis on system security. His experience spans automotive IoT devices through x86 server platforms. At Mercury, he is responsible for new product definition, architecture, and implementation including digital and software design for FPGAs and SoCs. He currently leads the system security team driving bid and proposal activities, direct customer interaction, and contract execution. Prior to joining Mercury Systems, he worked as a Product Development Engineer at AMD focusing on automated testing, characterization, and failure analysis of x86 microprocessors. He completed his undergraduate and graduate studies at Purdue University with a focus on quantum and probabilistic computing architectures leveraging spintronics.