FROM ELECTRON GLASSES TO QUANTUM MACHINE LEARNING: A STUDY IN QUADRATIC UNCONSTRAINED BINARY OPTIMIZATION

A Dissertation

by

AMIN BARZEGAR

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Helmut G. Katzgraber
Committee Members,	Joseph H. Ross
	Winfied Teizer
	Stephen A. Fulling
Head of Department,	Grigory V. Rogachev

May 2020

Major Subject: Physics

ABSTRACT

Quadratic unconstrained binary optimization (QUBO) problems are of paramount importance in scientific and industrial applications as many interesting non-deterministic polynomial (NP)-hard problems can be mapped to them. QUBO is a subset of combinatorial optimization in which one seeks the global minimum of an objective function within a finite by occasionally towering set of possible configurations. Many simplistic algorithms, such as the gradient descent, fail to solve hard QUBO problems within a reasonable time frame as they find themselves lingering through the myriad of valleys and hills, hardly ever advancing toward the true ground state. Hence, designing new heuristics that can efficiently find solutions to such problems, as well as studying new instances of them can be extremely fruitful. QUBO problems have also shown great utility in quantum computing applications via quantum annealing, which has proven to be a promising endeavor to demonstrate the superiority of quantum devices over their classical counterparts. This work is, therefore, dedicated to studying quadratic optimization problems from several perspectives. We develop and benchmark physics-inspired algorithms such as population annealing Monte Carlo and thermal cycling. Using the designed algorithms, we study electron glasses, an instance of hard QUBO problems. Here, we show numerically that a transition to a spin-glass phase occurs at extremely low temperatures, which previous numerical studies have not been able to capture. We also study another case of QUBO problems, where we investigate the distribution of spin avalanches in systems with quench disorder. We establish new quantities similar to the concept of natural time in Seismology used as a potential measure for predicting large earthquakes. Finally, we turn to another exciting application of QUBO, namely quantum machine learning, an effort to use quantum devices in order to perform artificial intelligence algorithms more efficiently. Here, in addition to showing the advantage of physics-inspired solvers over the conventional heuristics such as Ridge regression, we propose a new way to uncover the power of quantum annealers in conducting machine learning tasks.

DEDICATION

To my beloved wife, Aysan, and to my parents.

ACKNOWLEDGMENTS

I would like to thank my advisor, Helmut G. Katzgraber, for his guidance, mentorship, and support throughout my Ph.D. I would also like to thank my wife, Aysan Bahari, for her unconditional love and support. I am also grateful to my committee members Winfied Teizer, Joseph H. Ross, and Stephen A. Fulling, for their time and consideration as well as their valuable critique of this work. I thank my collaborators Salvatore Mandrà, Wenlong Wang, Juan Carlos Andresen, Moshe Schechter, Arnulf Möbius, John Ferre, Richard Scalettar, Christopher Pattison, and Anuj Kankani, for their contributions to my work. I am grateful to my cat, Rory, for being a reliable companion and a great source of amusement during many sleepless nights and long days of graduate school. I also acknowledge Jonathan Machta, Martin Weigel, Alexis Giguere, Molly Luginbuhl, John Rundle, Don Turcotte, Mario DAndrea, and Mezcal Reposado for indirectly helping me with useful discussions. In addition, I thank my colleagues Andrew Ochoa, Chao Fang, Dilina Perra, and Katja Biswas, for their great insights. Finally, I would also like to acknowledge the staff at the Department of Physics and Astronomy, the Office of Graduate and Professional Studies, and Texas A&M High-Performance Research Computing for their assistance and support throughout my graduate studies.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Helmut G. Katzgraber (advisor), Joseph H. Ross, and Winfried Teizer from the Department of Physics and Astronomy and Professor Stephen A. Fulling from the Department of Mathematics.

The data in Chapter 3 was collected and analyzed in collaboration with Christopher Pattison from the Department of Physics and Astronomy of Texas A&M University and Wenlong Wang from the Department of Theoretical Physics of the KTH Royal Institute of Technology, Stockholm, Sweden. This work was published in the Journal of Physical Review E in 2018.

The work described in Chapter 4 was conducted in collaboration with Juan Carlos Andresen and Moshe Schechter from the Department of Physics of Ben Gurion University of the Negev, Israel and Helmut G. Katzgraber from the Department of Physics and Astronomy of Texas A&M University. The work was published in the Journal of Physical Review B in 2019.

The analyses depicted in Chapter 5 were conducted in part by John Ferre and Richard Scalettar from the Department of Physics of the University of California, Davis and were published in the Journal of Physical Review B in 2019.

The work Chapter 6 was conducted in collaboration with Arnulf Möbius from the Leibniz Institute for Solid State and Materials Research, Dresden, Germany, Salvatore Mandrà from the Quantum Artificial Intelligence Laboratory (QuAIL) at the National Aeronautics and Space Administration (NASA) as well as Anuj Kankani and Helmut G. Katzgraber from Texas A&M University .

The work in Chapter 7 was conducted in collaboration with Salvatore Mandrà from the Quantum Artificial Intelligence Laboratory (QuAIL) at the National Aeronautics and Space Administration (NASA) and Helmut G. Katzgraber from the Department of Physics and Astronomy of Texas A&M University.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by funding from the National Science Foundation (Grant No. DMR-1151387) and in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright annotation thereon.

NOMENCLATURE

L	Linear system size
D	Dimension
T	Temperature
β	Inverse temperature $1/T$
E	Energy
$\mathcal{H}$	Hamiltonian
$\mathcal{L}$	Lagrangian
s_i	Ising spin on the i'th lattice site
h_i	Local magnetic field of spin s_i
B	External magnetic field
J_{ij}	Interaction between spins s_i and s_j
k_B	Boltzmann constant
Z	Partition function
F	Helmholtz Free energy
Q_{ab}	Overlap auxiliary field
S_N	Symmetric group of N objects
μ_i	Classical state
$\|\psi\rangle$	Quantum State
$\Pi(\mu_i \to \mu_j)$	Transition probability from state μ_i to μ_j
A	Acceptance probability
Γ	Transverse field
$\boldsymbol{\sigma}_i$	Pauli matrices

ϵ_0	Permittivity of free space
N_{S}	Number of sweeps
N_{T}	Number of temperatures
N_{C}	Number of cycles
N_{P}	Pool size
R	Replica population size
S_{f}	Family entropy
ρ_{s}	Entropic family size
ρ_{f}	equilibrium population size
C_{V}	Heat capacity
T_{c}	Critical temperature
m_{s}	Staggered magnetization
W_{c}	Critical Disorder strength
ξ	Correlation Length
χ	Susceptibility
$\mathbf{k}$	Reciprocal lattice momentum
$\mathbf{r}$	Position vector
g	Binder cumulant
$q_{\alpha\beta}$	Spin overlap between replicas α and β
ΔA	Inter-event time
ω_i	Weight variable
$\mathcal{D}$	Data set
$\|\|\vec{\omega}\|\|_p$	$\left(\sum_{i=1}^{N} \|\omega_i\|^p\right)^{1/p}$ L^p-norm
erfc	Complementary error function
var	Variance

TSP	Traveling Salesman Problem
NP	Non-Deterministic Polynomial
AQC	Adiabatic Quantum Computing
MH	Metropolis-Hastings
REMC	Replica Exchange Monte Carlo
MCMC	Markov Chain Monte Carlo
PAMC	Population Annealing Monte Carlo
PT	Parallel Tempering
SA	Simulated Annealing
PIMC	Path Integral Monte Carlo
SQA	Simulated Quantum Annealing
ICM	Isoenergetic Cluster Move
LB	Linear in Beta
LBLT	Linear in Beta and Linear in Temperature
TSPL	Two Stage Power Law
MPI	Message Passing Interface
CPU	Central Processing Unit
DOS	Density of States
DC	Direct Current
CG	Coulomb Glass
CO	Charge Order
SG	Spin Glass
AT	de Almeida Thouless
RS	Replica Symmetric
RSB	Replica Symmetry Breaking

FSS	Finite-Size Scaling
PDF	Probability Distribution Function
CDF	Cumulative Distribution Function
RFIM	Random Field Ising Model
SKM	Sherrington-Kirkpatrick Model
VBM	Viana-Bray Model
SOC	Self-Organized Criticality
RPM	Return Point Memory
D-Wave	Commercial Quantum Annealer
QML	Quantum Machine Learning
AI	Artificial Intelligence
gcPBM	genomic-context Protein Binding Microarray
DNA	Deoxyribonucleic Acid
LASSO	Least Absolute Shrinkage and Selection Operator
AUPRC	Area Under Precision-Recall Curve
SVMC	Spin-Vector Monte Carlo
TCA	Thermal Cycling Algorithm
TTS	Time to Solution
DCL	Deceptive Cluster Loop
GEO	Gene Expression Omnibus
NCBI	National Center for Biotechnology Information

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENTS iv

CONTRIBUTORS AND FUNDING SOURCES v

NOMENCLATURE vii

TABLE OF CONTENTS xi

LIST OF FIGURES xiv

LIST OF TABLES xvii

1. INTRODUCTION 1

2. FUNDAMENTAL CONCEPTS 4

   2.1 A Brief Overview of Complexity 4
   2.2 Spin Glasses 6
   2.3 Monte Carlo Methods 11
      2.3.1 Markov Chains 11
      2.3.2 Parallel Tempering 15
      2.3.3 Simulated Annealing 17
      2.3.4 Quantum Annealing 18
      2.3.5 Spin-Vector Monte Carlo 23

3. OPTIMIZATION OF POPULATION ANNEALING MONTE CARLO FOR LARGE-SCALE SPIN-GLASS SIMULATIONS 25

   3.1 Introduction 25
   3.2 Preliminaries 27
      3.2.1 Case Study: Spin Glasses 28
      3.2.2 Outline of Population Annealing Monte Carlo 28
      3.2.3 Outline of the Cluster Updates Used 31
   3.3 Implementation Optimizations 33
      3.3.1 Comparison of the Spin Selection Methods 34
      3.3.2 Optimization of the Annealing Schedules 35

3.3.3 Optimization of the Number of Temperatures 39
3.3.4 Dynamic Population Sizes 41
3.4 Algorithmic Accelerators 43
3.4.1 Isoenergetic Cluster Updates 44
3.4.2 Wolff Cluster Updates 45
3.5 Parallel Implementation 47
3.5.1 Massively Parallel MPI Implementation 47
3.6 Conclusions and Future Challenges 49

4. NUMERICAL OBSERVATION OF A GLASSY PHASE IN THE THREE-DIMENSIONAL COULOMB GLASS 51

4.1 Introduction 51
4.2 Model 54
4.3 Simulation Details 54
4.4 Equilibration 56
4.5 Results 59
4.5.1 Charge-Ordered Phase 59
4.5.2 Coulomb Glass Phase 65
4.6 Conclusion 72

5. DISTRIBUTION OF INTER-EVENT AVALANCHE TIMES IN DISORDERED AND FRUSTRATED SPIN SYSTEMS 74

5.1 Introduction 74
5.2 Model and Methods 76
5.3 Statistics of All Avalanche Inter-event Times 80
5.4 Statistics of Above Threshold Inter-event Times 86
5.5 Effect of Longer-Range Couplings 90
5.6 Return Point Memory 94
5.7 Conclusions 96

6. OPTIMIZATION AND BENCHMARKING OF THE THERMAL CYCLING ALGORITHM 99

6.1 Introduction 99
6.2 Thermal Cycling Algorithm 100
6.3 Details of Analysis 101
6.4 Results 103
6.5 Conclusion 106

7. QUANTUM MACHINE LEARNING USING BIOLOGICAL PROBLEMS 108

7.1 Introduction 108
7.2 Classical Machine Learning 108
7.3 Quantum Machine Learning 110

7.4 Details of Analysis 112
    7.4.1 Data Preparation 112
    7.4.2 Cost Function 113
7.5 Learning Process 115
    7.5.1 Calibration 116
    7.5.2 Evaluation 117
    7.5.3 Scoring 117
7.6 Results 119
7.7 Conclusion 121

8. SUMMARY AND CONCLUSIONS 123

8.1 Summary 123
8.2 Impacts and Applications 125
8.3 Future Directions 126

REFERENCES 127

LIST OF FIGURES

FIGURE Page

2.1 Computational complexity classes 4

2.2 The traveling salesman problem 6

2.3 Phase diagram of the Sherrington-Kirkpatrick model.......... 10

2.4 Random walk of the first 100 billion digits of π in base 4 12

2.5 Schematic diagram of the parallel tempering algorithm 16

2.6 Quantum versus thermal fluctuations 19

2.7 Quantum annealing by adiabatic evolution of a quantum state 20

3.1 Diagram outlining the different optimizations we have implemented for population annealing Monte Carlo.......... 26

3.2 Comparison of the entropic population size ρ_{s} for different spin selection methods .. 34

3.3 Annealing schedules experimented in two and three dimensions 36

3.4 Comparison of the systematic errors for various annealing schedules 38

3.5 Energy density distribution of the LBLT annealing schedule for $L = 8$ in three space dimensions.......... 39

3.6 Optimization of the number of annealing steps N_T in two space dimensions and three space dimensions 40

3.7 Instance-by-instance comparison for a PAMC simulation with fixed and dynamic population sizes 42

3.8 Population annealing with ICM updates in 2D 43

3.9 Population annealing with ICM updates in 3D 44

3.10 Mean normalized cluster size as a function of β for the Wolff algorithm as well as the performance of the algorithm in both 2D and 3D 46

3.11 Scaling of the total wall time as a function of the number of processors N for two system sizes $L = 8$ and $L = 12$ 48

4.1 Phase diagram of the three-dimensional Coulomb glass model 52

4.2 Equilibration of a PAMC simulation 57

4.3 Finite-size correlation length per system size ξ/L versus temperature T for various disorder strengths 60

4.4 Zero-temperature simulation results for the plasma-CO phase transition 62

4.5 Finite-size scaling analysis for the plasma-CO phase transition at $W=0.05$ 63

4.6 Spin-glass correlation length divided by system size $\xi_{\rm SG}/L$ calculated using two replicas at $W = 0.8$ versus temperature T 66

4.7 Spin-glass finite-size correlation length $\xi_{\rm SG}/L$ as a function of temperature T at various disorder strengths W 67

4.8 Importance of proper thermalization in observing a CG phase transition 68

4.9 Process of estimating the critical exponents, as well as the critical temperature $T_{\rm c}$ of the plasma-CG phase transition for $W = 0.5$ 70

5.1 Hysteresis loop of the random-field Ising model of size $N = 100^3$ with a distribution of fields of widths $R = 3$, 4, and 5 77

5.2 Distribution of avalanche sizes for field width $R = 3$ and 5 78

5.3 Scaling collapse of the distribution of inter-event times for the RFIM at field width $R = 2.3$ and the VBM for various system sizes N 81

5.4 Distribution of inter-event times for the SKM. The distributions largely overlap, apart from a modest separation at small ΔB 82

5.5 Hysteresis loop for the RFIM with $R = 3$ and for the VBM for different system sizes N 83

5.6 Hysteresis loop for the SKM 84

5.7 Distribution of inter-event times for the RFIM for various field widths R 85

5.8 Cumulative distribution of the inter-event times for the RFIM at $R = 3$ 86

5.9 Cumulative distribution of the inter-event times for the RFIM at $R = 2.3$ on a double-logarithmic scale 87

5.10 Cumulative distribution of the inter-event times for the RFIM at $R = 2.3$ on a double-logarithmic scale 88

5.11 Cumulative distribution of the inter-event times for the SKM with lattice size $N = 10000$ 89

5.12 Cumulative distribution of the inter-event times for the SKM with lattice size $N = 10000$ 90

5.13 Hysteresis loops for the Hamiltonian in Eq. (5.10) for various J' values 91

5.14 Distribution of inter-event times for various values of J' 92

5.15 $C(R)$ plotted against the disorder strength R for various J' values 93

5.16 Distribution of avalanche sizes plotted for various J' values for $N = 50^3$ and $R = 7$ 94

5.17 Cumulative distribution of the inter-event times for the RFIM across an internal hysteresis loop 95

5.18 Cumulative distribution of the inter-event times for the RFIM across an internal hysteresis loop 96

6.1 Parameter tuning for various algorithms 103

6.2 Comparison of the scaling results for the studied algorithms 105

7.1 Classical k-means clustering algorithm 111

7.2 Learning curves for the estimator LASSO 115

7.3 The procedure of determining the regularization parameter λ by 100-fold Mote-Carlo cross validation 116

7.4 Schematic process of training and testing. Multiple instances are prepared from the training data set at random 117

7.5 Schematic diagram showing precision and recall 118

7.6 Classification performance of the classical and quantum-inspired estimators 120

LIST OF TABLES

TABLE Page

3.1 Simulation parameters for various experiments to optimize PAMC 33

4.1 PAMC simulation parameters used for the finite-temperature simulations in the CO phase 56

4.2 PAMC simulation parameters used for the finite-temperature simulations in the CG phase 56

4.3 Critical parameters of the plasma-CO phase transition at different disorder values ... 64

4.4 Critical parameters of the plasma-CG phase transition for various values of the disorder W 72

6.1 Optimal parameters of the thermal cycling algorithm for different problem sizes L .. 104

7.1 An example of the DNA sequencing data after preprocessing 113

1. INTRODUCTION

Optimization is ubiquitous in many areas of science and industry; search for the ground state of exotic states of matter such as high temperature superconductors in physics, topology optimization in material science, lead optimization in pharmaceutical drug discovery, spacecraft trajectory optimization, active filter design in electric engineering, portfolio optimization in finance, scheduling in transportation, and speech recognition in artificial intelligence are only a few examples of the numerous applications of optimization. One important category in this plethora is combinatorial optimization, which is the search for the minima of an objective function within a finite set of solutions, for instance, the traveling salesman problem and graph coloring.

An important subset of combinatorial optimization are quadratic unconstrained binary optimization (QUBO) problems which in general have the following form:

$$\zeta(Q) = \min \left\{ \frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q}\, \boldsymbol{x} + \boldsymbol{h}^T \boldsymbol{x} \quad | \quad \boldsymbol{x} \in \{0,1\}^N \right\}, \tag{1.1}$$

where $\boldsymbol{Q}$ is a real symmetric matrix called *interaction*, and $\boldsymbol{h}$ is a real column vector often referred to as *field*. As a matter of fact, it can be shown [1–3] that many hard optimization problems such as number partitioning, satisfiability problems, set packing, vertex cover, job sequencing, and Hamiltonian cycles can be mapped to this format. Hence, it would be beneficial to study these problems and, at the same time, design new heuristics that can solve them efficiently.

The keen observer will notice the close similarity between Eq. (2.1) and the Hamiltonian for an Ising spin glass [4–6]. More than fifty years of research in the area of systems with quenched disorder, of which spin glasses are a subcategory, equip us with broad knowledge regarding the structure of such systems, much of which can be readily applied to the QUBO problems. A critical understanding that we gain from the spin-glass physics is the reason for the numerical hardness of such problems; disorder causes frustration and gives rise to competing interactions, which in turn contorts the energy landscape into a rugged manifold of many metastable states. Naturally, any attempt to extrapolate the locally attained information to global statements about the system will be in vain. That is why the time required to solve many QUBO problems often grows exponentially with the problem size rendering them intractable for even the best available algorithms. Such problems are often referred to as non-deterministic polynomial (NP) in the complexity theory jargon.

Another significant area where QUBO plays a crucial role is its utility in quantum computing applications. Since 1994 when Peter Shor proposed the first quantum algorithm involving integer factorization [7], there has been considerable interest in building noise-robust quantum devices as well ad finding problems where *quantum advantage* can be demonstrated. A promising endeavor in this area is adiabatic quantum computing (AQC) [8–10] that has lead to the advent of quantum annealing (QA). In QA, the ground state is found by slowly deforming the energy landscape from a trivially solvable problem to a target problem of often NP complexity by taking advantage of quantum phenomena such as tunneling and entanglement. Quantum annealers like D-Wave [11–14] processors are commercially available now. The qubits on a D-Wave processor form a Chimera graph architecture [14] that consists of a two-dimensional lattice of fully-connected $K_{4,4}$ cells inside which all qubits are coupled together ferromagnetically. Most optimization problems can be embedded on the Chimera graph in the QUBO format, which further highlights the importance of these problems.

Motivated by the above discussion, this dissertation is dedicated to studying quadratic binary optimization problems in various, yet complementary settings. From a purely algorithmic point of view, we develop and test the performance of new heuristics inspired by statistical physics such as Population Annealing Monte Carlo (PAMC) after presenting some elementary concepts in Chapter 2. Using the developed PAMC algorithm, we investigate an instance of a quadratic binary model in the context of a real physical system, i.e., Coulomb glasses in chapter 4, where we successfully identify a glass phase within the phase diagram, which has been elusive to numerical studies for many years. Later in Chapter 5, we study a QUBO problem with geophysical applications. There we show that the distribution of spin avalanches, following the hysteresis loop of a long-range spin glass, resembles that of the number of small earthquakes occurring between large seismic events often called *natural time*, which can be useful in predictions of major earthquakes.

We then embark on developing and optimizing another algorithm called thermal cycling, an effective tool in solving many hard QUBO problems. Thermal cycling takes advantage of periodic heating and cooling of a system to help it in escaping from local minima within an often jagged energy landscape teeming with meta-stable states. Our results show that thermal cycling performs better than conventional heuristics, such as simulated annealing and competitive with state-of-the-art algorithms, for instance, parallel tempering.

Finally, in chapter 7, we turn to another interesting practice in quadratic binary optimization, namely quantum machine learning, where we use techniques in artificial intelligence to study biological data involving DNA transcription factors of the human genome. Our results show that physics-inspired algorithms such as simulated quantum annealing outperform the commonly used models in machine learning, for instance, Ridge regression. We also show that the introduction of so-called “soft” spins can improve the predictive accuracy of the above algorithms during the learning process. This effectively implies that the true quantum states of an adiabatic quantum annealer, which are a linear combination of classical states due to entanglement, can indeed be used to show *quantum advantage*.

2. FUNDAMENTAL CONCEPTS

2.1 A Brief Overview of Complexity

Combinatorial optimization constitutes an important area of optimization in which one seeks to globally minimize or maximize a cost function with finite but often exponentially large configuration space. Some examples include the traveling salesman problem (TSP), the geometrical packing problem, graph coloring, the cutting stock problem, and integer quadratic programming. Many of these problems are considered non-deterministic polynomial (NP) in the sense that their solutions can be verified efficiently; however, the time required to solve them does not scale as a polynomial function of the problem size, but rather exponential. Note that for any NP problem,

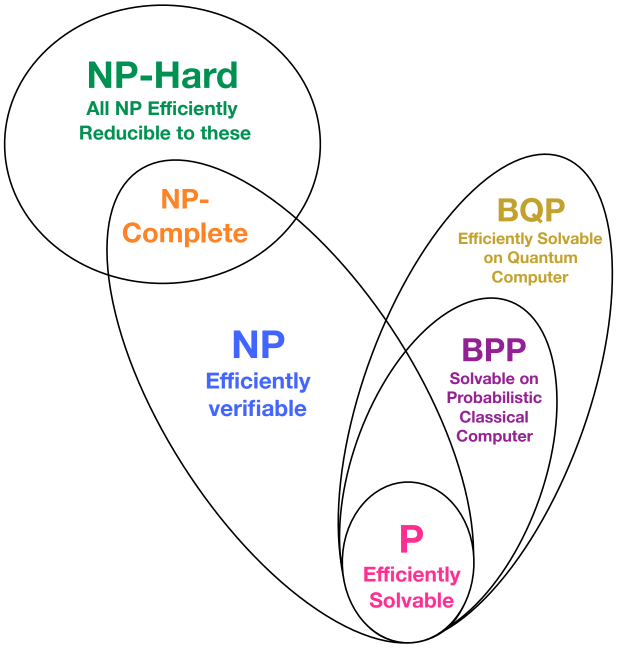

Figure 2.1: Computational complexity classes. P problems are efficiently solvable in polynomial time. The NP class includes all decision problems which cannot be solved efficiently, but their solution can be verified in polynomial time. NP-complete problems are the hardest problems in the entire NP class. The NP-hard class consists of the problems that are at least as hard as the hardest problem in NP to which all other NP problems can be reduced in polynomial time. BPP and BQP are decision problems that can be efficiently solved on a probabilistic classical and quantum computer, respectively.

there might be a yet unknown deterministic heuristic that can solve it efficiently. Therefore, it is uncertain whether or not P = NP, a conjecture that is listed as one of the millennium prize problems [15]. Some other problems in combinatorial optimization, such as the TSP, are shown to be NP-complete, which constitute the hardest problems in the NP class. NP-hard problems, which are not necessarily decision problems, are at least as hard as the hardest NP problem. It can be shown [16] that all NP problems can be reduced to NP-hard problems in polynomial time. Hence, an efficient solution to any of the NP-hard problems will automatically guarantee a polynomial-time solution to the entire NP class. As we show on the complexity diagram in Fig. 2.1 there are additional problems that can be solved efficiently using non-deterministic algorithms that make up the bounded-error probabilistic polynomial time (BPP) class. On the other hand, problems such as integer factorization are shown [7] to have polynomial time solution on quantum computers.

An important subset of combinatorial optimization are quadratic unconstrained binary optimization (QUBO) problems, which in general involve minimizing the following cost function:

$$\mathcal{H} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} Q_{ij}x_i x_j + \sum_{i=1}^{N} h_i x_i, \quad x_i \in \{0,1\}, \tag{2.1}$$

where Q_{ij} are the elements of a a symmetric real matrix and h_i are also real-valued. As a matter of fact, it can be shown [1–3] that many NP optimization problems such as number partitioning, satisfiability problems, set packing, vertex cover, job sequencing, Hamiltonian cycles, etc., can be mapped to this format. As an example, here we demonstrate how the traveling salesman problem can be set up in the QUBO format. The TSP is a subset of the Hamiltonian cycle problems which seeks the shortest route between a number of cities identified by their coordinates, such that each city is be visited only once and the path returns to its starting point in the end. This is shown schematically in Fig. 2.2. In the graph theory language, this problem can be thought as a directed graph $G = (V, E)$ where the edge $(uv) \in E$ is weighted by the Euclidean distance d_{uv} between the vertices $u, v \in V$. So the problem becomes finding the Hamiltonian cycle on which the sum of

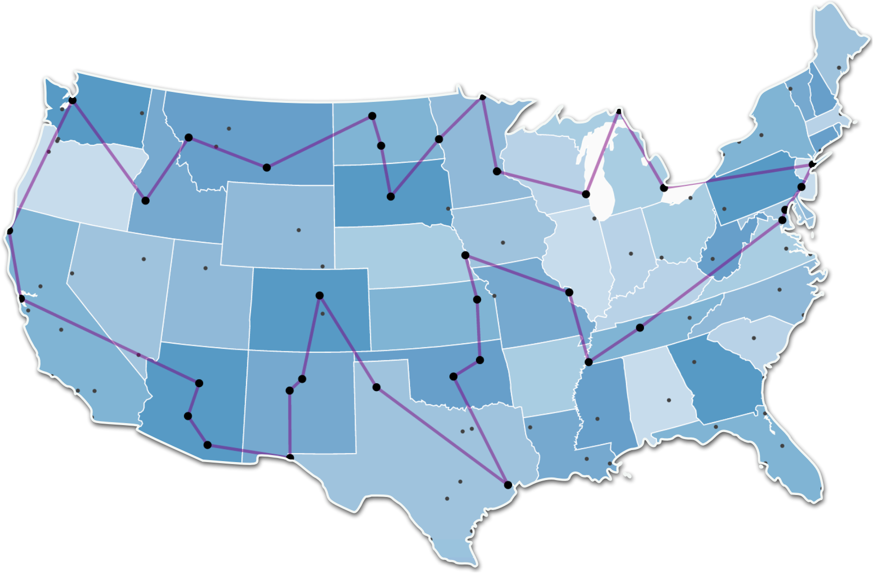

Figure 2.2: The traveling salesman problem (picture credit https://blog.essaycorp.com). The goal is to find a shortest closed route between N cities such that each city is visited exactly once. The configuration space grows as $N!$, which makes TSP an NP-complete problem.

weights is minimized. Using the variables $x_{v,j} \in \{0,1\}$ in which the index j shows that the vertex v appears j'th in the cycle, we can present the cost function as follows:

$$\mathcal{H}_{\text{TSP}} = \sum_{v=1}^{N}\left(1 - \sum_{j=1}^{N} x_{v,j}\right)^2 + \sum_{j=1}^{N}\left(1 - \sum_{v=1}^{N} x_{v,j}\right)^2 + \sum_{(uv)\in E}\sum_{j=1}^{N} d_{uv} x_{u,j} x_{v,j+1}, \tag{2.2}$$

where the first two terms ensure that each vertex appears in the cycles exactly once and the last term is simply the total length of the path.

2.2 Spin Glasses

It is straightforward to show that with some small modifications, Eq. (2.1) can also be shown in the following form:

$$\mathcal{H} = \frac{1}{2}\sum_{i}^{N}\sum_{j\in\mathcal{N}_i} J_{ij} s_i s_j - \sum_{i}^{N} h_i s_i, \tag{2.3}$$

where the spin variables s_i can take the values $s_i = \pm 1$, and the neighbor set $\mathcal{N}_i$ depends on the topology of the lattice. The above equation can be recognized as the Hamiltonian for an Ising spin glass [4–6, 17, 18]. Spin glasses are magnetic materials with *quenched* disorder meaning that the couplings J_{ij} are constant on the time scales over which the spins fluctuate. The introduction of such a disorder has far-reaching implications such as frustration and competing interactions. In the presence of frustration, it is impossible to simultaneously satisfy all of the bonds, which gives rise to spatially random configurations that are frozen in time. This results in a rugged energy landscape teeming with metastable states, making it extremely difficult to travel around in an ergodic fashion, which may explain the underlying reason for the intractability of many QUBO problems. A conceptual difficulty that poses itself here is that in the presence of disorder, any thermodynamic quantity will naturally depend on the disorder because the partition function is calculated based on a particular disorder instance.

$$Z(\mathbf{J}) = \sum_{\{s_i\}} e^{-\beta \mathcal{H}[\mathbf{s},\mathbf{J}]}. \tag{2.4}$$

When the system size N is sufficiently large, extensive quantities such as the free energy, can be generated by accumulating the contributions of smaller and relatively independent subsystems due to the additivity of such quantities. The central limit theorem [19] then predicts that the variance of the extensive quantity will fall off proportional to $1/N$, i.e., it will be independent of $\mathbf{J}$ in the thermodynamic limit $N \to \infty$. This phenomena is called *self averaging*. Therefore, the true disorder-independent value of extensive quantities can be obtained by averaging over the disorder instances. For instance, the free energy density can be calculated as follows:

$$f = -\frac{1}{\beta N}\overline{\log Z(\mathbf{J})} = -\frac{1}{\beta N}\int \prod_{i,j}^{N} dJ_{ij} P(J_{ij}) \log Z(\mathbf{J}). \tag{2.5}$$

Note that the disorder average is over the logarithm of the partition function, where a seemingly impasse is reached for the analytic calculation of the thermodynamic quantities. What often referred to as the *replica method* [4] is a clever way to circumvent this problem in which the following mathematical representation is used:

$$\log Z = \lim_{n\to 0} \frac{Z^n - 1}{n}. \tag{2.6}$$

Using this technique, the disorder average over the logarithm is reduced to the average over the product of n replicas of the partition function.

$$\overline{\log Z} = \lim_{n\to 0} \frac{\log \overline{Z^n}}{n} = \lim_{n\to 0} \frac{1}{n} \log \sum_{\{s_i^a\}} \overline{e^{-\beta \sum\limits_a^n \mathcal{H}[\mathbf{s}^a, \mathbf{J}]}}. \tag{2.7}$$

For instance, in the case of a fully connected lattice, i.e., $\mathcal{N}_i = \{1, \cdots, N\}$ with Gaussian disorder, namely, $P(J_{ij}) \sim e^{-NJ_{ij}^2/2J^2}$ and uniform external field $h_i = h$, which is called the Sherrington-Kirkpatrick model [20], it is easy to show that after disorder average we find:

$$\begin{aligned}
\overline{Z^n} =& \sum_{\{s_i^a\}} \exp\left[\frac{\beta^2 J^2}{8N} \sum_i^N \sum_{j\neq i}^N \left(\sum_a^n s_i^a s_j^a\right)^2 - \beta h \sum_a^n \sum_i^N s_i^a\right] \\
=& \sum_{\{s_i^a\}} \exp\left[\frac{\beta^2 J^2}{2N} \sum_{a>b}^n \left(\sum_i^N s_i^a s_i^b\right)^2 - \beta h \sum_a^n \sum_i^N s_i^a + \frac{nN\beta^2 J^2}{8}\right].
\end{aligned} \tag{2.8}$$

Note that the terms proportional to quadratic and higher powers of n can be dropped because the limit $n \to 0$ will be applied at the end. Using the Hubbard-Stratonovich transformation [21, 22], we can introduce the continuous *overlap* variables Q_{ab}.

$$\begin{aligned}
\overline{Z^n} = \sum_{\{s_i^a\}} \int \prod_{a\neq b}^n \left(\sqrt{\frac{N\beta^2 J^2}{4\pi}} dQ_{ab}\right) \exp\Bigg[-\frac{N\beta^2 J^2}{4} \sum_{a\neq b}^n Q_{ab}^2 + \frac{\beta^2 J^2}{2} \sum_{a\neq b}^n Q_{ab} \sum_i^N s_i^a s_i^b \\
-\beta h \sum_a^n \sum_i^N s_i^a + \frac{nN\beta^2 J^2}{8}\Bigg] = \frac{e^{-nN\beta^2 J^2/8}}{[N\beta^2 J^2]^{\frac{n}{2}}} \int \prod_{a\neq b}^n \left(\frac{dQ_{ab}}{\sqrt{4\pi}}\right) e^{\mathcal{L}[\mathbf{Q}]},
\end{aligned} \tag{2.9}$$

in which the Lagrangian $\mathcal{L}[\mathbf{Q}]$ has the following definition:

$$\mathcal{L}[\mathbf{Q}] = -\frac{N\beta^2 J^2}{4} \sum_{a\neq b}^{n} Q_{ab}^2 + N \log \sum_{\{s^a\}} e^{-\beta \mathcal{H}_{\text{eff}}[\mathbf{s},\mathbf{Q}]}, \tag{2.10}$$

where the effective Hamiltonian is defined as:

$$\mathcal{H}_{\text{eff}}[\mathbf{s},\mathbf{Q}] = -\frac{\beta^2 J^2}{2} \sum_{a\neq b}^{n} Q_{ab} s^a s^b + \beta h \sum_{a}^{n} s^a. \tag{2.11}$$

Using the above Lagrangian, it is straightforward to determine the saddle point (most probable) solution for the overlap variables.

$$\frac{\partial \mathcal{L}}{\partial Q_{ab}} = 0 \quad \Rightarrow \quad \tilde{Q}_{ab} = \left\langle s^a s^b \right\rangle = \frac{\sum\limits_{\{s^a\}} s^a s^b e^{-\beta \mathcal{H}_{\text{eff}}[\mathbf{s},\tilde{\mathbf{Q}}]}}{\sum\limits_{\{s^a\}} e^{-\beta \mathcal{H}_{\text{eff}}[\mathbf{s},\tilde{\mathbf{Q}}]}}, \tag{2.12}$$

From Eq. (2.12), it becomes clear why the auxiliary variables Q_{ab} were referred to as “overlap” earlier. We notice that the saddle point solution is symmetric in the replica indices, which motivates the so called replica symmetric (RS) solution where the overlap is assumed to be a constant and traceless matrix.

$$Q_{ab} = (1 - \delta_{a,b}) Q. \tag{2.13}$$

With the RS solution, the Hamiltonian in Eq. (2.46) reduces to the one dimensional ferromagnetic Ising model, which can be solved exactly using the transfer matrix method. After taking the limit $n \to 0$, Eq. (2.12) gives:

$$\tilde{Q} = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \tanh^2 \left(\beta J \sqrt{\tilde{Q}} z + \beta h \right). \tag{2.14}$$

The above equation can then be solved to give the RS overlap function. At zero external field

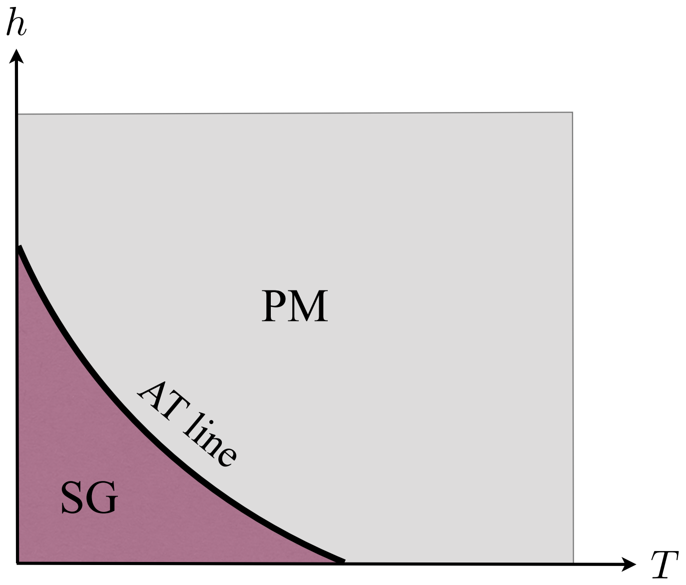

Figure 2.3: Phase diagram of the Sherrington-Kirkpatrick model. De Almeida-Thouless (AT) line separates the paramagnetic (PM) phase from the spin-glass (SG) phase. Below the AT line, the replica symmetric solution becomes unstable and is replaced by the replica symmetry breaking ansatz.

$h = 0$ and to the lowest orders, we find:

$$\tilde{Q} = (\beta J)^2 \tilde{Q} - \frac{2}{3}(\beta J)^4 \tilde{Q}^2 + \mathcal{O}(Q^4) \tag{2.15}$$

We observe that depending on the temperature, two distinct cases emerge.

$$\beta J < 1, \quad \tilde{Q} = 0 \tag{2.16}$$

$$\beta J > 1, \quad \tilde{Q} \neq 0, \tag{2.17}$$

which indicates a continuous phase transition with the overlap function as the order parameter. For $\beta J < 1$ the system is in a paramagnetic phase, while for $\beta J > 1$ it enters a spin glass (SG) phase. It can be shown that below the phase transition point, the replica symmetric solution becomes unstable to infinitesimal perturbations and gives a negative values for the entropy. This phenomenon is called *replica symmetry breaking* (RSB). Indeed, if we keep the external field h, we get a phase diagram like what is shown in Fig. 2.3. The line separating the the paramagnetic phase from the glass phase is called de Almeida-Thouless (AT) [23] line below which the replica symmetry is lost. Since in the SG phase the RS solution is no longer valid, one must think of an alternative anzats. This was first done by Blandin [24] and later generalized by Parisi [25–27] to what is known as the one-step and multiple-step replica symmetry breaking solutions. In the one-step ansatz, replicas are grouped into blocks with the overlap function Q_{ab} taking two different values on the diagonal blocks and the off-diagonal ones. For a more comprehensive analysis, we refer the reader to Ref. [18].

2.3 Monte Carlo Methods

In the quest for the solution of NP-hard QUBO problems, the exact enumeration of states is impractical as the number of states increases exponentially with the problem size. Heuristics such as gradient descent and branch and bound [28] tend to become equally inefficient due to the innumerable local minima within the rugged energy landscapes of these problems often caused by disorder and frustration, as we saw in the previous section. A better approach in this regard is stochastic optimization, where some randomness is introduced in the process to allow for escaping from the local minima and eventually reaching the global optimum. The prime example of such stochastic sampling are Monte Carlo methods. Monte Carlo methods use random sampling to obtain numerical estimates of deterministic quantities, for instance, a definite multidimensional integral or to find probabilistic expectation values in applications such as weather forecasting.

2.3.1 Markov Chains

When the probability distribution is known in advance, Markov Chain Monte Carlo (MCMC) techniques are often used. A Markov chain [29] is a memoryless random process in which the probability of future events depends solely on the state that the system is currently in, regardless of what its history has been so far. An example of such a process is a random walk like Brownian motion [30], where the momentum of each particle is determined based on its imminent collisions

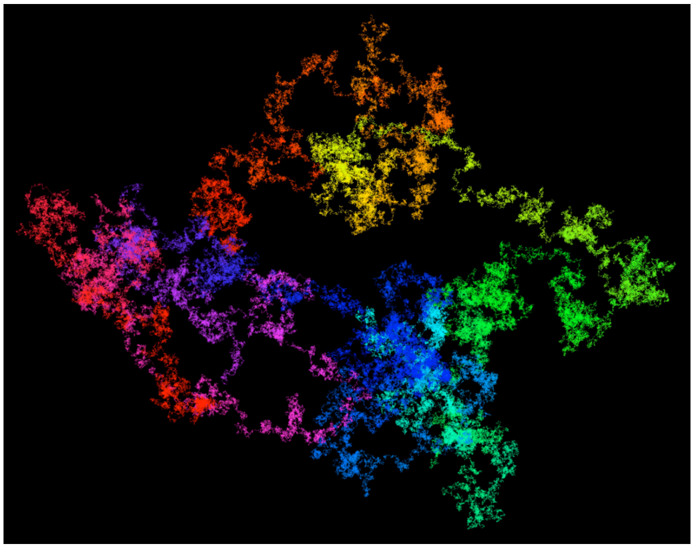

Figure 2.4: Random walk on the first 100 billion digits of π in base 4 (picture credit: F. J. Aragón, *et al.*, 2013). The walker takes equal strides to the right, left, up, or down depending if the next digit is 0, 1, 2, or 3, respectively. The walker returns to the starting point at the end of the walk. The color spectrum shows the relative spatial distance between different regions in the walk.

with the neighboring particles. In Fig. 2.4 we show a random walk on the first 100 billion digits of number π in base 4 [31], where the walker, with uniform probability, takes an equal step to the right, left, up, or down if the current digit is 0, 1, 2 or 3, respectively. The color map shows the relative spatial distance between different parts of the walk. Formally a discrete Morkov chain is a sequence of random variables $x_0, x_1, \dots$ with the possible states $\mu_0, \mu_1, \dots$ which satisfy the following property:

$$\mathcal{P}(x_n = \mu_n \,|\, x_{n-1} = \mu_{n-1}) = \mathcal{P}(x_n = \mu_n \,|\, x_0 = \mu_0, x_1 = \mu_1, \dots, x_{n-1} = \mu_{n-1}), \tag{2.18}$$

meaning that knowledge of the previous state is all that is necessary to determine the probability distribution of the current state. As such, we can define the marginal distribution $\mathcal{P}(x_n = \mu_j)$, i.e, the probability that the system acquires the state μ_j at the step k in the sequence.

$$\mathcal{P}(x_n = \mu_j) = \sum_i \mathcal{P}(x_n = \mu_j \,|\, x_{n-1} = \mu_i)\mathcal{P}(x_{n-1} = \mu_i). \tag{2.19}$$

From the above equation, we may define the transition probability matrix for a time-homogeneous Markov process as follows:

$$\Pi_{ij} = \Pi(\mu_i \to \mu_j) = \mathcal{P}(x_{n+1} = \mu_j \,|\, x_n = \mu_i), \quad \forall\, n. \tag{2.20}$$

One can then use Eq. (2.19) recursively to show that

$$\mathcal{P}(x_n = \mu_j) = \sum_{k_1,\ldots k_{n-1}} \Pi_{ik_1}\Pi_{k_1k_2}\cdots\Pi_{k_{n-1}j} = \Pi^n_{ij} \tag{2.21}$$

in which we have assumed that $\mathcal{P}(x_0 = \mu_i) = 1$ for the initial state μ_i. If after a sufficiently long time $\mathcal{P}(x_n = \mu_j)$ converges, then the Markov chain reaches its steady state and the limiting probabilities $p(\mu_j) = \lim\limits_{n\to\infty} \Pi^n_{ij}$ are called the stationary or equilibrium distribution of the Morkov process. The transition probabilities can be chosen such that a desired distribution is generated by the stationary distribution of the Markov chain. Note that $p(\mu_j)$ is independent of the initial state, meaning that it is possible to eventually get from every state to every other state if $p(\mu_j) > 0$ for all μ_j. In that case, the Markov chain is called *ergodic* after the ergodic hypothesis of thermodynamics [32], which says that all possible states are equiprobable over a long period of time. Ergodicity ensures that the time average of any quantity is equal to its average over the statistical ensemble. In other words, statistical properties can be deduced from a single and sufficiently long instance of the process. Returning to Eq. (2.19), it is straightforward to see that

$$\begin{aligned}\mathcal{P}(x_{n+1} = \mu_j) - \mathcal{P}(x_n = \mu_j) = &\sum_i \mathcal{P}(x_{n+1} = \mu_j \,|\, x_n = \mu_i)\mathcal{P}(x_n = \mu_i) \\ &- \sum_i \mathcal{P}(x_{n+1} = \mu_i \,|\, x_n = \mu_j)\mathcal{P}(x_n = \mu_j),\end{aligned} \tag{2.22}$$

where we have used the identity $\sum\limits_{i} \mathcal{P}(x_{n+1} = \mu_i \,|\, x_n = \mu_j) = 1$ in the second term. When the steady state is reached $\mathcal{P}(x_{n+1} = \mu_j) = \mathcal{P}(x_n = \mu_j)$, we find the equilibrium condition as follows:

$$\sum_i p(\mu_i)\Pi(\mu_i \to \mu_j) = \sum_i p(\mu_j)\Pi(\mu_j \to \mu_i). \tag{2.23}$$

If the above equation holds for each state μ_j, the Morkov chain is called reversible,

$$p(\mu_i)\Pi(\mu_i \to \mu_j) = p(\mu_j)\Pi(\mu_j \to \mu_i). \tag{2.24}$$

This expression is also referred to as *detailed balance*. It can be shown from the Boltzmann's H-theorem [33] that for an isolated physical system, detailed balance is the sufficient condition for strict increase of entropy, as stipulated by the second law of thermodynamics.

It is customary to separate the transition probability into the proposal $Q(\mu_j \,|\, \mu_i)$ and acceptance $A(\mu_i \to \mu_j)$ parts:

$$\Pi(\mu_i \to \mu_j) = Q(\mu_j \,|\, \mu_i) A(\mu_i \to \mu_j), \tag{2.25}$$

which upon substitution in Eq. (2.24) gives the acceptance ratio,

$$r = \frac{A(\mu_i \to \mu_j)}{A(\mu_j \to \mu_i)} = \frac{p(\mu_j)Q(\mu_i \,|\, \mu_j)}{p(\mu_i)Q(\mu_j \,|\, \mu_i)}. \tag{2.26}$$

Now, the goal is to choose the above probabilities such that Eq. (2.26) is satisfied and at the same time the acceptance probability $A(\mu_i \to \mu_j)$ is as high as possible, if not rejection free. One possibility is the Metropolis-Hastings (MH) [34, 35] approach:

$$A_{\mathrm{MH}}(\mu_i \to \mu_j) = \min\left[1, \frac{p(\mu_j)Q(\mu_i \,|\, \mu_j)}{p(\mu_i)Q(\mu_j \,|\, \mu_i)}\right]. \tag{2.27}$$

We can apply the MH algorithm at a finite and fixed temperature $T = 1/\beta$, where the equilibrium configurations of a system are represented by the canonical ensemble, and the probability for each state is given by the Boltzmann distribution:

$$p(\mu_j) = \frac{1}{Z} e^{-\beta \mathcal{H}[\mu_j]} = \frac{e^{-\beta E_j}}{\sum\limits_k e^{-\beta E_k}} \tag{2.28}$$

in which $\mathcal{H}[\mu_j]$ is the Hamiltonian of the system at microstate μ_j with energy E_j, and Z is the partition function. Since the proposition of the update $\mu_i \to \mu_j$ is equally probable as $\mu_j \to \mu_i$, we have $Q(\mu_j \,|\, \mu_i) = Q(\mu_i \,|\, \mu_j)$. Therefore, using Eqs. (2.27) and (2.28) we find:

$$A_{\mathrm{MH}}(\mu_i \to \mu_j) = \min\left[1, e^{-\beta(E_j - E_i)}\right]. \tag{2.29}$$

We observe that the MH acceptance probability becomes excessively small, namely the algorithms is rendered inefficient, when the energy difference $\Delta E = E_j - E_i$ is large relative to the ambient temperature. This problem is occasionally encountered at low temperatures or in systems with tall barriers within their energy landscapes, for instance, spin glasses. This shortcoming can be remedied by using multiple Markov chain techniques such as parallel tempering or annealing-based algorithms like simulated annealing and path integral Monte Carlo. In the next sections, we will explain these methods in greater detail.

2.3.2 Parallel Tempering

As we discussed earlier, the Metropolis dynamics tend to critically slow down when the height of the energy barriers is considerably larger than the temperature at which the Markov chain operates. A natural way to get around this problem is to use a so-called replica-exchange Monte Carlo (REMC) method [36] in which one has a number of the above Markov chains at different temperatures, whose configurations are regularly exchanged. Since the dynamics are slow at lower temperatures, the stochastic processes are confined to a small region of the phase space. In practice, the above exchange acts as a driving force that propels such impeded processes to other parts of the phase space, which could not otherwise be possible. A modern variant of the replica-exchange

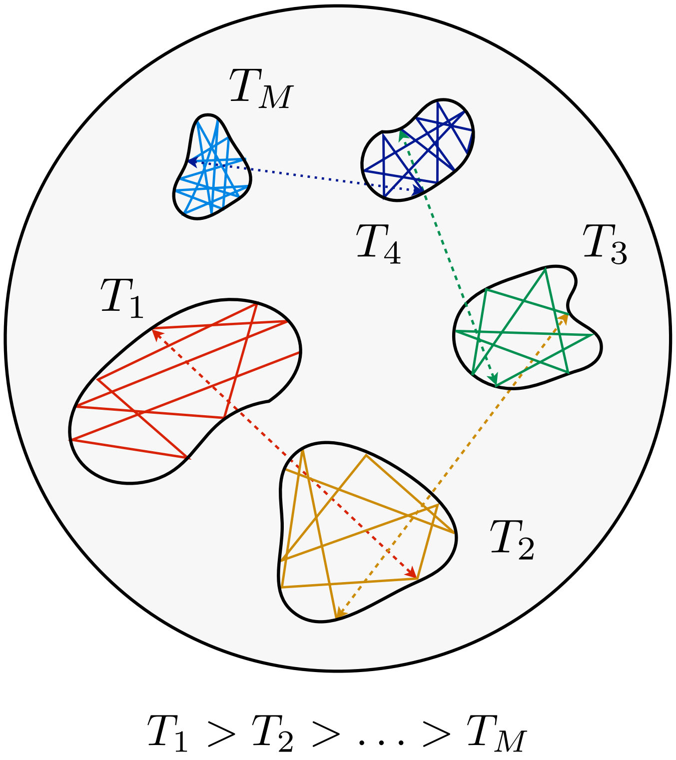

Figure 2.5: Schematic diagram of the parallel tempering algorithm. The enclosing circle represents the phase space, and the blobs show the Markov chains at different temperatures. The Markov chains at high temperatures explore larger areas of the phase space in a coarser fashion, wheres at low temperatures that the Metropolis dynamics become critically slow, the walkers are confined to small regions, which are searched in greater detail. By exchanging the configurations of the neighboring temperatures, we ensure that such processes have access to other regions as well.

Monte Carlo algorithms is parallel tempering (PT) [37–39] in which M copies of the system, initialized with random configurations, are subject to the Metropolis updates at a set of temperatures $T_1 > T_2 > \ldots > T_M$ that usually form a geometric progression, i.e,

$$T_k = T_1 \prod_{i=1}^{k-1} \sqrt[M-1]{\frac{T_M}{T_1}}. \tag{2.30}$$

Once in a while, the configurations of the neighboring temperatures are exchanged based on a probability distribution that preserves the detailed balance relation. This is shown schematically in Fig. 2.5. Assuming that we are exchanging the replicas at inverse temperatures $\beta' > \beta$ with energy

E' and E, respectively. Because the exchange is symmetric, we have:

$$Q\left(\mu,\beta;\mu',\beta' \,|\, \mu',\beta;\mu,\beta'\right) = Q\left(\mu',\beta;\mu,\beta' \,|\, \mu,\beta;\mu',\beta'\right). \tag{2.31}$$

Substituting the above equation in Eq. (2.26) and using $p(\mu,\beta) = e^{-\beta E}/Z(\beta)$ we find:

$$r = \frac{A\left(\mu,\beta;\mu',\beta' \to \mu',\beta;\mu,\beta'\right)}{A\left(\mu',\beta;\mu,\beta' \to \mu,\beta;\mu',\beta'\right)} = \frac{p(\mu',\beta;\mu,\beta')}{p(\mu,\beta;\mu',\beta')} = \exp\left[(\beta'-\beta)(E'-E)\right], \tag{2.32}$$

which can be used to find the acceptance probability as follows:

$$A_{\mathrm{PT}}\left(\mu,\beta;\mu',\beta' \to \mu',\beta;\mu,\beta'\right) = \min\left[1, e^{(\beta'-\beta)(E'-E)}\right]. \tag{2.33}$$

Comparing the PT acceptance ratio to that of the simple Metropolis algorithm in Eq. (2.27), we notice that the temperature prefactor in the exponent appears as a difference, which can be tuned to be as small as desired, and therefore the algorithm continues to the be efficient even if ΔE is large. The configuration swaps allow the replicas to continuously move up and down in temperature in a round trip. In this way, the system is able to avert metastable states when it visits higher temperatures and, subsequently, sample low-lying states once it has relaxed at lower temperatures again.

2.3.3 Simulated Annealing

Here we give a brief overview of the annealing-based algorithms that are sequential in temperature, meaning that the configurations at different temperatures are not independent. The cornerstone of such algorithms is simulated annealing (SA) [40]. Simulated annealing is a Markov Chain Monte Carlo process where equilibration is continuously maintained following an *annealing* schedule during which the temperature of the system is gradually lowered to a target temperature, starting from a sufficiently high temperature. The goal is to guide the stochastic process through an occasionally complex energy landscape toward the low-lying states. At high temperatures where the random "walker" can take long strides across the phase space, many energy minima can be visited. As the system is cooled down, the exploration domain of the walker is shrunk to the most probable, i.e., the lowest energy states, such that it eventually lands in the ground state. In many cases, the temperature schedule must be sufficiently slow in order for the annealing process to function because if the temperature is decreased too quickly, the system may become trapped in a local minimum. For a generic combinatorial optimization problem, it has been shown [41] that if the annealing schedule is $T \sim 1/\log(t)$ or slower, the ground state is asymptotically reached by a simulated annealing process.

The equilibration in simulated annealing is usually achieved by the Metropolis algorithm. As we pointed out in Sec. 2.3, the acceptance rates of the Metropolis updates become exponentially small at low temperatures; therefore, simulated annealing will also suffer from such slow-down. Similar to the replica-exchange algorithms, one can accelerate the dynamics in SA by using multiple copies of the system as semi-independent sequences that are annealed separately but replicated according to their relative Boltzmann weight. This improved algorithm is called Population Annealing Monte Carlo (PAMC), which we will study in great depth in Chapter. 3.

2.3.4 Quantum Annealing

We saw in the previous sections that thermal fluctuation could be used to prevent a stochastic process from stalling due to being trapped in metastable states. However, if it happens that the height of the energy barriers is much larger than the amplitude of the above fluctuations, the dynamics can come to an indefinite stop. An alternative is to introduce quantum fluctuation such that the system can tunnel through the barriers as opposed to thermally jumping over them. This is shown schematically in Fig. 2.6. Note that the tunneling is suppressed when the barrier width is large; one can use the WKB semi-classical approximation to show that the tunneling probability decays exponentially with the barrier width. Kadowaki and Nishimori [42] were first to show that the addition of a quantum transverse field to a simulated annealing process of the Ising model can indeed improve the chance of finding the ground state, a method which they named *quantum an-*

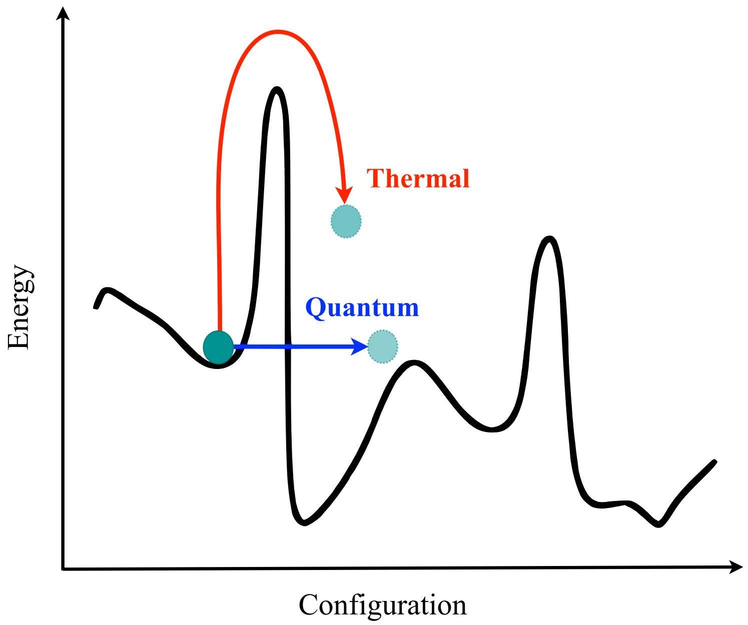

Figure 2.6: Quantum versus thermal fluctuations. Thermal fluctuations cause the system to jump over the barriers, which can get inefficient if the barrier height is large. On the other hand, with quantum fluctuations, the system is able to tunnel through the barrier directly. The tunneling probability decays exponentially with the barrier width.

nealing (QA). Quantum annealing owes its success to the adiabatic theorem of quantum mechanics [43] which states that:

A physical system remains in its instantaneous eigenstate if a given perturbation acts on it slowly enough and if there is a gap between that eigenstate and the rest of the Hamiltonian's spectrum.

QA uses the above theorem in the following way: suppose we have a gapped Hamiltonian $\mathcal{H}_0$ whose ground state $\mathcal{H}_0|\psi\rangle = E_0|\psi\rangle$ we would like to find. We add to this a driving perturbation $\mathcal{H}_1$ with a known ground state $\mathcal{H}_1|\phi\rangle = E_1|\phi\rangle$ as follows:

$$\mathcal{H}(\lambda) = \lambda\mathcal{H}_0 + (1-\lambda)\mathcal{H}_1, \quad \mathcal{H}|\Psi(\lambda)\rangle = E(\lambda)|\Psi(\lambda)\rangle. \tag{2.34}$$

We start from $\lambda = 0$ when $\mathcal{H}(\lambda = 0) = \mathcal{H}_1$ and prepare the system at the ground state, namely $|\Psi(\lambda = 0)\rangle = |\phi\rangle$. We then increase λ slowly to $\lambda = 1$, where the adiabatic theorem can be invoked to ensure that the system remains in the ground state of the instantaneous Hamiltonian.

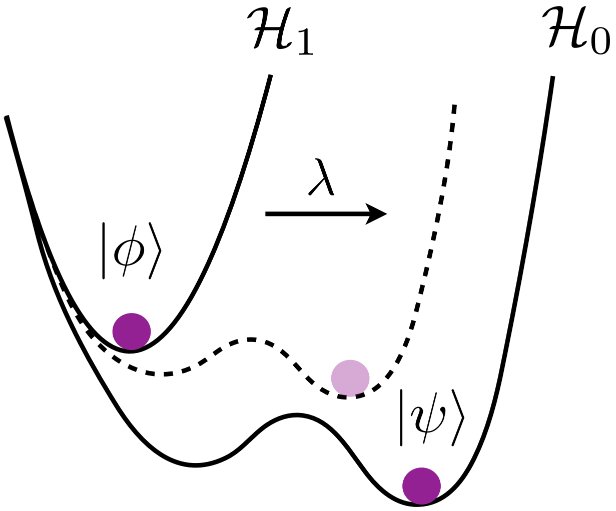

Figure 2.7: Quantum annealing by adiabatic evolution of a quantum state. The system is initialized at the known ground states of $\mathcal{H}_1$, and the Hamiltonian is slowly deformed to $\mathcal{H}_0$, whose ground state $|\psi\rangle$ we are interested in. The adiabatic theorem guarantees that the system remains in the ground state throughout the annealing process.

Therefore, at the end of the process, when $\mathcal{H}(\lambda = 1) = \mathcal{H}_0$, the system will settle in the ground state of the target Hamiltonian, i.e., $|\Psi(\lambda = 1)\rangle = |\psi\rangle$. We have shown this in Fig. 2.7.

Numerical simulations of quantum annealing were initially done by solving the time-dependent shrödinger equation. In an alternative approach, the full quantum Hamiltonian is approximated by a classical system comprised of many ferromagnetically interacting copies of the original system. This method is often called path integral Monte Carlo (PIMC) or simulated quantum annealing (SQA) [44–47]. PIMC uses the Trotter-Suzuki formalism [48–50], which we will briefly explain below. Suppose we want to find the solution to a QUBO problem which, as we discussed in Sec. 2.1, can be mapped to a classical Ising spin glass Hamiltonian, i.e.,

$$\mathcal{H}_0 = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} J_{ij} s_i s_j. \tag{2.35}$$

Now, we elevate this Hamiltonian to the quantum transverse Ising spin glass [51] as follows:

$$\mathcal{H} = \lambda \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} J_{ij} \boldsymbol{\sigma}_i^z \boldsymbol{\sigma}_j^z - (1-\lambda)\Gamma \sum_{i=1}^{N} \boldsymbol{\sigma}_i^x = \lambda \mathcal{H}_0 + (1-\lambda)\mathcal{H}_1 \tag{2.36}$$

in which Γ is the transverse field strength, and λ is the annealing parameter. $\boldsymbol{\sigma}_i^z$ and $\boldsymbol{\sigma}_i^x$ are the Pauli matrices that satisfy the commutation relation $[\boldsymbol{\sigma}_n^z, \boldsymbol{\sigma}_m^x] = i\delta_{n,m}\boldsymbol{\sigma}_n^y$ with $\boldsymbol{\sigma}_i^z|\sigma_i\rangle = s_i|\sigma_i\rangle$. The quantum partition function is calculated as the trace of the Boltzmann exponential, namely,

$$Z = Tr\left(e^{-\beta\mathcal{H}}\right) = Tr\left[e^{-\beta\lambda\mathcal{H}_0 - \beta(1-\lambda)\mathcal{H}_1}\right]. \tag{2.37}$$

Because the two terms in the Hamiltonian (2.36) are non-commuting, we cannot separate the exponentials. Instead, we use the Lie-Trotter formula [48] :

$$e^{A+B} = \lim_{M\to\infty} \left(e^{A/M} e^{B/M}\right)^M. \tag{2.38}$$

Using Eqs. (2.37) and (2.38) we find:

$$\begin{aligned} Z &= \lim_{M\to\infty} Tr\left[e^{-\beta\lambda\mathcal{H}_0/M} e^{-\beta(1-\lambda)\mathcal{H}_1/M}\right]^M \\ &= \lim_{M\to\infty} \sum_{\{\sigma_i\}} \langle \sigma_1, \dots, \sigma_N| \left[e^{-\beta\lambda\mathcal{H}_0/M} e^{-\beta(1-\lambda)\mathcal{H}_1/M}\right]^M |\sigma_1, \dots, \sigma_N\rangle. \end{aligned} \tag{2.39}$$

We now make $2M-1$ insertions the identity operator $I = \sum_{\{\sigma_i\}} |\sigma_1, \dots, \sigma_N\rangle\langle\sigma_1, \dots, \sigma_N|$ to get:

$$\begin{aligned} Z = \lim_{M\to\infty} \sum_{\{\sigma_i^1, \dots, \sigma_i^M\}} \prod_{\alpha \in 2\mathbb{N}+1}^{M-1} &\langle \sigma_1^\alpha, \dots, \sigma_N^\alpha | e^{-\beta\lambda\mathcal{H}_0/M} | \sigma_1^{\alpha+1}, \dots, \sigma_N^{\alpha+1}\rangle \\ &\langle \sigma_1^{\alpha+1}, \dots, \sigma_N^{\alpha+1} | e^{-\beta(1-\lambda)\mathcal{H}_1/M} | \sigma_1^{\alpha+2}, \dots, \sigma_N^{\alpha+2}\rangle, \end{aligned} \tag{2.40}$$

where periodic boundary conditions, i.e., $\sigma_i^1 = \sigma_i^{M+1}$ are assumed for all i values. Using Eq. (2.36), it is straightforward to see that

$$\prod_{\alpha=1}^{M}\langle\sigma_1^\alpha,\ldots,\sigma_N^\alpha|e^{-\beta\lambda\mathcal{H}_0/M}|\sigma_1^{\alpha+1},\ldots,\sigma_N^{\alpha+1}\rangle=\exp\left(-\frac{\beta\lambda}{M}\sum_{\alpha=1}^{M}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)}J_{ij}s_i^\alpha s_j^\alpha\right). \tag{2.41}$$

As for the $\mathcal{H}_1$ part in Eq. (2.40), we use the identity $\exp(\zeta\boldsymbol{\sigma}^x)=\cosh\zeta+\boldsymbol{\sigma}^x\sinh\zeta$ which gives:

$$\prod_{\alpha=1}^{M}\langle\sigma_1^\alpha,\ldots,\sigma_N^\alpha|e^{-\beta(1-\lambda)\mathcal{H}_1/M}|\sigma_1^{\alpha+1},\ldots,\sigma_N^{\alpha+1}\rangle=C^{\frac{M}{2}}\exp\left(\frac{\beta J'}{M}\sum_{\alpha=1}^{M}\sum_{i=1}^{N}s_i^\alpha s_i^{\alpha+1}\right) \tag{2.42}$$

in which C and J' are defined as following:

$$C=\frac{1}{2}\sinh\left[\frac{2(1-\lambda)\Gamma\beta}{M}\right],\quad J'=\frac{M}{2\beta}\ln\coth\left[\frac{(1-\lambda)\Gamma\beta}{M}\right]. \tag{2.43}$$

Substituting Eqs. (2.41) and (2.42) in Eq. (2.40), we find:

$$Z=\lim_{M\to\infty}C^{\frac{M}{2}}\sum_{\{\sigma_i^1,\ldots,\sigma_i^M\}}\exp\left(-\frac{\beta\lambda}{M}\sum_{\alpha=1}^{M}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)}J_{ij}s_i^\alpha s_j^\alpha+\frac{\beta J'}{M}\sum_{\alpha=1}^{M}\sum_{i=1}^{N}s_i^\alpha s_i^{\alpha+1}\right) \tag{2.44}$$

that can be written in terms of the classical temperature $\beta_c=\frac{\beta}{M}$ as

$$Z=\lim_{M\to\infty}C^{\frac{M}{2}}\sum_{\{\sigma_i^1,\ldots,\sigma_i^M\}}e^{-\beta_c\mathcal{H}_{\text{eff}}[\sigma_i^\alpha]}, \tag{2.45}$$

where the effective classical Hamiltonian $\mathcal{H}_{\text{eff}}$ has the following definition:

$$\mathcal{H}_{\text{eff}}[\sigma_i^\alpha]=\lambda\sum_{\alpha=1}^{M}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)}J_{ij}s_i^\alpha s_j^\alpha-\frac{1}{2\beta_c}\ln\coth\left[(1-\lambda)\Gamma\beta_c\right]\sum_{\alpha=1}^{M}\sum_{i=1}^{N}s_i^\alpha s_i^{\alpha+1}. \tag{2.46}$$

What we have done here is that we have encoded all the information in the original quantum Hamiltonian into the effective classical system at the expense of adding an extra dimension with ferromagnetic coupling J' to the system. Loosely speaking, we have mimicked the *quantum parallelism* by infinite classical replication and quantum entanglement by establishing interactions between those replicas. From Eq. (2.46), we see that the ferromagnetic coupling between the neighboring copies grows unboundedly as $\lambda \to 1$, showing that they become strongly coupled in the latter stages of the anneal. The above coupling ensures that the collective "knowledge" of the replicas, as they move around the phase space independently, is used to continuously push the system toward the ground state.

2.3.5 Spin-Vector Monte Carlo

As we saw in the previous section, quantum fluctuations can be induced in an Ising spin glass Hamiltonian by introducing a transverse field Γ that causes the classical *pure* states to become mixed owing to quantum *entanglement*. Therefore, the most general state of a system of N qubits is as follows:

$$|\psi\rangle = \frac{1}{\sqrt{N!}} \sum_{p \in S_N} C_p(q_1, \ldots, q_N) |q_{p(1)}\rangle \otimes \ldots \otimes |q_{p(N)}\rangle \tag{2.47}$$

in which the sum is over $N!$ permutations of the symmetric group S_N and $|q_i\rangle = \alpha_i|0\rangle + \beta_i|1\rangle$ is the state of the i'th qubit. Let us now take a transverse Ising Hamiltonian with local fields, i.e.,

$$\mathcal{H} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} J_{ij} \boldsymbol{\sigma}_i^z \boldsymbol{\sigma}_j^z - \sum_{i=1}^{N} h_i \boldsymbol{\sigma}_i^z - \Gamma \sum_{i=1}^{N} \boldsymbol{\sigma}_i^x. \tag{2.48}$$

A semi-classical approximation can be made by assuming no entanglement between the qubit states but still preserving the quantum superposition of the computational basis:

$$|\tilde{\psi}\rangle = \left(\cos\frac{\theta_1}{2}|0\rangle + \sin\frac{\theta_1}{2}|1\rangle\right) \otimes \ldots \otimes \left(\cos\frac{\theta_N}{2}|0\rangle + \sin\frac{\theta_N}{2}|1\rangle\right). \tag{2.49}$$

We my now use the above state to find the expectation value of the Hamiltonian in Eq. (2.48):

$$\mathcal{H}_{\text{SVMC}} = \langle\tilde{\psi}|\mathcal{H}|\tilde{\psi}\rangle = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} J_{ij} \langle\boldsymbol{\sigma}_i^z\rangle\langle\boldsymbol{\sigma}_j^z\rangle - \sum_{i=1}^{N} h_i \langle\boldsymbol{\sigma}_i^z\rangle - \Gamma \sum_{i=1}^{N} \langle\boldsymbol{\sigma}_i^x\rangle. \tag{2.50}$$

Now using $\langle \boldsymbol{\sigma}_i^z \rangle = \cos\theta_i$ and $\langle \boldsymbol{\sigma}_i^x \rangle = \sin\theta_i$ we find:

$$\mathcal{H}_{\text{SVMC}} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} J_{ij} \cos\theta_i \cos\theta_j - \sum_{i=1}^{N} h_i \cos\theta_i - \Gamma \sum_{i=1}^{N} \sin\theta_i, \tag{2.51}$$

which is precisely the spin-vector Monte Carlo (SVMC) Hamiltonian. Therefore, we observe that SVMC is the semi-classical limit of a quantum Ising model that naturally inherits some quantum properties. The transverse field Γ controls how "soft" the variables are; if $\Gamma = 0$, the energy is minimized by $\theta_i \in \{0, \pi\}$, and the SVMC Hamiltonian reduces to the classical spin glass Hamiltonian with "hard" spins, namely, $s_i = \pm 1$. On the other hand when $\Gamma > 0$, any real value is possible for the angles θ_i, and hence the corresponding spin variables are continuous.

3. OPTIMIZATION OF POPULATION ANNEALING MONTE CARLO FOR LARGE-SCALE SPIN-GLASS SIMULATIONS[†]

3.1 Introduction

Monte Carlo algorithms are widely used in many areas of science, engineering, and mathematics. These approaches are of paramount importance for problems where no analytical solutions are possible. For example, the class of Ising-like Hamiltonians can only be solved analytically in a few exceptionally rare cases. The vanilla Ising model can only be solved analytically in one, two, as well as infinite space dimensions. A solution in three space dimensions remains elusive to date [51, 53]. Therefore, simulations are necessary to understand these systems in three space dimensions. The situation is far direr when more complex interactions–such as k-local terms rather than the usual quadratic or 2-local terms–are used. Similarly, the inclusion of disorder allows for analytical solutions only in the mean-field regime [4, 5, 20, 25, 54]. These spin-glass problems, a subset of frustrated and glassy systems, represent the easiest 2-local Hamiltonian that is computationally hard. A combination of diverging algorithmic time scales (with the size of the input) due to rough energy landscapes and the need for configurational (disorder) averages to compute thermodynamic quantities makes them the perfect benchmark problems to study novel algorithms. Finally, computing ground states of spin glasses on non-planar graphs is an NP-hard problem, where Monte Carlo methods have been known to be efficient heuristics [55–57], whereas only a few efficient exact methods exist for small system sizes.

It is, therefore, of much importance to design or improve efficient algorithms either to save computational effort or to have better quality data with the same computational effort, when studying these complex systems. Two popular algorithms that are currently in use (for both thermal sampling as well as optimization) are parallel tempering (PT) Monte Carlo [37, 38] and population annealing Monte Carlo (PAMC) [58–61].

Although both PT and PAMC are extended ensemble Monte Carlo methods, PAMC is a se-

---

[†] Reprinted with permission from [52].

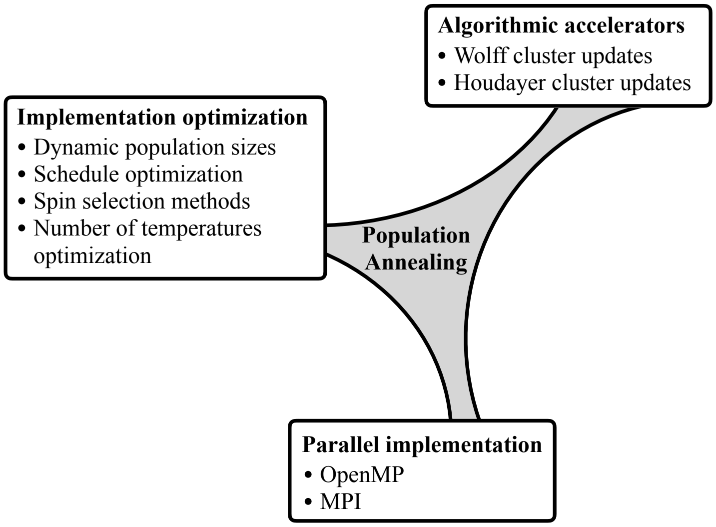

Figure 3.1: Diagram outlining the different optimizations we have implemented for population annealing Monte Carlo. These range from optimizations in the implementation, such as efficient spin selection techniques, to algorithmic accelerators (e.g., the inclusion of cluster updates), as well as parallel implementations. See the main text for details. Reprinted with permission from [52].

quential Monte Carlo algorithm, in contrast to PT that is a replica-exchange Markov-Chain Monte Carlo method. PAMC is a population-based Monte Carlo method and thus well-suited for implementations on multi-core high-performance computing machines. PAMC is similar to simulated annealing [40] with an extra resampling step, where the temperature is reduced to maintain thermal equilibrium. PT has been intensively optimized and has been the workhorse in statistical physics to date. It has also shown [61] to be equally efficient in simulating spin glasses, when compared to PAMC. PAMC, on the other hand, remains a relatively new simulation method. Although careful systematic studies of PAMC [61, 62] exist, and the method has been applied broadly [56, 63–66], little effort has been made to thoroughly optimize the algorithm. Here, we focus on this problem and study various approaches to improve the efficiency of PAMC for large-scale simulations. While some approaches improve PAMC, others have little to no effect. Note that related optimization ideas are explored in Ref. [67].

Our strategy to optimize PAMC is three-pronged, as illustrated in diagram 3.1. First, we study different implementation optimizations, where we discuss dynamic population sizes that vary with the temperature during the anneal, as well as the optimization of different annealing schedules. We also investigate different spin selection methods (order of spin updates in the simulation), such as random, sequential, and checkerboard. While for disordered systems sequential updates are commonplace, random updates are needed for nonequilibrium studies. In the case of bipartite lattices, a checkerboard spin-update technique can be used, which is perfectly suited for parallelization. Furthermore, we discuss how to determine the optimum number of temperatures for a given simulation. Second, we analyze the effects of algorithmic accelerators by adding cluster updates to PAMC. We have studied Wolff cluster updates [68] as well as Houdayer cluster updates [69], and isoenergetic cluster moves [70]. Third, we discuss different parallel implementations using both OpenMP [*] (ideal for shared-memory machines [61, 62]) and MPI [†] with load balancing (ideal for scalable massively-parallel implementations). Note that the implementation of PAMC on graphics processing units (GPUs) has been discussed extensively in Refs. [63, 71].

The chapter is structured as follows. We first introduce in Sec. 3.2 some concepts needed in this study, such as the case study Hamiltonian, and outline the PAMC algorithm. Implementation optimizations are presented in Sec. 3.3, algorithmic accelerators via cluster updates in Sec. 3.4, and parallel implementations are discussed in Sec. 3.5, followed by concluding remarks.

3.2 Preliminaries

In this section, we introduce some concepts needed for the PAMC optimization in the subsequent section. In particular, we introduce the Ising spin-glass Hamiltonian (our case study) as well as PAMC and different algorithmic accelerators.

---

[*] See http://www.openmp.org

[†] See, for example, https://www.open-mpi.org

3.2.1 Case Study: Spin Glasses

We study the zero-field two-dimensional (2D) and three-dimensional (3D) Edwards-Anderson Ising spin-glass [4] given by the Hamiltonian

$$H = -\sum_{\langle ij \rangle} J_{ij}^{N} s_i s_j, \tag{3.1}$$

where $s_i = \pm 1$ are Ising spins, and the sum is over the nearest neighbors on a D-dimensional lattice of linear size L with $N = L^D$ spins. The random couplings J_{ij} are chosen from a Gaussian distribution with mean zero and variance one. We refer to each disorder realization as an "instance". The model has no phase transition to a spin-glass phase in 2D [72], while in 3D there is a spin-glass phase transition at $T_{\mathrm{c}} \approx 0.96$ [73] for Gaussian disorder.

3.2.2 Outline of Population Annealing Monte Carlo

One of the commonly used algorithms for sampling the equilibrium states of a physical system as well as finding solutions to binary optimization problems is simulated annealing (SA) [40]. As we discussed in Chapter 2, simulated annealing is a Markov-Chain Monte Carlo (MCMC) process in which equilibration is achieved via simple Metropolis-Hastings algorithm [34, 35], as the temperature is continuously lowered. Population Annealing Monte Carlo (PAMC) introduced by Hukushima and Iba [58] and further developed by Machta and Wang [60, 61] has simulated annealing as its core, yet borrows some additional features from multi Markov-chain algorithms, such as parallel tempering and histogram reweighting [74]. Similar to simulated annealing, PAMC is sequential in the sense that it follows an annealing schedule, which is often linear in inverse temperature β. PAMC, however, outperforms SA by introducing a population of replicas of the system, which are simultaneously annealed toward the target temperature. PAMC [61] starts with a large population of R_0 replicas at a high temperature, where thermalization is easy. In our simulations, we initialize replicas randomly at the inverse temperature $\beta = 1/T = 0$. The population traverses an annealing schedule with N_T temperatures and maintains thermal equilibrium to a low target temperature of $T_{\mathrm{min}} = 1/\beta_{\mathrm{max}}$. Physical measurements can then be carried out by simple averaging of the desired observables at the replica level. When the temperature is lowered from β to β', the population is resampled. The mean number of the copies of replica i is proportional to the appropriate reweighting factor, $\exp[-(\beta' - \beta)E_i]$. The constant of proportionality is chosen such that the expectation value of the population size at the new temperature is $R(\beta')$. Note that $R(\beta')$ is usually kept close to R although this is not a necessary condition. Indeed, in our dynamical population size implementation, we let R change as a function of β and seek better algorithmic efficiency in the number of spin updates. The resampling is followed by $N_{\mathrm{S}} = 10$ Monte Carlo sweeps (one Monte Carlo sweep represents N attempted spin updates) for each replica of the new population using the Metropolis algorithm. We keep $N_{\mathrm{S}} = 10$ without loss of generality because the performance of PAMC is mostly sensitive to the product of $N_{\mathrm{S}}N_T$ near optimum. For example, two PAMC simulations with $\{N_{\mathrm{S}} = 10, N_T\}$ and $\{N_{\mathrm{S}} = 1, 10N_T\}$ are similar in efficiency, if N_T is reasonably large. The amount of work of a PAMC simulation in terms of sweeps is $W = RN_{\mathrm{S}}N_T$, where R is the average population size.

One of the unique benefits of population annealing is the convenient access to the free energy, which is often an arduous task, if not impossible, in many other algorithms. Another advantage of PAMC is the inherent parallelizability of the algorithm, which can considerably reduce the runtime of a simulation. Below, we present a brief outline of the PAMC algorithm.

Algorithm 1 Population Annealing Monte Carlo

```
Initialize $R_0$ replicas with random configurations at $\beta_0 = 0$.
for $N_\mathrm{T}$ steps until $\beta = \beta_{\mathrm{max}}$ do
    Increase $\beta \to \beta + \Delta\beta$ in which the step size $\Delta\beta$ depends on the annealing schedule.
    Perform $N_\mathrm{S}$ Metropolis sweeps on each replica independently at inverse temperature $\beta$.
    for $n = 1$ to $R$, i.e., current population size do
        Copy the $n$'th replica with energy $E_n$, $r_n = \frac{R_0}{Q} e^{-\Delta\beta E_n}$ times, where $Q = \sum_{i=1}^{R} e^{-\Delta\beta E_n}$.
    end for
    Make measurements of the physical observables by simple averaging over the replicas.
end for
```

As shown in Ref. [61], the quality of thermalization of any thermodynamic observable is in direct correlation with the family entropy S_{f} and the entropic family size ρ_{s}. The systematic errors, on the other hand, are controlled by the equilibrium population size ρ_{f}. What we here refer to as “efficiency” or “speed-up” relates to reducing the statistical as well as the systematic errors, while keeping the computational effort constant. Thus, it would be reasonable to use these quantities as measures of optimality for various PAMC implementations. S_{f}, ρ_{s} and ρ_{f} are defined as

$$\begin{aligned} S_{\mathrm{f}} &= -\sum_i \nu_i \ln \nu_i, && (3.2)\\ \rho_{\mathrm{s}} &= \lim_{R\to\infty} R/e^{S_{\mathrm{f}}}, && (3.3)\\ \rho_{\mathrm{f}} &= \lim_{R\to\infty} R \times \mathrm{var}(\beta \mathrm{F}), && (3.4) \end{aligned}$$

where ν_i is the fraction of the population that have descended from replica i in the initial population, and β and F are the inverse temperature and the free energy of the system, respectively. The free energy is measured using the free energy perturbation method. Intuitively, $\exp(S_{\mathrm{f}})$ characterizes the number of surviving families, and therefore ρ_{s} shows the average surviving family size. For a set of simulation parameters, the larger ρ_{s} and ρ_{f}, or the smaller S_{f}, the computationally harder the instance. Keep in mind that ρ_{f} is computationally more expensive to measure because many independent runs (at least 10) are needed to measure the variance of the free energy. Note that S_{f} is “extensive” and asymptotically grows as $\log(R)$, while both ρ_{s} and ρ_{f} are “intensive” quantities, growing asymptotically independent of $R$for a sufficiently large R. In our simulations, these metrics are estimated using finite but large enough R values such that the systematic errors are negligible.

It can be shown [61, 67] that the systematic errors in any population annealing observable at the limit of large R are proportional to $\mathrm{var}(\beta F)$. Therefore, in order to ensure that the simulations are not affected by the systematic errors, one needs to make certain that the quantity ρ_{f}/R is sufficiently small. When well defined, ρ_{s} is strongly correlated with ρ_{f} [61], as it is the case for the majority of the spin-glass instances that we study in this chapter. Hence, we may alternatively minimize $\rho_{\rm s}/R$ or equivalently maximize $S_{\rm f}$ as a proxy for the quality of equilibration. In our simulations, we ensure that $S_{\rm f} \gtrsim 2$ for all the instances.

3.2.3 Outline of the Cluster Updates Used

Having outlined PAMC, we now briefly introduce the different cluster algorithms we have experimented with in order to speed up thermalization.

Wolff cluster algorithm: the Wolff algorithm [68] greatly speeds up simulations of Ising systems without frustration near the critical point. It is well known that the Wolff algorithm does not work well for spin glasses in 3D [75] because the cluster size grows too quickly with β. Nevertheless, we revisit this algorithm systematically in both 2D and 3D. The idea is that even if the cluster size grows too quickly when β is still relatively small, the mean cluster size (normalized by the number of spins N) is still a continuous function in the range $[0, 1]$, when β grows from $\beta = 0$ to $\beta = \infty$. Therefore, it is a reasonable question to ask if there would be some speed-up when restricting the algorithm to the temperature range where the normalized mean cluster size is neither too larger nor too small, for example, in the range $[0.1, 0.9]$.

In the ferromagnetic Ising model, where $J_{ij} = J = 1$, one adds a neighboring spin s_j when it is parallel to a spin s_i in the cluster with probability $p_{\rm c} = 1 - \exp(-2J\beta)$. In spin glasses, this is generalized as follows: One adds a neighboring spin s_j to s_i when the bond between the two spins is satisfied, and with probability $p_{\rm c} = 1 - \exp(-2|J_{ij}|\beta)$. This can be compactly written as $p_{\rm c} = \max[0, 1 - \exp(-2\beta J_{ij} s_i s_j)]$ [75]. Note that from the above $p_{\rm c}$ expression, there are two interesting limits in order for the mean cluster size . In the limit $\beta \to 0$, the average cluster size is clearly 0, and in the limit $\beta \to \infty$, the normalized cluster size tends to 1 because in the ground state, each spin has at least one satisfied bond with its neighbors, and all the spins would be added to the cluster. One can also see that frustration actually makes the cluster size grow *slower* as a function of β. However, frustration significantly reduces the transition temperature, which is the primary reason why the Wolff algorithm is less efficient for spin glasses. Finally, note that the Wolff algorithm is both ergodic and satisfies detailed balance.

Houdayer cluster algorithm: designed for spin glasses, the Houdayer cluster algorithm [69] or its generalization, the isoenergetic cluster moves (ICM) [70] greatly improves the sampling for parallel tempering in 2D, while less so in 3D. ICM in 3D, like the Wolff algorithm, is restricted to a temperature window where the method is most efficient [70]. ICM works by updating two replicas at the same time. First, an overlap between the two replicas is constructed, which naturally forms positive and negative islands. One island is selected, and the spin configurations of the island in both replicas are flipped.

In its original implementation, the spin-down sector is always used to construct the cluster. In the implementation of Zhu *et al.*, a full replica is flipped if the chosen island is in the positive sector to make it negative [70] and therefore reduce the size of the clusters. Here, we improve on this implementation by allowing the chosen island to be either positive or negative and flipping the spins of the island in both replicas. Therefore, we never flip a full replica. This saves computational time and also has the advantage that it does not artificially make the spin-overlap function symmetric. ICM satisfies detailed balance, but its is not ergodic. Therefore, the algorithm is usually combined with an ergodic method, such as the Metropolis algorithm. ICM greatly improves the thermalization time and also slightly the autocorrelation time in parallel tempering. Because PAMC is a sequential method, there is no thermalization stage. We, therefore, focus on whether the algorithm reduces correlations, i.e., systematic and statistical errors.

Our implementation of PAMC with ICM is as follows: First, after each resampling step, we do regular Monte Carlo sweeps and ICM updates alternately. We first do $N_{\rm S}/2$ lattice sweeps for each replica, followed by R ICM updates done by randomly pairing two replicas in the population, followed by another $N_{\rm S}/2$ lattice sweeps. Second, for each ICM update, we choose an island from the spin sector with the smaller number of spins. Then, the spin configurations of the island in both replicas are flipped. This effectively means that the spin configurations associated with the selected island are either exchanged or flipped depending on the sign of the island being negative in the former or positive in the latter case. Note that the combined energy of the two replicas is conserved in both cases, therefore making the algorithm rejection free.

Table 3.1: Simulation parameters for various experiments to optimize PAMC. Spin selection methods (SSM), annealing schedules (AS), number of temperatures tuning (NT), dynamic population size experiment (DPS), and cluster algorithms (CA). D is the space dimension, L is the linear system size, R is the population size, $T_{\mathrm{min}} = 1/\beta_{\mathrm{max}}$ is the lowest temperature simulated, N_T is the number of temperatures, and M is the number of disorder realizations studied. The label "Schedule" refers to the annealing schedule used, such as the linear-in-β (LB) or the linear-in-β linear-in-T (LBLT) schedules. $N_{\mathrm{S}} = 10$ sweeps are applied to each replica at each temperature. Note that in the case of dynamic population sizes (DPS), R is the mean population size. See the text for more details. Reprinted with permission from [52].

Technique	D	L	R	T_{min}	N_T	Schedule	M
SSM	3	4	5×10^4	0.2	101	LB	1000
SSM	3	6	2×10^5	0.2	101	LB	1000
SSM	3	8	5×10^5	0.2	201	LB	1000
SSM	3	10	1×10^6	0.2	301	LB	1000
AS	3	8	5×10^5	0.2	201	All	1000
AS	3	10	1×10^6	0.2	301	All	1000
NT	2	8	5×10^4	0.2	variable	LBLT	100
NT	2	16	2×10^5	0.2	variable	LBLT	100
NT	2	25	5×10^5	0.2	variable	LBLT	100
NT	2	32	1×10^6	0.2	variable	LBLT	100
NT	3	4	5×10^4	0.2	variable	LBLT	100
NT	3	6	2×10^5	0.2	variable	LBLT	100
NT	3	8	5×10^5	0.2	variable	LBLT	100
NT	3	10	1×10^6	0.2	variable	LBLT	100
DPS	3	6	2×10^5	0.2	101	LB	1000
DPS	3	8	5×10^5	0.2	201	LB	1000
DPS	3	10	1×10^6	0.2	301	LB	1000
CA	2	8	5×10^4	0.2	101	LB/LBLT	1000
CA	2	16	2×10^5	0.2	101	LB/LBLT	1000
CA	2	25	5×10^5	0.2	201	LB/LBLT	1000
CA	2	32	1×10^6	0.2	301	LB/LBLT	1000
CA	3	4	5×10^4	0.2	101	LB/LBLT	1000
CA	3	6	2×10^5	0.2	101	LB/LBLT	1000
CA	3	8	5×10^5	0.2	201	LB/LBLT	1000
CA	3	10	1×10^6	0.2	301	LB/LBLT	1000

3.3 Implementation Optimizations

In this section, we present our implementation improvement to the population annealing algorithm. We first present spin selection methods, followed by experiments using different annealing

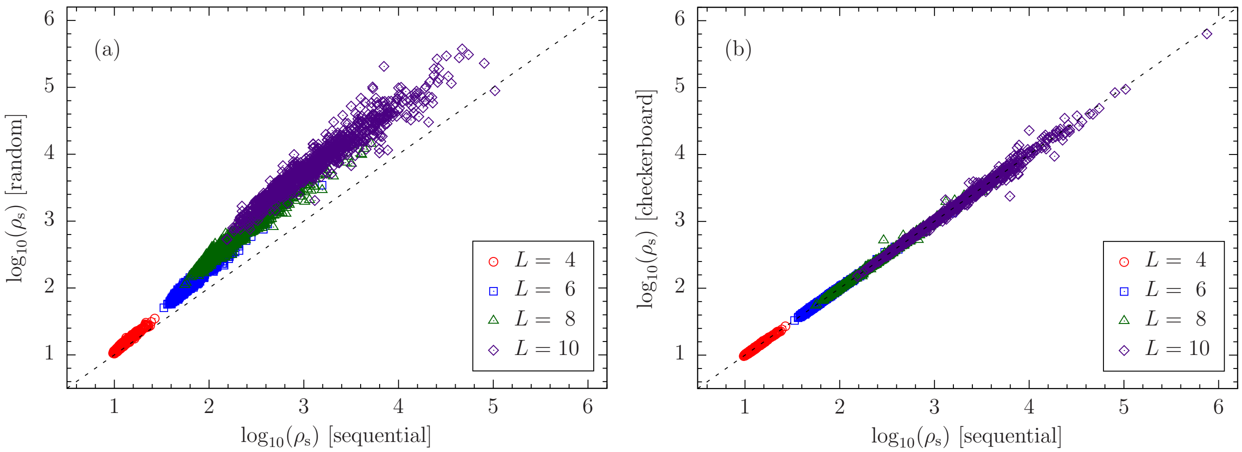

Figure 3.2: Comparison of the entropic population size ρ_s for different spin selection methods. Random, sequential and checkerboard updates in three space dimensions. Sequential and checkerboard updates have similar efficiency [panel (b)], and both are more efficient than random updates [panel (a)]. Reprinted with permission from [52].

schedules, numbers of temperatures, and the use of a dynamic population. The simulation parameters are summarized in Table 3.1.

3.3.1 Comparison of the Spin Selection Methods

We have studied three spin selection methods: sequential, random, and checkerboard. We have carried out large-scale simulations in 3D to compare these methods for $L = 4$, 6, 8, and 10, with 1000 instances for each system size. We first run the simulations using the parameters in Table. 3.1. To measure S_f or ρ_s reliably, we require $S_f \gtrsim 2$ [61]. When this is not satisfied for a particular instance, we rerun it with larger population size. We then compare ρ_s at the lowest temperature between different spin selection methods. Figure 3.2 shows scatter plots comparing ρ_s instance by instance for different system sizes and using different spin selections methods. Figure 3.2(a) compares random to sequential updates, whereas Fig. 3.2(b) compares checkerboard to sequential updates. Interestingly, sequential and checkerboard updates have similar efficiency (the data lie on the diagonal), whereas both sequential and checkerboard are more efficient than random updates. This is particularly visible for the larger system sizes, e.g., $L = 10$.

The random selection method is, therefore, the least efficient update technique for disordered Boolean problems, keeping in mind that it requires the computation of an additional random number for each attempted spin update, thus slowing down the simulation. We surmise that sequential updating of the spins accelerates the mobility of domain walls in most cases. However, in some pathological examples, such as the one-dimensional Ising chain, random updating is needed for Monte Carlo to be ergodic.

3.3.2 Optimization of the Annealing Schedules

Most early population annealing simulations used a simple linear-in-β (LB) schedule, where the change in β in the annealing schedule is constant as a function of the temperature index. This, however, is not necessarily the most optimal schedule to use. We use the following two approaches to optimize the annealing schedules and the number of temperatures: one approach uses a mathematical model with free parameters to be optimized, and the other includes adaptive schedules based on a guiding function, e.g., the energy fluctuations or the specific heat. For the parametric schedules, we introduce a linear-in-β linear-in-T (LBLT) and a two-stage power-law schedule (TSPL). For the LBLT schedule, there is one parameter to tune, namely, a tuning temperature T_N. In this schedule, half of the temperatures above T_N are linear in β, while the other half below T_N are linear in T. It is worth noting that optimizing $T_{\rm min}$ for various annealing schedules is not necessary because $\rho_{\rm s}$ is a monotonically-increasing function of temperature. Thus, a higher $T_{\rm min}$ with the same number of temperature steps trivially results in a better thermalization.

For the TSPL schedule, we define a rescaled annealing time $\tau = k_\beta/(N_T - 1) \in [0, 1]$, where $k_\beta = 0, \ldots, N_T - 1$ is the annealing step or temperature index. The TSPL schedule is modeled as

$$\beta(\tau) = a\tau^{\alpha_1}\theta(\tau_0 - \tau) + b\tau^{\alpha_2}\theta(\tau - \tau_0), \tag{3.5}$$

where θ is the Heaviside step-function. Here α_1 and α_2 are free parameters, while a and b enforce continuity and fix the final annealing temperature. In addition, τ_0 determines a switch-over temper-

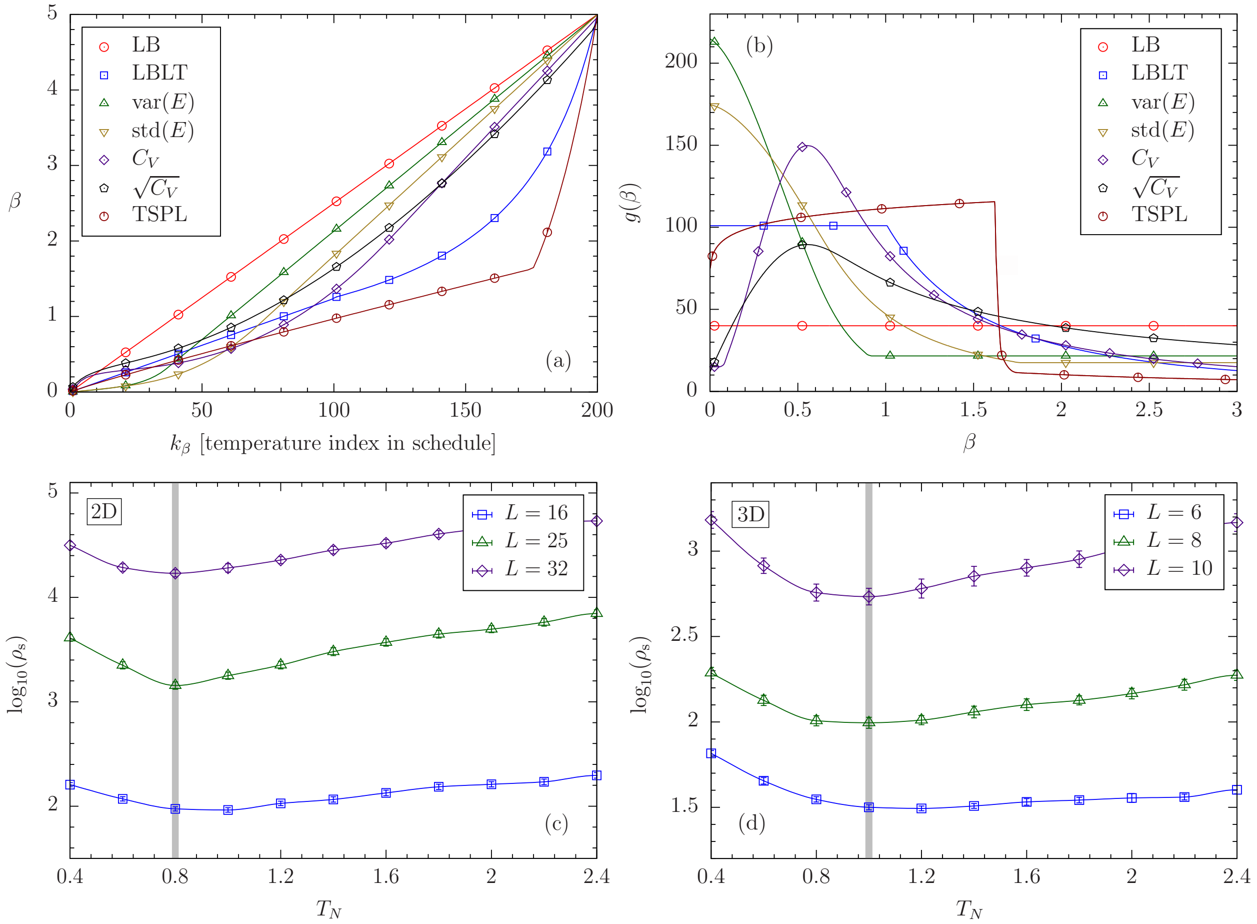

Figure 3.3: Annealing schedules experimented in two and three dimensions. Panel (a) shows the β values as a function of the inverse temperature index k_β for the different schedules, and panel (b) shows the resulting β-densities, $g(\beta)$ (the data is cut off at $\beta = 3$ for clarity). Note that both TSPL and LBLT schedules have more temperatures at high T. Panels (c) and (d) show $\rho_{\rm s}$ as a function of T_N for two-dimensional (2D) and three-dimensional (3D) simulations, respectively. The vertical shaded line marks the optimum. See the main text for details. Reprinted with permission from [52].

ature β_0. We optimize the LBLT schedule with a simple scan of the parameter T_N. The optimum value of T_N (where $\rho_{\rm s}$ is minimal) is shown in Fig. 3.3(c) for 2D ($T_N \approx 0.8$, marked with a vertical shaded area) and Fig. 3.3(d) for 3D ($T_N \approx 1.0$, marked with a vertical shaded area). The TSPL schedule, however, has more parameters that must be tuned. Therefore, we have used the Bayesian optimization package Spearmint [76, 77] rather than a full grid scan in the entire parameter space. We find numerically that the parameters $\alpha_1 = \exp(-0.0734)$, $\alpha_2 = \exp(2.15)$, and $\beta_0 = 1.63$

work well although there is no guarantee of global optimality. We optimize the adaptive schedules using information provided by energy fluctuations because energy is directly related to the resampling of the population. We, therefore, define a density of inverse temperature β, $g(\beta)$, and study the following adaptive schemes:

- var(E) schedule with $g(\beta) \sim \mathrm{var}(E)$,
- std(E) schedule with $g(\beta) \sim \sqrt{\mathrm{var(E)}}$,
- C_V schedule with $g(\beta) \sim C_V(\beta)$,
- $\sqrt{C_V}$ schedule with $g(\beta) \sim \sqrt{C_V(\beta)}$,

where C_V is the specific heat of the system. Note that the functions are disorder averaged, and the proportionality is determined by the number of temperatures. Because $g(\beta)$ may become extremely small, we have replaced all the function value less than 10% of $\max(g)$ with $0.1 \times \max(g)$ to prevent large temperature leaps. With this small modification, we generate N_T temperatures according to the above density functions. The shapes and β-densities of all schedules are shown in Figs. 3.3(a) and 3.3(b), respectively. There are clear distinctions between different schedules, especially in comparison to the traditionally-used LB schedule. We compare the efficiency of the above schedules in Fig. 3.4 by analyzing the systematic errors in a number of observables. We have studied the internal energy)E), free energy (F), and the spin-glass Binder cumulant (g_{SG}) [78] for the system size $L = 10$. To overcome the scale difference when showing the systematic errors of different observables in one plot, we have normalized the errors with respect to the schedule that has the greatest error. Therefore, all the errors will be relative to that of the worst schedule. In panel (a) of Fig. 3.4, we show the normalized systematic errors for a randomly-chosen and extremely hard instance. In panel (b), we show the disorder averaged systematic errors calculated from 100 of the hardest instances. It can be readily seen from the plots that the LBLT and TSPL schedules yield the best efficiencies among all of the experimented schedules, with TSPL slightly more efficient. Both LBLT and TSPL schedules place more temperatures on the high-temperature

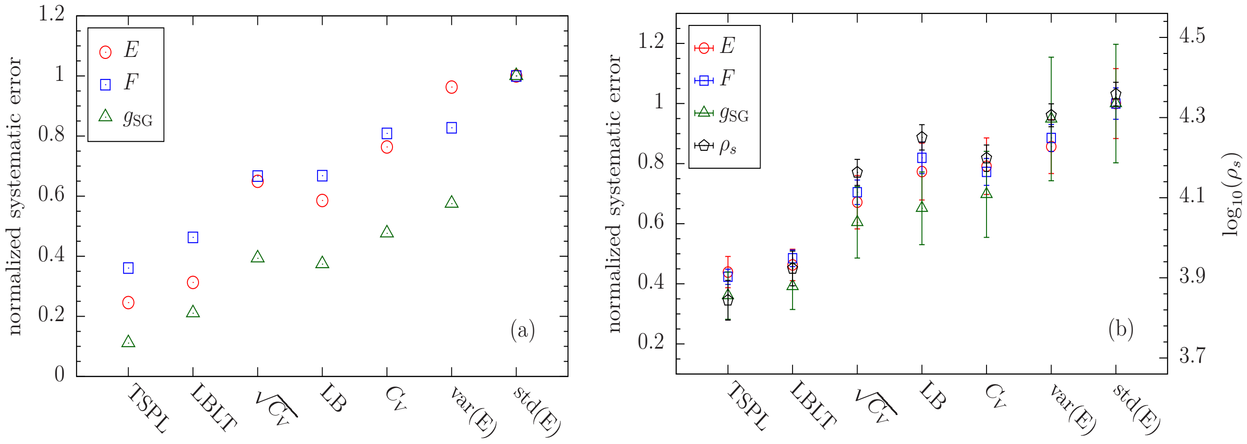

Figure 3.4: Comparison of the systematic errors for various annealing schedules. The studied observables are energy (E), free energy (F) and the spin-glass Binder cumulant ($g_{\rm SG}$) for the system size $L = 10$. Panel (a) shows the systemic errors fora randomly chosen hard instance, whereas panel (b) illustrates the systematic errors averaged over 100 of the hardest instances. Systematic errors of different observable often have magnitudes largely apart. For this reason, the errors in each observable have been normalized relative to the maximum error across all schedules. For instance, in the top panel, the $\mathrm{std}(E)$ schedule that has the greatest systematic error is normalized to 1, while the rest of the schedules lie below 1. It is seen from the plots that the TSPL schedule is the most efficient. The LBLT schedule, although conveniently simple, competes well with the optimal schedule. Note that we also show ρ_s (as a dual y-axis) in panel (b). We observe that ρ_s greatly correlates with the systematic errors justifying the use of it as an effective optimization criterion. Reprinted with permission from [52].

side (smaller β values), presumably because the Metropolis dynamics are more effective at high temperatures. Additionally, in panel (b) of Fig. 3.4, we have shown ρ_s for various schedules. We observe a great correlation between ρ_s and the systematic errors, which corroborates the use of ρ_s as a good measure of efficiency.

We stress that the optimal schedule depends on the choice of the number of sweeps at each anneal step, $N_{\rm S}$, because N_T and $N_{\rm S}$ are exchangeable for large enough N_T. In our approach, we have fixed $N_{\rm S}$. It is, therefore, possible that other techniques may result in different optimal schedules. For instance, one may use the energy distribution overlaps at two temperatures to define the optimal schedule [67, 71] that which depends only on the thermodynamic properties of the system. As an example, in Fig. 3.5, we show the energy distributions of the LBLT schedule

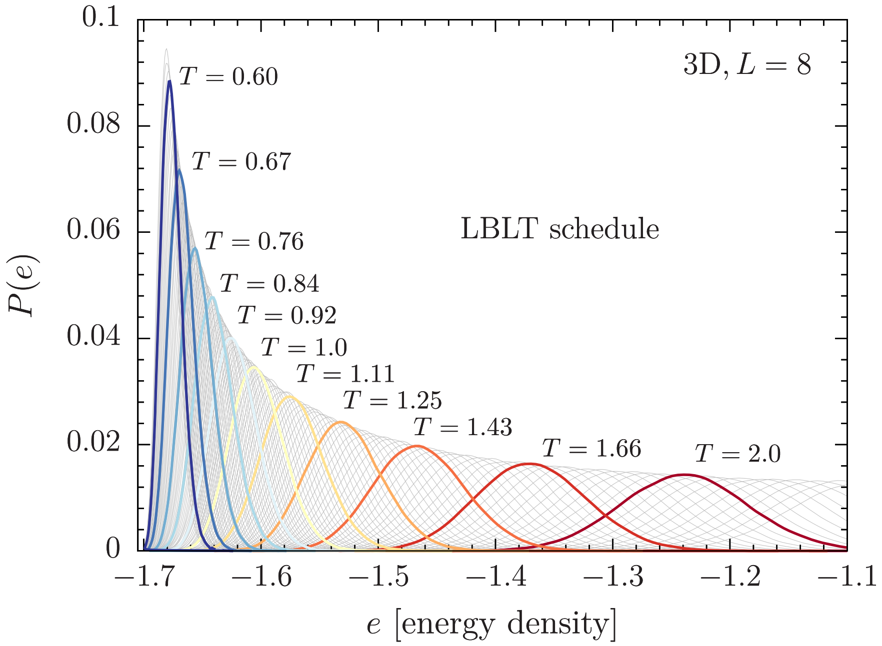

Figure 3.5: Energy density distribution of the LBLT annealing schedule for $L = 8$ in three space dimensions. Thinner curves show the histograms at all temperatures, whereas the thicker ones are drawn at every 10 temperature steps. There are 200 temperature steps in total. The histograms overlap considerably. Reprinted with permission from [52].

for $L = 8$ in 3D. The energy histograms overlap considerably up to several temperature steps. Within this framework, the optimization is transferred to the distribution of sweeps. However, the density of work (the product of the density of β and density of sweeps) should be similar in both approaches. In our implementation, as the number of sweeps is constant, the density of work is the same as the density of β.

3.3.3 Optimization of the Number of Temperatures

To optimize the number of temperatures and their range, we use the LBLT schedule as it is easy to implement and very close to optimal. Our figure of merit is to maximize the number of independent measurements R/ρ_{s} for constant work $W = RN_{\mathrm{S}}N_T$. We define efficiency as $\gamma = R/(\rho_{\mathrm{s}}W)$ by tuning N_T for a constant W. Because $N_{\mathrm{S}} = 10$ is fixed, we need to maximize $1/(\rho_{\mathrm{s}}N_T)$ by tuning N_T. In the limit $R \to \infty$, ρ_{s} and the efficiency γ are independent of the

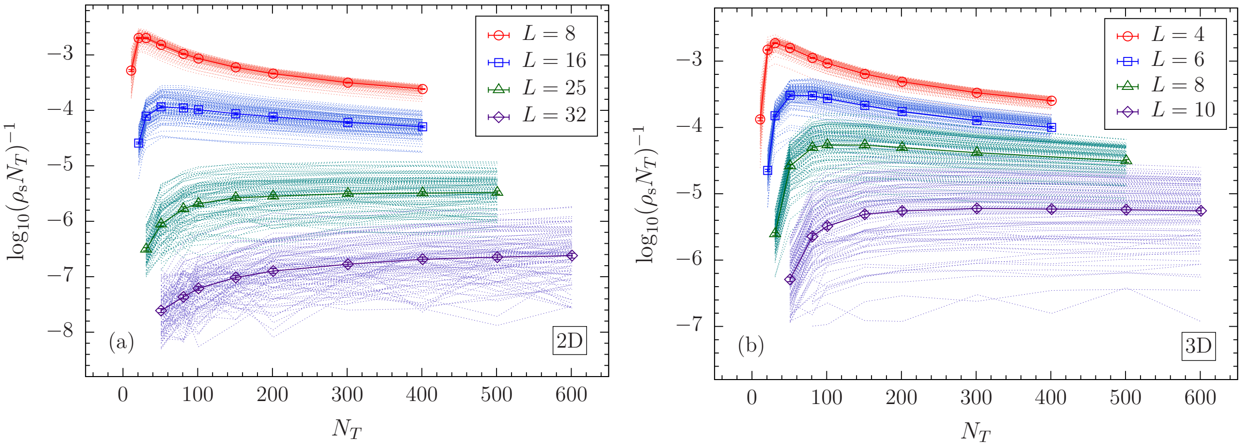

Figure 3.6: Optimization of the number of annealing steps N_T in two space dimensions [2D, panel (a)] and three space dimensions [3D, panel (b)]. To maximize sampling efficiency, one needs to optimize $1/(\rho_{\mathrm{s}} N_T)$ with respect to N_T. In both panels, the points and the solid curves show the disorder average, while the dashed envelopes display all 100 studied instances. For smaller system sizes, the peak (optimum) is sharp, whereas for systems with more than approximately 1000 spins, the peak is broadened, especially in two dimensions. The reason for this broadening can be understood by noticing the increase in the density of chaotic samples as the system grows in size (wiggly lines). Reprinted with permission from [52].

population size. This is expected as γ is an intensive quantity. Therefore, to measure γ, we only need to make sure that R is sufficiently large, such that ρ_{s} has converged. It is not necessary to use the same W for different N_T.

The results for both two and three-dimensional systems are shown in Figs. 3.6(a) and 3.6(b), respectively. The solid curves show the disorder average, while the dashed envelopes are the instance-by-instance results. It is interesting to note that for relatively smaller system sizes, we observe a pronounced peak. The existence of an optimal number of temperatures can be intuitively understood in the following way: for a fixed amount of computational effort, if N_T is too small, the annealing or resampling would become too stochastic, which is inefficient. On the other hand, if the annealing is too slow (N_T is too large), keeping a larger population size is more efficient. Therefore, optimality comes from a careful balance between N_T and R. As the system size grows, the optimal peak starts to flatten out due to the onset of temperature chaos[79–85]. This can be seen in Fig. 3.6 as a discernible increase in the density of instances with irregular oscillatory behavior. Thus, we conclude that the optimization presented here, although capturing the bulk of the instances, might not be reliable in the case of extremely hard (chaotic) instances. Instead, one may consider performing more Metropolis sweeps rather than merely increasing the temperature steps or the population size. This is especially relevant if memory (which correlates with R) becomes a concern for the hardest instances.

3.3.4 Dynamic Population Sizes

The reason the LBLT schedule is more efficient than a simple LB schedule is that the Metropolis dynamics are less effective at low temperatures, and therefore using more "hotter" temperatures is more efficient. Here, we investigate another technique, namely, a variable number of replicas that depends on the annealing temperature, thus having a similar effect to more temperatures at higher values. Regular PAMC is designed to have an approximately uniform population size as a function of temperature. Here, we allow the population size to change with β. Because most families are removed at a relatively early stage of the anneal, transferring some replicas from low temperatures to high temperatures may increase the diversity of the final population even though the final population size would be smaller. Note that the uniform population size is a special case of this generalized population schedule. We study a simple clipped exponential population schedule in which the population starts at a constant value R_0 until $\beta = \beta_0$, and then decreases exponentially to $R_\mathrm{f} = rR_0$ at $\beta = \beta_\mathrm{max}$:

$$R(\beta) = \begin{cases} R_0 & \beta \leq \beta_0, \\ R_0 \frac{e^{a\beta_0}+b}{e^{a\beta}+b} & \beta > \beta_0, \end{cases} \tag{3.6}$$

where $b = (re^{a\beta_\mathrm{max}} - e^{a\beta_0})/(1-r)$. The free parameters to tune are a, β_0 and r in which r is the ratio of the final to initial population size, and a characterizes the slope of the curve. Once the parameters are optimized, we can scale the full function to have a comparable average population size to that of the uniform schedule. Again, the optimization is done using Bayesian statistics, and we obtain $\beta_0 = 0.9$, $r = 0.03$, and $a = 0.08$. It is noteworthy to mention that there are two different measures to detect efficiency, when the population size is allowed to change. For the same average

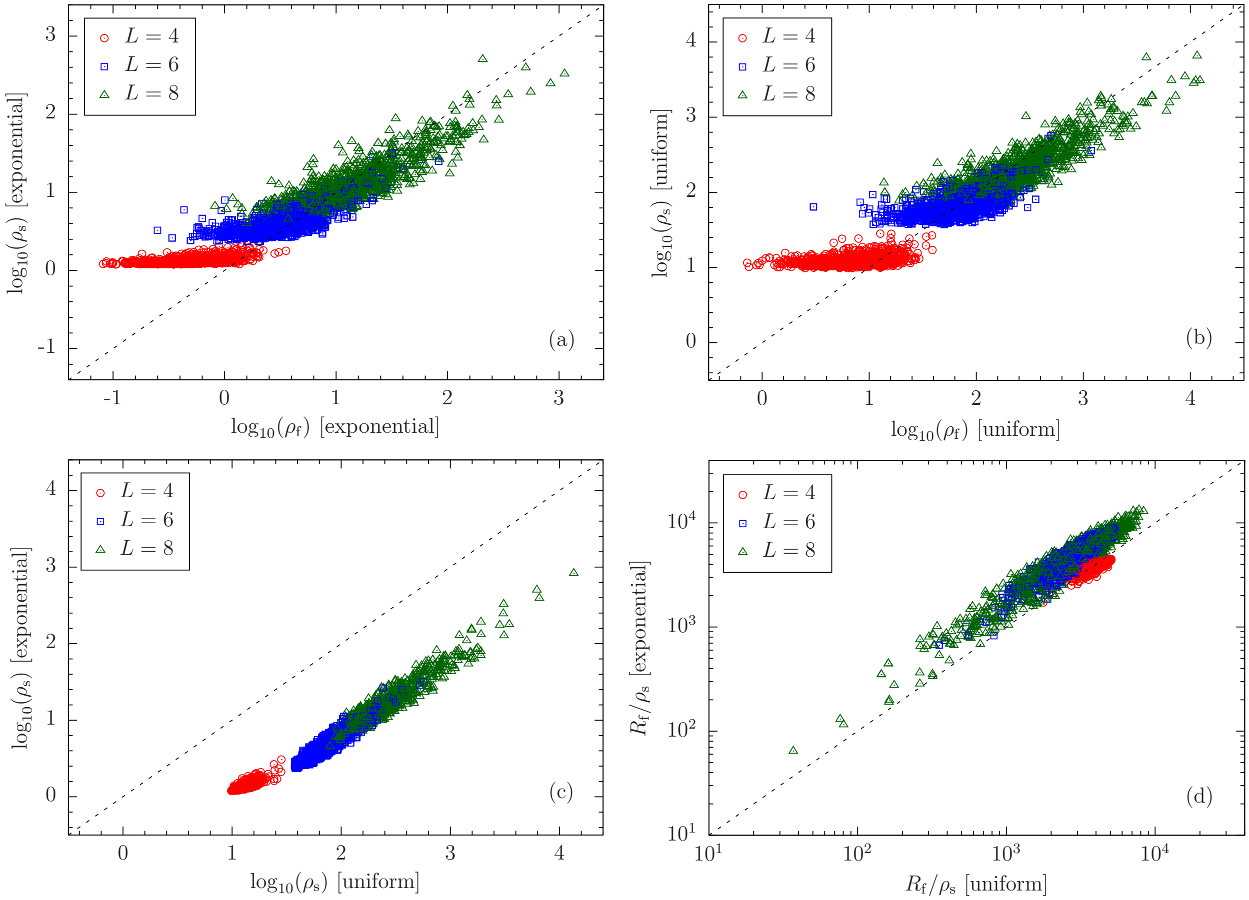

Figure 3.7: Instance-by-instance comparison for a PAMC simulation with fixed and dynamic population sizes. With a dynamic population size, ρ_s and ρ_f are well-correlated, similar to the case of uniform population. ρ_s is greatly reduced, suggesting that the simulation is much better at the level of averaging over all temperatures. The dynamic population size is also more efficient than the uniform one using the worst-case measure. Here, R_f is the final population size. Reprinted with permission from [52].

population size, the dynamic population schedule is *always* better at high temperatures. However, at low temperature, a smaller ρ_s does not justify that the number of independent measurements is larger because R is also smaller. Thus, It is reasonable to optimize the parameters using ρ_s, and then also compare to R/ρ_s. Note that we use the local population size R at each temperature to compute ρ_s. The correlations and comparisons between ρ_s and ρ_f are also studied. With the optimal parameters, we compare the efficiency of the dynamic and uniform population sizes. The results are shown in Fig. 3.7. We see that ρ_s and ρ_f are well-correlated for the dynamic population size.

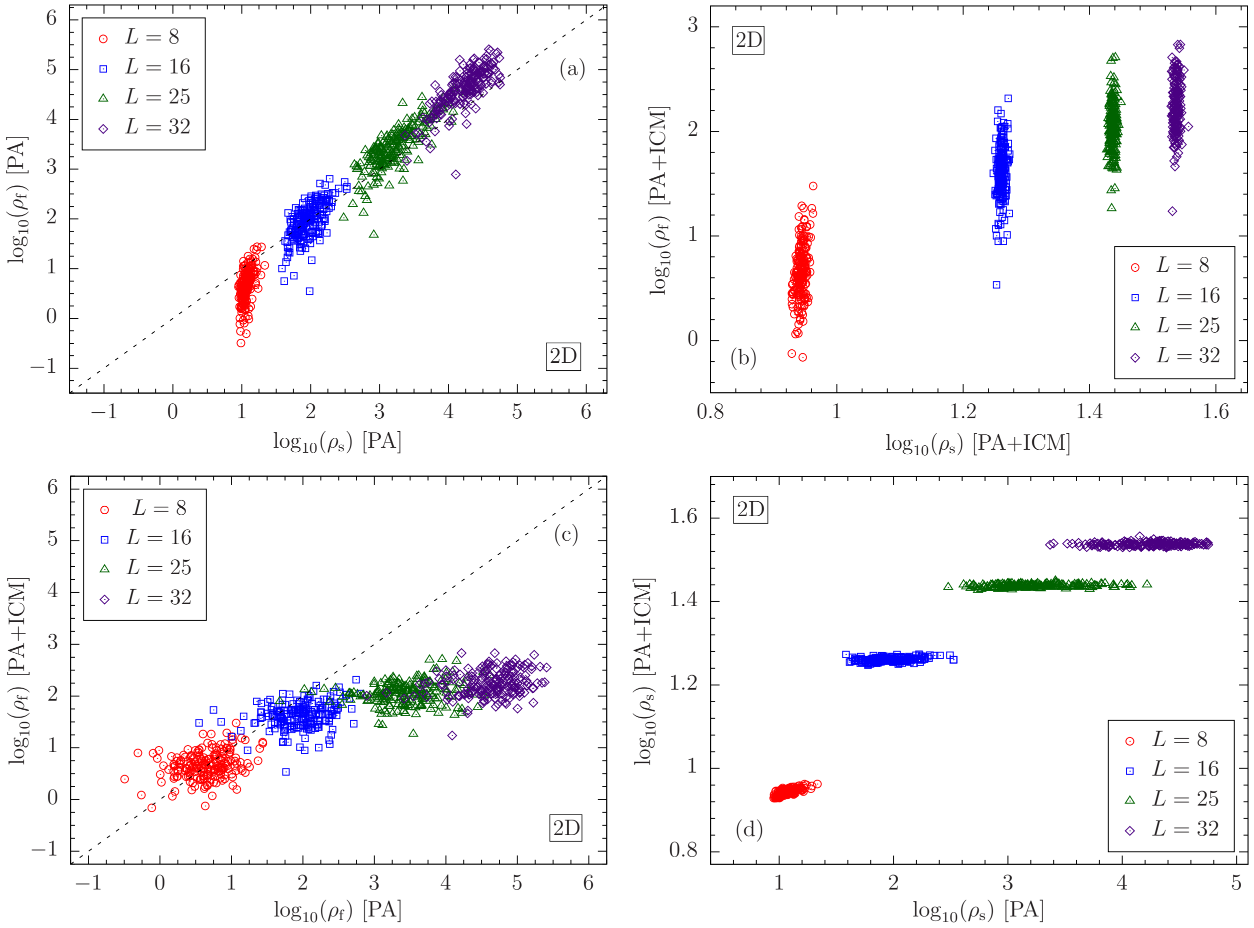

Figure 3.8: Population annealing with ICM updates in 2D. Note that replica family is not well defined when ICM updates are included. Therefore, we use ρ_f to characterize speed-up. Significant speed-up is observed in 2D. Reprinted with permission from [52].

ρ_s is greatly reduced, suggesting that the simulation is much better at the level of averaging over all temperatures. We also see that even using the worst-case measure, the dynamic population size is more efficient than the uniform one. Note, however, that the peak memory use of the dynamic population size is larger due to the non-uniformity of the number of replicas as a function of β.

3.4 Algorithmic Accelerators

We now turn our attention to algorithmic accelerators by including cluster updates in the simulations. The simulation parameters are summarized in Table 3.1.

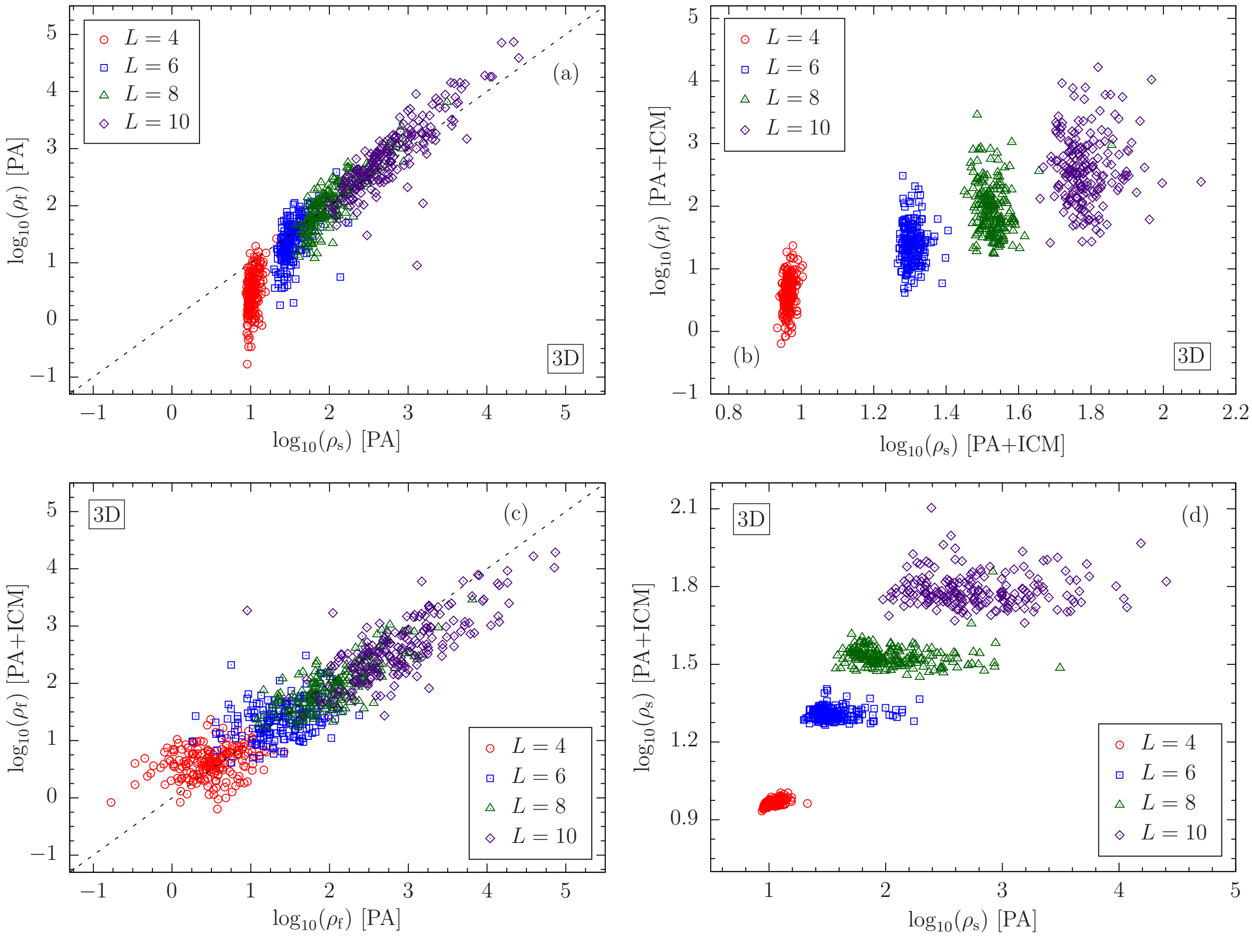

Figure 3.9: Population annealing with ICM updates in 3D. Note that replica family is not well defined when ICM updates are included. Therefore, we use ρ_f to characterize speed-up. Modest speed-up is observed in 3D. Reprinted with permission from [52].

3.4.1 Isoenergetic Cluster Updates

Here, we study PAMC with ICM updates. In 3D, similar to the Wolff algorithm, there is an effective temperature range where ICM (see Ref. [70] for more details) is efficient. In ICM, two replicas are updated simultaneously. This process uses the detailed structure of the two replica configurations, so it is natural to question if the family of a replica is still well defined. For example, occasionally, two replicas may merely exchange their configurations. This is equivalent to exchanging their family names, which potentially increases the diversity of the population *at little cost*. To resolve and investigate this issue, we have, therefore, measured the computationally more expensive equilibration population size ρ_{f} as well, which unlike ρ_{s}, does not depend on the definition of the families. Our results are shown in Fig. 3.8 and Fig. 3.9 for 2D and 3D, respectively. We find that ρ_{s} is, indeed, artificially reduced by the cluster updates. In both 2D and 3D, ρ_{f} has a wide distribution, while ρ_{s} is almost identical for all instances. Furthermore, ρ_{s} and ρ_{f} are strongly correlated for regular PAMC, but the correlation is poor when ICM is turned on. Therefore, we conclude that ρ_{s} is no longer a good equilibration metric for PAMC when combined with ICM. Using ρ_{f}, we find that similar to PT [70], there is a clear speed-up in 2D. In 3D, however, the speed-up becomes marginal again. This is in contrast to the discernible speed-up for PT with the inclusion of ICM in 3D. The results suggest that ICM is most efficient in 2D and likely quasi-2D lattices, reducing both thermalization times (PT) and correlations (PAMC and PT). In 3D, ICM merely reduces thermalization times, while marginally influencing correlations.

3.4.2 Wolff Cluster Updates

Wolff cluster updates are not effective in spin-glass simulations. We, nevertheless, have revisited this type of cluster update in the context of PAMC for the sake of completeness. For the Wolff algorithm, we first measure the mean cluster size per spin, as shown in panels (a) and (c) of Fig. 3.10 in 2D and 3D cases, respectively. Note the smooth transition of the mean cluster size from 0 to 1. We identify a temperature range where the mean cluster size is in the window $[0.1, 0.9]$. We have also experimented with other ranges such as $[0.2, 0.8]$ or $[0.3, 0.7]$. They consistently gave a similar or slightly worse speed-up. We perform Wolff updates in this temperature range, i.e., we perform 10 Wolff updates in addition to the 10 regular Metropolis lattice sweeps for each replica. The comparison of ρ_{s} with regular PAMC is shown on the panels (b) and (d) of Fig. 3.10. While the Wolff algorithm speeds up ferromagnetic Ising model simulations in 2D, the speed-up is marginal for 2D spin glasses because of the zero-temperature phase transition.

In 3D, the Gaussian spin glass has a phase transition near $T_{\mathrm{c}} \approx 0.96$, but the temperature window where the Wolff algorithm is effective is much higher than T_{c}. The speed-up is, therefore, almost entirely eliminated, presumably because the Metropolis algorithm is already sufficient for

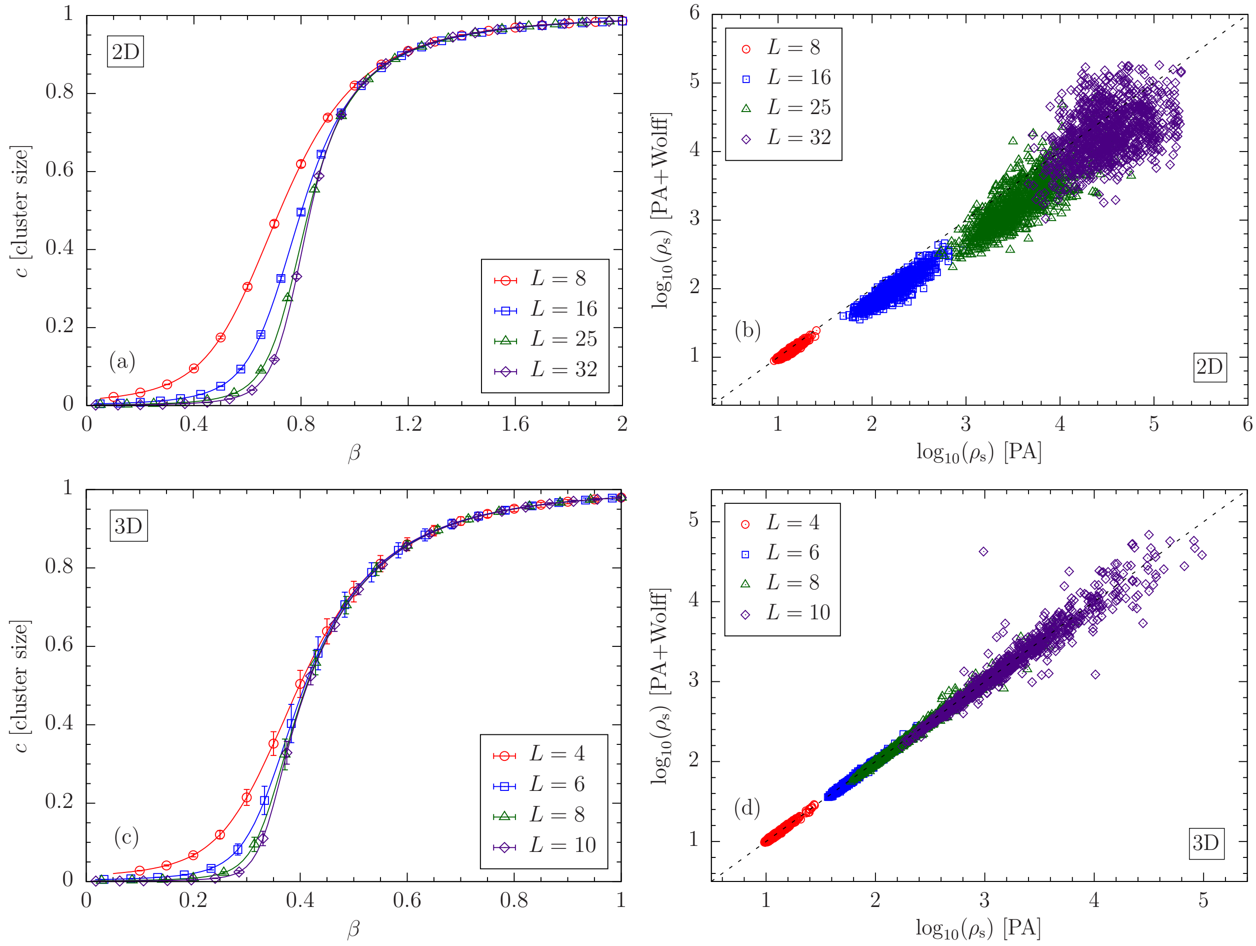

Figure 3.10: Mean normalized cluster size as a function of β for the Wolff algorithm [panels (a) and (c)] as well as the performance of the algorithm in both 2D and 3D. There is marginal speed-up in 2D [panel (b)] and no discernible speed-up in 3D [panel (d)]. Reprinted with permission from [52].

these high temperatures. The fact that the Wolff algorithm is more efficient in 2D than 3D is because clusters percolate faster in 3D, again rendering the effective temperature range higher in 3D. Therefore, Wolff updates constitute unnecessary overhead in the simulations of spin glasses in conjunction with PAMC. Even though PAMC with the Wolff algorithm does not appear to work very well for spin glasses, this does not mean they cannot be used together. For example, in two-dimensional spin glasses, adding the Wolff algorithm still has marginal benefits. The combination of PAMC and the Wolff cluster updates can be used for ferromagnetic Ising models for the purpose of parallel computing because parallelizing the Wolff algorithm, while doable, is challenging. In population annealing, however, this can be easily parallelized at the level of replicas and not within the Wolff algorithm itself.

3.5 Parallel Implementation

Population annealing is especially well suited for parallel computing because operations on the replicas can be carried out independently, and communication is minimal. Since OpenMP is a shared-memory parallelization library, it is limited to the resources available on a *single node* of a high-performance computing system. Although modern compute nodes have many cores and large amounts of RAM, these are considerably smaller than the number of available nodes by often several orders of magnitude. To benefit from machines with multiple compute nodes and therefore simulate larger problem sizes, we now present an MPI implementation of PAMC which can utilize resources up to the size of the cluster. While for typical problem sizes single-node OpenMP implementations might suffice for the bulk of the studied instances, hard-to-thermalize instances could then be simulated using a massively-parallel MPI implementation with extremely large population sizes. Although the exact run-time depends on many variables such as the simulation parameters, architecture, code optimality, compiler, etc., here we show some example of typical simulation time with the parameters listed in Table 3.1. On a 20-core node with Intel® Xeon® E5-2670 v2 2.50GHz processors, it takes approximately 1.3, 12, and 75 minutes to simulate an instance in 3D with $N = 216$, 512, and 1000 spins, respectively.

3.5.1 Massively Parallel MPI Implementation

The performance and scaling of our MPI implementation for 3D Edwards-Anderson spin glasses is shown in Fig. 3.11. Note that the wall time scales as $1/N$ with N the number of cores for less than 1000 cores. In our implementation, the population is partitioned equally between MPI processes (ranks). Each rank is assigned an index k with I/O operations occurring on the 0'th rank. A rank has a local population on which the Monte Carlo sweeps and resampling are carried out.

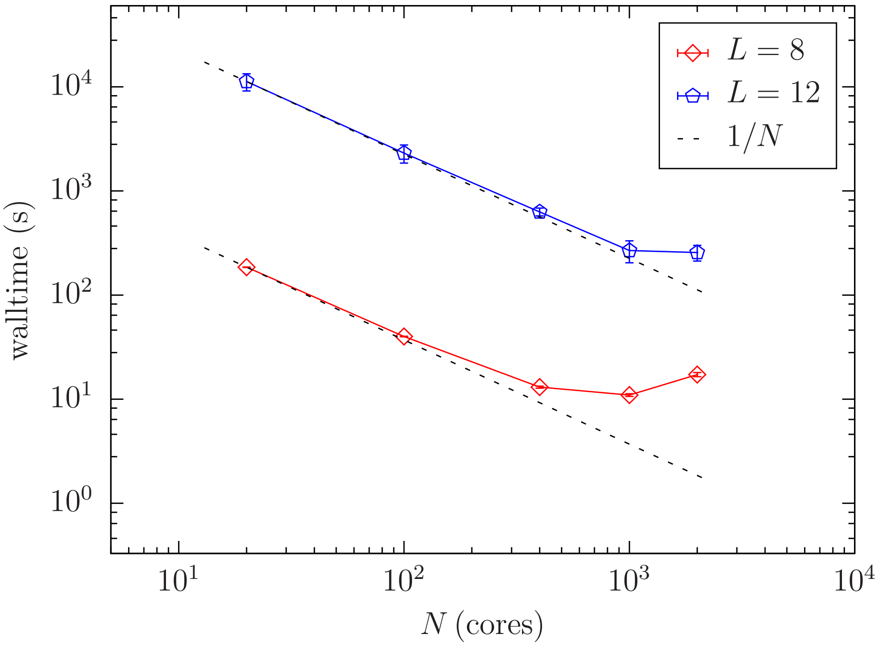

Figure 3.11: Scaling of the total wall time as a function of the number of processors N for two system sizes $L = 8$ and $L = 12$. Launching and initialization times are not included. Note that the efficiency becomes better for larger and harder problems. For $L = 12$, the scaling remains $1/N$ up to about 1000 processors. The efficiency then decreases when the time for collecting observables becomes dominant. Note that resampling still takes a relatively short time. Reprinted with permission from [52].

We also define a global index G which is the index of a replica as if it were in a single continuous array. In practice, the global index G of a replica j on a rank k is computed as the sum of the local populations r_i on the preceding ranks plus the local index j, i.e.,

$$G = j + \sum_{i=0}^{k-1} r_i. \tag{3.7}$$

The global index for a particular replica varies as its position in the global population changes.

Load balancing is carried out when a threshold percentage between the minimum and maximum local populations is exceeded. In our implementation, all members of a family must be in a continuous range of global indices to allow for efficient computation of the family entropy and the overlap function of the replicas. Therefore, load balancing must maintain adjacency. The destination rank k of a replica is determined by evenly partitioning the global population such that each rank has approximately the same number of replicas, i.e.,

$$k = \lfloor G/\left(\tfrac{R}{N}\right)\rfloor, \tag{3.8}$$

where N is the number of ranks (cores). Measurement of most observables is typically an efficient accumulation operation, i.e.,

$$\langle \mathcal{A} \rangle = \frac{1}{R} \sum_{k}^{N} \sum_{j}^{r_k} \mathcal{A}_{j,k}. \tag{3.9}$$

On the other hand, measuring observables such as the spin-glass overlap is more difficult and only done at select temperatures. Sets of replicas are randomly sampled from a rank's local population, and copies are sent to the range of ranks $[(k + N/4) \bmod N, (k + 3N/4) \bmod N]$ with periodic boundary conditions to ensure that the overlap is not computed between correlated replicas. The resulting histograms are merged in an accumulation operation similar to regular observables. Improving scaling with process count will require a lower overhead implementation of the spin-overlap measurements–a problem we intend to tackle in the near future.

3.6 Conclusions and Future Challenges

We have investigated various ways to optimize PAMC, ranging from optimizations in the implementation to the addition of accelerators, as well as massively parallel implementations. Many of these optimizations lead to often considerable speed-ups. We do emphasize that these approaches and even the ones that showed only marginal performance improvements for spin glasses in 2D and 3D might used to simulate statistical physics problems, potentially generating sizable performance boosts. The reduction in thermal error studied in this work can most directly be applied to the study of spin glasses by providing more CPU time for disorder averaging.

For the study of spin glasses, our results show that the best performance for PAMC is obtained by selecting the spins in a fixed order, i.e., sequentially or from a checkerboard pattern. Similarly, LBLT and TSPL schedules yield the best performance with LBLT having the least parameters to tune, and thus easier to implement. The number of temperatures needed for annealing is remarkably robust for large system sizes. Hence, in order to tackle hard instances, it is often convenient to increase the number of sweeps rather than merely using more temperatures. Dynamic population sizes are desirable albeit at the cost of a larger memory footprint. However, this can be easily mitigated via massively-parallel MPI implementations. In conjunction with Ref. [67], this study represents the first analysis of PAMC from an implementation point of view.

It has been shown [86] that the equilibration population size ρ_{f} can be measured in a single run using a blocking method. It would be interesting to investigate and test this idea thoroughly in the future. With an optimized PAMC implementation, it would be interesting also to perform large-scale spin-glass simulations to answer some of the unresolved problems in the field. In particular, in the next chapter, we will study the phase diagram of the so-called "Coulomb glass" model, where our developed PAMC algorithm will be crucial in allowing us to access extremely low temperatures.

4. NUMERICAL OBSERVATION OF A GLASSY PHASE IN THE THREE-DIMENSIONAL COULOMB GLASS[†]

4.1 Introduction

The existence of disorder in strongly interacting electron systems–which can be realized by introducing random impurities within the material, e.g., a strongly doped semiconductor–plays a significant role in understanding transport phenomena in imperfect materials and bad metals, as well as in condensed matter in general. When the density of impurities is sufficiently large, electrons become localized via the Anderson localization mechanism [87], and the long-range Coulomb interactions are no longer screened. This, in turn, leads to the depletion of the single-particle density of states (DOS) near the Fermi level, as first proposed by Pollak [88] and Srinivasan [89], thus forming a pseudogap. Later, Efros and Shklovskii [90] (ES) solidified this observation by describing the mechanisms involved in the formation of this pseudogap. The ES theory explains how the hopping (DC conductivity) within a disordered insulating material is modified in the presence of a pseudogap, also referred to as the "Coulomb gap". Numerous analytic studies have predicted [91–100] as well as experimental studies observed [101–115] the emergence of glassy properties in such disordered insulators, leading to the so-called "Coulomb glass" (CG) phase. Experimentally, to date, none of the aforementioned studies have observed a true thermodynamic transition into a glass phase, but rather have found evidence of nonequilibrium glassy dynamics, i.e., dynamic phenomena that are suggestive of a glass phase, such as slow relaxation, aging, memory effects, and alterations in the noise characteristics. Theoretically, more recent seminal mean-field studies by Pankov and Dobrosavljević [98] as well as Müller and Pankov [116] have shown that there exists a marginally stable glass phase within the CG model, whose transition temperature $T_{\rm c}$ decreases as $T_{\rm c} \sim W^{-1/2}$ for large enough disorder strength W, and is closely related to the formation of the Coulomb gap. Whether the results of the mean-field approach can be readily generalized to lower space dimensions is still uncertain. However, as we show in this work, the mean-field re-

---

[†] Reprinted with permission from [66].

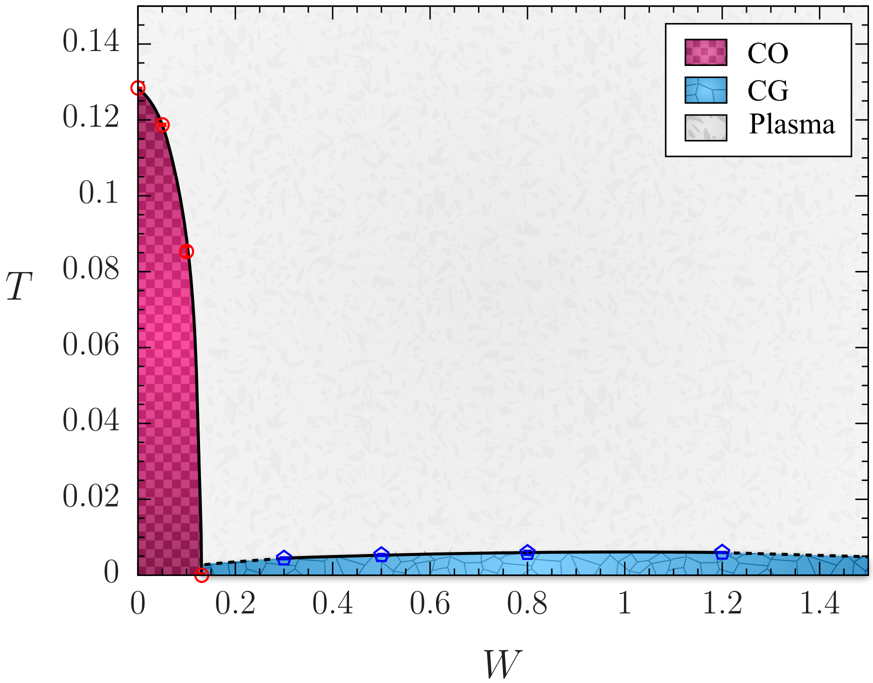

Figure 4.1: Phase diagram of the three-dimensional Coulomb glass model. There is a charge-ordered (CO) phase for $W \lesssim 0.131$, where electrons and holes form a checkerboard-like crystal. For $W \gtrsim 0.131$ the system undergoes a glassy transition into the Coulomb glass (CG) phase, albeit at considerably lower temperatures than in the CO phase. The dashed lines indicate extrapolations where numerical simulations are not available. Reprinted with permission from [66].

sults of Ref. [98] quantitatively agree with our numerical simulations in the charge-ordered regime (see Fig. 4.1) with similar values for the critical disorder $W_{\rm c}$ where the charge-ordered phase is suppressed. The critical temperatures $T_{\rm c}$ for the glassy phase, on the other hand, are substantially smaller than in the mean-field predictions. This, in turn, suggests that the mean-field approach of Ref. [98] includes the fluctuations of the uniform charge order collective modes, but not of the glassy collective modes. There have been multiple numerical studies that attempt to understand the DOS as well as the nature of the transitions of the CG model. In fact, there has even been some slight disagreement as to what the theoretical model to simulate should be, with some arguing for lattice disorder to introduce randomness into the model [117, 118], and others suggesting that the disorder should be introduced via random biases. Numerically, a Coulomb gap in agreement with the ES theory, has been observed in multiple studies. However, there is no consensus in the vast numerical work [117, 119–138] on the existence of a thermodynamic transition into a glassy phase. Nonequilibrium approaches suggest the existence of glassy behavior; however, thermodynamic simulations have failed to detect a clear transition.

In this work we investigate the phase diagram of the CG model using Monte Carlo simulations in three spatial dimensions. For the finite-temperature simulations, we make use of the population annealing Monte Carlo (PAMC) algorithm [52, 58, 60, 61, 67], which enables us to thermalize for a broad range of disorder values down to unprecedented low temperatures previously inaccessible. In addition, we argue that the detection of a glass phase requires a four-replica correlation length, as commonly used in spin-glass simulations in a field [139, 140]. Our main result is shown in Fig. 4.1. Consistently with previous numerical and analytical studies [98, 133, 141], we find a charge-ordered (CO) phase for disorders lower than $W_{\rm c} = 0.131(2)$, where electrons and holes form a checkerboard-like crystal. This is in close analogy with the classical Wigner crystal [142] that happens at low electron densities, where the potential energy dominates the kinetic energy, resulting in an ordered arrangement of the charges. It should, however, be noted that at $W = 0$ the lattice model, unlike in the continuum case, is not a standard Wigner crystal [143] because the system exhibits a pseudogap in the excitation spectrum (unrelated to the Coulomb gap) prior to entering the charge-ordered phase. For disorders larger than $W_{\rm c}$, we find strong evidence of a thermodynamic glassy phase restricted to temperatures that are approximately one order of magnitude smaller compared to the CO temperature scales. This, in turn, suggests that a thermodynamic glassy phase can, indeed, exist in experimental systems, where typically off-equilibrium measurements are performed. It also resolves the long-standing controversy where numerical simulations were unable to conclusively detect a thermodynamic glassy phase, while mean-field theory predicted such a phase. We note that for the disorder strength values studied, we are unable to discern a monotonic decrease in the critical temperature, as suggested by mean-field theory.

This chapter is structured as follows. In Sec. 4.2, we introduce the CG model, followed by the details of the simulation in Sec. 4.3. Section 4.5 is dedicated to the results of the study. Concluding remarks are presented in Sec. 4.6.

4.2 Model

The CG model in three spatial dimensions is described by the following Hamiltonian:

$$\mathcal{H} = \frac{e^2}{2\kappa} \sum_{i \neq j} (n_i - \nu) \frac{1}{|\mathbf{r}_{ij}|} (n_j - \nu) + \sum_i n_i \phi_i, \tag{4.1}$$

where $\kappa = 4\pi\epsilon_0$, $n_i \in \{0, 1\}$, and ν is the filling factor. The disorder ϕ_i is an on-site Gaussian random potential, i.e., $\mathcal{P}(\phi_i) = (2\pi W^2)^{-1/2} \exp(-\phi_i^2/2W^2)$. At half filling ($\nu = 1/2$), the CG model can conveniently be mapped to a long-range spin model via $s_i = (2n_i - 1)$. The Hamiltonian can be made dimensionless by choosing the units such that $e^2/\kappa = 1$ and $a = 1$ in which a is the lattice spacing. We thus simulate

$$\mathcal{H} = \frac{1}{8} \sum_{i \neq j} \frac{s_i s_j}{|\mathbf{r}_{ij}|} + \frac{1}{2} \sum_i s_i \phi_i, \tag{4.2}$$

where $s_i \in \{\pm 1\}$ represent Ising spins.

4.3 Simulation Details

In order to reduce the finite-size effects, we use periodic boundary conditions. Special care has to be taken to deal with the long-range interactions. We make infinitely many periodic copies of each spin in all spatial directions such that each spin interacts with all other spins infinitely many times. We use the Ewald summation technique [144, 145] so the double summation in Eq. (4.2) can be written in the following way:

$$\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} s_i s_j \left[f_{ij}^{(1)} + f_{ij}^{(2)} + f_{ij}^{(3)} + f_{ij}^{(4)} \right], \tag{4.3}$$

where the terms f_{ij} are defined as:

$$f_{ij}^{(1)} = \frac{1}{4}\sum_{\mathbf{n}}{}' \frac{\mathrm{erfc}\,(\alpha|\mathbf{r}_{ij} + \mathbf{n}L|)}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \tag{4.4}$$

$$f_{ij}^{(2)} = \frac{\pi}{N}\sum_{\mathbf{k}\neq 0} \frac{\mathrm{e}^{-\mathbf{k}^2/4\alpha^2}}{\mathbf{k}^2}\cos(\mathbf{k}\mathbf{r}_{ij}), \tag{4.5}$$

$$f_{ij}^{(3)} = \frac{\pi}{3N}\mathbf{r}_i.\mathbf{r}_j, \tag{4.6}$$

$$f_{ij}^{(4)} = -\frac{\alpha}{2\sqrt{\pi}}\delta_{ij}. \tag{4.7}$$

Here, erfc is the complimentary error function [146], α is a regularization parameter, and $\mathbf{k} = 2\pi\mathbf{n}/L$ is the reciprocal lattice momentum. The vector index $\mathbf{n}$ in Eq. (4.4) runs over the lattice copies in all spatial directions, and the prime indicates that $\mathbf{n} = 0$ is not taken into account in the sum when $i = j$. For numerical purposes, the real and reciprocal space summations, i.e., Eqs. (4.4) and (4.5) are bounded by $|\mathbf{r}_{ij} + \mathbf{n}L| < r_c$ and $k < 2\pi n_c/L$, respectively. The parameters α, r_c, and n_c are tuned to ensure a stable convergence of the sum. We find that $2 < \alpha < 4$, $n_c \gtrsim 4L$, and $r_c = L/2$ are sufficient for the above purpose.

We use Population Annealing Monte Carlo (PAMC) [52, 58, 60, 61, 67] to thermalize the system down to extremely low temperatures. In PAMC, similarly to simulated annealing (SA) [40], the system is equilibrated toward a target temperature, starting from a high temperature, following an annealing schedule. PAMC, however, outperforms SA by introducing many replicas of the same system and thermalizing them in parallel. Each replica is subjected to a series of Monte Carlo moves, and the entire pool of replicas is resampled according to an appropriate Boltzmann weight. This ensures that the system is equilibrated according to the Gibbs distribution at each temperature. For the simulations, we use particle-conserving dynamics to ensure that the lattice half-filling is kept constant, together with a hybrid temperature schedule linear in β and linear in T [52]. We use the family entropy of population annealing [61] as an equilibration criterion. Hard samples are resimulated with larger population size and number of sweeps until the equilibration criterion is met. Note that we have independently examined the accuracy of the results as well

Table 4.1: PAMC simulation parameters used for the finite-temperature simulations in the CO phase ($W \leq 0.131$). L is the linear system size, R_0 is the initial population size, M is the number of Metropolis sweeps, T_0 is the lowest temperature simulated, N_T is the number of temperatures, and $N_{\rm sa}$ is the number of disorder realizations. Note that the the values in the table vary slightly for different values of the disorder W. Reprinted with permission from [66].

L	R_0	M	T_0	N_T	$N_{\rm sa}$
4	2×10^4	10	0.05	401	5000
6	5×10^4	10	0.05	601	5000
8	1×10^5	20	0.05	801	2000
10	2×10^5	20	0.05	1001	1000
12	5×10^5	30	0.05	1201	500

Table 4.2: PAMC simulation parameters used for the finite-temperature simulations in the CG phase ($W > 0.131$). For details see the caption of Table 4.1. Note that the the values in the table vary slightly for different values of the disorder W. Reprinted with permission from [66].

L	R_0	M	T_0	N_T	$N_{\rm sa}$
4	2×10^4	20	0.004	401	100000
6	5×10^4	30	0.004	601	50000
8	1×10^5	40	0.004	801	30000
10	2×10^5	60	0.004	1001	20000

as the quality of thermalization for system sizes up to $L = 8$, using parallel tempering Monte Carlo [38]. Data from PAMC and parallel tempering Monte Carlo agrees within error bars. We investigate the phase diagram of the CG model using fixed values of the disorder width, i.e., vertical cuts on the W-T plane. Further details of the simulation parameters can be found in Tables 4.1 and 4.2 for the CO and CG phases, respectively.

4.4 Equilibration

In this section, we outline the steps taken to guarantee thermalization. The data for this work is predominantly generated using population annealing Monte Carlo (PAMC). In order to ensure that the states sampled by a Monte Carlo simulation are in fact in thermodynamic equilibrium, i.e., weighted according to the Boltzmann distribution, one needs to strive against bias by controlling the systematic errors intrinsic to the algorithm due to finite population size.

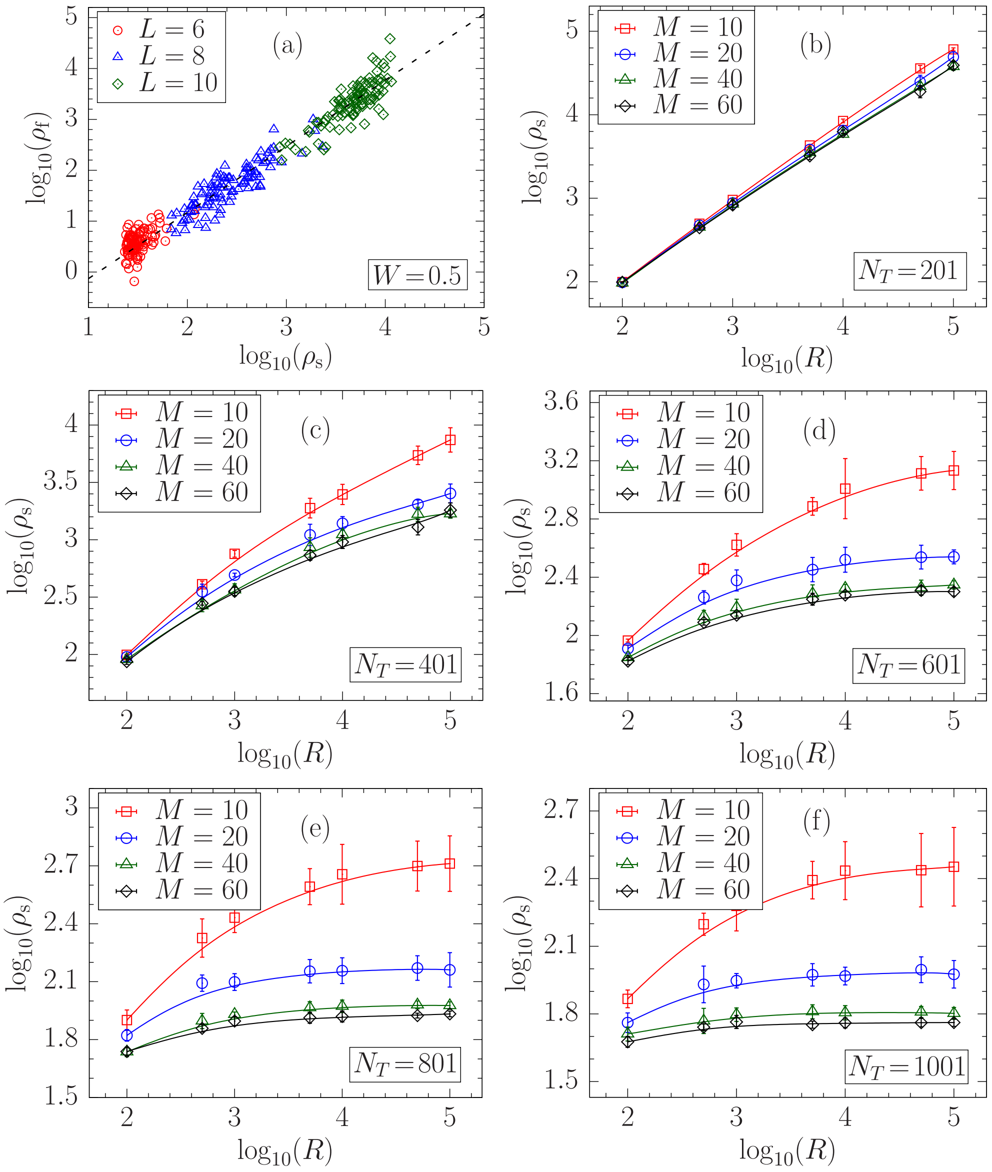

Figure 4.2: Equilibration of a PAMC simulation. (a): Equilibration population size ρ_{f} versus entropic family size ρ_{s} for a CG simulations at $W = 0.5$. 100 instances have been studied for each system size. Evidently, ρ_{s} is greatly correlated to ρ_{f}, which controls the systematic errors in thermodynamic quantities. Because ρ_{f} is computationally expensive to measure, one may instead use ρ_{s} as the measure of thermalization. (b)-(f): ρ_{s} versus the population size R for system size $L = 8$ at various number of temperatures N_T and Metropolis sweeps M. When ρ_{s} converges, the system is guaranteed to be in thermal equilibrium. As seen from the plots, convergence is achieved faster as the number of temperatures and sweeps is increased. However, for extremely large values of N_T and M, marginal improvement in equilibration is gained at the cost of extended run time of the simulation. Reprinted with permission from [66].

Fortunately, PAMC offers a convenient way to study and tune the systematic errors to a desired accuracy. It can be shown [61] that the systematic errors in a PAMC simulation are directly proportional to the *equilibration population size* ρ_{f}, which has the following definition:

$$\rho_{\mathrm{f}} = \lim_{R\to\infty} R\,\mathrm{var}(\beta F). \tag{4.8}$$

Here, R is the population size, and F is the free energy. ρ_{f} is an extensive quantity defined at the thermodynamic limit although, in reality, it converges at a large but finite R. Because ρ_{f} is computationally expensive to measure, as it requires multiple independent runs, one may alternatively study the *entropic family size* ρ_{s} defined as:

$$\rho_{\mathrm{s}} = \lim_{R\to\infty} R\,e^{-S_{\mathrm{f}}}, \tag{4.9}$$

where S_{f} is the family entropy of PAMC. As shown in of Fig. 4.2(a), ρ_{s} is well correlated with ρ_{f}, which is why we can reliably use ρ_{s} as the measure of equilibration. ρ_{s} similarly to ρ_{f} converges at a finite R. The population size at which the convergence is achieved is a function of the number of temperatures N_T as well as the number of Metropolis sweeps M. Optimization of PAMC is studied in great detail in the context of spin glasses [52, 67] much of which can be carried over to the CG simulations. As an example, we show in Figs. 4.2(b)-4.2(f) how we choose the optimal values of the PAMC parameters. We observe that the convergence of ρ_{s} is attained faster as the number of temperatures and sweeps is increased. However, beyond a certain point, any further increase solely prolongs the simulation time, while contributing negligibly to lowering the convergent value of ρ_{s}. A good rule of thumb for checking thermalization, as seen in Fig. 4.2, is that ρ_{s}, and as a result ρ_{f}, converges when $\rho_{\mathrm{s}}/R = \exp(-S_{\mathrm{f}}) < 0.01$. We ensure that the above criterion is met for every instance that we have studied. This matter has been investigated thoroughly in Ref. [61].

4.5 Results

4.5.1 Charge-Ordered Phase

To characterize the CO phase, we measure the specific heat capacity $c_{\mathrm{v}} = C_{\mathrm{v}}/N$ (only used to extract critical exponents), staggered magnetization

$$m_{\mathrm{s}} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i, \tag{4.10}$$

where $\sigma_i = (-1)^{x_i+y_i+z_i}s_i$ and $N = L^3$ the number of spins, as well as the disconnected and connected susceptibility

$$\bar{\chi} = N[\langle m_{\mathrm{s}}^2\rangle], \tag{4.11}$$

$$\chi = N[\langle m_{\mathrm{s}}^2\rangle - \langle|m_{\mathrm{s}}|\rangle^2]. \tag{4.12}$$

In addition, we measure the Binder ratio g [78]

$$g = \frac{1}{2}\left(3 - \frac{[\langle m_{\mathrm{s}}^4\rangle]}{[\langle m_{\mathrm{s}}^2\rangle]^2}\right), \tag{4.13}$$

and the finite-size correlation length ξ/L [147–149] defined as:

$$\xi = \frac{1}{2\sin\left(|\mathbf{k}_{\mathrm{min}}|/2\right)}\left(\frac{\chi(0)}{\chi(\mathbf{k}_{\mathrm{min}})} - 1\right)^{1/2}, \tag{4.14}$$

in which $\mathbf{k}_{\mathrm{min}} = (2\pi/L, 0, 0)$ is the smallest nonzero wave vector, and

$$\chi(\mathbf{k}) = \frac{1}{N}\sum_{ij}[\langle\sigma_i\sigma_j\rangle]\exp(i\mathbf{k}.\mathbf{r}_{ij}) \tag{4.15}$$

is the Fourier transform of the susceptibility. Furthermore, $\langle\cdots\rangle$ represents a thermal average and $[\cdots]$ is an average over disorder.

According to the scaling ansatz, in the vicinity of a second-order phase transition temperature T_{c}, any dimensionless thermodynamic quantity such as the Binder ratio and the finite-size corre-

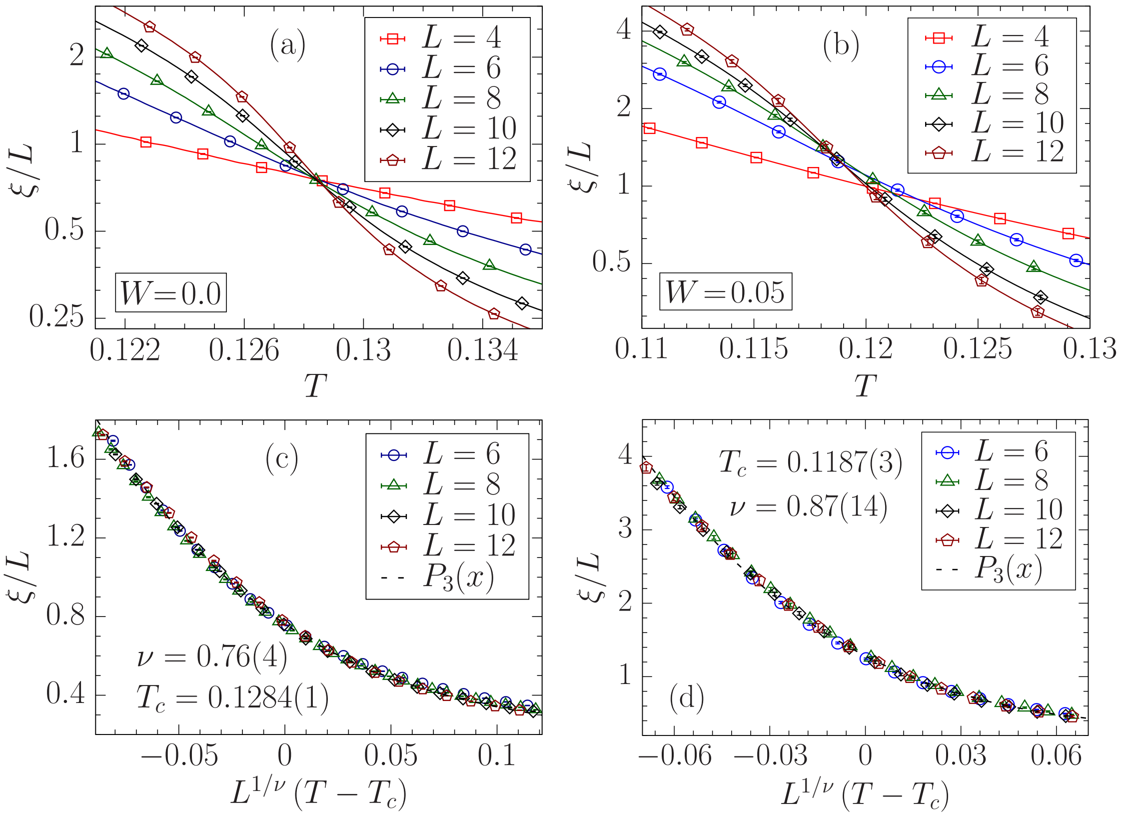

Figure 4.3: Finite-size correlation length per system size ξ/L versus temperature T for various disorder strengths. (a) no disorder, (b) small disorder ($W = 0.05$). In both cases we observe a crossing of the data for different system sizes, suggesting a phase transition between a disordered electron plasma and a CO phase. (c), (d): finite-size scaling analysis used to determine the best estimates for the critical temperature T_{c} as well as the critical exponent ν at the aforementioned disorder values. Note that the smallest system size is left out of the analysis for better accuracy. The transition temperature T_{c} of the CO phase decreases as the disorder grows. Reprinted with permission from [66].

lation length divided by linear system size will be a universal function of $x = L^{1/\nu}(T - T_{\mathrm{c}})$, i.e., $g = \tilde{F}_g(x)$ and $\xi/L = \tilde{F}_\xi(x)$, where ν is a critical exponent. Therefore, an effective way of probing a phase transition is to search for a point where g or ξ/L data intersect. Given the universality of the scaling functions $\tilde{F}_g$ and $\tilde{F}_\xi$, if one plots g or ξ/L versus $x = L^{1/\nu}(T - T_{\mathrm{c}})$, the data for all system sizes must collapse onto a common curve. Because we are dealing with temperatures close to T_{c}, we may approximate this universal curve by an appropriate mathematical function such as a third-order polynomial $f(x) = P_3(x)$ in the case of ξ/L or a complimentary error function

$f(x) = \frac{1}{2}\mathrm{erfc}(x)$ when studying the Binder cumulant. Hence, by fitting $f(x)$ to the data with T_c and ν as part of the fit parameters, we are able to determine their best estimates. The statistical error bars of the fit parameters are calculated by bootstrapping over the disorder realizations. In Fig. 4.3, we show the simulation data as well as the finite-size scaling (FSS) plots for ξ/L at two different disorder values. Crossings can clearly be observed, which signals a phase transition into the CO phase. Simulating multiple values of W, we observe a phase transition between a disordered electron plasma and a CO phase for $W < 0.131(2)$, consistent with previous studies [98, 133, 141]. The CO phase is a checkerboard-like crystal [142], where electrons and holes form a regular lattice as the potential energy dominates the kinetic energy at low temperatures.

We have also conducted zero-temperature simulations using simulated annealing to determine the zero-temperature critical disorder W_c that separates the CO from the CG phase. We average over $N_\mathrm{sa} = 2048$ different disorder realizations for disorders $W > 0.10$ and $N_\mathrm{sa} = 512$ for $W \leq 0.10$. Each disorder realization is restarted at least at 20 different initial random spin configurations and at each temperature step equilibrated N_eq Monte Carlo steps. If at least 15% of the runs reach the same minimal energy configuration, we assume that the chosen N_eq was large enough, and that the reached configuration is likely the ground state. If less than 15% of the configurations reach the minimal state, we increase N_eq and re-run the simulation until the 15% threshold is achieved. For the largest simulated system size ($L = 8$) and large disorders, typical equilibration times are $N_\mathrm{eq} = 2^{27}$ Monte Carlo sweeps.

To estimate W_c, we use the Binder ratio defined in Eq. (4.13) which, by definition, quickly approaches 1 when $T \to 0$ within the CO phase. Therefore, in order to retain a good resolution of a putative transition, we use an alternative quantity Γ that is defined in the following way [135]:

$$\Gamma = -\ln(1-g). \tag{4.16}$$

Close to W_c, we may assume the following finite-size scaling behavior for Γ:

$$\Gamma = \tilde{F}_\Gamma\left[L^{1/\nu}(W - W_\mathrm{c})\right]. \tag{4.17}$$

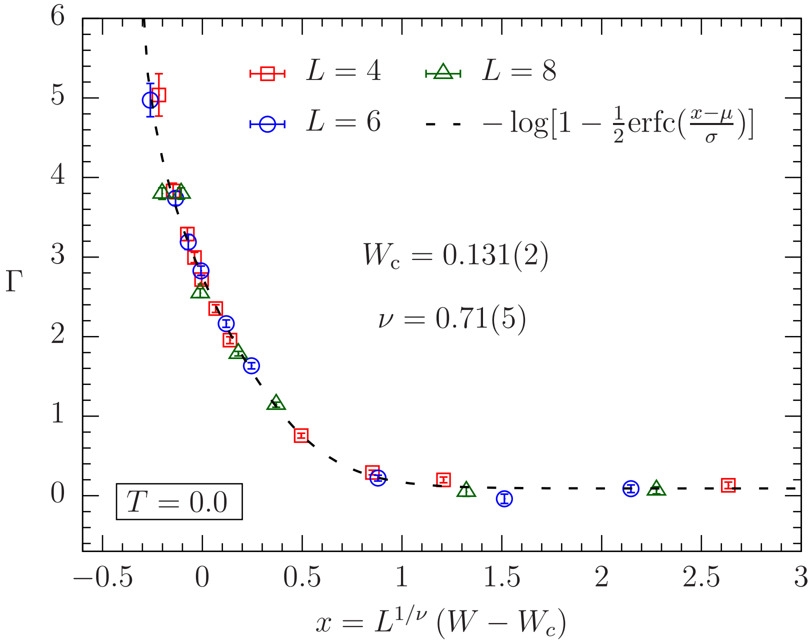

Figure 4.4: Zero-temperature simulation results for the plasma-CO phase transition. The quantity Γ defined in Eq. (4.16) is used to perform a finite-size scaling analysis. We conclude that the CO phase terminates at $W_c = 0.131(2)$. The statistical error bars are estimates by bootstrapping over disorder instances. $\mathrm{erfc}(x)$ is the complimentary error function, which is used to fit the Binder ratio data (see the main text). Reprinted with permission from [66].

As g is restricted to $0 \leq g \leq 1$ with a step-function like shape, we may use a complimentary error function $\frac{1}{2}\mathrm{erfc}(\frac{x-\mu}{\sigma})$ to represent the universal scaling function $\tilde{F}_\Gamma$ in which $x = L^{1/\nu}(W - W_\mathrm{c})$ and W_c, ν, μ, σ are the fit parameters. The fit is shown in Fig. 4.4 where we obtain $W_\mathrm{c} = 0.131 \pm 0.002$ and $\nu = 0.71 \pm 0.05$. Critical exponents such as α, γ, and β can be estimated by performing a FSS analysis using the peak values of the specific heat $c_\mathrm{v} = C_\mathrm{v}/N$, connected susceptibility χ, and the disconnected susceptibility $\bar{\chi}$, as well as the inflection point value of the staggered magnetization m_s that scale as following:

$$c_\mathrm{v}^\mathrm{max} \sim L^{\alpha/\nu}, \quad m_\mathrm{s}^\mathrm{inflect} \sim L^{-\beta/\nu}, \tag{4.18}$$

$$\chi^\mathrm{max} \sim L^{\gamma/\nu}, \quad \bar{\chi}^\mathrm{max} \sim L^{\bar{\gamma}/\nu}. \tag{4.19}$$

As we can see in Fig. 4.5, the above scaling behaviors are satisfied very well. The best estimates of the critical parameters for various values of the disorder are listed in Table 4.3. Note that with

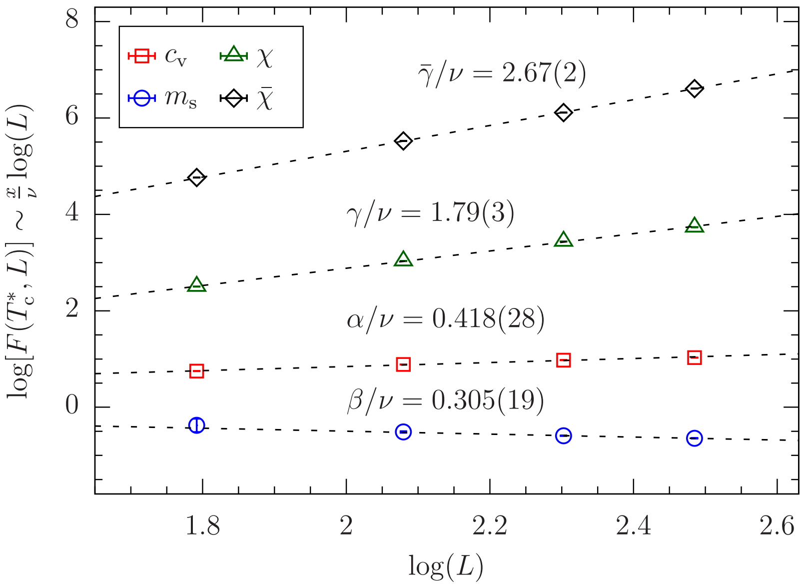

Figure 4.5: Finite-size scaling analysis for the plasma-CO phase transition at $W = 0.05$. The peak values of the specific heat capacity c_{v}, connected and disconnected susceptibilities χ and $\bar{\chi}$, as well as the inflection point value of the staggered magnetization are used to estimate the critical exponents α, β, γ, and $\bar{\gamma}$, respectively. According to Eqs. (4.18) and (4.19), the above quantities scale as a power law in the linear system size L, as clearly seen from the figure. Reprinted with permission from [66].

the exception of the universal exponent ν, other critical exponents vary with the disorder, which can be due to the trade-off between large-scale thermal and random-field fluctuations. Because at $T = 0$ the system has settled in the ground state, one cannot use thermal sampling to measure the variance of energy and staggered magnetization that are proportional to the heat capacity and susceptibility, respectively. Instead, we have used the techniques developed by Hartmann and Young in Ref. [150]. An important observation one can promptly make is that the exponents–except for ν that is universal–vary with the disorder. This can be attributed to the fact that the perturbations at large length scales are contested between random field fluctuations, which have static nature and dynamic thermal fluctuations [151–153]. At $W = 0$, the perturbations are purely thermal, while at $T = 0$, the random field completely dominates. At such large length scales, the interactions within the charge-ordered phase resemble the random-field Ising model (RFIM) [154–157] with short-range bonds, namely, screening takes place. This can be understood by

Table 4.3: Critical parameters of the plasma-CO phase transition at different disorder values. The exponents, except for ν, change with the disorder. Note that at $T = 0$, the exponents α and γ have been calculated in a different way (see text). Reprinted with permission from [66].

W	T_c	ν	α/ν	β/ν	$\bar{\gamma}/\nu$	γ/ν
0.000	$0.1284(1)$	$0.76(4)$	$0.550(2)$	$0.42(1)$	$2.41(1)$	$2.05(2)$
0.050	$0.1187(3)$	$0.87(14)$	$0.418(25)$	$0.305(19)$	$2.67(2)$	$1.79(3)$
$0.131(2)$	0.000	$0.71(5)$	$0.006(31)$	$0.154(5)$	$2.88(1)$	$1.55(4)$

remembering that the dynamics of the system is constrained by charge conservation. In the spin language, excitations are no longer spin flips but spin-pair flip-flops, owing to the conservation of total magnetization. For instance, one can create a local excitation while preserving charge neutrality by moving a number of electrons out of a subdomain in the CO phase. The excess energy of such a domain scales like its surface, similarly to the short-range ferromagnetic Ising model. It is worth mentioning that the Imry-Ma [158] picture gives a lower critical dimension of 2 for discrete spins with short-range interactions. Hence three-dimensional Ising spins, such as in the RFIM, are stable to small random fields as we also find here.

Returning to the discussion of the critical exponents, we note that scaling relations such as

$$\gamma = \beta(\delta - 1) = (2 - \eta)\nu, \tag{4.20}$$

as well as the modified hyperscaling relation

$$(d - \theta)\nu = 2 - \alpha = 2\beta + \gamma \tag{4.21}$$

can be utilized to obtain estimates for the critical exponents η, θ, and δ. For instance, using the values in Table 4.3, we see that $\eta(W = 0.0) = -0.05(2)$ and $\eta(W = 0.05) = 0.22(1)$. Near criticality, the correlation functions decay as a power of distance, i.e., $G(\mathbf{x}) \sim 1/|\mathbf{x}|^{d-2+\eta}$. The fact that the exponent η is slightly negative for $W = 0.0$ shows that correlation between the spins remains in effect over a much longer distance in the absence of disorder. Physically this is plausible as disorder tends to decorrelate the spins.

4.5.2 Coulomb Glass Phase

To examine the existence of a glassy phase in the CG model, we measure the spin-glass correlation length defined in Eq. (4.14), however, for a spin-glass order parameter, namely,

$$\xi_{\mathrm{SG}} = \frac{1}{2\sin\left(|\mathbf{k}_{\min}|/2\right)}\left(\frac{\chi_{\mathrm{SG}}(0)}{\chi_{\mathrm{SG}}(\mathbf{k}_{\min})} - 1\right)^{1/2}. \tag{4.22}$$

Here, the spin-glass susceptibility χ_{SG} has the following definition [149]:

$$\chi_{\mathrm{SG}}(\mathbf{k}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(\langle s_i s_j\rangle - \langle s_i\rangle\langle s_i\rangle)^2\right]\mathrm{e}^{i\mathbf{k}.(\mathbf{r}_i-\mathbf{r}_j)}. \tag{4.23}$$

It is important to note that $\langle s_i\rangle \neq 0$ because the Hamiltonian [Eq. (4.2)] is not symmetric under global spin flips. Therefore, at least four replicas are needed to compute the connected correlation function in Eq. (4.23). We start with the partition function of the system using Eq. (4.2):

$$Z = \sum_{\{s_i\}}\exp\left[-\beta\left(\frac{1}{8}\sum_{i\neq j}\frac{s_i s_j}{|\mathbf{r}_{ij}|} + \frac{1}{2}\sum_i s_i\phi_i\right)\right]. \tag{4.24}$$

We may now expresses any combination of the spin moments in terms of the replicated spin variables s_i^α in the following way:

$$\begin{aligned}\langle s_{1_1}\ldots s_{1_{k_1}}\rangle^{l_1}\ldots\langle s_{m_1}\ldots s_{m_{k_m}}\rangle^{l_m} &= \frac{1}{Z^n}\sum_{\{s_i^\alpha\}}\mathrm{e}^{-\beta\sum\limits_{\alpha=1}^{n}\mathcal{H}[\{s_i^\alpha\}]}s_{1_1}^1\ldots s_{1_{k_1}}^1\cdots s_{m_1}^n\ldots s_{m_{k_m}}^n\\ &= \frac{1}{n!}\sum_{\alpha_1\ldots\alpha_n}^{n}\langle s_{1_1}^{\alpha_1}\ldots s_{1_{k_1}}^{\alpha_1}\cdots s_{m_1}^{\alpha_n}\ldots s_{m_{k_m}}^{\alpha_n}\rangle,\end{aligned} \tag{4.25}$$

where $n = l_1 + \cdots + l_m$ is the total number of replicas, and replica indices $\alpha_1, \ldots \alpha_n$ are all distinct. As a special case, one can show that

$$(\langle s_i s_j\rangle - \langle s_i\rangle\langle s_j\rangle)^2 = \frac{2}{4!}\sum_{\alpha,\beta}^{4}\langle s_i^\alpha s_j^\alpha s_i^\beta s_j^\beta\rangle - \frac{2}{4!}\sum_{\alpha,\beta,\gamma}^{4}\langle s_i^\alpha s_j^\alpha s_i^\beta s_j^\gamma\rangle + \frac{1}{4!}\sum_{\alpha,\beta,\gamma,\lambda}^{4}\langle s_i^\alpha s_i^\beta s_j^\gamma s_j^\lambda\rangle. \tag{4.26}$$

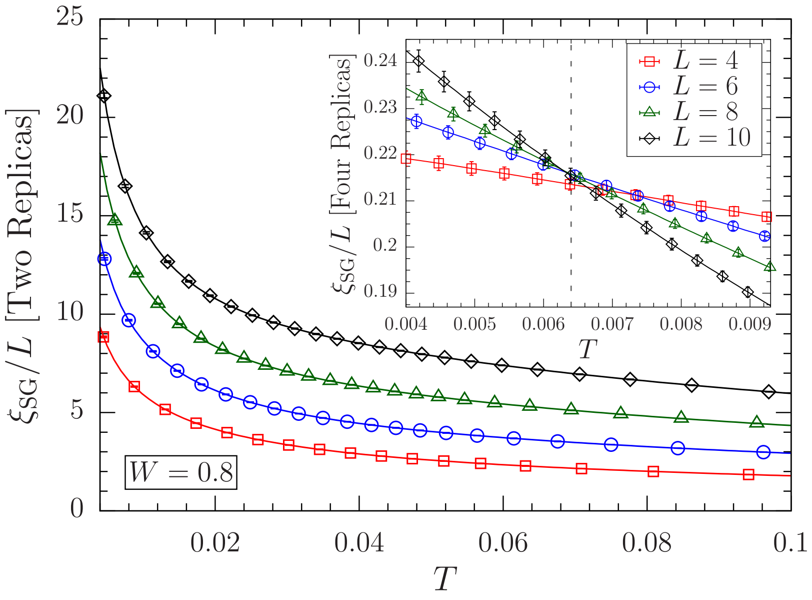

Figure 4.6: Spin-glass correlation length divided by system size ξ_{SG}/L calculated using two replicas at $W = 0.8$ versus temperature T. No crossing is observed down to very low temperatures. The inset shows the same quantity using four replicas, where a transition is clearly visible. Here, data points for different system sizes cross approximately at the temperature indicated by the dashed line. This suggests that in the presence of external fields four-replica quantities need to be used to characterize phase transitions in glassy systems. Reprinted with permission from [66].

Using the above expression, the spin-glass susceptibility [Eq. (4.23)] can be written in terms of the replica overlaps as follows:

$$\begin{aligned}\chi_{\mathrm{SG}}(\mathbf{k}) = \frac{N}{6}\sum_{\alpha<\beta}^{4}\left[\langle q_{\alpha\beta}(\mathbf{k})q^{*}_{\alpha\beta}(\mathbf{k})\rangle\right] - \frac{N}{6}\sum_{\alpha}^{4}\sum_{\beta<\gamma}^{4}\left[\langle q_{\alpha\beta}(\mathbf{k})q^{*}_{\alpha\gamma}(\mathbf{k})\rangle\right] \\ + \frac{N}{3}\sum_{\alpha<\beta}^{4}\sum_{\gamma<\lambda}^{4}\left[\langle q_{\alpha\beta}(\mathbf{k})q^{*}_{\gamma\lambda}(\mathbf{k})\rangle\right]. \qquad (4.27)\end{aligned}$$

Once again, the indices α, β, γ, and λ must be distinct. Here, $q^{*}_{\alpha\beta}(\mathbf{k})$ represents the complex conjugate of $q_{\alpha\beta}(\mathbf{k})$, and $q_{\alpha\beta}(\mathbf{k})$ is the Fourier transformed spin overlap, i.e.,

$$q_{\alpha\beta}(\mathbf{k}) = \frac{1}{N}\sum_{i=1}^{N} s_i^{\alpha} s_i^{\beta} \mathrm{e}^{i\mathbf{k}.\mathbf{r}_i}. \qquad (4.28)$$

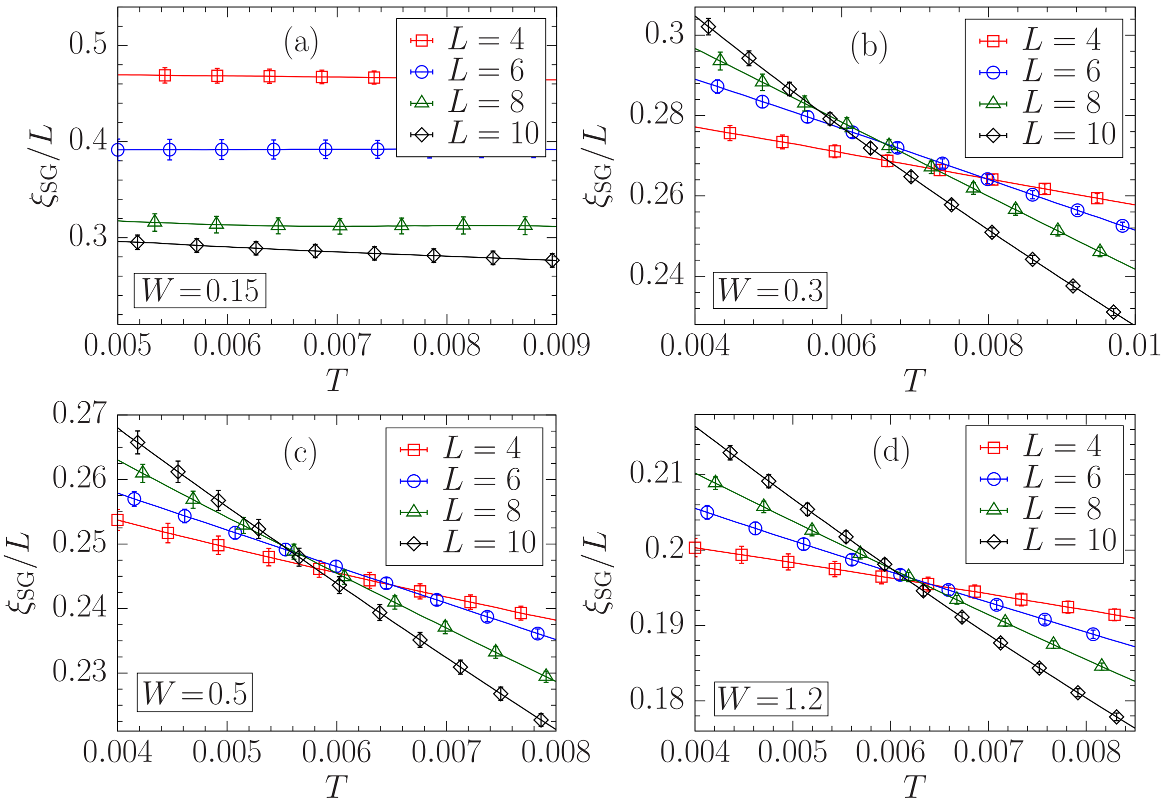

Figure 4.7: Spin-glass finite-size correlation length ξ_{SG}/L as a function of temperature T at various disorder strengths W. (a) $W = 0.15$, (b) $W = 0.30$, (c) $W = 0.50$, and (d) $W = 1.20$. For $W \gtrsim 0.15$, the data for different system sizes cross, indicating a plasma-CG phase transition. Corrections to scaling must be considered to reliably estimate the value of the critical temperature T_{c} (see the main text for details). Reprinted with permission from [66].

To underline the significance of this matter, we have shown in Fig. 4.6 the spin-glass correlation length calculated using two replicas, as has been done in some previous numerical studies of the CG [117, 159]. The inset shows the same quantity computed using four replicas. While the two-replica version of the finite-size correlation length shows no sign of a CG transition, the four-replica expression captures the existence of a phase transition into a glassy phase.

We have performed equilibrium simulations for $W \in \{0.15, 0.30, 0.50, 0.80, 1.2\}$. In Fig. 4.7, we plot the four-replica spin-glass correlation length as a function of temperature at select disorder values. Our results strongly suggest that there is a transition to a glassy phase that persists for relatively large values of the disorder. This is significant in the sense that it confirms the phase

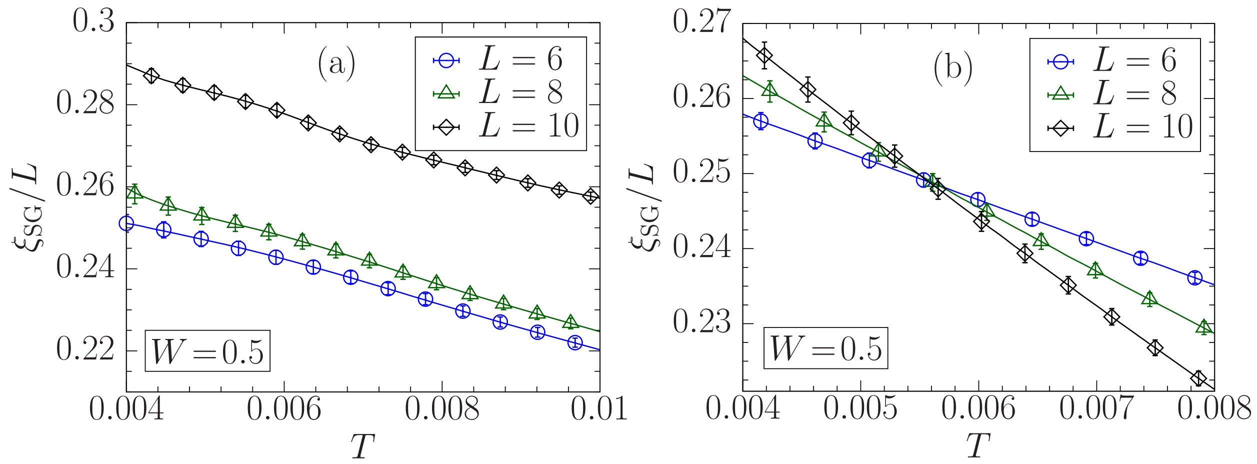

Figure 4.8: Importance of proper thermalization in observing a CG phase transition. Panel (a) shows a simulation where some instances have not reached thermal equilibrium, whereas panel (b) illustrates the same simulation in which all of the instances have been thoroughly thermalized. Reprinted with permission from [66].

transition via replica symmetry breaking as predicted by mean-field theory. The nontriviality of our findings can be better understood if one juxtaposes the CG case with that of finite-dimensional spin glasses lacking time-reversal symmetry due to an arbitrarily small external field, where the existence of de Almeida-Thouless [160] transition, except for a few rare cases [161, 162], has been ruled out by numerous studies [139, 163–167]. For the random-field Ising model, the droplet picture of Fisher and Huse [163, 164] can be invoked to show the instability of the glass phase to infinitesimal random fields. Yet, the CG model is different in two significant ways: typical compact domains are not charge-neutral, and therefore can not be flipped; and the long-range of the interactions, while it does not affect the domain wall formation energy in the ordered phase, may be significant in the more complex domain formation of the glass phase.

It is worth mentioning here that proper equilibration is crucial in observing phase transitions, especially in subtle cases like the CG model. We have illustrated this matter in Fig. 4.8. Figure 4.8(a) shows a simulation where the system has been poorly thermalized in which $\rho_{\mathrm{s}}/R \sim 0.1$ on average across the studied instances. By contrast in Fig. 4.8(b), the same simulation is done with careful equilibration; that is to say, the criterion $\rho_{\mathrm{s}}/R < 0.01$ is strictly enforced for every instance. It is clear that the observation of a crossing is contingent upon ensuring that every instance has reached thermal equilibrium. This, in turn, could explain why simulations using parallel tempering Monte Carlo, e.g., Ref. [133], see no sign of a transition.

Some corrections to scaling must be considered in the analysis in order to estimate the position of the critical temperature and the values of the critical exponents. In the vicinity of the critical temperature $T_{\rm c}$ and to leading order in corrections to scaling, we may consider the following FSS expressions for the spin-glass susceptibility $\chi_{\rm SG}$ and the finite-size two-point correlation length divided by the linear size of the system, $\xi_{\rm SG}/L$:

$$\chi_{\rm SG} \sim C_\chi L^{2-\eta}\left[1 + A_\chi L^{-\omega} + B_\chi L^{1/\nu}(T - T_{\rm c})\right], \tag{4.29}$$

$$\xi_{\rm SG}/L \sim C_\xi + A_\xi L^{-\omega} + B_\xi L^{1/\nu}(T - T_{\rm c}), \tag{4.30}$$

where A_χ, B_χ, C_χ, A_ξ, B_ξ, and C_ξ are constants. In order to find the critical temperature $T_{\rm c}$ as well as the critical exponents ν, η, ω, we perform the following procedure.

(i) Estimation of $T_{\rm c}$: Given any pair of system sizes (L_1, L_2), we have

$$L_1 = \bar{L} - \Delta L/2, \quad L_2 = \bar{L} + \Delta L/2, \tag{4.31}$$

in which $\Delta L = L_2 - L_1$ and $\bar{L} = (L_1 + L_2)/2$. Using Eq. (4.30), to the leading order in $\Delta L/\bar{L}$ we find

$$\frac{\xi_{\rm SG}(L_i, T)}{L_i} \sim \frac{\xi_{\rm SG}(\bar{L}, T)}{\bar{L}} - (-1)^i \frac{\Delta L}{2\bar{L}}\left[\omega A_\xi \bar{L}^{-\omega} - \frac{B_\xi}{\nu}\bar{L}^{1/\nu}(T - T_{\rm c})\right], \tag{4.32}$$

where the index i can take values $i = 1, 2$. One can now use Eq. (4.32) to determine the temperature $T^*(L_1, L_2)$ at which the curves of $\xi_{\rm SG}/L$ cross; in other words, $\xi_{\rm SG}(L_1, T^*)/L_1 = \xi_{\rm SG}(L_2, T^*)/L_2$ and

$$T^*(L_1, L_2) \sim T_{\rm c} + \Theta_\xi \bar{L}^{-\omega-1/\nu} = T_{\rm c} + \Theta_\xi \bar{L}^{-\phi}. \tag{4.33}$$

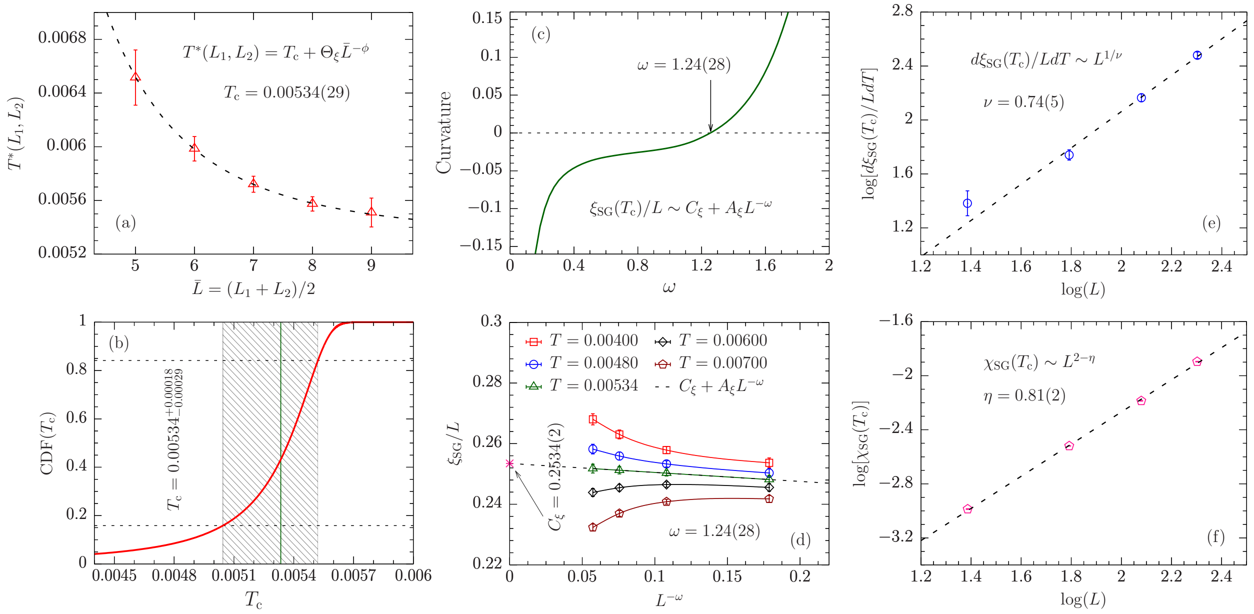

Figure 4.9: Process of estimating the critical exponents, as well as the critical temperature T_c of the plasma-CG phase transition for $W = 0.5$. Other values of W are analyzed using the same procedure. (a): The temperatures where ξ_{SG}/L curves of different systems sizes cross are used to determine the critical temperature T_c. The crossing temperatures decay toward the thermodynamic limit T_c. (b): The cumulative distribution function (CDF) is constructed by minimizing χ^2 with respect to Θ_ξ and ϕ while holding T_c constant. The shaded region shows the 68% confidence interval and the green vertical line indicates the best estimate of T_c. (c): The value of T_c obtained in the previous step is used to determine ω. At $T = T_c$ and optimal ω, ξ_{SG}/L is linear as a functions of $L^{-\omega}$; i.e., it has zero curvature as demonstrated in panel (d). (e): The critical exponent ν is estimated using the derivative of ξ_{SG}/L with respect to temperature, which scales as $L^{1/\nu}$ when evaluated at T_c. Some deviations are evident for the smallest system size. (f): The spin-glass susceptibility χ_{SG} at $T = T_c$ which scales as $L^{2-\eta}$ is used to determine the best estimate of the exponent η. Reprinted with permission from [66].

Here, T_c is the true critical temperature in the limit $L \to \infty$, and Θ_ξ is a constant. In Fig. 4.9(a), we show the T_c estimate for the case $W = 0.50$. The best fit curve is obtained by minimizing the sum of the square of the residuals, namely,

$$\chi^2 = \sum_{i=1}^{N} \left(T_i^* - T_c - \Theta_\xi \bar{L}_i^{-\phi} \right)^2 , \tag{4.34}$$

where i runs over all pairs of linear system sizes. Now we vary T_c, minimizing χ^2 along the way with respect to the remaining parameters. Since Θ_ξ appears linearly in the model, it can be eliminated [168] to reduce the optimization task to one free parameter, i.e., ϕ:

$$\left(\frac{\partial \chi^2}{\partial \Theta_\xi}\right)_{T_\mathrm{c}} = 0 \ \Rightarrow\ \tilde{\Theta}_\xi(T_\mathrm{c}, \phi) = \frac{\sum\limits_{i=1}^{N} (T_i^* - T_\mathrm{c}) \bar{L}_i^{-\phi}}{\sum\limits_{i=1}^{N} \bar{L}_i^{-2\phi}}. \tag{4.35}$$

Because there are five data points with three parameters in the original model, we have two degrees of freedom. Therefore, the probability density function (PDF) is proportional to $e^{-\chi^2/2}$. To determine the confidence intervals, we calculate the cumulative distribution function (CDF) [169]:

$$Q(T_\mathrm{c}) = \int^{T_\mathrm{c}} e^{-\frac{1}{2}\chi^2(T_\mathrm{c}')} dT_\mathrm{c}'. \tag{4.36}$$

As an example, in Fig. 4.9(b) we have shown the 68% confidence interval as well as the best estimate for the critical temperature.

(ii) Estimation of ω: From Eq. (4.30), we observe that

$$\xi_\mathrm{SG}(T_\mathrm{c})/L \sim C_\xi + A_\xi L^{-\omega}. \tag{4.37}$$

Thus, using the best estimate of T_c from the previous step, we expect the data points of $\xi_\mathrm{SG}(T_\mathrm{c})/L$ as a function of $L^{-\omega}$ to follow a straight line when ω is chosen correctly. We can therefore vary ω and measure the curvature until it vanishes at the optimal value. We have demonstrated this in Figs. 4.9(c) and 4.9(d). Note that the error bar for ω is calculated using the bootstrap method.

(iii) Estimation of ν and η: It is straightforward to show from Eqs. (4.29) and (4.30) that to the leading order in corrections

$$\chi_\mathrm{SG}(T_\mathrm{c}) = C_\chi L^{2-\eta}(1 + A_\chi L^{-\omega}), \tag{4.38}$$

$$\frac{d}{dT}(\xi_\mathrm{SG}/L)(T_\mathrm{c}) = B_\xi L^{1/\nu}(1 + D_\xi L^{-\omega}), \tag{4.39}$$

in which the best estimates obtained for T_c and ω are used. We see that the above quantities simply scale as $\chi_\mathrm{SG}(T_\mathrm{c}) \sim L^{2-\eta}$ and $\frac{d}{dT}(\xi_\mathrm{SG}/L)(T_\mathrm{c}) \sim L^{1/\nu}$ for large enough L. Therefore, a linear fit

Table 4.4: Critical parameters of the plasma-CG phase transition for various values of the disorder W. The exponent ν and ω are independent of W within error bars, highlighting their universality, whereas the exponent η varies as the disorder strength increases. Reprinted with permission from [66].

W	T_{c}	ν	ω	η
0.300	0.00446(25)	0.62(5)	1.26(7)	0.56(1)
0.500	0.00534(29)	0.74(5)	1.24(28)	0.82(5)
0.800	0.00590(56)	0.64(2)	1.28(20)	0.97(5)
1.200	0.00600(16)	0.65(3)	1.33(21)	1.09(1)

in the logarithmic scale will yield the exponents ν and ω. This is shown in Figs. 4.9(e) and 4.9(f), respectively. The above procedure has been repeated for all other values of the disorder W. The results are summarized in Table 4.4. We observe that within the error bars, the critical exponents ν and ω are robust to the disorder, which underlines the universality of these exponents. Nevertheless, larger system sizes–currently not accessible via simulation–would be needed to determine the universality class of the model conclusively. The fact that we observe stronger corrections to scaling for smaller disorder shows that the energy landscape is rougher due to competing interactions where finite-size effects are accentuated. For larger values of W, on the other hand, the system becomes easier to thermalize as the disorder dominates the electrostatic interactions.

4.6 Conclusion

We have shown that, using the four-replica expressions for the commonly-used observables, the CG model displays a transition into a glassy phase for the studied system sizes, provided that large enough disorder and sufficiently low temperatures are used in the simulations (see Fig. 4.1 for the complete phase diagram of the model). Previous numerical studies–including a work [134] by a subset of us–have failed to observe the glassy phase. In this study, we are able to present strong numerical evidence for the validity of the mean-field results in three space dimensions, which predicts transition to a glassy phase at large disorder via replica symmetry breaking. Moreover, we corroborate the results of previous studies for the low-disorder regime where a CO phase, similar to the ferromagnetic phase in the RFIM, is observed. Interestingly for large disorder values, the CG and the RFIM are different because the RFIM does not exhibit a transition into a glassy phase (see, for example, Ref. [157] and references therein). A possible reason is the combination of the constrained dynamics (magnetization-conserving dynamics) and the long-range Coulomb interactions not present in the RFIM. These two factors can increase frustration such that a glassy phase can emerge. Our findings open the possibility of describing electron glasses through an effective CG model both theoretically and numerically. Because most of the electron glass experiments are performed in two-dimensional materials, it would be desirable to investigate these results in two-dimensional models. We intend to visit this problem in the near future.

5. DISTRIBUTION OF INTER-EVENT AVALANCHE TIMES IN DISORDERED AND FRUSTRATED SPIN SYSTEMS[†]

5.1 Introduction

Many physical systems, when perturbed, respond in discrete jumps between metastable states. The earth's tectonic plates provide an example of such behavior in the form of earthquakes, which release a large amount of energy before being pinned again [171]. Similarly, a sheet of paper creases and tears in jerky movements, resulting in crackling sounds [172], the vortex lines of type-II superconductors depin when the electric current becomes large enough [173], and the magnetic dipoles of ferromagnets align with a changing external magnetic field in individual steps [174, 175].

In these situations and many others, the system waits in its new configuration until further changes in a driving field induce the next jump. The history of the sample is of great importance. The configuration of the system is not just a function of the instantaneous value of the drive, but depends on the path followed. In this manuscript, we study this phenomenon from a relatively new perspective which focuses on the distribution of inter-event times, and how that distribution is affected by the introduction of a threshold in the definition of an event. Our goal is not only to gain additional insight into the detailed mechanism of hysteresis, but also to examine the idea of *natural time* [176] in a more simple context than the geophysical applications that have mainly been considered up to now.

Klein *et al.* have suggested [171] that an alternate approach to the prediction of large earthquakes is to use as a clock the number of smaller earthquakes rather than quantifying intervals via a traditional counting of days and years. Varotsos *et al.* first introduced the term *natural time* to describe this procedure [177, 178]. Recent investigations have studied this concept in complex stochastic nonlinear processes, including its use in characterizing the current state of a system as it progresses between events [179]. Investigating natural time with geophysical data is difficult

---

[†] Reprinted with permission from [170].

owing to the absence or incompleteness of historical, and even modern, data on small earthquakes; large earthquakes are, fortunately, not excessively common. Additionally, controlled experiments are out of the question.

Here we use numerical simulations to analyze hysteresis and natural time in the context of several simple disordered and frustrated Ising spin models [5, 6, 25, 163, 180–182] exhibiting magnetic hysteresis: the three-dimensional random-field Ising model, the Sherrington-Kirkpatrick model, and the Viana-Bray model. Our key results are: (i) The distribution of inter-event times between all avalanches scales with the number of lattice sites for the random-field Ising model and the Viana-Bray model, but not for the Sherrington-Kirkpatrick model. (ii) The pseudo-gap exponent θ, which characterizes the behavior of the inter-event distribution for vanishing inter-event time is zero. (iii) The addition of long-range interactions decreases the number of small inter-event times, but does not affect the statistics of the intervals between large events, nor do they alter θ. (iv) Despite exploring various models and parameter regimes, we fail to find a situation where the predictive capability of the natural time method is strong for spin avalanches in magnetic hysteresis. (v) By imposing a minimum avalanche size threshold, different models can be classified by their inter-event distribution. (vi) Finally, at a sufficiently large minimum avalanche size threshold, the inter-event time in the Sherrington-Kirkpatrick model follows a Weibull distribution with shape factor $k \sim 1$, i.e., a Poisson distribution.

The use of simulations allows us to generalize to higher dimensions a recent analytical study by Nampoothiri *et al.* [183] on the inter-event time distribution of the one-dimensional random-field Ising model. The central result of that work was the computation of the distribution of times $P(\Delta B)$ of the magnetic field change ΔB between spin avalanches. (If the magnetic field is increased at a constant rate, ΔB is proportional to time.) It was found that $P(\Delta B) \sim (\Delta B)^{\theta}$ as $\Delta B \to 0$ [183], with $\theta = 0$ for the short-range ferromagnetic random-field Ising model, whereas $\theta = 0.95$ for the long-range antiferromagnetic case. Other studies of the distribution of gaps between events have been conducted in the context of amorphous solids and hard frictionless spheres [184–188]. There, when the strain γ is sufficiently increased, a corresponding stress drop follows.

The distribution of gaps $\Delta\gamma$ highlights differences between the yielding and depinning processes, and reveals information on mechanical stability.

The paper is structured as follows. Section 5.2 introduces the models studied, as well as the algorithmic approach and analysis methods used. In Sec. 5.3 we present results on the statistics of inter-event times for all avalanches and in Sec. 5.4 we repeat the analysis with the introduction of an event threshold. Section 5.5 discusses the effects of additional small-world bonds between the variables, followed by a study of return point memory and concluding remarks in Sec. 5.7.

5.2 Model and Methods

We first consider the random-field Ising model (RFIM) defined by the Hamiltonian

$$\mathcal{H}_{\mathrm{RFIM}} = -J\sum_{\langle ij\rangle} s_i s_j - \sum_i h_i s_i - B\sum_i s_i. \tag{5.1}$$

Here $s_i = \pm 1$ is a discrete degree of freedom at site i of a cubic lattice with N sites, and $\langle ij\rangle$ represents a sum over nearest neighbors. J is the exchange constant between nearest-neighbor sites, and units are set so that $J = 1$. Each site i is assigned a random magnetic field h_i, drawn from a Gaussian distribution

$$P(h_i) = \frac{1}{\sqrt{\pi}R}\exp(-h_i^2/R^2). \tag{5.2}$$

R controls the width of the distribution, and thereby the strength of the disorder. A spatially uniform field B is used to drive the hysteresis loop.

Unlike the ferromagnetic Ising model where $h_i = 0\ \forall i$, the RFIM does not exhibit ferromagnetic order in $d = 2$ (or below). As the temperature is lowered in higher space dimensions, however, a freezing transition occurs. This is followed by a ferromagnetic transition [189–193]. Here we are not concerned with these aspects of the equilibrium finite-temperature phase diagram, but instead focus on the evolution of the magnetization as the external field B is sequentially changed, with dynamics defined by each spin remaining parallel to its local environment at each step, that is, effectively $T = 0$ (see below).

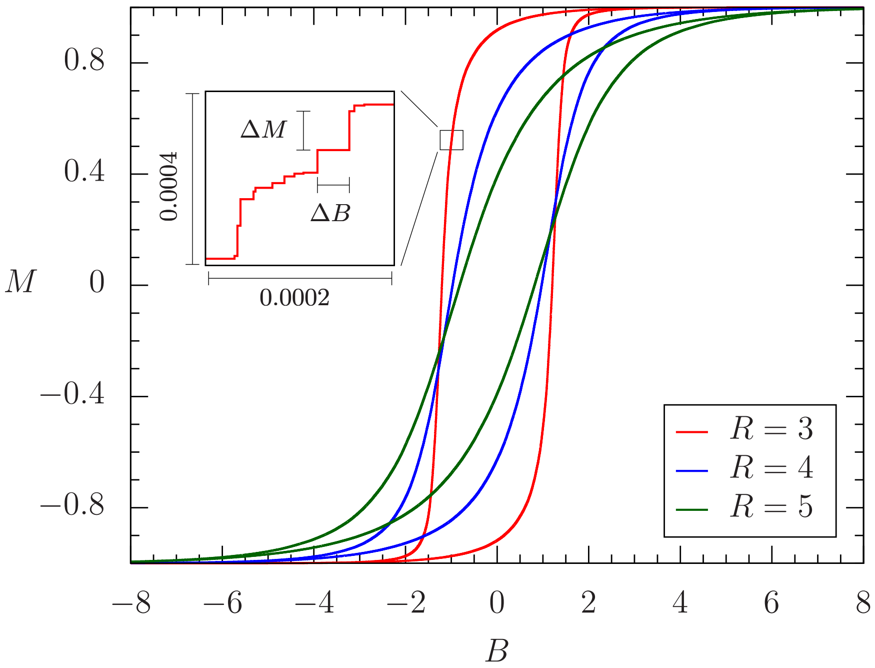

Figure 5.1: Hysteresis loop of the random-field Ising model of size $N = 100^3$ with a distribution of fields of widths $R = 3$, 4, and 5. On a large scale, the curves appears to represent a smooth evolution of the magnetization $M(B)$. However, as can be seen in the inset, the magnetization $M(B)$ is composed of a series of discrete jumps. ΔM is the change in magnetization. ΔB is the inter-event time. Reprinted with permission from [170].

Typical RFIM hysteresis curves are shown in Fig. 5.1. From them, we can extract the size of all the individual magnetization jumps S and hence their distribution. It is well-known [194] that this distribution has power-law behavior at low disorder, which we reproduce in Fig. 5.2. We can also extract the distribution of time (as measured by the change in external field ΔB) between events. The latter quantity has been much less studied than the former.

The Sherrington-Kirkpatrick model (SKM) [20] is given by the Hamiltonian

$$\mathcal{H}_{\text{SKM}} = -\sum_{i<j} J_{ij} s_i s_j - B \sum_i s_i. \tag{5.3}$$

In the SKM, every site $i \in \{1, \ldots, N\}$ interacts with every other site j via J_{ij}. That is, the interaction is infinite range. The exchange constants J_{ij} are disordered, and in our study are given by a Gaussian distribution with zero mean and a standard deviation of J_0. We set $J_0 = 1$ as our

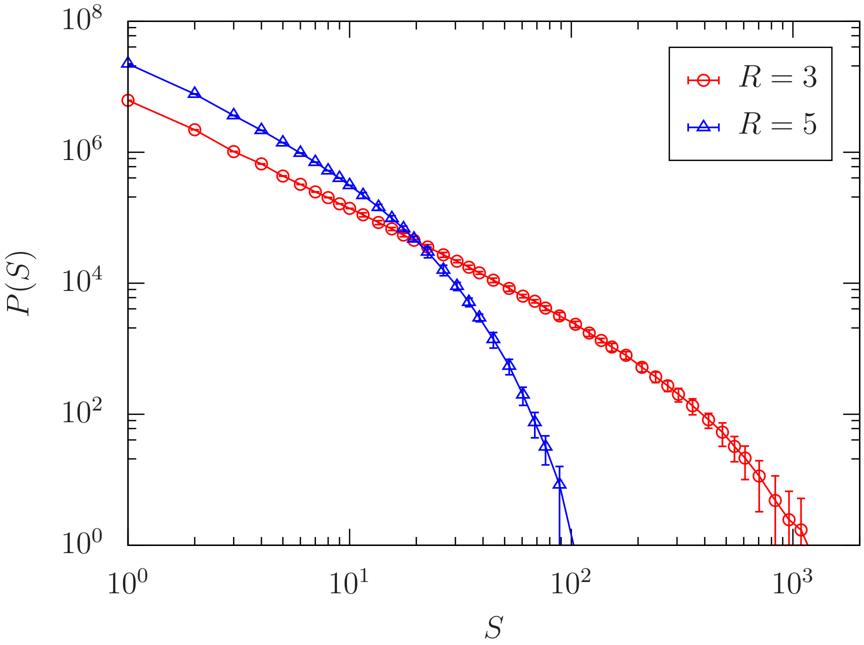

Figure 5.2: Distribution of avalanche sizes for field width $R = 3$ and 5. The distribution is known to have power-law behavior for $R = R_c \approx 2.16$. A critical region, for which power law behavior persists for several decades of avalanche size, occurs up to $R \approx 4$. For details see Ref. [194]. Reprinted with permission from [170].

unit of energy. The SKM shows self-organized criticality (SOC) for all disorder [195, 196].

Finally, we also study the Viana-Bray model (VBM) [197] in which each spin is randomly connected to $z = 6$ other spins. Thus, the VBM is still long-ranged, but with a finite coordination number. Unlike the SKM, the VBM does not have SOC [195]. Note that the RFIM has an explicit parameter R with which the disorder strength can be tuned, whereas the SKM and VBM do not.

In order to generate a hysteresis loop for the RFIM, we compute the local fields,

$$B_i = -h_i - J \sum_{j \in \mathcal{N}(i)} s_j \, , \tag{5.4}$$

and for the SKM and VBM,

$$B_i = - \sum_{j \in \mathcal{N}(i)} J_{ij} s_j \, . \tag{5.5}$$

For the RFIM, $\mathcal{N}(i)$ includes the nearest neighbor spins, whereas $\mathcal{N}(i)$ for the SKM consists of all spins, while the six randomly chosen neighbors define $\mathcal{N}(i)$ for the VBM.

A hysteresis loop is generated as follows: Starting at $B = \infty$ we reduce B to a value $B_k = \max\{B_i\}$. This is the external field at which the spin S_k becomes unstable. S_k is then reversed and the re-configuration of the lattice and of the collection of local fields $\{B_i\}$ is computed based on greedy dynamics [198]. Once S_k flips, and the local fields are recomputed, the next most unstable spin l is flipped, i.e., its (updated) local field is now greater than the external field: $B_l > S_l \cdot B$. This process is continued until all unstable spins are reversed. The total count of spins flipped is recorded as the size of the associated avalanche. The avalanche size determines the change in magnetization ΔM, which is twice the total fraction of spins that flip. At this point, the external field B is reduced once again to the next largest $\{B_i\}$, and the process is repeated until all spins are flipped and the system reaches saturation, but with the opposite sign of the magnetization. The inter-event times are the values ΔB that the external field jumps between each completed avalanche.

We begin, in Sec. 5.3, by analyzing the distribution of time intervals $P(\Delta B)$, which results from using the broadest definition of an avalanche, i.e., by including even the smallest possible $\Delta M = 2/N$, resulting from a single spin flip. We also calculate the pseudo-gap exponent θ, given by $P(\Delta B) \sim (\Delta B)^{\theta}$ as $\Delta B \to 0$. This follows the procedure described in recent literature [183] on the one-dimensional RFIM.

Next, in Sec. 5.4 we use a minimal threshold ΔM, only above which a change in a spin configuration is considered an event. We analyze how the distribution in inter-event times is affected by making $\Delta M > 2/N$. The introduction of such a threshold ΔM allows us to consider alternative measures of the *interval* between events. Specifically, we can define the *natural time* ΔA between large avalanches by counting the number of small avalanches (those with ΔM less than the threshold) which occur between large ones. We also define the *total natural time* ΔF between large avalanches to be the total number of flipped spins (i.e., the change in magnetization) that has accumulated. This latter procedure weights each small avalanche by the number of spins which turned over. These different approaches to the inter-event time are chosen to parallel analogous definitions in the geophysics community [199] where natural time employs only earthquakes exceeding a certain size as events and the number of small earthquakes between the large ones is recorded. The current state of an earthquake cycle is analyzed by constructing a cumulative distribution function of inter-event times between large earthquakes, which shows a Weibull form [176],

$$f(t) = 1 - e^{-(t/\lambda)^k}, \tag{5.6}$$

Here λ is the scale parameter, and k is the shape parameter. For $k < 1$ the cumulative probability function of the Weibull distribution has an initial rapid rise, while if $k > 1$ the initial slope is small. If $k = 1$, Eq. (5.6) becomes the inter-event distribution of a Poisson process. The Weibull distribution is commonly used in the materials science community to characterize the time to failure, where $k < 1$ corresponds to a failure rate which decreases with time. In contrast, $k > 1$ corresponds to a failure rate which increases with time. Motivated by the geophysics problem, we perform a similar fit to the cumulative distribution of inter-event times in our spin model hysteresis loops. By taking the natural logarithm of both sides twice, Eq. (5.6) becomes

$$\log[-\log(1 - f(t))] = k\log(t) - k\log(\lambda), \tag{5.7}$$

so that a plot of the data in the form $\log[-\log(1 - f(t))]$ vs $\log(t)$ yields a linear relation if $f(t)$ has a Weibull form.

5.3 Statistics of All Avalanche Inter-event Times

We start by analyzing the distributions of time intervals that occur between every avalanche, including avalanches of a single flip, i.e., $\Delta M = 2/N$. When an avalanche occurs, we mark its magnetic field value. Then, we can define ΔB as the difference between any two consecutive avalanches and accumulate the distribution of inter-event times as $P(\Delta B)$.

In order to compare distributions of different parameters properly, $P(\Delta B)$ is normalized. This

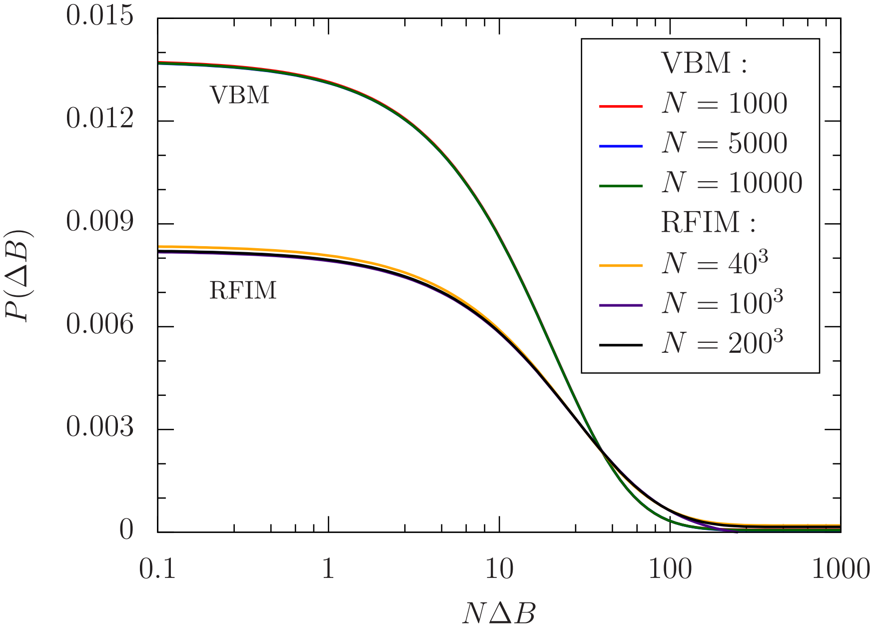

Figure 5.3: Scaling collapse of the distribution of inter-event times for the RFIM at field width $R = 2.3$ and the VBM for various system sizes N. If the field interval axis is scaled by N, the distributions for different N coincide. That is, $P(N, \Delta B) \sim \tilde{P}(N\Delta B)$. The vanishing slope at small ΔB indicates the pseudo-gap exponent defined by $P(\Delta B) \sim (\Delta B)^\theta$ as $\Delta B \to 0$ obeys $\theta = 0$, in agreement with analytic results in one space dimension ($d = 1$). Reprinted with permission from [170].

is done by dividing by the total number of intervals N_{int}. As R increases to large values, the spins feel a wide range of random fields, and their local fields B_i [Eq. (5.4)] become widely separated. In the limit $R \to \infty$, all events become single flips because the contribution to B_i from the exchange interactions J is negligible in comparison. The total number of intervals N_{int} approaches the number of lattice sites (spins) N. Similarly, as R decreases, the total number of intervals becomes small. At $R = 0$, the hysteresis loop becomes completely square, $N_i = 1$, and the entire lattice flips from up to down at the single external field value $B = -2dJ$, where $d = 3$ is the space dimension. Normalizing $P(\Delta B)$ to N_{int} eliminates this trivial effect. The sum of all avalanches in the distribution equals unity, independent of the choice of parameters.

We begin by analyzing the RFIM, plotting the distribution of inter-event times in Fig. 5.3. Distributions of varying lattice sizes N are seen to collapse if the event intervals ΔB are scaled by

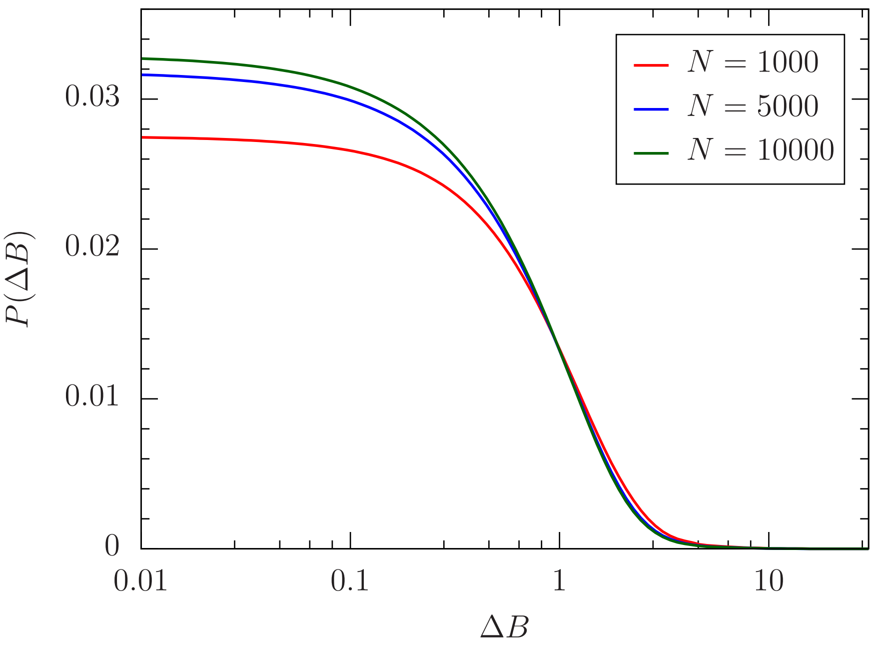

Figure 5.4: Distribution of inter-event times for the SKM. The distributions largely overlap, apart from a modest separation at small ΔB. In sharp contrast to the RFIM and VBM data of Fig. 5.3, $P(N, \Delta B)$ does not show a scaling collapse to $P(N\Delta B)$. Reprinted with permission from [170].

N, that is

$$P(N, \Delta B) \sim \tilde{P}(N\Delta B)\,. \tag{5.8}$$

The inter-event distribution for the VBM is similar to the RFIM (see Fig. 5.3) and scales with the number of variables N. Due to the fundamental difference between finite and diverging number of neighbors [195], we expect the same scaling behavior for differing finite coordination numbers of the VBM.

As can be seen in Fig. 5.4, the SKM distributions for different lattice sizes collapse with an *unscaled* ΔB. Other than that, the shape of the inter-event distribution of all avalanches is similar to the RFIM (Fig. 5.3). Because the VBM is long-ranged, and the RFIM is short-ranged, but both have similar scaling forms for $P(\Delta B)$, we conclude that the connections between distant spins do not by themselves give rise to a change from the RFIM collapse with $N\Delta B$. Instead, the most likely cause of scaling differences is the presence of SOC at all disorder strengths in the SKM due to the fully-connected topology.

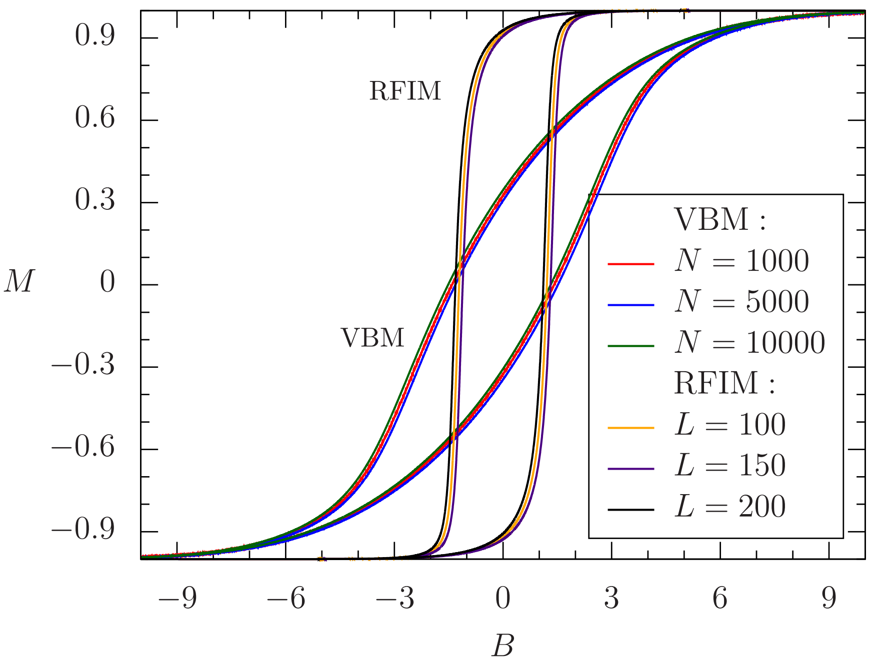

Figure 5.5: Hysteresis loop for the RFIM with $R = 3$ and for the VBM for different system sizes N. The shape is similar for various lattice sizes for all disorders. As argued in the text, this independence of lattice size underlies the scaling form $P(N, \Delta B) \sim \tilde{P}(N\Delta B)$. Reprinted with permission from [170].

Another way to gain insight into the scaling of inter-event times is to observe the behavior of the hysteresis loops. While hysteresis loops have been extensively studied [195, 200–203], we focus on the width of the loop in relation to the lattice size. For the RFIM and VBM, the loops are the same width across all lattice sizes (see Fig. 5.5). This means that the total time T for traversal of the loop is constant. In the limit where events are small and fairly isolated spatially, the number of events grows linearly with lattice size N, and the time ΔB between events is proportional to $1/N$. This picture offers a qualitative explanation of the dependence of P on $N\Delta B$.

The hysteresis loop for the SKM (Fig. 5.6) is different from the VBM and the RFIM. Its width grows as the lattice size is increased – the total time across the loop increases with N. If we again consider a limit where events are small and fairly isolated spatially, so that the number of events grows linearly with lattice size N, the interval between individual events ΔB is expected to be roughly independent of lattice size.

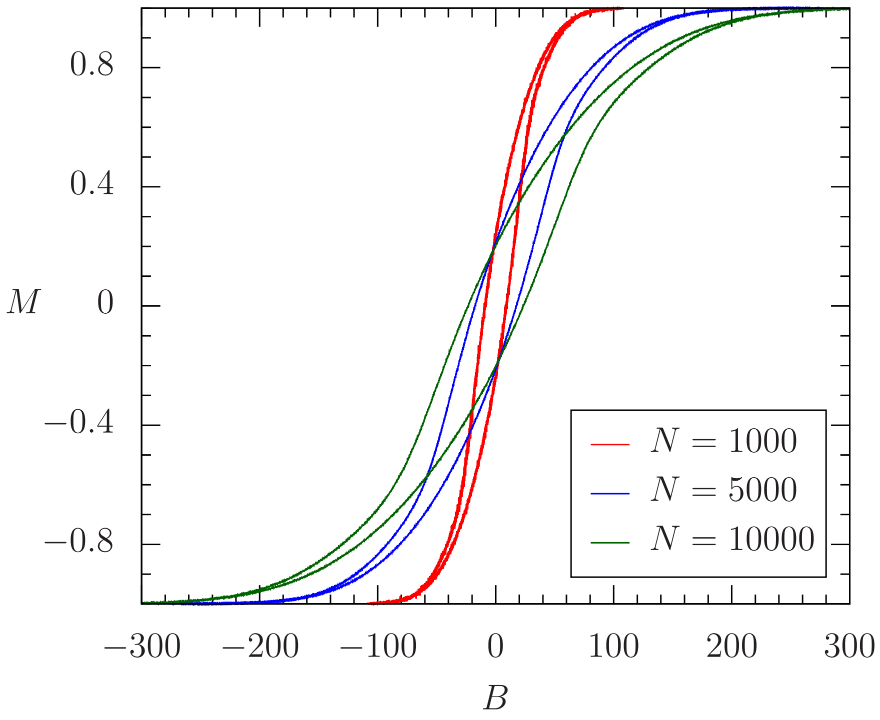

Figure 5.6: Hysteresis loop for the SKM. The width of the loop (e.g., at $M = 0$) increases with lattice size: $\Delta B \sim 9$, 17, and 23 for $N = 1000$, 5000, and 10000 respectively. As argued in the text, this dependence on lattice size underlies the difference in scaling behavior from the RFIM and VBM cases. Reprinted with permission from [170].

Returning to the RFIM, we consider the dependence of the limit of inter-event time distribution at small ΔB on disorder width R. Figure 5.7 shows the distributions for different R values. The quantity

$$C(R) = \lim_{N\Delta B \to 0} \tilde{P}(N\Delta B), \tag{5.9}$$

characterizes the value of the distribution at the smallest interval sizes. The inset to Fig. 5.7, showing $C(R)$, exhibits a peak for $R \approx 3.7$. Note that for the analytical calculation of the one-dimensional case, $C(R)$ peaks at $R \approx 1$ [183].

The flatness of $P(\Delta B)$ as $\Delta B \to 0$ implies that the pseudo-gap exponent $\theta = 0$ for all R in the RFIM. It likewise vanishes for the SKM and VBM. This value for the exponent is the same as the one-dimensional RFIM [183]. It has been argued that this is a consequence of the mapping between the RFIM and a depinning process when the dimensionality is less than 5 [204]. The

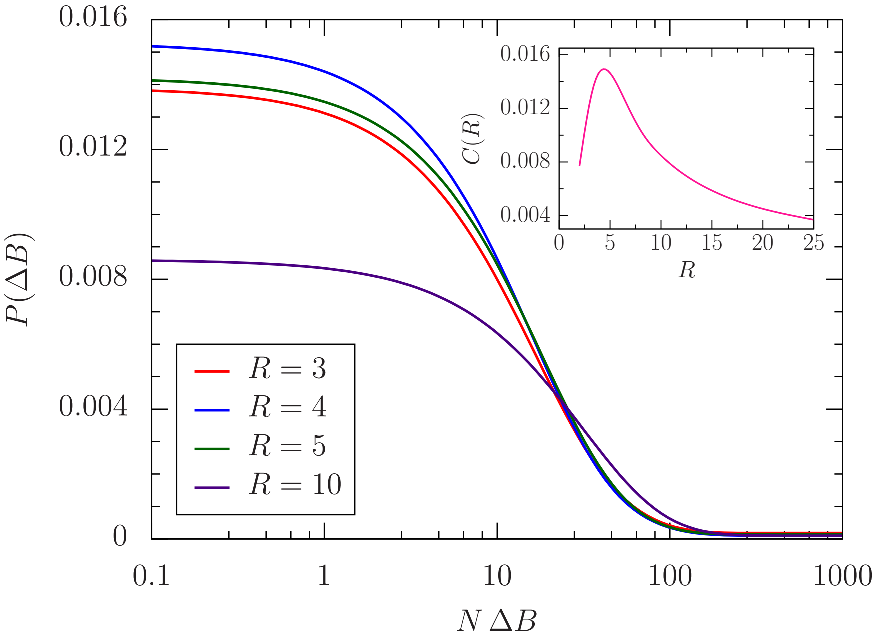

Figure 5.7: Distribution of inter-event times for the RFIM for various field widths R. $C(R)$ [the value of the inter-event distribution at small ΔB, Eq. (5.9)] is highest for $R = 4$. The behavior of $C(R)$ is nonmonotonic. The slope of $P(\Delta B)$ is zero for small ΔB, confirming that $\theta = 0$ for all R. Inset: $C(R)$ [Eq. (5.9)] for a fine mesh of R values. This allows for a more refined determination of the peak position, $R \approx 3.7$. This is in qualitative agreement with the $d = 1$ analytical result, which has a similar peak at $R = 1$ [183]. Reprinted with permission from [170].

depinning process is known to have $\theta = 0$ [184]. Thus our results confirm previous conjectures on the nature of the gap statistics.

The inter-event time distributions $P(\Delta B)$ of Figs. 5.3 and 5.4 illustrate the unpredictability of avalanche occurrences. The distributions have significant weight over several orders of magnitude of ΔB. The time between avalanches, as measured by the traditional definition ΔB (which is proportional to the conventional time interval Δt if B is swept at a constant rate), takes a very wide range of values. In Sec. 5.4 we explore whether alternate definitions might yield a narrower distribution, and hence more predictable intervals.

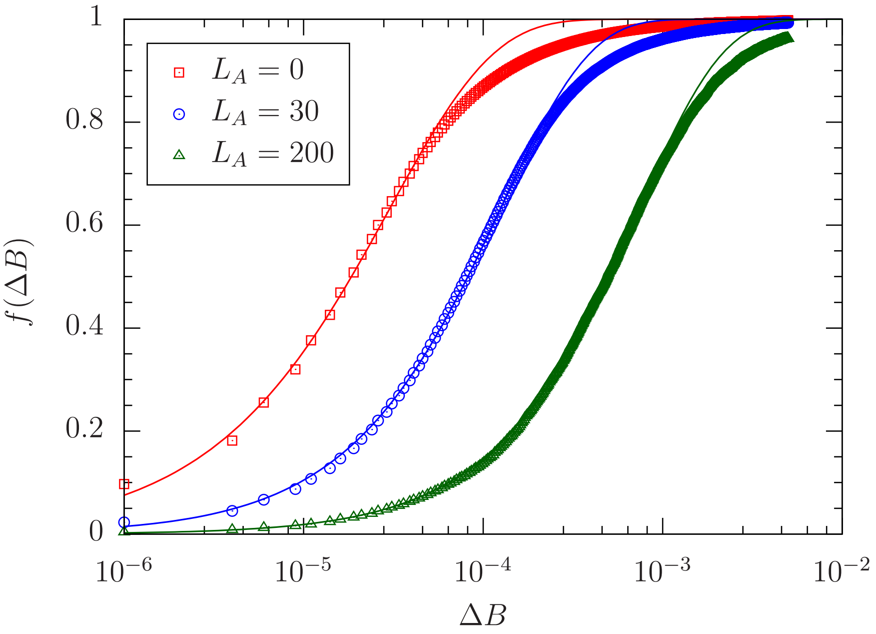

Figure 5.8: Cumulative distribution of the inter-event times for the RFIM at $R = 3$. The lattice size is $N = 100^3$. The data (points) are plotted along with the Weibull fit (lines). Several large avalanche thresholds are imposed. The Weibull cumulative distribution function appears to fit reasonably well to the cumulative distribution of the data. This occurs for ΔA and ΔF as well. The deviations are examined more critically in Fig. 5.9, which better emphasizes deviations from the Weibull form. Reprinted with permission from [170].

5.4 Statistics of Above Threshold Inter-event Times

In Sec. 5.3, we have seen that the distribution of inter-event times $P(\Delta B)$ is very broad when all avalanches are considered as events. We now re-analyze the distribution but impose an event threshold of L_A. This both eliminates the small (and therefore presumably more random) magnetization jumps and opens the door to counting the number of jumps as an alternate definition of inter-event time. This latter procedure follows suggestions in the geophysics community where including an avalanche threshold was argued to help determine where a certain geographic region is located in the earthquake cycle.

In geophysics studies, the imposition of a threshold was shown to lead to the inter-event distribution following a Weibull process [176]. Here we use a similar approach and verify if the statistics obey the same distribution. We examine several definitions of the inter-event time: ΔA character-

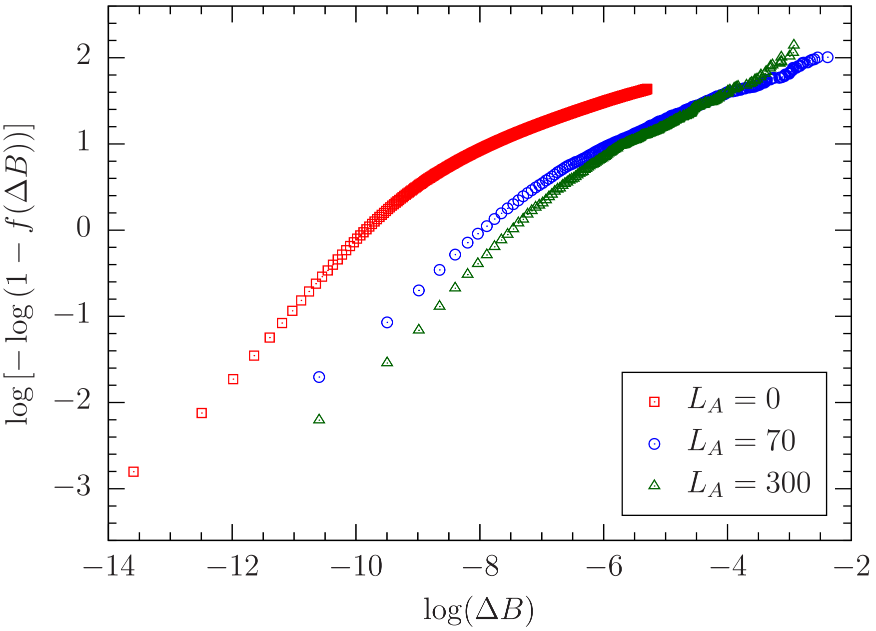

Figure 5.9: Cumulative distribution of the inter-event times for the RFIM at $R = 2.3$ on a double-logarithmic scale. For the distribution to be fit by a Weibull distribution, the data must be linear. Instead, the curves show a significant downward concavity for all values of L_A, The same trend is found for ΔA (not shown). Reprinted with permission from [170].

izes the number of small avalanches, ΔF the total number of individual flips. This complements the use of ΔB, the change in the magnetic field between events (see Sec. 5.3).

If large avalanches tended to occur after relatively *constant* numbers of small avalanches ΔA, then a plot of the cumulative distribution function $f(\Delta A)$ would take the form of an abrupt step, reflecting a sharply peaked probability $P(\Delta A)$, i.e., large events separated by *one specific* ΔA. Figure 5.8 shows the cumulative distribution functions for different disorder strengths R and for different choices of the threshold L_A for counting small avalanches. We see no significant tendency for the cumulative distribution to become more step-like than when plotted as a function of ΔB.

Although the continued broadness of the distributions of Fig. 5.8—despite the replacement of ΔB by ΔA and ΔF—suggests that natural time does not sharpen the distribution of inter-event spacing, we can still ask whether the underlying distributions of inter-event times are similar to those found in geophysics applications. Figure 5.8 shows the Weibull fits to the distributions f in addition to the raw data.

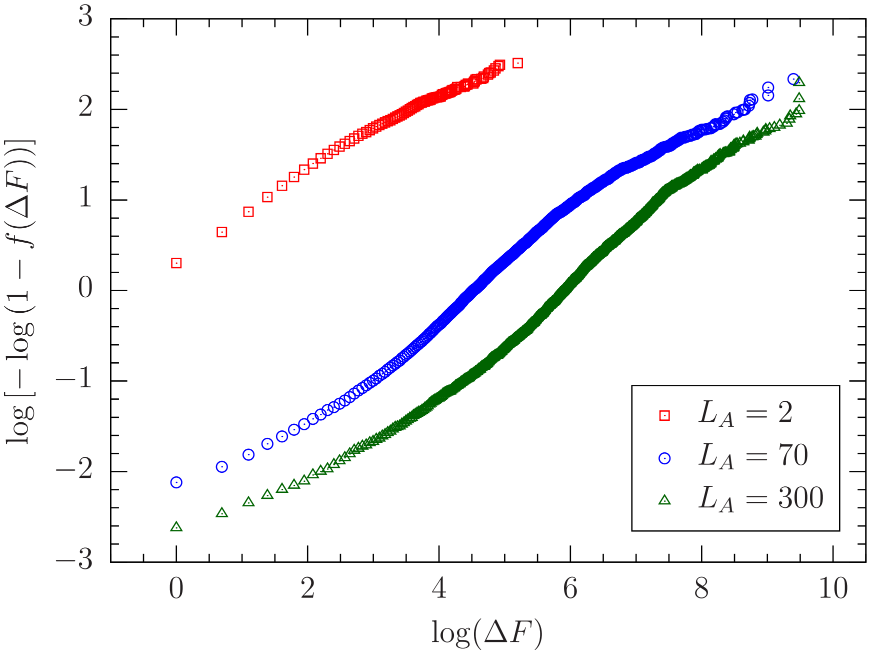

Figure 5.10: Cumulative distribution of the inter-event times for the RFIM at $R = 2.3$ on a double-logarithmic scale. The data are closest to linear for the smallest threshold. Reprinted with permission from [170].

While naively it appears that the data of Fig. 5.8 might follow a Weibull inter-event distribution, a more discerning check is made by plotting the data with modified axes: $\log[-\log(1 - f(t))]$ vs $\log(t)$. On these axes, the data should form a straight line, as discussed earlier [see Eq. (5.7)].

First, we analyze the RFIM, where the distributions do not appear to be well fit by a Weibull distribution by any definition of time (see Figs. 5.9 and 5.10). We focus on $R > R_\mathrm{c} \sim 2.16$, the critical value of R below which an infinite avalanche occurs in which a macroscopic fraction of the spins all flip at once [194]. Above R_c, ΔA and ΔB yield a curve that is concave down, while ΔF yields a curve with an inflection point. The fact that the distribution is concave down implies that there is a scarcity of large-time intervals for the distributions to be Weibull. The quality of the fits of RFIM inter-event times ΔB to a Weibull distribution (Fig. 5.9) does not appear to be very sensitive to the value of the avalanche threshold. In the case of natural time ΔF (Fig. 5.10) smaller threshold gives a somewhat better fit. In either case, as L_A increases, the average time between events increases, which leads to a lower intercept with the vertical axis; this follows from

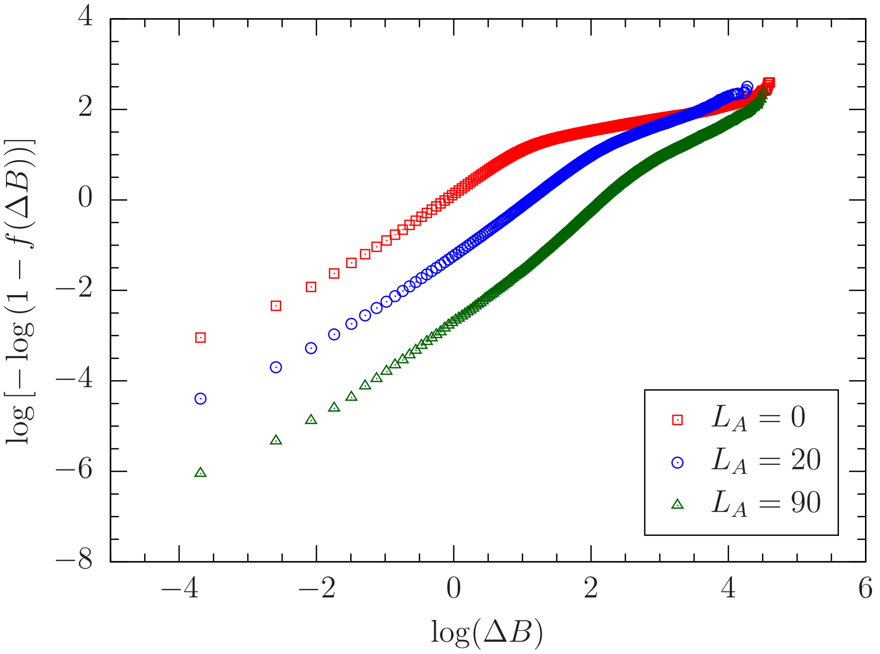

Figure 5.11: Cumulative distribution of the inter-event times for the SKM with lattice size $N = 10000$. As L_A is increased, the data become more linear, which shows that the distribution of ΔB can be approximated as Weibull at sufficiently large L_A. The slope for $L_A = 90$ is $k \sim 1$, which implies that the Weibull distribution simplifies into a Poisson distribution. The same phenomenon occurs in the VBM with a slope of $k \sim 0.8$. Reprinted with permission from [170].

Eq. (5.7). The situation is rather different for the SKM when using the inter-event time ΔB. As L_A is increased, the Weibull fit improves significantly, as shown in Fig. 5.11. Interestingly, the shape parameter k is close to unity, i.e., the distribution is Poissonian. The same fit improvement with larger L_A occurs for the VBM (not shown), but $k \sim 0.8$ in that case. We are unable to determine if this value of k varies with the VBM's coordination number z, and we might perform simulations of varying z in a future paper. The SKM also provides interesting results for the natural time methods. $f(\Delta A)$ is rather close to a Weibull distribution for all L_A (Fig. 5.12). ΔF, however, provides a fit that worsens as L_A is increased. The VBM shows similar results in that both methods of natural time provide the best fit at low L_A.

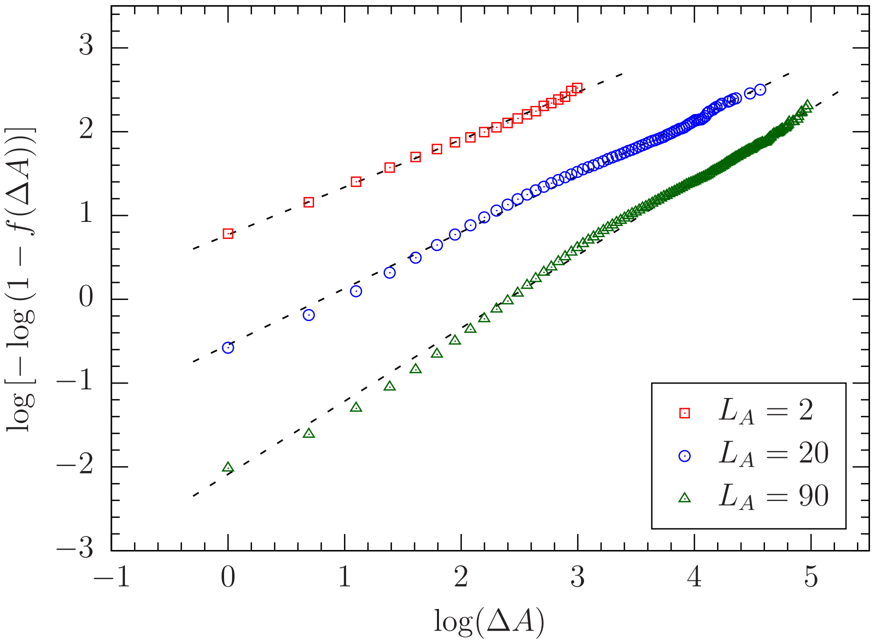

Figure 5.12: Cumulative distribution of the inter-event times for the SKM with lattice size $N = 10000$. The data are approximately linear for all L_A, i.e., the distribution is Weibull. Reprinted with permission from [170].

5.5 Effect of Longer-Range Couplings

On large lattices, small avalanches amongst clusters of spins which are far from each other are likely to occur in a rather independent manner. This might be problematic for periodic large-event intervals, because avalanches which are decoupled are unlikely to provide a predictive countdown to an above-threshold event. In order to introduce a more collective behavior of the entire cubic lattice, we introduce long-range couplings by dividing the entire cubic lattice into randomly selected pairs of sites. At the algorithmic level, this is accomplished by starting with site 0, and then randomly selecting one of the other sites p_0 of the lattice as a partner to site 0. Note that p_0 is not allowed to be one of the existing six nearest neighbor sites. After this is done, both sites 0 and p_0 are eliminated as potential partners and one proceeds to site 1 (assuming $p_0 \neq 1$) and randomly assigns it a partner p_1. This process is continued until all sites in the cubic lattice have a seventh neighbor. When assigned in this way, the probability of a site being part of a pair is independent of the geometric proximity between the two sites (as long as they are not nearest neighbors). These

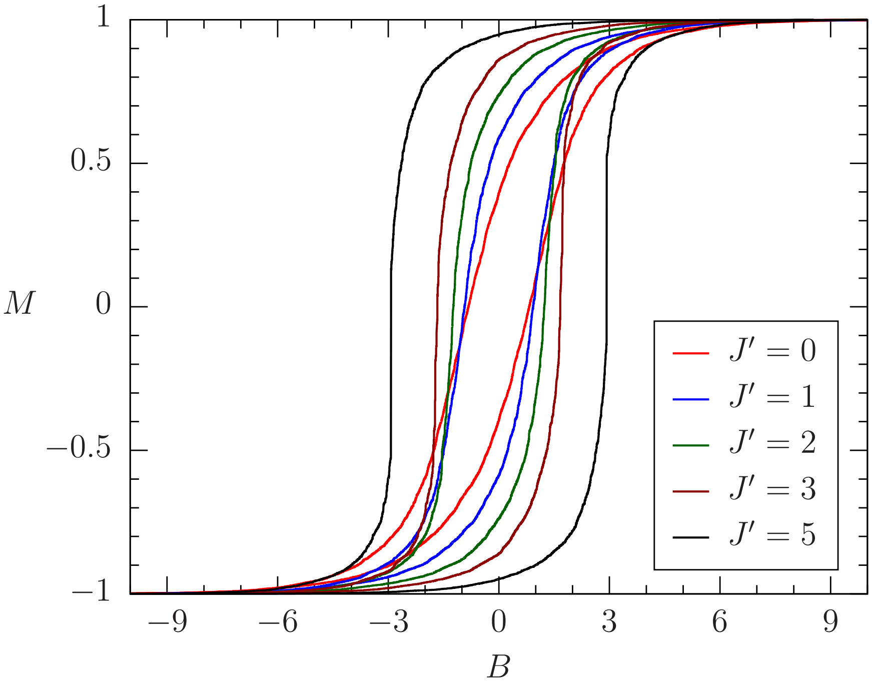

Figure 5.13: Hysteresis loops for the Hamiltonian in Eq. (5.10) for various J' values. Here, $R = 5$, and the lattice size is $N = 30^3$. As the strength of J' is increased, the hysteresis loop becomes steeper and wider. This is due to the increased correlation of lattice sites, which causes avalanches to occur at the same magnetic field value. Reprinted with permission from [170].

longer-range neighbors are coupled by an exchange constant J', so that in the computation of the local field, and hence the determination of whether to flip s_i, Eq. (5.4) is generalized to include p_i as part of $\mathcal{N}(i)$. Setting $J' = 0$ recovers the original nearest neighbor only model, and increasing J' allows us, in a smooth manner, to increase the long-range interactions across the lattice. Our model Hamiltonian thus becomes

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - J' \sum_i s_i S_{p_i} - \sum_i h_i s_i - B \sum_i s_i, \tag{5.10}$$

where p_i is the long-range site connected to site i. A similar procedure has previously been introduced in Ref. [205] to study finite-temperature phase transitions in Ising models with long-range interactions (which prove to be of mean-field character) and, more generally, are considered in the context of small-world networks [206].

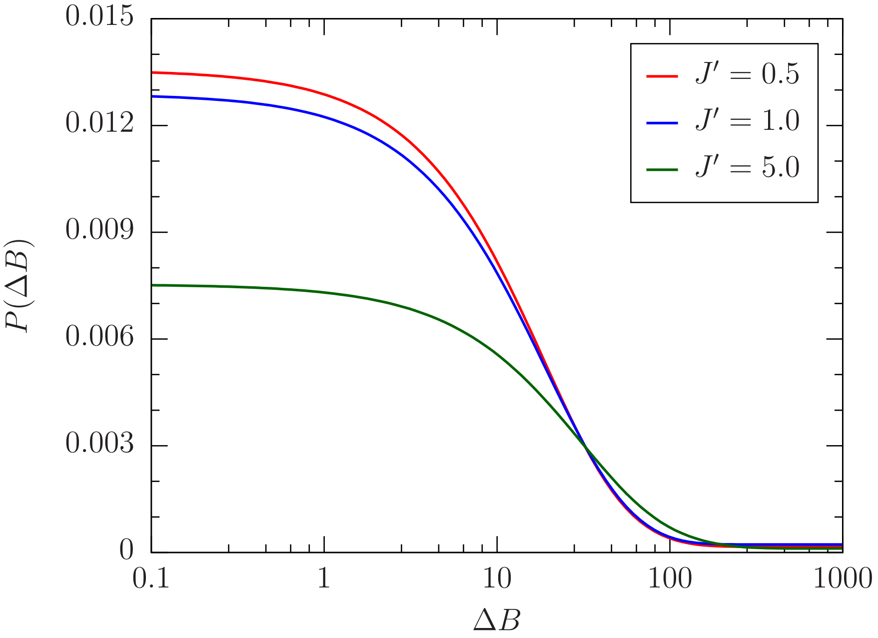

Figure 5.14: Distribution of inter-event times for various values of J'. The distribution is always monotonic, and there are fewer small inter-event times when J' is increased. This is due to the additional correlation between avalanches. Reprinted with permission from [170].

Figure 5.13 shows hysteresis loops for the Hamiltonian presented in Eq. (5.10). The loops become steeper, and the width is increased with J'. This occurs because different lattice sites become correlated, which causes avalanches to combine. The distribution of inter-event times ΔB is shown in Fig. 5.14. For all J' values, the distribution is monotonically decreasing, and larger J' values suppress the frequency of small avalanches. The pseudo-gap exponent is zero for any strength of the long-range connections.

$C(R)$, given by Eq. (5.9), is shown in Fig. 5.15 for different values of J'. $C(R)$ is non-monotonic; the value of R for which $C(R)$ is largest grows with the strength of the long-range connection. The overall curve is lowered when J' is increased. As is to be expected, once the ratio of R to J' becomes large enough that the local fields dominate the system, the curves for $C(R)$ collapse. The lowering of $C(R)$ with J' means there are fewer small inter-event times for larger J' values at a given disorder R. However, despite the shift in the distribution from smaller to larger inter-event times, $P(\Delta B)$ remains monotonically decreasing. These trends are present as well in

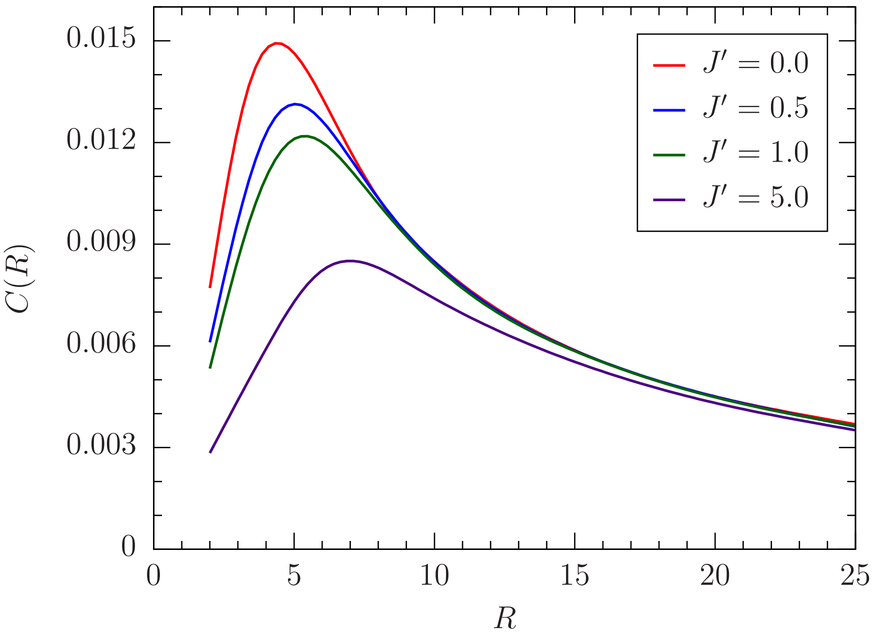

Figure 5.15: $C(R)$ [Eq. (5.9)] plotted against the disorder strength R for various J' values. $C(R)$ is lowered as J' increases, and the curve collapses onto the $J' = 0$ case when R becomes large enough. Reprinted with permission from [170].

Fig. 5.14, where the distribution becomes visibly flatter as J' is increased.

In order to explain this phenomenon, consider $J' = 0$ and two avalanches that occur separately in space and nearly simultaneously in time, i.e., at a very similar global field value B. At this moment, the two events are uncorrelated, and they have a very small inter-event time ΔB. If long-range connections are included, it is plausible that the two avalanches might now be correlated and occur simultaneously as J' grows. The probability of small avalanches decreases and of large ones increases, as seen in Fig. 5.16. This also implies that the frequency of small inter-event times is reduced. The long-range correlations do not ever result in a peak of $P(\Delta A)$ or $P(\Delta F)$, which would indicate specific most probable *natural time spacings*, which predict when a large avalanche would be imminent.

Although J' affects the quantitative value of $P(\Delta B)$, as well as $P(\Delta F)$ and $P(\Delta A)$, it does not significantly change their width. In short, the conjecture that J' might make avalanches constant in *natural time* seems to be false.

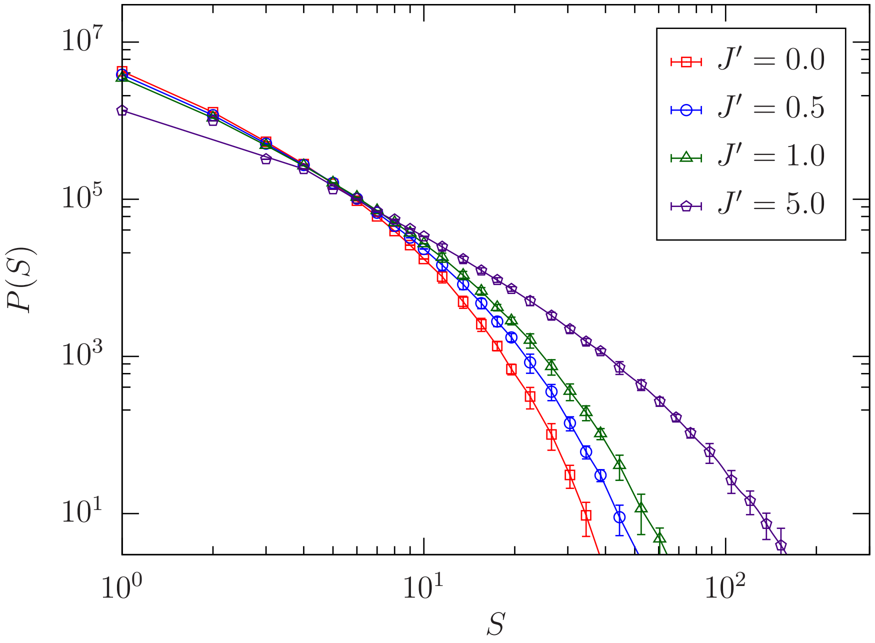

Figure 5.16: Distribution of avalanche sizes plotted for various J' values for $N = 50^3$ and $R = 7$. As J' increases, there are larger avalanche sizes, and the distribution approaches the critical region, as described in Ref. [194]. This is due to the correlation of lattice sites. Note that two avalanches may occur at the same moment, which leads to a greater amount of large avalanche sizes. Reprinted with permission from [170].

5.6 Return Point Memory

In the study of hysteresis loops, "return point memory" is a central concept [207–209]. Instead of driving the system to saturation, the external field B is lowered from infinity until some intermediate field B_0 is reached. At this point, B is raised to B_1 and then lowered back to B_0, which creates an "internal" hysteresis loop. If the system exhibits return point memory (RPM), then the state (that is, the magnetization) of the system is the same at both instances of B_0.

It is natural to ask how the inter-event times along an internal hysteresis loop are distributed. It is known that both the RFIM and the SKM exhibit RPM [200, 207]. While the distribution of inter-event times in the SKM has been shown to be well approximated by a Weibull distribution for sufficiently large L_A (see Fig. 5.11), the RFIM inter-event times are not as well described by a Weibull distribution (see Fig. 5.9). Therefore, we simulate the RFIM for both $R = 2.3$ and $R = 3$ on a lattice size of $N = 200^3$ to study the inter-event times for internal hysteresis loops. This helps

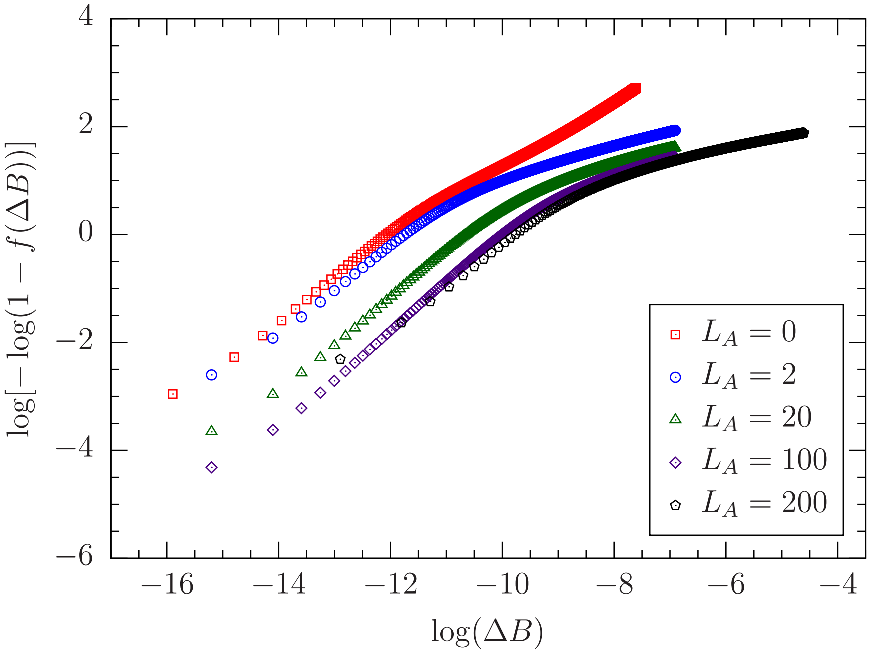

Figure 5.17: Cumulative distribution of the inter-event times for the RFIM across an internal hysteresis loop. The loop exhibits return point memory for $R = 2.3$ and $N = 200^3$. The data are more linear in a log-log plot for $L_A = 0$ than $L_A = 2$. Both ΔA and ΔF are similar to the results for the whole hysteresis loop. Reprinted with permission from [170].

identify whether RPM can help the RFIM inter-event distributions become Weibull-like, and the results can be directly compared to Figs. 5.8 and 5.9.

The results for internal hysteresis loops are shown in Figs. 5.17 and 5.18. The data for $L_A = 0$ approach a Weibull distribution even though the data are not Weibull distributed for the smallest nonzero large avalanche. While the data in a log-log plot are not exactly linear for $R = 2.3$, there is still a striking difference between $L_A = 0$ and $L_A = 2$. A more detailed analysis of RPM is needed to fully understand the behavior of small L_A. However, we expect the same trend to occur with alternate choices of disorder and RPM turning points. Note that ΔA and ΔF are the same as the previous case when the system is driven to saturation (see Fig. 5.10).

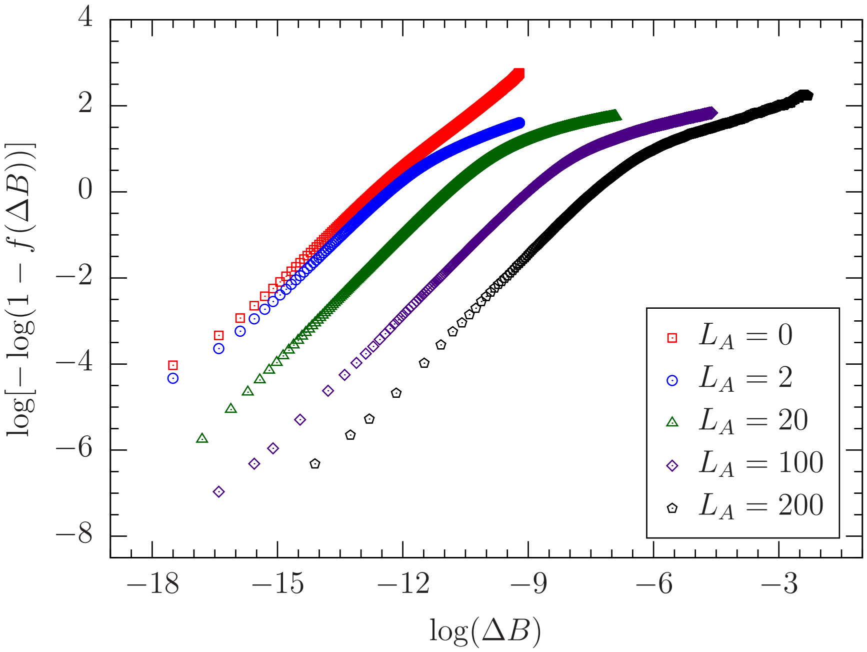

Figure 5.18: Cumulative distribution of the inter-event times for the RFIM across an internal hysteresis loop. The loop exhibits return point memory for $R = 3$ and $N = 200^3$. The data are Weibull distributed for $L_A = 0$. Reprinted with permission from [170].

5.7 Conclusions

We have evaluated the distribution of inter-event times of the three-dimensional random-field Ising, Sherrington-Kirkpatrick, and Viana-Bray models. Our motivation was two-fold: First, to extend the analytic results for this distribution—which have been obtained in one space dimension [183]— to higher space dimensions, providing complementary numerical results to the well-studied distributions of the avalanche amplitudes. Second, to explore the idea of natural time to study if the distribution is more sharply peaked when measured by counting the number of small avalanches rather than the change in the field itself.

Our conclusions regarding the first point are summarized in Figs. 5.3 and 5.4, which provide explicit forms for $P(\Delta B)$ for the different models. A central feature of our results is the scaling relation, $P(N, \Delta B) \sim \tilde{P}(N\Delta B)$ obeyed by both the RFIM and VBM, whose validity we trace to a hysteresis loop width which is nearly independent of the lattice size N (Fig. 5.5). In contrast, the SKM hysteresis loop width changes significantly with N (Fig. 5.6), and $P(N, \Delta B)$ does not scale. We also observe a pseudo-gap exponent $\theta \approx 0$, which is the same as in the one-dimensional case [183]. By examining the dependence of the inter-event distribution on the disorder, one finds nonmonotonic behavior where $C(R)$ [Eq. (5.9)] peaks at ~ 3.7, similar to the one-dimensional case.

Regarding the second point, we have added a large avalanche threshold, similar to the large earthquake threshold used in geophysics [199]. This leads to a distribution of inter-event times for several methods of counting time. Counting the number of individual spin flips, counting the small avalanches, and the original measurement in terms of the change in the magnetic field. We see no evidence for a sharpening of the inter-event time distribution function which would be a confirmation that large events occur at a specific ΔA or ΔF. The clock-time fit to a Weibull distribution, however, is improved by the introduction of an event threshold L_A (Fig. 5.11) in the SKM. The distribution of the natural time ΔA can also be fitted to a Weibull distribution in the SKM (Fig. 5.12).

Finally, we applied the same analysis to inter-event distributions for a model system with added small-world bonds, i.e., bonds between random pairs of lattice sites. As the strength of these small-world bonds is increased, there are fewer small inter-event times and fewer small avalanches. This is due to the increased correlation of lattice sites. As the correlation increases, avalanches coalesce into large avalanches, which reduces the number of small inter-event times and avalanche sizes. In some sense, the strength of the long-range bonds could be thought of as a tuning parameter between the distribution of avalanche sizes and inter-event times.

By adding the long-range bonds to the lattice, the inter-event distributions of large avalanches do not change. There is one main difference when long-range bonds are added. Namely, the value of the critical disorder increases when the strength of the long-range interactions are increased. As long as the disorder relative to the critical region is the same, the statistics of large avalanche inter-event times is the same for any strength of the long-range Interactions. Thus this modification of the model to make the natural time a more effective clock is seen not to be effective.

There are several possible objectives for a quantitative evaluation of inter-event times. One goal is the determination of their distribution function. In such an investigation, it is possible that alternate definitions of time, ΔB, ΔA, or ΔF in the work reported here, might lead to more simple or well-understood distributions. We have shown that in the SKM, the use of a finite event threshold and ΔA simplifies the nature of $P(\Delta A)$ to a Weibull distribution.

A second goal concerns the *prediction* of the next (large) event. That requires not only finding the distribution function but also through the use of an appropriate redefinition of time, acquiring a distribution function that is sharply peaked, so the separation between events is known. This is, obviously, a holy grail for earthquake prediction. We have not succeeded in finding such a transformation for interacting spin models. Nevertheless, we suggest that further exploring the idea within simple models might be a useful, more controllable, complement to the analysis of observational data.

6. OPTIMIZATION AND BENCHMARKING OF THE THERMAL CYCLING ALGORITHM

6.1 Introduction

As we discussed in Chapter 2, many combinatorial optimization problems are considered NP-hard, often associated with rough energy landscapes, which consist of numerous metastable states. Therefore, heuristics based on local search such as the greedy algorithm [210] tend to perform poorly on these types of problems as they can easily get stuck in deep local minima [211]. One way to circumvent this difficulty is to use a Markov-chain Monte Carlo method like simulated annealing (SA) [40], where the system is cooled down from a high temperature, enabling the system to visit many regions of the phase space. Since SA is stochastic in nature, running many such processes in parallel can increase the chance of arriving at the true ground state. Nevertheless, without establishing a way for the phase space information gathered by the random walkers to be shared, mere replication of a simulated annealing process will not yield any meaningful speedup. Multiple Morkov-chain algorithms such as path-integral Monte Carlo (PIMC) [44–46, 48–50], parallel tempering (PT) [37, 38] and population annealing (PA) [52, 58, 60, 61, 67] take advantage of such "collective knowledge" to efficiently probe the solution space of a problem.

Thermal cycling algorithm (TCA) [212] is another heuristic that integrates the power of parallel annealing processes with the utility of local search methods. The annealing part of this algorithm ensures that the phase space can be visited egodically, whereas the local search part biases the dynamics toward the lower-energy states. When introduced almost twenty years ago, thermal cycling was shown to outperform simulated annealing in solving some limited instances of the traveling salesman problem. Despite the early indications that TCA might be a useful tool in dealing with such hard optimization problems, it has not been widely adopted by the community. As such, in this chapter, we reintroduce the thermal cycling algorithm and explain it in clear terms. Additionally, we conduct a comprehensive parameter optimization of TCA using synthetic planted problems, where we compare the performance of TCA to several modern solvers, including simulated quantum annealing (SQA). To quantify the efficiency of the aforementioned heuristics, we study how their time to solution (TTS) [213, 214] scales with the problem size. Our results show that when optimized properly, TCA can indeed be a serious contender.

6.2 Thermal Cycling Algorithm

The thermal cycling algorithm works by periodic heating and cooling of an ensemble of states, while continuously reducing the temperature. The ensemble is prepared by selecting N_{p} lowest energy states among N_0 quenched random configurations. Starting from a relatively high temperature $T_{\mathrm{i}} = 1/\beta_{\mathrm{i}}$, the above pool of states is annealed toward a lower temperature of $T_{\mathrm{f}} = 1/\beta_{\mathrm{f}}$ in N_{T} steps. At a given temperature, the states in the pool are thermalized using N_{s} Metropolis updates (heating) and immediately quenched via a greedy search (cooling). If any of the resulting states are lower in energy than the original one, it is replaced in the pool. The heating-cooling cycle is repeated N_{c} times at a fixed temperature. In practice, the above process steers the ensemble toward the low-lying states, while ensuring that metastable configurations do not hinder the dynamics. The temperate is then slightly reduced, and the cycles start over.

A potential pitfall of the cycling mechanism is that it might cause oscillatory transitions between a handful of configurations without ever allowing the pool states to approach the ground state. As we mentioned earlier, this problem can be alleviated by establishing an interaction between the pool states. One way to do this is to freeze the variables that are common among all of the states. We can justify this reduction by realizing that if the pool states have a feature in common, it is very likely that the feature will also appear in the ground state configuration. Note that this step is closely related to metaheuristics like tabu search [215, 216], search for backbones [217–220]–often used in genetic type algorithms–and sample persistence [221, 222] which has been used in conjunction with algorithms such as simulated annealing as well as simulated quantum annealing. Below we present a concise outline of the thermal cycling algorithm.

Algorithm 2 Thermal Cycling Algorithm

```
Randomly Initialize $N_0$ configurations of the problem.
Quench each of the $N_0$ states using the greedy algorithm.
Construct a pool of states by selecting $N_\mathrm{p}$ states with the lowest energy from the above
  quenched states.
Build a list of lattice sites by comparing the spins on a given site between all pool sates. If all
  aligned, add the site to the list.
for $N_\mathrm{T}$ steps starting from $\beta = 0$ until $\beta = \beta_\mathrm{f}$ do
    for $N_\mathrm{c}$ cycles do
        Pick a random state from the pool.
        Add heat to the pool state using $N_\mathrm{s}$ Metropolis sweeps at
        $\beta$ avoiding the spins in the site list.
        Quench the selected state.
        if lower energy is achieved then
            Replace the old state in the pool with the new one.
            Rebuild the site list by comparing the spins between all pool states.
        end if
    end for
    Increase $\beta \to \beta + \Delta\beta$ in which the step size $\Delta\beta$ is usually
    constant, i.e, linear schedule.
end for
Identify the pool state with the lowest energy as the solution of the problem.
```

6.3 Details of Analysis

As we saw in Sec. 6.2, the thermal cycling algorithm has many tunable parameters, which include the final inverse temperature β_f, pool size N_p, number of anneal steps N_T, number of cycles N_c, and finally, the number of Metropolis sweeps N_s. For benchmarking, we use synthetic problems whose ground state is unique and known beforehand. As the measure of performance, we use *time to solution* (TTS) [213, 214], which is defined in the following way:

$$\mathrm{TTS}(\lambda) = n(\lambda)\,\tau_\mathrm{run} \tag{6.1}$$

in which $n(\lambda)$ is the number of times that the algorithm must be repeated for a given parameter set λ to find the ground state at least once with a desired probability of p_{d}, and τ_{run} is the average run time conventionally measured in microseconds. If we assume that the *success probability*, i.e., the chance of hitting the ground state in a single run of the algorithm is $p_{\mathrm{s}}(\lambda)$, then it is straightforward to show from the binomial distribution that

$$p_{\mathrm{d}} = \sum_{k\geq 1}^{n} \binom{n}{k} p_{\mathrm{s}}^{k}(1-p_{\mathrm{s}})^{n-k} = 1-(1-p_{\mathrm{s}})^{n}. \tag{6.2}$$

We may now use the above expression to find $n(\lambda)$ in Eq. (6.1):

$$n(\lambda) = \frac{\log[1-p_{\mathrm{d}}]}{\log[1-p_{\mathrm{s}}(\lambda)]}. \tag{6.3}$$

It is customary to set the desired probability in Eq. (6.2) to a high confidence value of $p_{\mathrm{d}} = 0.99$. Since TTS is a function of the algorithm tunables, a thorough optimization of the parameters must be performed in order to reliably compare heuristics based on TTS. Note that the optimization is often multidimensional that makes the benchmarking a relatively laborious task.

For each set of parameters λ and each problem instance, we repeat the runs 100 times and calculate the success probability $p_{\mathrm{s}}(\lambda)$ as the percentage of the ground state hits. This process is repeated for all instances, in this case 100, to calculate the median TTS, and the error bars are estimated using the bootstrap method. The above procedure is carried out for many other parameter set values, and the optimal parameters are identified as the global minimum point of the TTS function. Having calculated the optimal TTS for all problem sizes (L), we can study the scaling behavior of an algorithm which is often a power law, i.e.,

$$\mathrm{TTS}_{\mathrm{opt}} \sim 10^{aL+b}. \tag{6.4}$$

The scaling exponent a determines the performance of an algorithm for realistic problem sizes, whereas b introduces a constant offset that depends on the elements nonintrinsic to the algorithm,

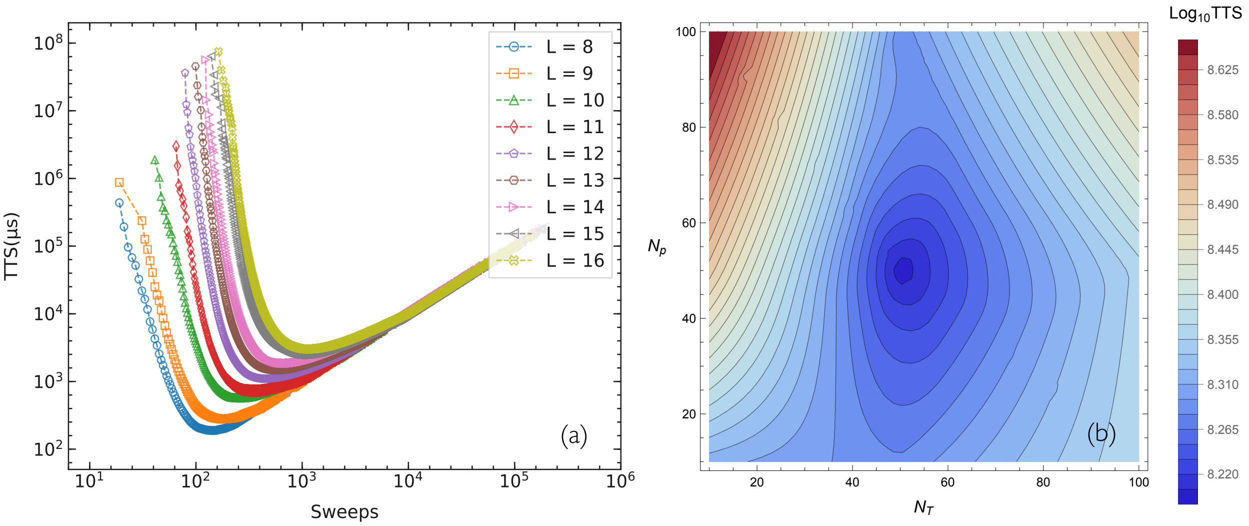

Figure 6.1: Parameter tuning for various algorithms. (a): Thermal cycling algorithm for the problem size $L = 11$. Solution pool size N_{p} versus annealing steps N_{T}, with the rest of the parameters fixed. The color map shows the TTS values. A global minimum corresponding to the optimal parameters is clearly observed. (b): Optimization of the number of sweeps for PT+ICM for various systems sizes. The minimum points of the TTS curves mark the optimal sweep values.

such as hardware speed, code efficiency, etc. Therefore, a relatively unbiased way to compare different algorithms in efficiency is to study their scaling exponents.

6.4 Results

In this study, we compare TCA to simulated annealing (SA), simulated quantum annealing (SQA), parallel tempering (PT), and parallel tempering with isoenergetic cluster moves (PT+ICM) [57]. SQA is the classical implementation of the quantum annealing process in which the system is initialized in the ground state of a simple Hamiltonian and adiabatically [43] deformed into a target Hamiltonian, whose ground state is difficult to find. PT is a Monte Carlo algorithm that efficiently samples the equilibrium configurations of a system using the replica-exchange technique. The ICM updates–which consist of rearranging large collection of variables by inspecting the overlap between two replicas of a system–are extremely effective for low connectivity graphs, where the cluster percolation threshold is small.

We use the deceptive cluster loop (DCL) problems [223] for benchmarking. The DCL problems

Table 6.1: Optimal parameters of the thermal cycling algorithm for different problem sizes L. N_{T} is the number of anneal steps, N_{p} is the pool size, N_{c} is the number of heating-cooling cycles per temperature, N_{s} is the number of Metropolis sweeps, and β_{f} is the target inverse temperature (lowest temperature).

L	N_{T}	N_{p}	N_{c}	N_{s}	β_{f}
8	100	10	100	10	0.2
9	100	10	200	10	0.2
10	100	50	300	10	0.2
11	150	50	300	10	0.25
12	150	50	300	10	0.25
13	150	50	500	10	0.25
14	200	100	500	10	0.3
15	200	100	500	10	0.3
16	200	100	500	10	0.3

are inspired by the original frustrated cluster loop (FCL) problems [224, 225] that are specially designed for testing the performance of D-Wave [11] quantum annealer against classical algorithms. The FCL's are planted problems with ferromagnetic ground states defined on the chimera graph [14] which consists of a two-dimensional lattice of fully-connected $K_{4,4}$ cells inside which all qbits are coupled together ferromagnetically–the entire $K_{4,4}$ unit cell can, therefore, be viewed as one virtual qbit. The cells are then connected via randomly chosen frustrated loops. By varying the density of the loops, the hardness of the problem can be tuned. In the DCL problems, the inter-cell couplers are multiplied by a real number of λ. Depending on the value of λ, the internal structure of the cells can be masked or accentuated, hence deceiving the annealers to spend more time optimizing the local structures rather than finding the global minimum.

For each studied algorithm, we optimize the parameters via a grid search within its parameter space. In Fig. 6.1 we show some examples of such optimization. Fig. 6.1(a) shows the optimization of sweeps for PT+ICM for various system sizes. The minimum TTS values correspond to the optimal sweep values. Fig. 6.1(b) illustrates the parameter optimization of TCA for the system size $L = 11$ in which the color map shows the TTS values in the logarithmic scale, and the axes are pool size N_{p} and the annealing steps N_{T}. Note that this is only a two-dimensional cross-section of the entire parameter space on which the rest of the parameters have fixed values of $N_{\mathrm{c}} = 100$,

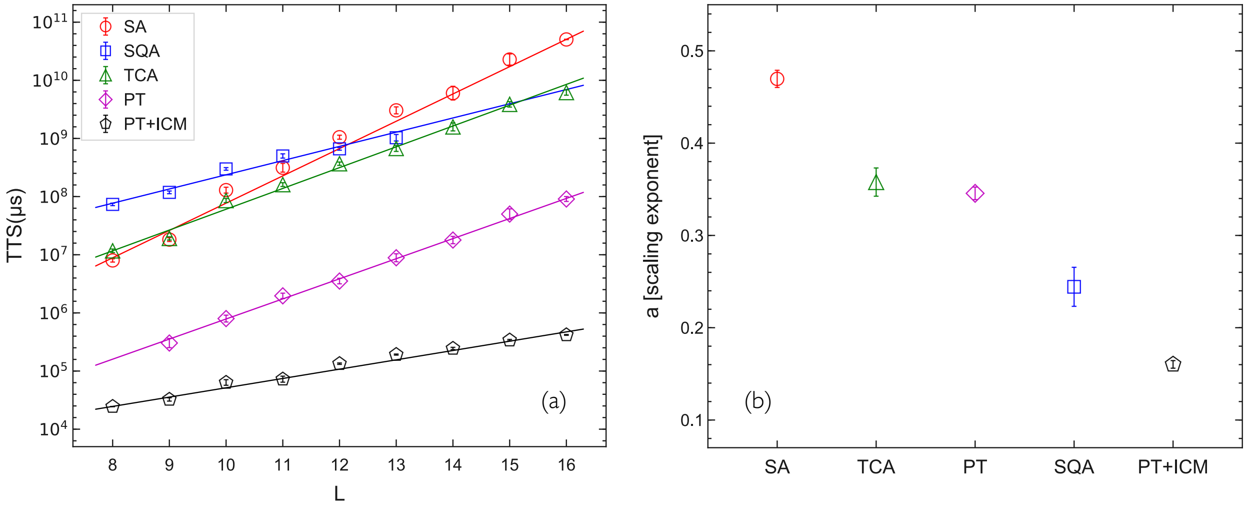

Figure 6.2: Comparison of the scaling results for the studied algorithms. (a): Time to solution (TTS) versus the problem size L for various algorithms. Note that TTS is given in a logarithmic scale. The sizable offsets in PT and PT+ICM are due to the use of a highly optimized code. (b): The scaling exponent a in Eq. (6.4) for various algorithms. TCA scales much better than SA and comparable to PT, while far worse than SQA and PT+ICM. The latter is due to the special structure of the DCL problems that makes them well-suited for SQA and ICM updates.

$N_{\mathrm{s}} = 10$ and $\beta_{\mathrm{f}} = 0.3$. Once again, we observe that there is a global minimum point marking the optimal parameters. This optimization has been done for all other algorithms as well.

In table. 6.1, we have listed the optimal parameters of the thermal cycling algorithm for the studied system sizes. We observe that the algorithm is less sensitive to N_{s} and more so to N_{p}, N_{c}, and β_{f}. In practice, the total effort in a TCA simulation is roughly proportional to $N_{\mathrm{T}}N_{\mathrm{p}}N_{\mathrm{c}}N_{\mathrm{s}}$ with some additional overhead caused by the greedy search. This being said, it is unlikely that any combination of the above parameters can be lumped together to make a reduced set of parameters. For instance, regardless of the number of annealing steps, we must perform a sufficient number of heat-quench cycles such that the pool states can visit many local minima. Also, during each cycle, there must be enough Metropolis updates to stop the local search from stalling the dynamics.

In Fig. 6.2, we show the scaling results. Fig. 6.2(a) demonstrates how the performance of different algorithms scale with the problem size. There is a considerable offset for PT and PT+ICM, which is due to the use of a highly optimized code. Fig. 6.2(b) shows the scaling exponent a–the slope of the linear fit in Fig. 6.2(a). We see that TCA scales much better than SA in agreement with the previous studies. Another interesting finding is that TCA scales almost as well as PT, which has been established as a powerful heuristic in many applications. However, SQA and PT+ICM are still far more efficient than TCA. This can be attributed to the structure of the DCL problems that involve tall but thin barriers that can be easily tunneled through using SQA. The Isoenergetic cluster moves are also well suited for the DCL problems as they cause large rearrangements of the variables, resulting in long leaps across the configuration space thus, efficiently avoiding deep metastable states.

6.5 Conclusion

In this chapter, we reintroduced the thermal cycling algorithm and comprehensively optimized it using the deceptive cluster loop problems for benchmarking and the time to solution as the measure of performance. Our results strongly suggest that TCA can indeed be a competitive heuristic in solving problems with complex structures, as it takes advantage of the repeated heating and cooling to push the system toward the lower energy states while ensuring that it does not get trapped in false ground states. By carefully tuning the parameters, we showed that TCA could be as effective as state-of-the-art algorithms such as PT, while overpowering SA by a great margin. We showed that by reducing the variables among the TCA replicas, the stochastic process could be further accelerated, and the system could be guided more effectively toward the global minimum using the collective memory of the solution pool.

Due to the special structure of the DCL problems, which involve tall yet narrow barriers, SQA and PT+ICM outperform TCA as they utilize quantum effects and cluster updates to bypass those barriers. The true advantage of TCA might be revealed when using dense graphs with broad barriers, where PT+ICM and SQA would naturally struggle. TCA also lends itself to being integrated with the ICM updates because it involves annealing many replicas of the system simultaneously. It has been shown by Ochoa *et al.* [226] that a lower energy state can be generated by overlapping two excited states via an ICM update. Therefore, one interesting addition to TCA can be trying to push the pool states further down in energy by performing a few ICM updates at the end of the algorithm. We hope that we can return to this matter in the future.

7. QUANTUM MACHINE LEARNING USING BIOLOGICAL PROBLEMS

7.1 Introduction

In recent years, there has been great interest in the utility of adiabatic quantum computing (AQC) [8–10] for solving hard optimization problems. One of the most paradigm-shifting developments in this direction is the advent of quantum annealers, such as D-Wave processors [11–14], which attempt to pave the way towards the ultimate quantum computers that are conjectured to solve certain important tasks, for example, factoring of integers [7] much more efficiently than any classical computer using quantum entanglement and tunneling. Therefore, great attention has been paid to problems where *quantum advantage* can be demonstrated [214, 224, 227–233] . One of the offshoots of this endeavor has been quantum machine learning (QML) [234–247], an effort to use quantum architecture to carry out various artificial intelligence (AI) practices hoping to surpass the performance of their classical counterparts. In a recent study by Li *et al.,* [248], machine learning models have been compared to the D-Wave quantum annealer using small scale biological problems. They show that in many cases, the quantum annealer performs as well as any classical machine learning algorithm. The goal of this study is to verify the findings of Li *et al.,* and in some instances, challenge their claims regarding the performance of classical solvers such as parallel tempering. Moreover, we will propose new ideas to examine the true quantum advantage in machine learning applications. First, we will review some preliminary concepts, followed by a thorough description of our analysis. Lastly, we will present the findings of this study and discuss their relevance.

7.2 Classical Machine Learning

Machine learning (ML) is the study of statistical models and algorithms that enables computers to carry out specific tasks without being directly programmed [249]. The term machine learning was first coined by Arthur Samuel [250] at IBM when he first designed an algorithm for playing the game of checkers. Machine learning has come a long way ever since and is an indispensable part of many modern applications such as computer vision, speech and text recognition, social network filtering, medical diagnosis, etc. Succinctly, the holy grail of any machine learning algorithm is to use data in order to build an appropriate mathematical model based on which future predictions can be made. In the first part of a machine learning application, which is referred to as "training", the parameters of the model are tuned by showing many examples to the algorithm such that the model "learns" to recognize patterns, symmetries, correlations, dependencies, etc., within the data. Once the model has been properly trained, the algorithm is 'tested" by performing a certain operation such as classification, using unseen instances of the same type of data. The critical part of the above process is to judiciously choose an appropriate model that can effectively capture as much information as possible from the data. Generally speaking, machine learning algorithms fall into one of the following main categories:

1. *Supervised learning*: in this type of the ML algorithms, the data consists of labeled inputs, and the goal is to train the algorithm such that it can predict the correct labels of the future instances. In other words , supervised learning is where one has input variables $\{X_1, \dots, X_N\}$ and the corresponding output variables $\{Y_1, \dots, Y_N\}$ and we would like to find a mapping function $Y_i = f_{\vec{\omega}}(X_i)$, with parameters $\vec{\omega}$, that can best predict the outputs when given new input variables. Classification and regression are types of supervised learning which are used when the output labels are restricted to a limited set of value, for instance, in speech recognition or when they belong to a continuous range such as in signal processing applications.

2. *Unsupervised learning*: in unsupervised learning, one has only input data $\{X_1, \dots, X_N\}$, and no corresponding output variables, i.e., a set of unlabeled data. The task is to identify symmetries, structures, or distributions within the training data and make predictions based on such commonalities. The most important unsupervised learning algorithm is clustering. Recommender systems, image compression, and anomaly detection (for instance, bank fraud) are some applications that use clustering.

3. *Reinforcement learning*: these types of algorithms are used when exact models are not avail-

able or impractical. In other words, instead of having a cost function, the "agent" is rewarded or punished by the environment on the fly based upon the decisions that it makes, and in this way, the algorithm learns to make adjustments. Reinforcement algorithms are usually used in autonomous vehicles or computer games against a human.

7.3 Quantum Machine Learning

In many of the classical machine learning applications, one deals with extremely high dimensional data, i.e., many features. This type of data is represented on a classical computer as vectors in a multidimensional space. Due to the so-called *curse of dimensionality* caused by the exponential increase in the accessible volume, the data quickly becomes sparse. Therefore, to achieve a reasonable statistical significance, one would need to dd more data to the model, which will eventually result in a memory overflow. The power of quantum computers for data processing is revealed by inspecting the generic state of a qubit which can be written as a linear combinations of the computational basis $|0\rangle$ and $|1\rangle$, namely $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with the normalization constraint $|\alpha|^2 + |\beta|^2 = 1$. Since $\alpha, \beta \in \mathbb{C}$ are continuous variables, an infinite number of states can be encoded to a single qubit. Therefore, we are, in principle, able to map any amount of data to a handful of qubits, which is not possible in a classical computer due to the deterministic nature of the classical memory.

With the above opportunity in hand, the challenge is to find ways to implement conventional machine learning algorithms such as decision trees, neural networks, and support vector machines on a quantum computer. One way is the quantum circuit approach in which one attempts to manipulate the data using a series of quantum gates, i.e., unitary transformations similar to the amplitude amplification algorithms like Grover's search [251]. Another possibility is adiabatic quantum machine learning in which one uses quantum annealing, as discussed in Chapter 2, in order to adiabatically deform an initial quantum state to a desired final state. As an example, we show how the quantum version of the k-means clustering algorithm can be constructed [240]. The goal of k-means clustering is to separate the data points into k distinct classes in which the data points

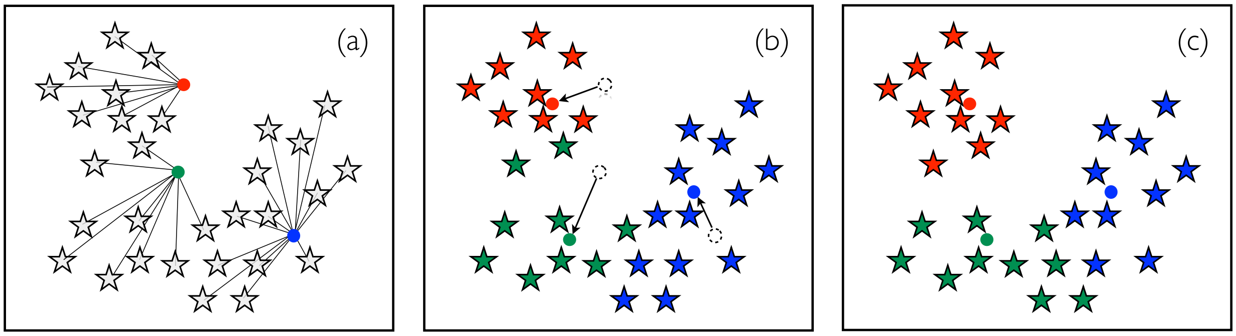

Figure 7.1: Classical k-means clustering algorithm. The goal is to separate the data into k (in this case 3) distinct classes. The spatial adjacency of the points is used as the measure of affinity. (a): k centroids are placed at arbitrary positions. The data points are assigned to k clusters according to their Euclidean distance from the corresponding centroids. (b): The centroids are moved the the "center of mass" of the clusters. (c): The clusters are reconstructed using the new centroids. The processes repeats until convergence.

are clustered based on their spatial adjacency within the feature space. The algorithm, as shown in Fig. 7.1, starts by choosing k centroids at arbitrary positions. The points are then assigned to the closest centroid in terms of their Euclidean distance. Next, the centroids are moved to the center of mass, i.e., $\vec{v}_c = \frac{1}{N_c}\sum\limits_{j\in c}\vec{v}_j$ of each formed cluster c with N_c points. The cycles repeat until no further improvement is possible. Lloyd, Mohseni and Rebentrost [236] first proposed the quantum version of the above clustering algorithm. The idea is to use quantum annealing to perform the following adiabatic reduction:

$$|\phi\rangle = \frac{1}{\sqrt{k}}\sum_{c}^{k}|c\rangle \rightarrow |\psi\rangle = \frac{1}{\sqrt{N}}\sum_{c}^{k}\sum_{j\in c}|c\rangle|j\rangle, \tag{7.1}$$

where k is the number of clusters and N is the total number data points. $|\phi\rangle$ and $|\psi\rangle$ are the ground states of the driver Hamiltonian $\mathcal{H}_1$ and the target Hamiltonian $\mathcal{H}_0$, respectively The instantaneous Hamiltonian $\mathcal{H}(\lambda) = \lambda\mathcal{H}_0 + (1-\lambda)\mathcal{H}_1$ is parameterized by λ which increases continuously from

$\lambda = 0$ to $\lambda = 1$. Lloyd *et al.*, use the following Hamiltonians:

$$\mathcal{H}_1 = 1 - \frac{1}{k}\sum_{c,c'}^{k} |c\rangle\langle c'|, \tag{7.2}$$

$$\mathcal{H}_0 = -\sum_{j_1 \dots j_k} \sum_{l,l'}^{k} |\vec{v}_{j_l} - \vec{v}_{j_{l'}}|^2 |j_1\rangle\langle j_1| \otimes \dots \otimes |j_k\rangle\langle j_k|, \tag{7.3}$$

in which the global minimum of Eq. (7.3) is reached when the points in each cluster are closest to one another and farthest from the points in the adjacent clusters, what a clustering algorithms is ideally supposed to do.

7.4 Details of Analysis

In this section, we lay out the details of our analysis. We use regression, which, as we alluded to earlier, is a supervised learning method. The machine learning studies have been implemented in the python library, Scikit-learn [252].

7.4.1 Data Preparation

The original data is from the Gene Expression Omnibus (GEO) repository of the National Center for Biotechnology Information (NCBI) and can be downloaded under accession number GSE59845 from https://www.ncbi.nlm.nih.gov/geo. The data consists of three genomic-context protein binding microarray (gcPBM) experiments [253] of human transcription factors, namely Myc, Max, and Mad. The DNA sequences such as $\mathrm{GAGCTCGTGAC}\dots$ are 36bp (base pair) long and are quantified by fluorescence intensity as a measure of binding affinity. The data is preprocessed as follows. First, the DNA sequences are trimmed to the central $L = 10$bp, and the binding values of the duplicate sequences are averaged over. Next, the sequences are transformed into a binary format using the following representation.

$$(\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}) = (1000, 0100, 0010, 0001) \tag{7.4}$$

Table 7.1: An example of the DNA sequencing data after preprocessing. The sequences are cut to central 10bp and transformed into binary.

DNA Sequnce	Binary Transform	Max	Mad	Myc
CCCACGCGTG	0100010001001000010000100100001000010010	10991	4423	4359
TCCACGCGCC	0001010001001000010000100100001001000100	8469	2920	2617
TGCATGCGGA	0001001001001000000100100100001000101000	7625	2349	2419
GCCACGCGTC	0010010001001000010000100100001000010100	9033	3550	2776
CCCACATGCA	0100010001001000010010000001001001001000	12277	4458	4534
⋮	⋮	⋮	⋮	⋮

Therefore, after preprocessing, we will have a data set of $\{(\vec{X}_n, y_n)\}$ with $n = 1, 2, \dots, N$ in which $\vec{X}_n$ is a binary array and y_n is the corresponding binding value. Table. 7.1 shows an example of such a data set. Now we split the preprocessed data into training set $\mathcal{D}_{\text{train}}$ with 90% of the original data and the test set $\mathcal{D}_{\text{test}}$ with the remaining 10%.

$$\mathcal{D} = \underbrace{\Big\{(\vec{X}_1, y_1), (\vec{X}_2, y_2), (\vec{X}_3, y_4), \dots,}_{\mathcal{D}_{\text{train}}} \underbrace{(\vec{X}_{N-1}, y_{N-1}), (\vec{X}_N, y_N)\Big\}}_{\mathcal{D}_{\text{test}}}. \tag{7.5}$$

7.4.2 Cost Function

Assuming a linear model $f_{\vec{\omega}} : \vec{X} \to \vec{X}.\vec{\omega} + \omega_0$ the cost function can simply be the sum of squares with a regularization term.

$$E = \sum_{n=1}^{M} \Big(y_n - \vec{X}_n.\vec{\omega} + \omega_0\Big)^2 + \lambda ||\vec{\omega}||_p^q, \tag{7.6}$$

where M is the number of sequences picked from the total of N available sequences. The last term in Eq. (7.6) is a penalty term that prevents overfitting or underfitting where $||\vec{\omega}||_p^q = \left(\sum_{i=1}^{N} |\omega_i|^p\right)^{q/p}$. The integers p and q are chosen based on the estimator used. The weights $\vec{\omega}$ and the intercept term ω_0 must be determined such that the cost function is minimized. It is straightforward to find ω_0

first.

$$\frac{\partial E}{\partial \omega_0} = 0 \quad \Rightarrow \quad \omega_0 = \frac{1}{M}\sum_{n=1}^{M} y_n - \frac{1}{M}\sum_{n=1}^{M} \vec{X}_n.\vec{\omega} = \bar{y} - \bar{\vec{X}}.\omega. \tag{7.7}$$

Substituting Eq. (7.7) in Eq. (7.6) we arrive at the normalized cost function.

$$E = \sum_{n=1}^{M} \left[(y_n - \bar{y}) - \left(\vec{X}_n - \bar{\vec{X}}\right).\vec{\omega}\right]^2 + \lambda||\vec{\omega}||_p^q. \tag{7.8}$$

We use continuous estimators such as Ridge regression and LASSO (least absolute shrinkage and selection operator) in which the weights are real valued. Ridge and LASSO regressions are L^2 and L^1, respectively meaning that

$$E_{\text{Ridge}} = \sum_{n=1}^{M} \left[(y_n - \bar{y}) - \left(\vec{X}_n - \bar{\vec{X}}\right).\vec{\omega}\right]^2 + \lambda||\vec{\omega}||_2^2, \tag{7.9}$$

$$E_{\text{LASSO}} = \sum_{n=1}^{M} \left[(y_n - \bar{y}) - \left(\vec{X}_n - \bar{\vec{X}}\right).\vec{\omega}\right]^2 + \lambda||\vec{\omega}||_1. \tag{7.10}$$

We also use discrete solvers such as SA, SQA, PT, and PT+ICM where binary weights are assumed. In this case, we may further transform the cost function into QUBO form:

$$E = \sum_{i=1}^{4L}\sum_{j=1}^{4L} J_{ij}\omega_i\omega_j + \sum_{i=1}^{4L} h_i\omega_i + E_0, \tag{7.11}$$

where the bonds J_{ij}, the external fields h_i and the constant offset E_0 have the following definitions.

$$\begin{aligned} J_{ij} &= \sum_{n=1}^{M} \left(X_{n,i} - \bar{X}_i\right)\left(X_{n,j} - \bar{X}_j\right), \\ h_i &= \lambda - 2\sum_{n=1}^{M} \left(y_n - \bar{y}\right)\left(X_{n,i} - \bar{X}_i\right), \\ E_0 &= \sum_{n=1}^{M} \left(y_n - \bar{y}\right)^2. \end{aligned} \tag{7.12}$$

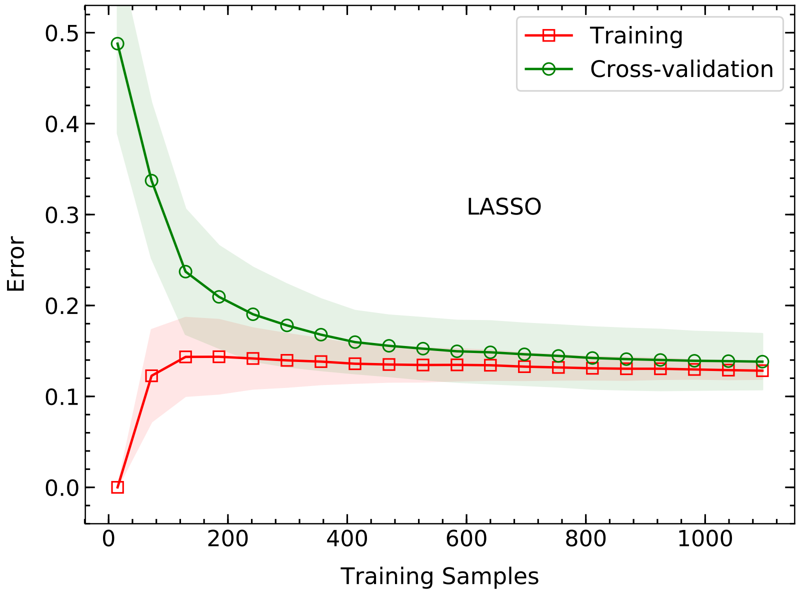

Figure 7.2: Learning curves for the estimator LASSO. The AUPRC scoring is used for classification. The shaded envelopes show the standard deviation. As more training data is used, the model can not fit all the points perfectly; therefore, the training error increases. On the other hand, the validation error decreases because the variation in features of the training and validation sets becomes smaller. The curves converge to a small error value meaning that the bias (underfitting) is small. Also, the gap between the curves tends to close, suggesting small variance (overfitting).

By choosing M sequences randomly form the available N sequences, an *instance* is constructed according to Eq. (7.12) . Note that when binary weights are used, the range of the linear model is bounded by the length of the DNA sequences L. Therefore, the binding values must be normalized in the following way:

$$y_n \to y_n \times \frac{L}{y_{\max}}. \tag{7.13}$$

7.5 Learning Process

Here, we will go over the learning process; first, we determine the regularization parameter via cross-validation, then we train the model over 50 randomly constructed instances, and finally, for each of those instances, we make predictions using the test data set.

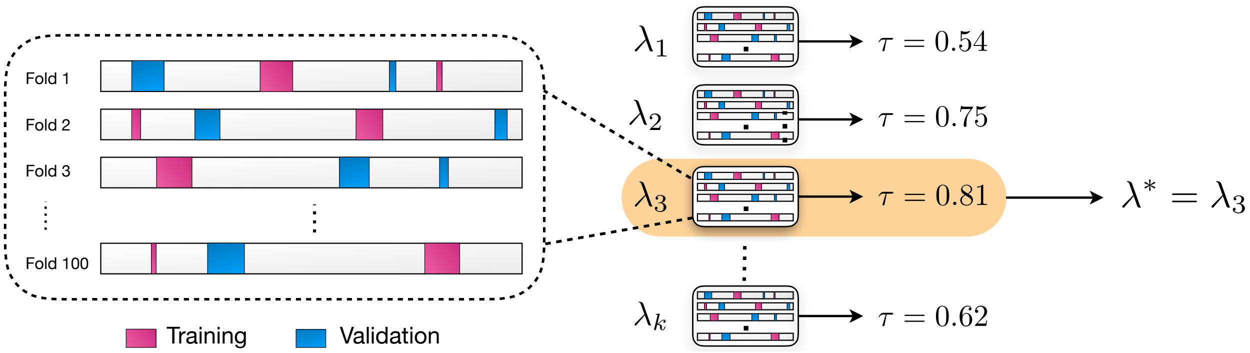

Figure 7.3: The procedure of determining the regularization parameter λ by 100-fold Mote-Carlo cross validation. For each fold, $p\%$ of $\mathcal{D}_{\text{train}}$ is selected randomly as the training and validation data. At a given value of λ, the average score over all 100 folds is calculated. The optimal value of the regularization parameter λ^* will then correspond to the maximum score.

7.5.1 Calibration

As we mentioned earlier, the regularization parameter λ in Eq. (7.8) is introduced to prevent the model from overfitting or underfitting the data. This is essentially done in order to control the complexity of the model. If the model is too simplistic, it can merely capture the features or patterns within the data partially. In this case, the model will be biased, and thus both training and validation scores will suffer as more training samples are used. On the other hand, if the model is too complex, it will seemingly fit the training data well, but in reality, it will be obscured by the noise. As a result, the model will perform poorly when presented with new data during validation. In Fig. 7.2, we show an example of the learning curves for LASSO. As we can see, the learning and validation curves converge to a low error value showing that the model complexity is optimal.

To determine the regularization parameter λ, we use 100-fold Monte-Carlo cross-validation. For each fold the model is trained on randomly chosen $p\%$ ($p = 2, 5, 10, \ldots$) of the training set and validated on another $p\%$. In other words, $M = \frac{p}{100} \times |\mathcal{D}_{\text{train}}|$ in Eqs. (7.8) and (7.12). Each fold gives a score (AUPRC or Kendall's τ), and the overall score is the average of all 100 folds. This process is repeated for various values of λ. The value with the highest score is chosen as the optimal. The schematic diagram in Fig. 7.5.1 shows the above calibration procedure.

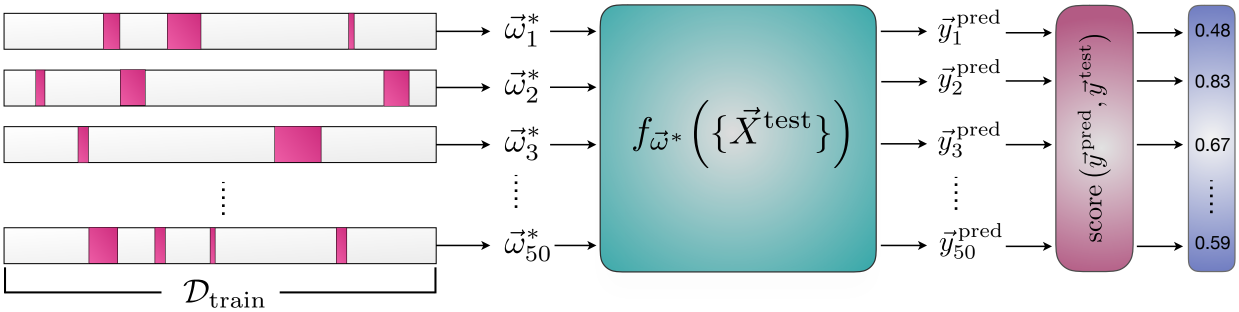

Figure 7.4: Schematic process of training and testing. Multiple instances are prepared from the training data set at random. The resulting ground state weights are used to obtain the prediction binding values $\vec{y}^{\text{pred}}$ given the test sequences $\{\vec{X}^{\text{test}}\}$. Finally, the test scores are obtained by comparing the predictions to the true binding values of the test set.

7.5.2 Evaluation

Having determined the optimal value of the regularization parameter, we now evaluate the predicting performance of the model against the test data set $\mathcal{D}_{\text{test}}$, which has not been seen before. To do the training, we split the training set 50 times with random chunks of size $\frac{p}{100} \times |\mathcal{D}_{\text{train}}|$ where $p = 2, 5, 10, \ldots$. For each resulting instance, the cost function is minimized to give the ground state weights $\vec{\omega}^*$. Next, the resulting ground state weights $\vec{\omega}^*$ and the test sequences $\{\vec{X}^{\text{test}}\}$ are used in the linear model to predict the binding affinity values $\vec{y}^{\text{pred}}$. Lastly, the predictions are compared to the true values $\vec{y}^{\text{test}}$ using the scoring scheme (AUPRC or Kendall's τ in this case) to yield the test scores. One may now gain statistical insights into the predictive power of the model by calculating the mean and the standard deviation of the test scores. The above procedure is shown schematically in Fig. 7.4

7.5.3 Scoring

Two scoring schemes are used in this study: area under precision and recall curve (AUPRC) and Kendall's rank coefficient.

1. *AUPRC:* This is a classification score. Precision-Recall is a useful measure of success of prediction when the classes are very imbalanced. Precision (P) which is a measure of result

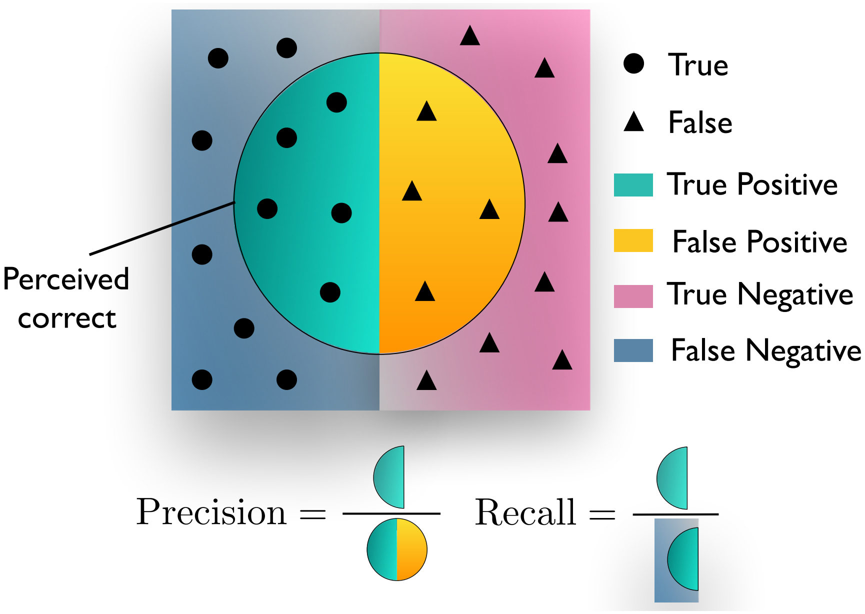

Figure 7.5: Schematic diagram showing precision and recall.

relevancy is defined as the number of true positives (T_P) over the number of true positives plus the number of false positives (F_P).

$$P = \frac{T_P}{T_P + F_P}. \tag{7.14}$$

Recall (R), on the other hand, is a measure of how many truly relevant results are returned and is defined as the number of true positives (T_P) over the number of true positives plus the number of false negatives (F_N).

$$R = \frac{T_P}{T_P + F_N}. \tag{7.15}$$

In Fig. 7.5 we have shown the above definitions in a schematic way. For binary classification, we label the binding values in the test set as 0 or 1 according to the following criterion.

$$\tilde{y}_n^{\text{test}} = \begin{cases} 0 & \text{if} \quad y_n^{\text{test}} < \theta, \\ 1 & \text{if} \quad y_n^{\text{test}} \geq \theta \end{cases} \qquad n = 1, 2, \ldots, |\mathcal{D}_{\text{test}}|, \tag{7.16}$$

in which θ is the binary classification threshold. Given the predictions $y_n^{\text{pred}} = f_{\vec{\omega}^*}(\vec{X}_n^{\text{test}})$ corresponding to the ground state weights $\vec{\omega}^*$ calculated in the training phase, one may perform a binary classification similar to Eq. (7.16).

$$\tilde{y}_n^{\text{pred}} = \begin{cases} 0 \quad \text{if} \quad y_n^{\text{pred}} < c, \\ 1 \quad \text{if} \quad y_n^{\text{pred}} \geq c \end{cases} \qquad n = 1, 2, \ldots, |\mathcal{D}_{\text{test}}|. \tag{7.17}$$

We may now construct the set $\left\{(\tilde{y}_n^{\text{pred}}, \tilde{y}_n^{\text{test}})\right\}$ and calculate precision and recall according to Eq. (7.14) and Eq. (7.15). By varying the threshold c in Eq. (7.17) we obtain a set of points $\{(R, P)\}$ which are then used to form the precision-recall curve. The area under this curve will give he AUPRC score.

2. *Kendall's τ:* Kendall's rank correlation coefficient, commonly referred to as Kendall's τ coefficient, is a statistic used to measure the ordinal association between two measured random quantities. Let $(p_1, q_1), (p_2, q_2), \ldots, (p_M, q_M)$ be a set of M observations of the joint random variables P and Q respectively, such that $\{p_i\}$ and $\{q_i\}$ are unique. Any pair of observations (p_i, q_i) and (p_j, q_j) where $i < j$, are said to be concordant if both $p_i > p_j$ and $q_i > q_j$ or if both $p_i < p_j$ and $q_i < q_j$. More precisely, if the sort order by p and by q agree. They are said to be discordant if $p_i > p_j$ and $q_i < q_j$ or if $p_i < p_j$ and $q_i > q_j$. The Kendall's τ coefficient is then defined as following.

$$\tau = \frac{(\text{number of concordat pairs}) - (\text{number of discordant pairs})}{M(M-1)/2}. \tag{7.18}$$

Here we use Eq. (7.18) to calculate the Kendall's rank coefficient for the set $\{(y_n^{\text{pred}}, y_n^{\text{test}})\}$ in which $y_n^{\text{pred}} = f_{\vec{\omega}^*}(\vec{X}_n^{\text{test}})$ for a given ground state solution $\vec{\omega}^*$ obtained in the training phase.

7.6 Results

In this study, we have investigated the performance of different solvers for various training data sizes. As shown in Fig. 7.6(a), discrete solvers such as SA and PT perform equally well and

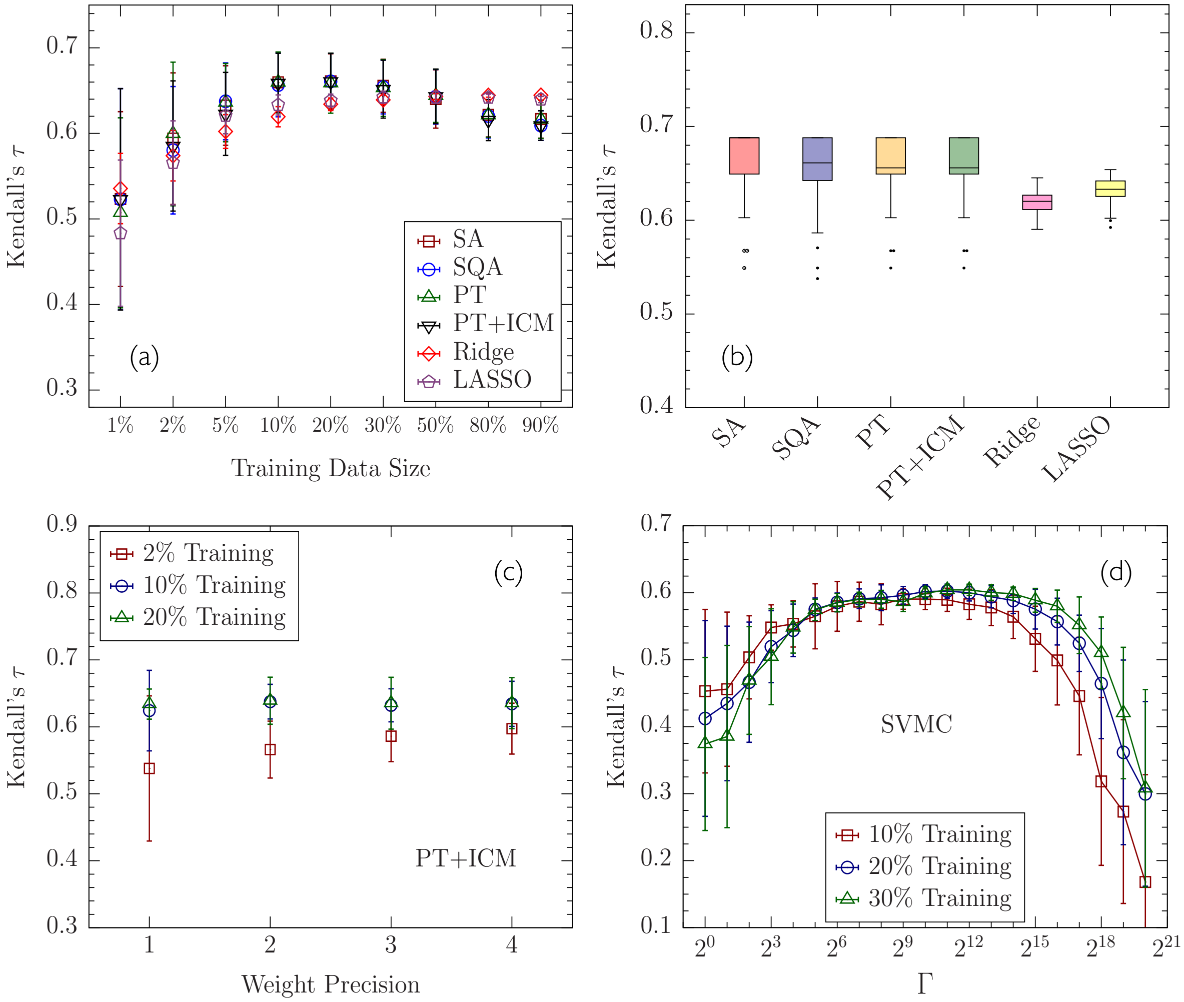

Figure 7.6: Classification performance of the classical and quantum-inspired estimators. (a): Kendall's τ coefficient versus training data size for different solvers. Discrete solvers perform the best for the training data size of 10%. (b): Kendall's τ coefficient for different solvers with a training data size of 10%. All discrete solvers perform similarly and better than continuous ones. (c): Kendall's τ coefficient versus weight precision. Increasing weight precision shows marginal improvement. (d): Kendall's τ coefficient versus the transverse field Γ of the spin-vector Monte Carlo for different training sizes. There is a region where the prediction score is greatest, showing that soft weights are better than hard ones.

better than continuous solvers, for instance, Ridge regression when the training data size is around 10%, as demonstrated more clearly in Fig. 7.6(b). We have also studied the effect of increasing the precision of the weights, which according to Fig. 7.6(c), yields no discernible improvement. We can explain this by noting that spin Hamiltonians such as Eq. (7.11) are always minimized by hard spins. Additionally, we have used the spin-vector Monte Carlo (SVMC) as a solver in the learning process. As we described in Chapter 2, in SVMC the weights are represented as a spin-vector, that is $\vec{\omega}_i = \mathrm{SO}(d,\, \mathbb{R})$ with $d \geq 2$. In two dimensions, the Hamiltonian simplifies to

$$\mathcal{H}_{\mathrm{SVMC}} = \sum_{i,j=1}^{N} J_{ij} \cos\theta_i \cos\theta_j + \sum_{i=1}^{N} h_i \cos\theta_i + \Gamma \sum_{i=1}^{N} \sin\theta_i, \tag{7.19}$$

in which the transverse field Γ determines the "softness" of the weights. When $\Gamma = 0$, the wights simply take values of $\{0, 1\}$ which means that $\mathcal{H}_{\mathrm{SVMC}}$ reduces to (7.11) with hard spins. When $\Gamma \neq 0$, the weights are soft, allowing for a better exploration of the energy landscape. As we see in Fig. 7.6(d), the optimal prediction score happens for a non-zero transverse field, showing that soft spins are favorable. In Ref. [248], it is argued that using the states produced by D-Wave, one finds prediction scores comparable to those of the classical solvers. However, we contend that since these states are taken at the end of the annealing when the system has settled in the ground state of the classical Hamiltonian, the spins are necessarily hard, which is in contradiction with our finding here. We showed in Chapter 2 that SVMC is the semiclassical approximation to a full quantum Hamiltonian where entanglement effects are ignored. This indicates that the introduction of some quantum fluctuations can indeed boost the performance, which can be a potential avenue to demonstrate quantum advantage. Based on our SVMC results, we suggest that sampling must be done during the quantum annealing process when quantum effects are still present.

7.7 Conclusion

In this study, we investigated have investigated the performance of classical as well as quantum-inspired solvers in machine learning applications. We use simple biological data consisting of DNA sequences of the human genome characterized by fluorescence intensity values. The machine learning process was comprised of binary classification using various estimators, including Ridge and LASSO regression, as well as discrete solvers such as SA, PT, and SQA. Our findings show that the latter can indeed perform as well as if not better than conventional estimators like Ridge regression. This can be understood by noting that regression-based estimators often use stochastic gradient descent methods for the optimization, which can become trapped in metastable states if the problem has a rugged energy landscape. This is especially the case for intermediate training data sizes of $10\% - 20\%$, where the constructed instances are sufficiently hard to solve.

Another striking finding that we presented here was the improvement in performance using spin-vector Monte Carlo (SVMC). We showed that the best prediction score is obtained for a non-zero value of the transverse field, which shows that soft variables are more favorable. Since SVMC inherits some quantum properties from the parent transverse Ising spin glass Hamiltonian, we conclude that the learning process might, in fact, benefit from quantum fluctuations. The new D-Wave processor has some novel features that allow the user to prematurely quench the anneal and immediately carry out a readout. Therefore, in light of our SVMC findings here, we propose that unlike the previous studies where sampling was done at the end of the anneal when the states are essentially classical, one should instead use the above quench capabilities to sample from the intermediate quantum states. We intend to do this experiment in the near future.

8. SUMMARY AND CONCLUSIONS

8.1 Summary

This dissertation was dedicated to studying the quadratic unconstrained binary optimization (QUBO) problems that have numerous scientific and industrial applications from finance and traffic routing to drug discovery and artificial intelligence. Many of these problems are considered computationally hard, often getting exponentially more difficult as the problem size grows. Therefore, there is a great demand for heuristics that can solve such problems efficiently. In this work, we have tried to design and optimize new algorithms alongside investigating novel cases of QUBO problems. In chapter 2, we introduced the basic concepts necessary for this study, including complexity, spin glasses, as well as various Monte Carlo techniques. The knowledge of spin-glass physics is instrumental in studying hard QUBO problems since many such problems can be formulated as a spin-glass Hamiltonian. Spin glasses have been scrutinized for over half a century, producing valuable insight into the underlying structure of these deceptively simple yet profoundly rich physical systems. We now know that competing interaction and frustration endow spin glasses with enormously complex energy landscapes thronged with metastable states. Local search heuristics, therefore, tend to spend a considerable amount of time wandering around this maze, rarely ever succeeding in reaching the global minimum.

As we discussed above, it is imperative for any effective QUBO algorithm to explore the phase space of these problems in a rather intelligent fashion. Population Annealing Monte Carlo (PAMC), which we studied in Chapter 3, is an example of such algorithms in which an ensemble of system replicas randomly and simultaneously visit different regions of the phase space, this way increasing the likelihood of finding the true ground state. The crucial part of this algorithm comes when the replicas are resampled according to their relative Boltzmann weight that favors the replicas with lower energy states. This ensures that the average energy of the population is constantly reduced, hence the replicas are collectively guided toward the low-lying states. We performed a fairly comprehensive optimization of the PAMC parameters, where we showed that the annealing schedule equally linear in temperature T and inverse temperature β works better than all other schedules. We also showed that considerable speedup could be obtained by dynamically reducing the population size during the annealing process. Moreover, we developed a massively parallel implementation of population annealing, which proved to be highly nontrivial due to continual communication between the replicas. PAMC, in the fully optimized form that we constructed here, can be a powerful tool in not only solving difficult QUBO problems but also in large-scale simulations of physical systems.

In Chapter 4, we used the optimized population annealing algorithm that was developed in the previous chapter to study an instance of QUBO problems, namely, the three dimensional electron glass, also known as the Coulomb glass (CG) model that is a concept of great interest to condensed matter physics because despite its utter simplicity, elegantly describes many physical properties of systems with quenched disorder. We conducted a comprehensive numerical study of the phase diagram of the CG model. By careful equilibration of the system, we were able to access unprecedented low temperatures, where we demonstrated the existence of a glass phase that was long predicted by the mean-field calculations and observed by numerous experiments yet elusive to numerical studies for many years. Using the four-replica formulation of the spin-glass correlation length as well as novel finite-size scaling analysis techniques, we were able to pinpoint the position of the critical temperature and estimate the critical exponents along the glass phase boundary. Consistent with the previous studies, we also found a charge-ordered (CO) phase in which electrons and holes form a regular antiferromagnetic lattice. This work was significant in the sense that it will settle a long-standing disagreement between the analytical, experimental, and numerical studies.

Next in Chapter 5, we investigated another interesting application of QUBO pertinent to an important problem in seismology, namely, earthquake prediction. By studying the distribution of spin-avalanche sizes in systems with quenched disorder, we showed that one could establish the concept of natural time in such disordered systems similar to that in seismic events. Due to the shortage of accurate measurements, the numerical study of earthquakes is not always possible.

The findings of this work present an alternative way to gain physical insight into the underlying mechanism of earthquakes.

In Chapter 6, we introduced and optimized an algorithm called thermal cycling, which uses repeated heating and cooling of a system to efficiently probe the energy landscape of an optimization problem. We conducted careful benchmarking of this algorithm against state-of-the-art solvers using the deceptive cluster loop (DCL) problems that are specially designed to test the performance of quantum annealers by misleading the classical solvers into metastable states. Our results show that thermal cycling overpowers simulated annealing and performs as well as parallel tempering, proving that it can be a competitive tool in solving hard QUBO problems.

Finally in Chapter 7, we studied another novel application of QUBO, i.e., quantum machine learning (QML). QML is an effort to process data faster and find nontrivial patters in it using the power of quantum architecture such as quantum annealers as a potential path to demonstrate quantum advantage. We used biological samples obtained from the human genome experiment and employed classical solvers such as SA and PT, as well as quantum-inspired algorithms such as SQA in the learning process. Our results show that these algorithms, indeed, perform better than conventional machine learning heuristics, for example Ridge regression, for binary classification. We also used spin vector Monte Carlo (SVMC) in this study, where we showed that the optimal prediction performance occurred for a finite value of the transverse field, meaning that soft variables are more favorable. Since SVMC is the semi-classical approximation of a full quantum Hamiltonian, we propose that it is likely that quantum advantage can be achieved if one samples form the true quantum states of a quantum annealer by quenching the anneal, a feature which is currently feasible on the D-Wave processors.

8.2 Impacts and Applications

In this work, we illustrated the broad utility of quadratic optimization; from electron glasses to quantum machine learning. The algorithms that we developed, such as population annealing and thermal cycling, are highly competitive implements not only in solving NP-hard problems with industry-wide applications but also for numerical simulations in scientific endeavors. By studying the CG model, we numerically showed that glassy properties emerge in strongly interacting electronic systems with a quenched disorder, which can be realized as doped semiconductors with many technological applications. Studying spin avalanche distributions in spin glasses yielded a striking correspondence to the concept of natural time commonly used in geological sciences as a potential criterion for predicting large seismic events. The work on quantum machine learning showed that QUBO solvers, such as the ones we developed here, can, in fact, be used to great effect in artificial intelligence as they are less susceptible to break down when the problem has an extremely complex structure. We also suggested a potential avenue to demonstrate quantum advantage–a matter of strategic interest to industry leaders as well as great curiosity to the scientific community–by using the intermediate states of a quantum annealer.

8.3 Future Directions

Our work on the Coulomb glass model was focused on three space dimensions in which the key to the emergence of a glass phase, as opposed to the random field Ising model, was the presence of long-range interactions and charge conservation. These constraints can also be applied to planar charges, where the Coulomb inverse square law is still valid. It would be interesting to see whether or not our findings can be extended to this case. We studied the thermal cycling algorithm using the DCL problems with thin and tall energy barriers, where, as we showed, SQA and PT+ICM were far more superior. Alternatively, one can use other problems such as four-dimensional spin-glass instances, where the above algorithms struggle due to the increased barrier widths and cluster percolation, respectively. With a benchmarking similar to what we did in this work, we propose to investigate the scaling of the thermal cycling algorithm using the new problem set. As we suggested in the last chapter, our SVMC results indicate that one might be able to observe improved performance by implementing machine learning algorithms using the quench feature of the D-wave quantum annealer. We plan to visit this conjecture in the near future. Should this prove to be correct, a strong argument can be put forth for a verifiable case of quantum advantage.

REFERENCES

[1] R. M. Karp, *Complexity of Computer Computations*, p. 85. New York: Plenum, 1972.

[2] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*. San Francisco: Freeman, 1979.

[3] A. Lucas, “Ising formulations of many NP problems,” *Front. Physics*, vol. 12, p. 5, 2014.

[4] S. F. Edwards and P. W. Anderson, “Theory of spin glasses,” *J. Phys. F: Met. Phys.*, vol. 5, p. 965, 1975.

[5] K. Binder and A. P. Young, “Spin Glasses: Experimental Facts, Theoretical Concepts and Open Questions,” *Rev. Mod. Phys.*, vol. 58, p. 801, 1986.

[6] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.

[7] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM J. Comp.*, vol. 26, p. 1484, 1997.

[8] 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, p. 472, 2001.

[9] A. Das and B. K. Chakrabarti, “Quantum Annealing and Analog Quantum Computation,” *Rev. Mod. Phys.*, vol. 80, p. 1061, 2008.

[10] T. Albash and D. A. Lidar, “Adiabatic Quantum Computing,” 2016. (arXiv:1611.04471).

[11] D-Wave Systems Inc. The D-Wave 2X Quantum Computer.

[12] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose, “Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor,” *Phys. Rev. B*, vol. 82, p. 024511, Jul 2010.

[13] 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, p. 194, 2011.

[14] P. Bunyk, E. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, and J. Whittaker, "Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor," *IEEE Trans. Appl. Supercond.*, vol. 24, p. 1, 2014.

[15] J. Carlson, A. Jaffe, and A. Wiles, "The millennium prize problems," *American Mathematical Society*, pp. 88–104, 2000.

[16] W. Dean, "Computational complexity theory," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, winter 2016 ed., 2016.

[17] T. Castellani and A. Cavagna, "Spin-glass theory for pedestrians," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, p. P05012, May 2005.

[18] C. de Dominicis and I. Giardina, *Random Fields and Spin Glasses*. Cambridge: Cambridge University Press, 2006.

[19] W. Feller, "The fundamental limit theorems in probability," *Bull. Amer. Math. Soc.*, vol. 51, pp. 800–832, 11 1945.

[20] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin glass," *Phys. Rev. Lett.*, vol. 35, p. 1792, 1975.

[21] R. L. Stratonovich, "On a Method of Calculating Quantum Distribution Functions," *Soviet Physics Doklady*, vol. 2, p. 416, Jul 1957.

[22] J. Hubbard, "Calculation of Partition Functions," *Phys. Rev. Lett.*, vol. 3, pp. 77–78, Jul 1959.

[23] J. R. L. de Almeida and D. J. Thouless, "Stability of the Sherrington-Kirkpatrick solution of a spin glass model," *J. Phys. A*, vol. 11, p. 983, 1978.

[24] Blandin, A., “Theories versus experiments in the spin glass systems,” *J. Phys. Colloques*, vol. 39, pp. C6–1499–C6–1516, 1978.

[25] G. Parisi, “Infinite number of order parameters for spin-glasses,” *Phys. Rev. Lett.*, vol. 43, p. 1754, 1979.

[26] G. Parisi, “The order parameter for spin glasses: a function on the interval 0–1,” *J. Phys. A*, vol. 13, p. 1101, 1980.

[27] G. Parisi, “A sequence of approximated solutions to the S-K model for spin glasses,” *J. Phys. A*, vol. 13, p. L115, 1980.

[28] A. H. Land and A. G. Doig, “An automatic method of solving discrete programming problems,” *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.

[29] P. A. Gagniuc, ed., *Markov Chains: From Theory to Implementation and Experimentation*. USA, NJ: John Wiley and Sons, 2017.

[30] A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen,” *Annalen der Physik*, vol. 322, pp. 549–560, Jan 1905.

[31] F. J. Aragón Artacho, D. H. Bailey, J. M. Borwein, and P. B. Borwein, “Walking on real numbers,” *The Mathematical Intelligencer*, vol. 35, pp. 42–60, Mar 2013.

[32] L. Boltzmann, “Vorlesungen über Gastheorie,” *Leipzig, J. A. Barth*, 1896.

[33] L. Boltzmann, *Lectures on gas theory*. Berkeley, CA, USA: University of of California Press, 1964.

[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *J. Chem. Phys.*, vol. 21, p. 1087, 1953.

[35] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, pp. 97–109, 04 1970.

[36] R. H. Swendsen and J.-S. Wang, “Replica Monte Carlo simulation of spin-glasses,” *Phys. Rev. Lett.*, vol. 57, p. 2607, 1986.

[37] C. Geyer, “Monte Carlo Maximum Likelihood for Dependent Data,” in *23rd Symposium on*

the Interface (E. M. Keramidas, ed.), (Fairfax Station, VA), p. 156, Interface Foundation, 1991.

[38] K. Hukushima and K. Nemoto, “Exchange Monte Carlo method and application to spin glass simulations,” *J. Phys. Soc. Jpn.*, vol. 65, p. 1604, 1996.

[39] D. J. Earl and M. W. Deem, “Parallel Tempering: Theory, Applications, and New Perspectives,” *Phys. Chem. Chem. Phys.*, vol. 7, p. 3910, 2005.

[40] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, p. 671, 1983.

[41] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *IEEE Trans. Pattern. Analy. Mach. Intell.*, vol. PAMI-6, p. 721, 1984.

[42] T. Kadowaki and H. Nishimori, “Quantum annealing in the transverse Ising model,” *Phys. Rev. E*, vol. 58, p. 5355, 1998.

[43] M. Born and V. Fock, “Beweis des adiabatensatzes,” *Zeitschrift für Physik*, vol. 51, pp. 165–180, Mar 1928.

[44] M. Suzuki, *Quantum Monte Carlo Methods in Condensed Matter Physics*. Singapore: World Scientific, 1993.

[45] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2000.

[46] M. Troyer, F. Alet, S. Trebst, and S. Wessel, “Non-local Updates for Quantum Monte Carlo Simulations,” in *AIP Conf. Proc. 690: The Monte Carlo Method in the Physical Sciences*, pp. 156–169, 2003.

[47] A. Das and B. K. Chakrabarti, *Quantum Annealing and Related Optimization Methods*. Edited by A. Das and B.K. Chakrabarti, Lecture Notes in Physics 679, Berlin: Springer, 2005.

[48] H. F. Trotter, “On the product of semi-groups of operators,” *Proc. Amer. Math. Soc*, vol. 10, pp. 545–551, 1959.

[49] M. Suzuki, “Relationship among Exactly Soluble Models of Critical Phenomena. I: 2D Ising

Model, Dimer Problem and the Generalized XY-Model," *Progress of Theoretical Physics*, vol. 46, pp. 1337–1359, 11 1971.

[50] 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, pp. 1454–1469, 11 1976.

[51] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Z. Phys.*, vol. 31, p. 253, 1925.

[52] A. Barzegar, C. Pattison, W. Wang, and H. G. Katzgraber, "Optimization of population annealing Monte Carlo for large-scale spin-glass simulations," *Phys. Rev. E*, vol. 98, p. 053308, 2018.

[53] K. Huang, *Statistical Mechanics*. New York: Wiley, 1987.

[54] D. L. Stein and C. M. Newman, *Spin Glasses and Complexity*. Primers in Complex Systems, Princeton NJ: Princeton University Press, 2013.

[55] H. G. Katzgraber, M. Körner, F. Liers, M. Jünger, and A. K. Hartmann, "Universality-class dependence of energy distributions in spin glasses," *Phys. Rev. B*, vol. 72, p. 094421, 2005.

[56] W. Wang, J. Machta, and H. G. Katzgraber, "Comparing Monte Carlo methods for finding ground states of Ising spin glasses: Population annealing, simulated annealing, and parallel tempering," *Phys. Rev. E*, vol. 92, p. 013303, 2015.

[57] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber, "Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension," *Phys. Rev. Lett.*, vol. 115, p. 077201, 2015.

[58] K. Hukushima and Y. Iba, "Population Annealing and Its Application to a Spin Glass," in *The Monte Carlo method in the physical sciences: celebrating the 50th anniversary of the Metropolis algorithm* (J. E. Gubernatis, ed.), vol. 690, (Los Alamos, New Mexico (USA)), p. 200, AIP, 2003.

[59] E. Zhou and X. Chen, "A new population-based simulated annealing algorithm," in *Proceedings of the 2010 Winter Simulation Conference (WSC)*, (New York), p. 1211, Springer, 2010.

[60] J. Machta, "Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes," *Phys. Rev. E*, vol. 82, p. 026704, 2010.

[61] W. Wang, J. Machta, and H. G. Katzgraber, "Population annealing: Theory and application in spin glasses," *Phys. Rev. E*, vol. 92, p. 063307, 2015.

[62] W. Wang, J. Machta, and H. G. Katzgraber, "Evidence against a mean-field description of short-range spin glasses revealed through thermal boundary conditions," *Phys. Rev. B*, vol. 90, p. 184412, 2014.

[63] Borovský, M. and Weigel, M. and Barash, Lev Yu. and Žukovič, M., "GPU-Accelerated Population Annealing Algorithm: Frustrated Ising Antiferromagnet on the Stacked Triangular Lattice," *EPJ Web of Conferences*, vol. 108, p. 02016, 2016.

[64] Barash, Lev Yu. and Weigel, M. and Shchur, Lev N. and Janke, W., "Exploring first-order phase transitions with population annealing," *Eur. Phys. J. Special Topics*, vol. 226, no. 4, p. 595, 2017.

[65] J. Callaham and J. Machta, "Population Annealing Simulations of a Binary Hard Sphere Mixture," 2017. (arXiv:1701.00263).

[66] A. Barzegar, J. C. Andresen, M. Schechter, and H. G. Katzgraber, "Numerical observation of a glassy phase in the three-dimensional coulomb glass," *Phys. Rev. B*, vol. 100, p. 104418, Sep 2019.

[67] C. Amey and J. Machta, "Analysis and optimization of population annealing," *Phys. Rev. E*, vol. 97, p. 033301, 2018.

[68] U. Wolff, "Collective Monte Carlo updating for spin systems," *Phys. Rev. Lett.*, vol. 62, p. 361, 1989.

[69] J. Houdayer, "A cluster Monte Carlo algorithm for 2-dimensional spin glasses," *Eur. Phys. J. B.*, vol. 22, p. 479, 2001.

[70] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber, "Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension," 2015. (cond-mat/1501.05630).

[71] Barash, Lev Yu. and Weigel, M. and Borovský, M. and Janke, W. and Shchur, Lev N., "GPU

accelerated population annealing algorithm," 2017. (arXiv:1703.03676).

[72] R. R. P. Singh and S. Chakravarty, "Critical behavior of an Ising spin-glass," *Phys. Rev. Lett.*, vol. 57, p. 245, 1986.

[73] H. G. Katzgraber, M. Körner, and A. P. Young, "Universality in three-dimensional Ising spin glasses: A Monte Carlo study," *Phys. Rev. B*, vol. 73, p. 224432, 2006.

[74] A. M. Ferrenberg and R. H. Swendsen, "New Monte Carlo technique for studying phase transitions," *Phys. Rev. Lett.*, vol. 61, p. 2635, 1988.

[75] D. A. Kessler and M. Bretz, "Unbridled growth of spin-glass clusters," *Phys. Rev. B*, vol. 41, p. 4778, 1990.

[76] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, NIPS'12, (Lake Tahoe, Nevada, USA), p. 2951, Curran Associates Inc., 2012.

[77] R. P. Adams, M. Gelbart, and J. Snoek, "Spearmint," *Git Repository, github.com/HIPS/Spearmint, commit ffbab66*, 2016.

[78] K. Binder, "Critical properties from Monte Carlo coarse graining and renormalization," *Phys. Rev. Lett.*, vol. 47, p. 693, 1981.

[79] M. Ney-Nifle and A. P. Young, "Chaos in a two-dimensional Ising spin glass," *J. Phys. A*, vol. 30, p. 5311, 1997.

[80] T. Aspelmeier, A. J. Bray, and M. A. Moore, "Why Temperature Chaos in Spin Glasses Is Hard to Observe," *Phys. Rev. Lett.*, vol. 89, p. 197202, 2002.

[81] P. E. Jönsson, H. Yoshino, and P. Nordblad, "Symmetrical Temperature-Chaos Effect with Positive and Negative Temperature Shifts in a Spin Glass," *Phys. Rev. Lett.*, vol. 89, p. 097201, 2002.

[82] H. G. Katzgraber and F. Krzakala, "Temperature and Disorder Chaos in Three-Dimensional Ising Spin Glasses," *Phys. Rev. Lett.*, vol. 98, p. 017201, 2007.

[83] L. A. Fernandez, V. Martin-Mayor, G. Parisi, and B. Seoane, "Temperature chaos in 3D

Ising spin glasses is driven by rare events," *Europhys. Lett.*, vol. 103, no. 6, p. 67003, 2013.

[84] W. Wang, J. Machta, and H. G. Katzgraber, "Chaos in spin glasses revealed through thermal boundary conditions," *Phys. Rev. B*, vol. 92, p. 094410, 2015.

[85] Z. Zhu, A. J. Ochoa, F. Hamze, S. Schnabel, and H. G. Katzgraber, "Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities," *Phys. Rev. A*, vol. 93, p. 012317, 2016.

[86] L. Y. Barash, M. Weigel, M. Borovský, W. Janke, and L. N. Shchur, "GPU accelerated population annealing algorithm," *Comp. Phys. Comm.*, vol. 220, p. 341, 2017.

[87] P. W. Anderson, "Absence of Diffusion in Certain Random Lattices," *Phys. Rev.*, vol. 109, p. 1492, 1958.

[88] M. Pollak, "Effect of carrier-carrier interactions on some transport properties in disordered semiconductors," *Disc. Faraday Soc.*, vol. 50, p. 13, 1970.

[89] G. Srinivasan, "Statistical Mechanics of Charged Traps in an Amorphous Semiconductor," *Phys. Rev. B*, vol. 4, p. 2581, 1971.

[90] A. L. Efros and B. I. Shklovskii, "Coulomb gap and low temperature conductivity of disordered systems ," *J. Phys. C*, vol. 8, p. L49, 1975.

[91] M. Grünewald, B. Pohlmann, L. Schweitzer, and D. Würtz, "Mean field approach to the electron glass," *J. Phys. C*, vol. 15, p. 1153, 1982.

[92] M. Pollak, "The Coulomb gap: a Review and New Developments," *Phil. Mag. B*, vol. 65, p. 657, 1975.

[93] T. Vojta, "Spherical random-field systems with long-range interactions: general results and application to the Coulomb glass," *J. Phys. A*, vol. 26, p. 2883, 1993.

[94] A. A. Pastor and V. Dobrosavljević , "Melting of the Electron Glass," *Phys. Rev. Lett.*, vol. 83, p. 4642, 1999.

[95] A. A. Pastor, V. Dobrosavljević, and M. L. Horbach, "Mean-field glassy phase of the random-field Ising model," *Phys. Rev. B*, vol. 66, p. 014413, 2002.

[96] V. Dobrosavljević, D. Tanasković, and A. A. Pastor, "Glassy Behavior of Electrons Near

Metal-Insulator Transitions," *Phys. Rev. Lett.*, vol. 90, p. 016402, 2003.

[97] M. Müller and L. B. Ioffe, "Glass Transition and the Coulomb Gap in Electron Glasses," *Phys. Rev. Lett.*, vol. 93, p. 256403, 2004.

[98] S. Pankov and V. Dobrosavljević, "Nonlinear Screening Theory of the Coulomb Glass," *Phys. Rev. Lett.*, vol. 94, p. 046402, 2005.

[99] E. Lebanon and M. Müller, "Memory effect in electron glasses: Theoretical analysis via a percolation approach," *Phys. Rev. B*, vol. 72, p. 174202, 2005.

[100] A. Amir, Y. Oreg, and Y. Imry, "Mean-field model for electron-glass dynamics," *Phys. Rev. B*, vol. 77, p. 165207, 2008.

[101] D. Monroe, A. C. Gossard, G. B. English, J. H., W. H. Haemmerle, and M. Kastner, "Long Lived Coulomb Gap in a Compensated Semiconductor - The Electron Glass," *Phys. Rev. Lett.*, vol. 59, p. 1148, 1987.

[102] M. Ben-Chorin, Z. Ovadyahu, and M. Pollak, "Nonequilibrium transport and slow relaxation in hopping conductivity," *Phys. Rev. B*, vol. 48, p. 15025, 1993.

[103] J. G. Massey and M. Lee, "Direct Observation of the Coulomb Correlation Gap in a Nonmetallic Semiconductor, Si:B," *Phys. Rev. Lett.*, vol. 75, p. 4266, 1995.

[104] Z. Ovadyahu and M. Pollak, "Disorder and Magnetic Field Dependence of Slow Electronic Relaxation," *Phys. Rev. Lett.*, vol. 79, p. 459, 1997.

[105] G. Martinez-Arizala, C. Christiansen, D. E. Grupp, N. Markovic, A. M. Mack, and A. M. Goldman, "Coulomb-glass-like behavior of ultrathin films of metals," *Phys. Rev. B*, vol. 57, p. R670, 1998.

[106] A. Vaknin, Z. Ovadyahu, and M. Pollak, "Aging Effects in an Anderson Insulator," *Phys. Rev. Lett.*, vol. 84, p. 3402, 2000.

[107] S. Bogdanovich and D. Popovic, "Onset of Glassy Dynamics in a Two-Dimensional Electron System in Silicon," *Phys. Rev. Lett.*, vol. 88, p. 236401, 2002.

[108] A. Vaknin, Z. Ovadyahu, and M. Pollak, "Nonequilibrium field effect and memory in the electron glass," *Phys. Rev. B*, vol. 65, p. 134208, 2002.

[109] V. Orlyanchik and Z. Ovadyahu, "Stress Aging in the Electron Glass," *Phys. Rev. Lett.*, vol. 92, p. 066801, 2004.

[110] H. E. Romero and M. Drndic, "Coulomb Blockade and Hopping Conduction in PbSe Quantum Dots," *Phys. Rev. Lett.*, vol. 95, p. 156801, 2005.

[111] J. Jaroszyński and D. Popović, "Nonexponential Relaxations in a Two-Dimensional Electron System in Silicon," *Phys. Rev. Lett.*, vol. 96, p. 037403, 2006.

[112] T. Grenet, J. Delahaye, M. Sabra, and F. Gay, "Anomalous electric-field effect and glassy behaviour in granular aluminium thin films: electron glass?," *Euro. Phys. J. B*, vol. 56, p. 183, 2007.

[113] Z. Ovadyahu, "Relaxation Dynamics in Quantum Electron Glasses," *Phys. Rev. Lett.*, vol. 99, p. 226603, 2007.

[114] I. Raičević, J. Jaroszyński, D. Popović, C. Panagopoulos, and T. Sasagawa, "Evidence for Charge Glasslike Behavior in Lightly Doped $\mathrm{La}_{2-x}\mathrm{Sr}_x\mathrm{CuO}_4$ at Low Temperatures," *Phys. Rev. Lett.*, vol. 101, p. 177004, 2008.

[115] I. Raičević, D. Popović, C. Panagopoulos, and T. Sasagawa, "Non-Gaussian noise in the in-plane transport of lightly doped $\mathrm{La}_{2-x}\mathrm{Sr}_x\mathrm{CuO}_4$: Evidence for a collective state of charge clusters," *Phys. Rev. B*, vol. 83, p. 195133, 2011.

[116] M. Müller and S. Pankov, "Mean-field theory for the three-dimensional Coulomb glass," *Phys. Rev. B*, vol. 75, p. 144201, 2007.

[117] E. R. Grannan and C. C. Yu, "Critical behavior of the Coulomb glass," *Phys. Rev. Lett.*, vol. 71, p. 3335, 1993.

[118] T. Vojta and M. Schreiber, "Comment on "Critical behavior of the Coulomb glass"," *Phys. Rev. Lett.*, vol. 73, p. 2933, 1994.

[119] S. D. Baranovskii, A. L. Efros, B. L. Gelmont, and B. I. Shklovskii, "Coulomb gap in disordered systems: computer simulation ," *J. Phys. C*, vol. 12, p. 1023, 1979.

[120] J. H. Davies, P. A. Lee, and T. M. Rice, "Electron glass," *Phys. Rev. Lett.*, vol. 49, p. 758, 1982.

[121] J. H. Davies, P. A. Lee, and T. M. Rice, “Properties of the electron glass,” *Phys. Rev. B*, vol. 29, p. 4260, 1984.

[122] W. Xue and P. A. Lee, “Monte Carlo simulations of the electron glass,” *Phys. Rev. B*, vol. 38, p. 9093, 1988.

[123] A. Möbius, M. Richter, and B. Drittler, “Coulomb gap in two- and three-dimensional systems: Simulation results for large samples,” *Phys. Rev. B*, vol. 45, p. 11568, 1992.

[124] Q. Li and P. Phillips, “Unexpected activated temperature dependence of the conductance in the presence of a soft Coulomb gap in three dimensions,” *Phys. Rev. B*, vol. 49, p. 10269, 1994.

[125] M. Sarvestani, M. Schreiber, and T. Vojta, “Coulomb Gap at Finite Temperatures,” *Phys. Rev. B*, vol. 52, p. R3820, 1995.

[126] T. Wappler, T. Vojta, and M. Schreiber, “Monte Carlo Simulations of the Dynamical Behavior of the Coulomb Glass,” *Phys. Rev. B*, vol. 55, p. 6272, 1997.

[127] A. Diaz-Sanchez, A. Moebius, M. Ortuno, A. Neklioudov, and M. Schreiber, “Non-ergodic effects in the Coulomb glass: Specific heat,” *Phys. Rev. B*, vol. 62, p. 8030, 2000.

[128] B. Sandow, K. Gloos, R. Rentzsch, A. N. Ionov, and W. Schirmacher, “Electronic Correlation Effects and the Coulomb Gap at Finite Temperature,” *Phys. Rev. Lett.*, vol. 86, p. 1845, 2001.

[129] D. Grempel, “Off-equilibrium dynamics of the two-dimensional Coulomb Glass,” *Europhys. Lett.*, vol. 66, p. 854, 2004.

[130] M. H. Overlin, L. A. Wong, and C. C. Yu, “Effect of increasing disorder on the critical behavior of a Coulomb system,” *Phys. Rev. B*, vol. 70, p. 214203, 2004.

[131] A. B. Kolton, D. R. Grempel, and D. Domínguez, “Heterogeneous dynamics of the three-dimensional Coulomb glass out of equilibrium,” *Phys. Rev. B*, vol. 71, p. 024206, 2005.

[132] A. Glatz, V. M. Vinokur, and Y. M. Galperin, “Statistics of Deep Energy States in Coulomb Glasses,” *Phys. Rev. Lett.*, vol. 98, p. 196401, 2007.

[133] M. Goethe and M. Palassini, “Phase Diagram, Correlation Gap, and Critical Properties of

the Coulomb Glass," *Phys. Rev. Lett.*, vol. 103, p. 045702, 2009.

[134] B. Surer, H. G. Katzgraber, G. T. Zimanyi, B. A. Allgood, and G. Blatter, "Density of States and Critical Behavior of the Coulomb Glass," *Phys. Rev. Lett.*, vol. 102, p. 067205, 2009.

[135] A. Möbius and U. K. Rössler, "Critical behavior of the Coulomb-glass model in the zero-disorder limit: Ising universality in a system with long-range interactions," *Phys. Rev. B*, vol. 79, p. 174206, 2009.

[136] M. Palassini and M. Goethe, "Elementary excitations and avalanches in the Coulomb glass," *J. Phys.: Conf. Ser.*, vol. 376, p. 012009, 2012.

[137] J. Rehn, R. Moessner, and A. P. Young, "Spin glass behavior in a random Coulomb antiferromagnet," *Phys. Rev. E*, vol. 94, p. 032124, 2016.

[138] M. Goethe and M. Palassini, "Avalanches in the Relaxation Dynamics of Electron Glasses." (arXiv:1808.01466), 2018.

[139] A. P. Young and H. G. Katzgraber, "Absence of an Almeida-Thouless line in Three-Dimensional Spin Glasses," *Phys. Rev. Lett.*, vol. 93, p. 207203, 2004.

[140] H. G. Katzgraber and A. P. Young, "Probing the Almeida-Thouless line away from the mean-field model," *Phys. Rev. B*, vol. 72, p. 184416, 2005.

[141] V. Malik and D. Kumar, "Thermodynamics and excitations of Coulomb glass," *Phys. Rev. B*, vol. 76, p. 125207, 2007.

[142] E. Wigner, "On the Interaction of Electrons in Metals," *Phys. Rev.*, vol. 46, p. 1002, 1934.

[143] Y. Pramudya, H. Terletska, S. Pankov, E. Manousakis, and V. Dobrosavljevic, "Nearly frozen coulomb liquids," *Phys. Rev. B*, vol. 84, p. 125120, 2011.

[144] P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[145] S. W. de Leeuw, J. W. Perram, and E. R. Smith, "Simulation of Electrostatic Systems in Periodic Boundary Conditions I. Lattice Sums and Dielectric Constants," *Proc. R. Soc. A*, vol. 373, p. 27, 1980.

[146] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas,*

Graphs, and Mathematical Tables. New York: Dover, 1964.

[147] F. Cooper, B. Freedman, and D. Preston, "Solving $\phi^4_{1,2}$ theory with Monte Carlo," *Nucl. Phys. B*, vol. 210, p. 210, 1982.

[148] M. Palassini and S. Caracciolo, "Universal Finite-Size Scaling Functions in the 3D Ising Spin Glass," *Phys. Rev. Lett.*, vol. 82, p. 5128, 1999.

[149] H. G. Ballesteros, A. Cruz, L. A. Fernandez, V. Martin-Mayor, J. Pech, J. J. Ruiz-Lorenzo, A. Tarancon, P. Tellez, C. L. Ullod, and C. Ungil, "Critical behavior of the three-dimensional Ising spin glass," *Phys. Rev. B*, vol. 62, p. 14237, 2000.

[150] A. K. Hartmann and A. P. Young, "Specific-Heat Exponent of Random-Field Systems via Ground-State Calculations," *Phys. Rev. B*, vol. 64, p. 214419, 2001.

[151] G. Grinstein, "Ferromagnetic Phase Transitions in Random Fields: The Breakdown of Scaling Laws," *Phys. Rev. Lett.*, vol. 37, p. 944, 1976.

[152] A. J. Bray and M. A. Moore, "Critical behavior of the three-dimensional Ising spin glass," *Phys. Rev. B*, vol. 31, p. 631, 1985.

[153] D. S. Fisher, "Scaling and critical slowing down in random-field Ising systems," *Phys. Rev. Lett.*, vol. 56, p. 416, 1986.

[154] A. A. Middleton and D. S. Fisher, "Three-dimensional random-field Ising magnet: Interfaces, scaling, and the nature of states," *Phys. Rev. B*, vol. 65, p. 134411, 2002.

[155] N. G. Fytas and V. Martin-Mayor, "Universality in the Three-Dimensional Random-Field Ising Model," *Phys. Rev. Lett.*, vol. 110, p. 227201, 2013.

[156] N. G. Fytas, P. E. Theodorakis, I. Georgiou, and I. Lelidis, "Critical aspects of the random-field Ising model," *Eur. Phys. J. B*, vol. 86, p. 268, 2013.

[157] B. Ahrens, J. Xiao, A. K. Hartmann, and H. G. Katzgraber, "Diluted antiferromagnets in a field seem to be in a different universality class than the random-field Ising model," *Phys. Rev. B*, vol. 88, p. 174408, 2013.

[158] Y. Imry and S.-K. Ma, "Random-Field Instability of the Ordered State of Continuous Symmetry," *Phys. Rev. Lett.*, vol. 35, p. 1399, 1975.

[159] P. Bhandari and V. Malik, “Finite temperature phase transition in the two-dimensional Coulomb glass at low disorders,” *Eur. Phys. J. B*, vol. 92, p. 147, 2019.

[160] J. R. L. de Almeida and D. J. Thouless, “Stability of the Sherrington-Kirkpatrick solution of a spin glass model,” *J. Phys. A*, vol. 11, p. 983, 1978.

[161] R. A. Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, M. Guidetti, D. Iñiguez, A. Maiorano, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, P. Tellez, R. Tripiccione, and D. Yllanes, “Thermodynamic glass transition in a spin glass without time-reversal symmetry,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, p. 6452, 2012.

[162] 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. Tripiccione, and D. Yllanes, “The three-dimensional Ising spin glass in an external magnetic field: the role of the silent majority,” *J. Stat. Mech.*, vol. P05014, 2014.

[163] D. S. Fisher and D. A. Huse, “Absence of many states in realistic spin glasses,” *J. Phys. A*, vol. 20, p. L1005, 1987.

[164] D. S. Fisher and D. A. Huse, “Equilibrium behavior of the spin-glass ordered phase,” *Phys. Rev. B*, vol. 38, p. 386, 1988.

[165] J. Mattsson, T. Jonsson, P. Nordblad, H. ArugaKatori, and A. Ito, “No Phase Transition in a Magnetic Field in the Ising Spin Glass $\mathrm{Fe_{0.5}Mn_{0.5}TiO_3}$,” *Phys. Rev. Lett.*, vol. 74, p. 4305, 1995.

[166] P. E. Jönsson, H. Takayama, H. Aruga Katori, and A. Ito, “Dynamical breakdown of the Ising spin-glass order under a magnetic field,” *Phys. Rev. B*, vol. 71, p. 180412(R), 2005.

[167] H. Takayama and K. Hukushima, “Field-shift aging protocol on the 3D Ising spin-glass model: dynamical crossover between the spin-glass and paramagnetic states,” *J. Phys. Soc. Jpn.*, vol. 73, p. 2077, 2004.

[168] W. H. Lawton and E. A. Sylvestre, “Elimination of Linear Parameters in Nonlinear Regres-

sion," *Technometrics*, vol. 13, p. 461, 1971.

[169] H. G. Katzgraber and A. P. Young, "Geometry of large-scale low-energy excitations in the one-dimensional Ising spin glass with power-law interactions," *Phys. Rev. B*, vol. 68, p. 224408, 2003.

[170] J. Ferre, A. Barzegar, H. G. Katzgraber, and R. Scalettar, "Distribution of interevent avalanche times in disordered and frustrated spin systems," *Phys. Rev. B*, vol. 99, p. 024411, Jan 2019.

[171] W. Klein, J. B. Rundle, and C. D. Ferguson, "Scaling and Nucleation in Models of Earthquake Faults," *Phys. Rev. Lett.*, vol. 78, p. 3793, 1997.

[172] L. I. Salminen, A. I. Tolvanen, and M. J. Alava, "Acoustic Emission from Paper Fracture," *Phys. Rev. Lett.*, vol. 89, p. 185503, 2002.

[173] B. B. Goodman, "Type II superconductors," *Reports on Progress in Physics*, vol. 29, p. 445, 1966.

[174] I. D. Mayergoyz, "Mathematical Models of Hysteresis," *Phys. Rev. Lett.*, vol. 56, p. 1518, 1986.

[175] I. D. Mayergoyz, *Mathematical Models of Hysteresis*. New York: Springer, 1991.

[176] J. B. Rundle, D. L. Turcotte, A. Donnellan, L. Grant Ludwig, M. Luginbuhl, and G. Gong, "Nowcasting earthquakes," *Earth and Space Science*, vol. 3, p. 480, 2016.

[177] P. A. Varotsos, N. V. Sarlis, H. K. Tanaka, and E. S. Skordas, "Some properties of the entropy in the natural time," *Phys. Rev. E*, vol. 71, p. 032102, 2005.

[178] P. A. Varotsos, N. V. Sarlis, and E. S. Skordas, *Natural Time Analysis: The New View of Time*. Berlin: Springer, 2011.

[179] Rundle, J. B. and Luginbuhl, M. and Giguere, A. and Turcotte, D. L., *Natural Time, Nowcasting and the Physics of Earthquakes: Estimation of Seismic Risk to Global Megacities*, p. 123. Cham: Springer International Publishing, 2019.

[180] A. J. Bray and M. A. Moore, "Scaling theory of the ordered phase of spin glasses," in *Heidelberg Colloquium on Glassy Dynamics and Optimization* (L. Van Hemmen and I. Morgenstern, eds.), p. 121, New York: Springer, 1986.

[181] W. L. McMillan, "Scaling theory of Ising spin glasses," *J. Phys. C*, vol. 17, p. 3179, 1984.

[182] C. M. Newman and D. L. Stein, "Multiple states and thermodynamic limits in short-ranged

Ising spin-glass models," *Phys. Rev. B*, vol. 46, p. 973, 1992.

[183] J. N. Nampoothiri, K. Ramola, S. Sabhapandit, and B. Chakraborty, "Gaps between avalanches in one-dimensional random-field Ising models," *Phys. Rev. E*, vol. 96, p. 032107, 2017.

[184] J. Lin, A. Saade, E. Lerner, A. Rosso, and M. Wyart, "On the density of shear transformations in amorphous solids," *Eur. Phys. Lett.*, vol. 105, p. 26003, 2014.

[185] S. Karmakar, E. Lerner, and I. Procaccia, "Statistical physics of the yielding transition in amorphous solids," *Phys. Rev. E*, vol. 82, p. 055103, 2010.

[186] M. Müller and M. Wyart, "Marginal Stability in Structural, Spin, and Electron Glasses," *Annual Review of Condensed Matter Physics*, vol. 6, p. 177, 2015.

[187] J. Lin, E. Lerner, A. Rosso, and M. Wyart, "Scaling description of the yielding transition in soft amorphous solids at zero temperature," *Proc. Natl. Acad. Sci.*, vol. 111, no. 40, p. 14382, 2014.

[188] M. Wyart, "Marginal Stability Constrains Force and Pair Distributions at Random Close Packing," *Phys. Rev. Lett.*, vol. 109, p. 125502, 2012.

[189] T. Nattermann, "Dipolar interaction in random-field systems," *J. Phys. A*, vol. 21, p. L645, 1988.

[190] D. P. Belanger, "Randomfield experiments in dilute antiferromagnets," *Phase Transitions*, vol. 11, p. 53, 1988.

[191] D. P. Belanger and A. P. Young, "The random field Ising model," *J. Magn. Magn. Mater.*, vol. 100, p. 272, 1991.

[192] H. Rieger, *Monte Carlo Studies of Ising Spin Glasses and Random Field Systems*, p. 295. World Scientific, 1995.

[193] T. Nattermann, "Theory of the Random Field Ising Model," in *Spin Glasses and Random Fields* (A. P. Young, ed.), p. 277, Singapore: World Scientific, 1998.

[194] O. Perkovic, K. A. Dahmen, and J. P. Sethna, "Avalanches, Barkhausen Noise, and Plain Old Criticality," *Phys. Rev. Lett.*, vol. 75, p. 4528, 1995.

[195] J. C. Andresen, Z. Zhu, R. S. Andrist, H. G. Katzgraber, V. Dobrosavljević, and G. T. Zimanyi, "Self-Organized Criticality in Glassy Spin Systems Requires a Diverging Number of Neighbors," *Phys. Rev. Lett.*, vol. 111, p. 097203, 2013.

[196] A. Sharma, J. Yeo, and M. A. Moore, “Self-organized critical behavior and marginality in Ising spin glasses,” *J. Stat. Mech.*, vol. 053302, 2018.

[197] L. Viana and A. J. Bray, “Phase diagrams for dilute spin glasses,” *J. Phys. C*, vol. 18, p. 3037, 1985.

[198] Schnabel, S. and Janke, W., “Dynamic greedy algorithms for the EdwardsAnderson model,” *Comp. Phys. Comm.*, vol. 220, p. 74, 2017.

[199] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, H. K. Tanaka, and M. S. Lazaridou, “Attempt to distinguish long-range temporal correlations from the statistics of the increments by natural time analysis,” *Phys. Rev. E*, vol. 74, p. 021123, 2006.

[200] F. Pázmándi, G. Zaránd, and G. T. Zimányi, “Self-organized criticality in the hysteresis of the sherrington-kirkpatrick model,” *Phys. Rev. Lett.*, vol. 83, p. 1034, 1999.

[201] Z. Zhu, J. C. Andresen, M. A. Moore, and H. G. Katzgraber, “Boolean decision problems with competing interactions on scale-free networks: Equilibrium and nonequilibrium behavior in an external bias.” (arXiv:cond-mat/1310.1139), 2013.

[202] A. Sharma, A. Andreanov, and M. Müller, “Avalanches and hysteresis in frustrated superconductors and XY spin glasses,” *Phys. Rev. E*, vol. 90, p. 042103, 2014.

[203] O. S. Sarıyer, A. Kabakcioglu, and A. N. Berker, “Deep Spin-Glass Hysteresis Area Collapse and Scaling in the $d = 3 \pm J$ Ising Model,” *Phys. Rev. E*, vol. 86, p. 041107, 2012.

[204] R. Bruinsma and G. Aeppli, “Interface Motion and Nonequilibrium Properties of the Random-Field Ising Model,” *Phys. Rev. Lett.*, vol. 52, p. 1547, 1984.

[205] R. T. Scalettar, “Critical properties of an Ising model with dilute long range interactions,” *Physica A*, vol. 170, p. 282, 1991.

[206] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Rev. Mod. Phys.*, vol. 74, p. 47, 2002.

[207] H. G. Katzgraber, F. Pázmándi, C. R. Pike, K. Liu, R. T. Scalettar, K. L. Verosub, and G. T. Zimányi, “Reversal-field memory in the hysteresis of spin glasses,” *Phys. Rev. Lett.*, vol. 89, p. 257202, 2002.

[208] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, “Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations,” *Phys. Rev. Lett.*, vol. 70, p. 3347, may 1993.

[209] J. M. Deutsch, A. Dhar, and O. Narayan, “Return to Return Point Memory,” *Phys. Rev. Lett.*, vol. 92, p. 227203, 2004.

[210] T. H. Cormen, T. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 1990.

[211] G. Bendall and F. Margot, “Greedy-type resistance of combinatorial problems,” *Discrete Optimization*, vol. 3, no. 4, pp. 288 – 298, 2006.

[212] A. Möbius, A. Neklioudov, A. Díaz-Sánchez, K. H. Hoffmann, A. Fachat, and M. Schreiber, “Optimization by thermal cycling,” *Phys. Rev. Lett.*, vol. 79, pp. 4297–4301, Dec 1997.

[213] 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,” *Nat. Phys.*, vol. 10, p. 218, 2014.

[214] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, “Defining and detecting quantum speedup,” *Science*, vol. 345, p. 420, 2014.

[215] F. Glover, “Tabu searchpart i,” *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.

[216] F. Glover, “Tabu searchpart ii,” *ORSA Journal on Computing*, vol. 2, no. 1, pp. 4–32, 1990.

[217] J. Schneider, C. Froschhammer, I. Morgenstern, T. Husslein, and J. M. Singer, “Searching for backbonesan efficient parallel algorithm for the traveling salesman problem,” *Computer Physics Communications*, vol. 96, no. 2, pp. 173 – 188, 1996.

[218] W. Zhang, “Configuration landscape analysis and backbone guided local search: Part I: satisfiability and maximum satisfiability,” *Artif. Intell.*, vol. 158, no. 1, pp. 1–26, 2004.

[219] Y. Wang, Z. Lü, F. Glover, and J.-K. Hao, “Effective variable fixing and scoring strategies for binary quadratic programming,” in *Evolutionary Computation in Combinatorial Optimization* (P. Merz and J.-K. Hao, eds.), (Berlin, Heidelberg), pp. 72–83, Springer Berlin Heidelberg, 2011.

[220] Y. Wang, Z. Lü, F. Glover, and J.-K. Hao, “Backbone guided tabu search for solving the ubqp problem,” *Journal of Heuristics*, vol. 19, pp. 679–695, Aug 2013.

[221] P. Chardaire, J. L. Lutton, and A. Sutter, “Thermostatistical persistency: A powerful improving concept for simulated annealing algorithms,” *European Journal of Operational Re-*

search, vol. 86, no. 3, pp. 565 – 579, 1995.

[222] H. Karimi, G. Rosenberg, and H. G. Katzgraber, "Effective optimization using sample persistence: A case study on quantum annealers and various Monte Carlo optimization methods," *Phys. Rev. E*, vol. 96, p. 043312, 2017.

[223] S. Mandrá and H. G. Katzgraber, "A deceptive step towards quantum speedup detection," *Quantum Sci. Technol.*, vol. 3, p. 04LT01, 2018.

[224] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, "Probing for quantum speedup in spin-glass problems with planted solutions," *Phys. Rev. A*, vol. 92, p. 042325, 2015.

[225] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, "Quantum annealing amid local ruggedness and global frustration," *Journal of the Physical Society of Japan*, vol. 88, no. 6, p. 061007, 2019.

[226] A. J. Ochoa, D. C. Jacob, S. Mandrà, and H. G. Katzgraber, "Feeding the multitude: A polynomial-time algorithm to improve sampling," *Phys. Rev. E*, vol. 99, p. 043306, Apr 2019.

[227] B. Altshuler, H. Krovi, and J. Roland, "Anderson localization makes adiabatic quantum optimization fail," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12446–12450, 2010.

[228] T. Jörg, F. Krzakala, G. Semerjian, and F. Zamponi, "First-order transitions and the performance of quantum algorithms in random optimization problems," *Phys. Rev. Lett.*, vol. 104, p. 207206, May 2010.

[229] N. G. Dickson and M. H. S. Amin, "Does adiabatic quantum optimization fail for np-complete problems?," *Phys. Rev. Lett.*, vol. 106, p. 050502, Feb 2011.

[230] I. Hen and A. P. Young, "Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems," *Phys. Rev. E*, vol. 84, p. 061152, 2011.

[231] E. Farhi, D. Gosset, I. Hen, A. W. Sandvik, P. Shor, A. P. Young, and F. Zamponi, "Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs," *Phys. Rev. A*, vol. 86, p. 052334, Nov 2012.

[232] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, "The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective,"

Physics Reports, vol. 523, no. 3, pp. 127 – 205, 2013. The Quantum Adiabatic Algorithm Applied to Random Optimization Problems: The Quantum Spin Glass Perspective.

[233] S. Mandrà, Z. Zhu, W. Wang, A. Perdomo-Ortiz, and H. G. Katzgraber, “Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches,” *Phys. Rev. A*, vol. 94, p. 022337, 2016.

[234] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, “Training a Binary Classifier with the Quantum Adiabatic Algorithm,” *arXiv e-prints*, p. arXiv:0811.0416, Nov 2008.

[235] K. L. Pudenz and D. A. Lidar, “Quantum adiabatic machine learning,” *Quantum Information Processing*, vol. 12, pp. 2027–2070, May 2013.

[236] S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum algorithms for supervised and unsupervised machine learning,” *arXiv e-prints*, p. arXiv:1307.0411, Jul 2013.

[237] S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum principal component analysis,” *Nature Physics*, vol. 10, p. 631633, Jul 2014.

[238] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Phys. Rev. Lett.*, vol. 113, p. 130503, Sep 2014.

[239] N. Wiebe, A. Kapoor, and K. M. Svore, “Quantum Deep Learning,” *arXiv e-prints*, p. arXiv:1412.3489, Dec 2014.

[240] M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics*, vol. 56, pp. 172–185, Apr 2015.

[241] N. Wiebe, A. Kapoor, and K. M. Svore, “Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning,” *Quantum Information & Computation*, vol. 15, pp. 316–356, 2014.

[242] S. Aaronson, “Read the fine print,” *Nature Physics*, vol. 11, pp. 291–293, Apr. 2015.

[243] S. H. Adachi and M. P. Henderson, “Application of Quantum Annealing to Training of Deep Neural Networks,” *arXiv e-prints*, p. arXiv:1510.06356, Oct 2015.

[244] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, 09 2017.

[245] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, “Quantum-assisted learning of hardware-embedded probabilistic graphical models,” *Phys. Rev. X*, vol. 7, p. 041052, Nov 2017.

[246] A. Mott, J. Job, J.-R. Vlimant, D. Lidar, and M. Spiropulu, “Solving a Higgs optimization problem with quantum annealing for machine learning,” *Nature*, vol. 550, p. 375379, 2017.

[247] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, “Quantum boltzmann machine,” *Phys. Rev. X*, vol. 8, p. 021050, May 2018.

[248] R. Y. Li, R. Di Felice, R. Rohs, and D. A. Lidar, “Quantum annealing versus classical machine learning applied to a simplified computational biology problem,” *NPJ Quantum Inf.*, vol. 4, no. 1, p. 14, 2018.

[249] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[250] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 44, pp. 206–226, Jan 2000.

[251] L. K. Grover, “Quantum mechanics helps in searching for a needle in a haystack,” *Phys. Rev. Lett.*, vol. 79, no. 2, p. 3258, 1997.

[252] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[253] R. Gordân, N. Shen, I. Dror, T. Zhou, J. Horton, R. Rohs, and M. L. Bulyk, “Genomic regions flanking e-box binding sites influence dna binding specificity of bhlh transcription factors through dna shape,” *Cell Reports*, vol. 3, no. 4, pp. 1093 – 1104, 2013.