Optimization with spin glass models

David Cirauqui Garcia

ADVERTIMENT La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del repositori institucional UPCommons (http://upcommons.upc.edu/tesis) i el repositori cooperatiu TDX (http://www.tdx.cat/) ha estat autoritzada pels titulars dels drets de propietat intel·lectual **únicament per a usos privats** emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei UPCommons o TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a UPCommons (*framing*). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del repositorio institucional UPCommons (http://upcommons.upc.edu/tesis) y el repositorio cooperativo TDR (http://www.tdx.cat/?locale-attribute=es) ha sido autorizada por los titulares de los derechos de propiedad intelectual **únicamente para usos privados enmarcados** en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio UPCommons No se autoriza la presentación de su contenido en una ventana o marco ajeno a UPCommons (*framing*). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the institutional repository UPCommons (http://upcommons.upc.edu/tesis) and the cooperative repository TDX (http://www.tdx.cat/?locale- attribute=en) has been authorized by the titular of the intellectual property rights **only for private uses** placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading nor availability from a site foreign to the UPCommons service. Introducing its content in a window or frame foreign to the UPCommons service is not authorized (framing). These rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

OPTIMIZATION WITH SPIN GLASS MODELS

David Cirauqui Garcia

Supervisors:
Prof. Dr. Maciej Lewenstein
Dr. José Ramón Martínez Saavedra

Co-supervisor:
Dr. Przemysław R. Grzybowski

This thesis has been carried out jointly in the Quantum Optics Theory group at ICFO - The Institute of Photonic Sciences, and the Computing Team at Quside Technologies S.L., in Castelldefels, Barcelona. It has been possible partially thanks to the funding of a predoctoral grant from the Doctorats Industrials program (2019 2n termini, ref. BDNS 443065) of the Agency for the Management of University and Research Grants (AGAUR), of the Generalitat de Catalunya.

ZE ILLNAMIKIZ NIN EZKAYO [1]

What will I keep before I depart?
Will I leave anything about me on this earth?
How shall my heart react?
Is it in vain that we come to live
and just spring above the earth?
Let's, at least, leave flowers behind
Let's, at least, leave chants behind

Nezahualcóyotl

---

[1] The poem was originally written in Nahuatl. The title translates to "A memory I leave".

Acknowledgments

Some time ago, during my undergraduate studies, I happened to be in someone else's kitchen with one of those people who exudes knowledge through every pore of their skin, next to whom any minimally curious person feels attracted by a force of unknown nature, but that is for sure spooky and at a distance. I remember as if it was yesterday that, holding a pan with the right hand and a cloth with the left one, that person told me that studying statistical physics was just like cooking a spanish omelet. And I remember, just as well, that I had no idea whatsoever about what that person intended to mean. It feels crazy to think that so many years later, I am writing the last lines of a Ph.D. thesis in which I have been diving into such physical systems.

Anyhow, that moment is something that got stuck in my mind and that has come to me again now that the time to write these lines has arrived, almost without noticing it. Time can flow in strange ways sometimes. However, these lines are perhaps the most important ones of one's thesis, as scientific works in general, but Ph.D. theses in particular, are not something that arise from scratch by spontaneous generation, but are rather built upon other people's ideas and works. Just like mushrooms are nothing but the visible and beautiful parts of a much larger, usually unseen, forgotten structure of mycelium, only thanks to the people that came before and those one relies and leans on, the scientific advances are possible. The whole body of scientific literature upon which this work intends to advance is properly cited and henceforth acknowledged in the bibliography section. The other ones, more personal acknowledgements are now to be given. Bear in mind that the work you have in your hands, dear reader, would not exist at all had it not been for the people I will list now (and possibly many others, who I hope will forgive my selectively elusive memory). Consequently, and in case you happen to enjoy or learn something by reading it, I would like you to remember that they are the ones to be ultimately thanked for.

To begin with, the scientific environment in which I have conducted my research these last few years is by no means something that I take for granted: not everyone is given the chance to work in ICFO under the guidance of Professor Maciej Lewenstein. Having had the opportunity to have him as a reference is a privilege for which I feel both profoundly humbled and honored, and for which I must give my most sincere

thanks. Moreover, I must also thank him for having met and worked with Przemek, as he has most probably been the person from whom I have learned the most during these years. From physics to history and everyday life topics, being by his side has definitely made me grow in so many different ways. I no longer run towards the light at the end of the tunnel, since I now know, thanks to him, that it's rather a train coming towards me. And finally, ICFO has also led me to work with Miguel Ángel, to whom, besides the laughs and the scientific collaboration, I must acknowledge his efforts to proof-read my works an innumerable amount of times.

It is true, nevertheless, that ICFO and everything it has brought me have only been a consequence of another place that priorly opened its doors to me: Quside Technologies. Quside is where I have worked all these years and where I have really felt supported during the voyage that this industrial Ph.D. has been. Thanks, Carlos, for answering that email: the fact that he and the rest of the (back then, such a small) team believed in me profoundly changed my life and ultimately brought me where I am today. It is crazy to think how some small, mostly unmeditated decisions can affect one's future. Life is without any trace of doubt a chaotic system.

Still, Quside would not have been the same without Jose, who I consider to be not only a good supervisor but a good friend too. Every time someone asks me about my experience as a Ph.D. student, I cannot help but stress how important this is, and how lucky I have been in this regard. I must admit that some months ago a colleague surprised me by making me realize that I had adopted for myself and was using some of Jose's typical expressions. I can only wish for this strange osmosis process to work in other areas as well, and thus to have been able to absorb a small fraction of his (sometimes (almost always) overwhelming) knowledge: whoever experienced it knows it, and I cannot read any further.

Besides, there are many other people at Quside I have to thank so many things for. To Ferran, for introducing me to Mansare (the true master of Ising models and TFGs, legends say), and for making me rethink, rewrite, reread, redo and relearn. And for encouraging me, of course: in the deepest and dirtiest of the valleys, the echo of his voice making me promise that I would show him my thesis once I had it finished was a true candle I had to handle more than once. To Héctor, now a friend as well, who I started helping with his master thesis and ended up proof-reading all these pages. To Javi, for always being available when I needed it. And many others, both present and past, with who I have shared moments, cookies and laughs during these years: Andrea, Sergio, Jaume, Álvaro, Marc, Pablo, Ismael, Tomás, Álex…

But, as I have said above, this thesis is not only the product of the last five years of work. I would not be writing these lines had it not been for some other people from whom I have learnt different things throughout my whole life. People who too often pass unseen but that are in fact the true mycelium of any Ph.D. thesis, and to whom I shall now take the opportunity to extend my most sincere thanks: Manu, Manolo, Isabel, Helena, Núria, Jordi, Eugene, Luca, Anders.

And last, but definitely not least, to my family and friends. There are no words I could use to express anything that minimally resembles what I would like to tell you and how this work would not have been possible without you, so I will not even dare to try. Thank you all for being there. No names needed.

To conclude, I would like to stress that I still do not know what statistical physics and omelets have in common. While I really think that, since that moment in that kitchen, I have learned things related to both of them, I have also realized that there is such a vast amount of things I still have to improve. Maybe this is it, or maybe not. Or maybe only that person in that precise moment knew.

Contents

Acknowledgments **i**

Contents **v**

List of Figures **ix**

List of Tables **xiii**

List of Publications **xv**

Abstract **1**

Resum **3**

1 Introduction **5**

1.1 The state of the art . . . 6

1.2 Specific problems looked at in this thesis . . . 8

2 Basics on solving spin glass models **11**

2.1 Spin systems . . . 12

2.1.1 The Ising model . . . 14

2.1.2 Spin glass models . . . 16

2.2 QUBO formulation of real world problems . . . 20

2.3 Heuristic optimization methods . . . 21

2.3.1 Monte Carlo methods . . . 22

2.3.2 The Metropolis-Hastings algorithm . . . 24

2.3.3 Simulated Annealing . . . 25

2.3.4 Parallellizability of the simulations . . . 26

3 Phase transitions of non-planar graphs **31**

3.1 Introduction . . . 32

3.2 The HLW approach . . . 33

3.3 Saddle point/steepest descent calculations . . . 35

3.4 Thouless argument . . . 38

3.5 Self-consistent SPSD and local mean field solutions 38
3.5.1 High $\tilde{\beta}$ regime 42
3.5.2 Phase transition at moderate $\tilde{\beta}$ 43
3.5.3 Numerical calculations 43
3.6 Conclusions 47

4 Correlated RN sequences in MC simulations 49
4.1 Introduction 50
4.2 Randomness and Technology 52
4.2.1 Pseudo Random Number Generators 53
4.2.2 Quantum Random Number Generators 54
4.3 The dynamic critical exponent 56
4.3.1 The characteristic autocorrelation time 56
4.3.2 Dependence on the temperature: autocorrelation and scale invariance 61
4.3.3 Universality and critical exponents 64
4.3.4 Values of the dynamic critical exponent in the literature 67
4.4 Results 68
4.4.1 Consumption of random numbers 69
4.4.2 Detecting correlations in PRNG 71
4.4.3 Computing z with a QRNG 76
4.4.4 Offset in the linear fit of the dynamic exponent 79
4.4.5 Comparison with other standard randomness tests 79
4.5 Conclusions 80

5 Defect-driven Population Annealing 83
5.1 Introduction 84
5.2 Population Annealing 85
5.2.1 Resampling protocol 86
5.2.2 Free energy estimation 86
5.2.3 Confidence on the solution found 88
5.3 Bond representation and TC topological defects 91
5.3.1 The 2d Edwards-Anderson model 91
5.3.2 Defects' dynamics in 2d: cluster updates 93
5.3.3 A proper three-dimensional generalization 95
5.3.4 Hardness of the RFW Lattice model 98
5.4 Defect-driven Population Annealing 105
5.4.1 Constant-fraction entropy-loss annealing 106
5.4.2 Remaining defects 107
5.4.3 Free energy estimation in DPA 110
5.4.4 Outline of the algorithm 111

5.5 Comparison between PA and DPA 112
  5.5.1 $L = 4$ lattices 113
  5.5.2 $L = 6$ lattices 116
  5.5.3 $L = 8$ lattices 120
  5.5.4 Constant-fraction entropy-loss annealing in PA 122
5.6 Conclusions 122

Conclusions **125**

A A real optimization problem **127**
A.1 The Maximum Covering Problem 127
A.2 A toy model example 129

B Values of the dynamic critical exponent **133**

List of Figures

2.1 An undirected, non-planar graph. 13
2.2 A two dimensional square Ising system. 15
2.3 Magnetization per spin in the 2d Ising model. 16
2.4 Spin glasses are characterized by rugged energy landscapes. 18
2.5 The Metropolis algorithm sometimes accepts new states with a higher energy. 25
2.6 Checkerboard scheme of a two-dimensional spin lattice with nearest neighbours interactions 27
2.7 Four-coloured parallelization scheme of a spin lattice with next-to-nearest neighbours connectivity 29

3.1 The temperature at which the system is simulated determines the maximum (within thermal fluctuations) energy that can be explored . 33
3.2 A spin lattice with nearest neighbours interactions in a cylinder 40
3.3 Magnetization over a cylinder with $L = 500$ layers for different temperatures and dimensions 44
3.4 Domain wall width for fixed L (left panel) and fixed $\tilde{\beta}$ (right panel) . 45
3.5 $A(\tilde{\beta})$ for all studied dimensions 45
3.6 $B_{\pm}(\tilde{\beta})$ for all studied dimensions 46
3.7 $\delta(\tilde{\beta})$ for all studied dimensions 46

4.1 Estimating the value of π with a Monte Carlo method 51
4.2 Scheme of Quside's PD-QRNG used in this work 56
4.3 Effects of the change in scale at temperatures away from the critical point 63
4.4 Scale invariance at the critical temperature 64
4.5 2d Ising model dynamics 65
4.6 Values of the dynamic critical exponent z in the literature 68
4.7 Histograms of the dynamic critical exponent z in the literature 69
4.8 Reduced variance of z 73

4.9 Normalized variance with low-frequency reseeding of PRNG3 as a function of lattice size . . . 74
4.10 Normalized variance with high-frequency reseeding of PRNG3 as a function of lattice size . . . 75
4.11 Fit of the reduced variance with κ . . . 76
4.12 Determination of the relaxation time as a function of lattice size . . . 77
4.13 Critical exponent calculation with a Quantum Random Numbers source 78
4.14 Violin plots of the randomness tests passed by all the PRNGs and the QRNG . . . 80

5.1 Quantum Annealing exploits quantum tunneling to reach the global ground state . . . 84
5.2 Physical states in the bond representation are characterized by $B_\square = 1$ 92
5.3 If periodic boundary conditions are assumed, physical states in the bond representation are characterized by $B_| = 1$ . . . 93
5.4 Cluster update in a 2d quadratic spin lattice . . . 94
5.5 The dynamics of the defects in 2d lattices induce cluster moves . . . 95
5.6 The 3d Random Field Wall (RFW) model (right) is a generalization of the 2d bond model (left) . . . 96
5.7 Graphical representation of an edge spin flip as the flipping of four walls 97
5.8 RFW models are computationally harder than their EA counterparts . 100
5.9 RFW and EA models have similar MC acceptance ratios when solved using the PA algorithm . . . 100
5.10 Annealing towards $\beta > 5$ does not yield lower energy states . . . 101
5.11 The hardness of RFW models against EA clearly emerges when varying the frustration level of the solved lattices . . . 102
5.12 RFW models are computationally harder than their EA counterparts . 104
5.13 Constant-fraction entropy-loss annealing . . . 107
5.14 DPA outperforms PA for L=4 lattices when MCMC process simulations are the computational bottleneck . . . 114
5.15 DPA provides on average higher family entropies than PA when solving $L = 4$ lattices . . . 115
5.16 Hard cases on PA require more adiabatic annealing processes to reach an equivalent family entropy to DPA . . . 117
5.17 DPA enables thermalization of $L = 6$ lattices, even when these are hard for the PA algorithm . . . 118
5.18 DPA provides on average higher family entropies than PA when solving $L = 6$ lattices . . . 119
5.19 Hard cases on PA require more adiabatic annealing processes to reach an equivaent family entropy to DPA . . . 120

5.20 Both algorithms have a low convergence ratio to the same minimum energy state among independent runs 121
5.21 DPA yields a larger spread of minimum energies 122
5.22 DPA finds lower energy states 123
5.23 The entropy-preserving steps are not optimal in standard PA 124

A.1 Map of the city and the covering range of the antennas 130
A.2 Graph of the QUBO formulation of the problem 131
A.3 Optimal distribution of the antennas 131

List of Tables

2.1 Order of magnitude of the number of possible paths n_p in a Traveling Salesman Problem of different sizes N. 22

4.1 Estimated randomness consumption. 70
4.2 Parameters used in the Linear Congruential Generators used. 71
4.3 Dynamic exponents z and relative errors with respect to theoretical estimate ϵ_r, obtained for each PRNG. 71
4.4 Dynamic exponents z and relative errors with respect to theoretical estimate ϵ_r, for different reseeding frequencies of PRNG3. 76
4.5 Results obtained for $log(\tau_0)$ by using different RNG. These values correspond to the fittings yielding the dynamic exponents shown in previous sections. 79

B.1 All calculated exponents from the literature in two dimensions, to the best of our knowledge. 134
B.2 All calculated exponents from the literature in three dimensions, to the best of our knowledge. 135

List of Publications

The present thesis is based on the following research papers:

- **Haake-Lewenstein-Wilkens approach to spin-glasses revisited.**
  M. Lewenstein, D. Cirauqui, M.Á. García-March, G. Guigó i Corominas, P. Grzybowski, J.R.M. Saavedra, M. Wilkens, and J. Wehr.
  Journal of Physics A: Mathematical and Theoretical, 55.45 (2022): 454002.
- **Quantum Random Number Generators: benchmarking and challenges.**
  D. Cirauqui, M.Á. García-March, G. Guigó i Corominas, T. Graß, P. Grzybowski, G. Muñoz-Gil, J.R.M. Saavedra, and M. Lewenstein.
  Submitted for publication, arXiv:2206.05328 (2022).
- **Population annealing with topological defect driven nonlocal updates for spin systems with quenched disorder.**
  D. Cirauqui, M.Á. García-March, J.R.M. Saavedra, M. Lewenstein and P. Grzybowski.
  Physical Review B, 109, 144202 (2024).

Other research papers not included in the thesis:

- **Universal representation by Boltzmann Machines with Regularised Axons.**
  P. Grzybowski, A. Jankiewicz, E. Piñol, D. Cirauqui, D.H. Grzybowska, P.M. Petrykowski, M.Á. García-March, M. Lewenstein, G. Muñoz-Gil and A. Pozas-Kerstjens.
  Submitted for publication, arXiv:2310.14395 (2023).

Abstract

With applicability on almost every aspect of our lives, optimization problems are ubiquitous to a broad range of fields within both scientific research and industrial environments. As such, these are growing in size and complexity at a fast pace, and are only expected to continue to do so. Accordingly, the urgency for better methods that can yield more optimal solutions in shorter times is increasing and, while the development of quantum computing technologies that are capable of tackling these problems evolves steadily, it does so too slowly for the challenges that today society's demands represent. Consequently, a lot of effort is being invested in further developing classical methods and machines that are specially designed to solve optimization problems of relevant enough sizes. The present thesis is framed within this paradigm: classical optimization techniques are studied from various different perspectives, with the goal of improving their efficiency.

To this end, we first dive into basic concerns related to the physical properties of the systems that allow for the convenient formulation of industrially-relevant optimization problems, namely spin glasses with quenched disorders. The understanding of such properties is of utmost importance for the correct designing of the annealing schedules used by thermally-based optimization methods. We then study the impact that the hidden correlations of the pseudo random number streams used in their simulations have in the results by comparing simulations using PRNGs of various qualities and perfectly random QRNGs. To conclude, we investigate novel ways, inspired by quantum-mechanical systems, to efficiently navigate the energy landscapes of spin glasses in classical algorithms, which has the potential of preventing the simulations getting stuck into local energy minima and thus reaching more optimal solutions.

Resum

Amb aplicacions en gairebé tots els aspectes de les nostres vides, els problemes d'optimització són omnipresents en un rang molt ample de camps pertanyents tant a la recerca científica com als ambients industrials. Com a tals, aquests problemes estan creixent en mida i en complexitat a un ritme accelerat, i no s'espera que deixin de fer-ho. Així doncs, la urgència per obtenir mètodes capaços de trobar solucions més òptimes en menys temps no para de créixer i, mentre que el desenvolupament de les tecnologies de computació quàntica capaces d'afrontar aquests problemes avança sense pausa, ho fa a un ritme massa lent per als reptes que les necessitats de la societat actual representen. Conseqüentment, actualment una gran quantitat d'esforços està essent invertida en millorar els mètodes de resolució clàssics i les plataformes específicament dissenyades per a executar-los, per tal de poder resoldre problemes d'optimització de grandàries suficientment rellevants. La present tesi està emmarcada en aquest paradigma: s'estudien tècniques d'optimització clàssiques des de diferents perspectives, amb l'objectiu d'incrementar-ne la seva eficiència.

Amb aquesta finalitat, en primer lloc ens endinsem en qüestions relacionades amb les propietats físiques dels sistemes que permeten formular problemes d'optimització rellevants des del punt de vista industrial en un llenguatge matemàtic convenient: els cristalls de spin amb acoblaments fixes. Entendre'n les seves propietats és de màxima importància per a un disseny apropiat dels procediments de refredament utilitzats en els algoritmes d'optimització basats en la simulació de processos tèrmics. Seguidament s'estudia l'impacte que les correlacions ocultes en les seqüències de nombres pseudoaleatoris utilitzats en aquestes simulacions tenen en els seus resultats, comparant simulacions que utilitzen generadors pseudoaleatoris de diverses qualitats i un generador de nombres aleatoris quàntic. Per finalitzar, s'investiguen nous mecanismes, inspirats en processos de caràcter quàntic, de millorar la forma en què els algoritmes clàssics naveguen els perfils d'energia dels cristalls d'espín, amb la finalitat d'evitar que les simulacions quedin atrapades en mínims d'energia locals i així obtenint solucions més òptimes.

Chapter 1

Introduction

In an era in which each and every individual holds a device in their pocket that is capable, within a fraction of a second, of accessing all the information that humankind has accumulated throughout millennia, it might sometimes be hard to justify that such powers are not enough. And, still, they are clearly and doubtlessly not. Some of the most urging and important challenges that our society as a whole will face in the short- and mid-term future (or that are already facing) demand even more powerful, faster, more efficient machines. Further, this obstacle is not something that lays solely on the sometimes esoteric world of pure scientific curiosity, but that traverses and permeates many aspects of our quotidianity.

In particular, solving optimization problems or, in other words, facing the challenge of choosing the best, most optimal option from a pool of possible candidates and subjected to a certain set of constraints [1, 2], is something completely undetachable from the inner core of most of not only humanity's, but all living creatures' most basic concerns. One can see this in different polar mammals convergently evolving towards having a higher volume-to-surface ratio (becoming more spherical) in order to *minimize* heat loss, and those inhabiting warm deserts doing exactly the contrary to *maximize* it [3]. In bees realizing that hexagonal nests *minimize* needed building materials while *maximize* its structural strengths, and fungi producing both toxic compounds and high-lighting colours to *minimize* their probability of being eaten. In some species of animals killing part of their own offspring in scarce times when they are too numerous, to ensure that there is enough food for everyone and therefore to *maximize* the number of surviving individuals, and even in some unicellular organisms being able to change its shape in order to *maximize* their food supplies while *minimizing* their exposure to external hazards [4, 5, 6]. Solving optimization problems is intimately tangled with nature.

And, when talking about humans, this fact only gets hugely exacerbated. From distributing the correct amount of water to *maximize* the crops' yield while *minimizing* at the same time the water consumption, to building the longest-lasting monuments while *minimizing* the amount of materials and human force, our entire history has been surrounded by decisions that try to *optimize* the result of a given process constrained to some needs or the availability of resources.

Besides the general and maybe abstract mathematical definition of an optimization problem, there are two main things that one should emphasize from all these examples:

- First, that given an optimization problem and a set of constraints to be fulfilled, there are many different strategies that can yield an optimal solution to the same problem, even though some of them can be clearly (or only very, very slightly) better than others.
- Second, and of course related to the fact that there are many different solutions, that optimization problems are no easy task, at all.

Nowadays, unfortunately, too often our problems are only growing in complexity and urgency in the need for solutions. We clearly want our firefighters to know the route to reach a forest wildfire in the *minimum* time possible. During the COVID19 pandemic, we demanded our politicians to resign if they were not able to *optimize* the distribution of vaccines among all the people, as well as to *minimize* our lockdown and curfew restrictions subjected to *minimizing* the spread of the virus. If we want to fight the current climate crisis, our engineers must distribute wind turbines so that the extracted energy is *maximum* [7, 8], and our scientists must discover the processes that will allow us to build the *most efficient* solar panels and batteries. And, when facing all such problems, first of all we would rather avoid suboptimal solutions, and secondly, to reach them we cannot spend the huge amount of time that evolution took for bees to learn that hexagonal nests were optimal. We need the best solutions, and we need them to be obtained as fast as possible.

1.1 The state of the art

In the matter of obtaining the best solutions in the minimum amount of time possible, the promise of the development of Quantum Technologies in general and Quantum Computers [9] in particular has sparkled a lot of (often too hyped) hope and excitement in the last decades, with some experiments even claiming the advent of the so-called quantum advantage [10]. Principally, given their intrinsic suitability to solve optimization problems [11], Quantum Annealers (QA) [12, 13, 14] often lift most of the expectations. On the one hand, these (already commercially available) machines are built with a topology that is specially designed to be able to embed and solve problems formulated in a Quadratic Unconstrained Binary Optimization (QUBO) form [15], in which, being NP-complete, any other NP-complete optimization problem can be mapped. On the other hand, (an oversimplified statement of) the adiabatic theorem guarantees that, given a certain set of conditions, the most optimal solution will always be found with probability 1. Always, no matter what. This somewhat magical theory may sound astonishing. Reality, as always, is nevertheless more complicated than that.

Although promising results have been obtained lately, quantum annealing machines still face many problems and challenges. First of all, they pose huge technological challenges themselves, with open problems such as quantum decoherence [16, 17], error correction [18], adiabaticity or energy consumption [19, 20, 21]. But secondly and maybe more importantly, in order to win the race and become the real go-to solvers for optimization problems, they also must overcome classical computing in the first place [22, 23, 24]. While it is true that quantum mechanics give QA some properties (Hilbert space's size, entenglement and tunneling, for example) that make them avoid some of the problems that its classical counterpart cannot, it is not less true that current classical computer's scalability and speed are (and will most probably keep being in the mid-term future) still many orders of magnitude above quantum's ones.

Even with such huge current available computational power, given the computational complexity of most of the optimization problems stated above, obtaining good solutions in an acceptable amount of time is usually far out of reach when considering big enough system sizes. This fact and their urgent need and industrial relevance has led, in the last few decades, to many groups from both the academic and the industrial worlds centering their attention on developing and improving classical machines and optimization algorithms.

On the algorithm side, heuristic algorithms have emerged as one of the main branches. Many strategies have been proposed, often taking ideas and inspiration from natural processes (hence the name Nature- or Physics-inspired computation), among which genetic algorithms [25, 26], neural networks [27, 28], and quantum-inspired algorithms [29, 30, 31] might be some of the most prominent and paradigmatic examples. Which one of these approaches yields the highest efficiency still remains a highly debated question, and might even depend on the inner structure of the problem at hand in the first place. This new type of computation takes advantage of the mathematical apparatus that was initially developed for quantum computers, to solve problems in platforms available today, which translates into a substantial improvement in both cost and time investments.

To understand how physics-inspired computing works, imagine yourself lying on the beach with a glass of soda. Even if it looks flat when observed from far away, the sand's surface is a chaotic and irregular set of innumerable valleys and peaks. In an attempt to break such an idyllic situation, you could ask yourself: which is the lowest valley, and what is its exact depth? The classical solution consists in measuring and writing down all of the valley's heights (which will probably differ by indistinguishable fractions of the precision of your measurement apparatus). However, this approach is by no means viable if we want to consider a sand extension larger than the towel we are lying on.

So if you want to solve this problem within a reasonable amount of time, a new strategy is mandatory. And here is where physics plays its role: if, according to thermodynamic principles, every system tends to evolve towards a steady state in which its energy is minimized, could we pour our soda over the sand in a careful enough way such that, by gravity's action, the liquid gets stored in the deepest valley, in which its potential energy is minimal? In this case, the physical evolution of the system would do the work for us, and we could continue enjoying our vacation. Obviously, the answer to this question is yes, even though this 'carefully enough' (adiabatically) might be a bit difficult to execute. Quantum annealing bases its functionality on this very same principle, and in classical platforms the idea is to simulate it with heuristic algorithms.

Heuristic algorithms compensate the lack of computational power with a trade-off between the time required to find a solution and the quality of the solution itself, by means of a random-number-assisted search of the solution space. The production and usage of such random numbers therefore plays a central role in such algorithms, and one must pay close attention to it. On the one hand, the production of true random numbers is often the most demanding part of the whole algorithm, ending up being the principal bottleneck in the simulation's velocity. However, overcoming this bottleneck by means of producing fast, slightly correlated pseudo-random numbers can lead to errors on the solutions obtained [32, 33]. In this regard, one of the currently more advanced quantum technologies, namely Quantum Random Number Generators (QRNGs), is already starting to yield considerable enough generation speeds [34] for them to be used in real world applications. The development of fast QRNGs that, given its quantum nature, are intrinsically and truly random and thus lack the drawbacks inherent to classical ones, is already a reality.

On the hardware side, special-purpose computing devices are appearing and evolving at a fast pace as well. From parallel computing in Graphical Processing Units (GPU) and Field Programmable Gate Arrays (FPGA), to huge clusters of extremely fast CPUs in supercomputers that can reach the Exaflop range, and even machines specifically designed to manage the randomized workloads, Random Processing Units (RPU), different platforms and computing paradigms are being developed with the aim of increasing the computation's velocity.

1.2 Specific problems looked at in this thesis

The present thesis is intended to be seen as a holistic approach to the general problem of optimization within a classical framework. This means that, from a broad point of view, the main objective will consist in improving the understanding and functionality of current heuristic optimization algorithms. To this end, we will not focus on the development of new algorithms alone, but will rather address aspects of different nature related to them. These will range from purely theoretical ones to direct applications on real-world problems. Specifically, the thesis is divided into four different chapters.

- In the second chapter we present a short review of the necessary concepts and tools upon which the following ones will be built. We navigate through the basic definitions of spin glasses and the heurisitc algorithms used to solve them as well. Finally, we briefly describe the QUBO formalism that enables to connect them to the industrially-relevant problems that we ultimately desire to solve, and discuss the parallelization scheme that allows us to efficiently solve such systems in special purpose hardware such as GPUs and FPGAs.
- In the third chapter we will focus on the study of one of the current hot topics of statistical physics: the spin glass phase transition. Understanding this phase transition (even whether it exists at all or not) is crucial for the optimal choice of the annealing schedules in heuristic algorithms, as they require to simulate the system at both above and below the critical temperature in order to guarantee a proper exploration of the configuration space. Specifically, we review the Haake-Lewenstein-Wilkens (HLW) approach to spin glasses and apply it to the phase transition of Ising spin systems with non-planar topologies such as the Edwards-Anderson (EA) lattices in various dimensions. This chapter is based on the results presented in [35].
- The random exploration of the phase space of the studied systems allows heuristic algorithms to trade off some accuracy on the solution found for a much greater efficiency. As such, random numbers constitute one of their most important parts, and correlations in the streams used can yield to sub-optimal solutions in optimization problems or to wrong values of a given estimated observable. The fourth chapter is devoted to study how this hidden, difficult-to-detect correlations can affect the results obtained by these algorithms. To this end we will measure the dynamic critical exponent of two-dimensional Ising spin lattices with pseudorandom number generators of various qualities, to finally compare them with results obtained when using completely uncorrelated, purely random number streams output by a quantum random number generator. This chapter is based on the results presented in [36].
- While most broadly used algorithms (such as Simulated Annealing, Parallel Tempering and Population Annealing) make use of thermal properties (annealing temperature schedules, thermal fluctuations, etc.) to efficiently explore the phase space of the studied systems, there is a growing tendency on using some other nature-inspired processes to this end. Paradigmatic examples are, on the one hand, genetic algorithms, in which a natural-selection process that

keeps and reproduces the fittest solutions is simulated; and quantum-inspired algorithms, on the other one, such as Simulated Quantum Annealing, that classically simulates the adiabatic transition of a wave function from the initial state to the ground state of the spin-glass model. In this fifth chapter we devise a quantum inspired modification of Population Annealing that effectively allows for the realization of non-local moves, which helps the system escape from local minima and therefore improve the efficiency on finding the global ground state. This chapter is based on the results presented in [37].

Chapter 2

Basics on solving spin glass models

As the holistic view of solving optimization problems that this thesis aspires to represent, many different topics are discussed through its chapters. As different some of them might appear to be, though, there is some fundamental knowledge that permeates all of them and that builds up the basic concepts needed in order to understand their foundations.

In this chapter we present the basic concepts and methodologies from which the present thesis will be built on, and that are more or less transversal to all chapters. Namely, we review concepts related to spin glass systems, heuristic methods to solve them, and the QUBO formalism that allows to use them to solve real-world optimization problems. More in depth theoretical tools needed for each individual chapter will be discussed in them.

2.1 Spin systems

Through this thesis we will work with a type of models that are usually referred to as spin systems, which have long been one of the cornerstones of statistical physics since its beginning. In particular, we will specifically focus on binary (Ising) spin systems with quenched disorders [38]. These are nothing but mathematical models describing a set of variables (called spins) that can take binary values (normally either ± 1 or $\{0, 1\}$) and that interact between each other with a certain, fixed strength (thus quenched). Since spins are only allowed to have two values, it is illustrative and usual to visualize them as arrows pointing either up or down. Further, we will restrict as well to quadratic models, so only single-spin and two-spin interactions will be considered. In general, not all spins forcefully need to interact with all of the rest, in which case a null interaction strength is set.

Mathematically, such systems are usually visualized as graphs in which their nodes represent the spins and their topology defines the interactions between them, Fig.2.1. For these graphs to represent realistic physical systems that can be described by a properly defined Hamiltonian function, we require them to be undirected, so that the weight J_{ij} of any edge coupling two given nodes i and j is the same in the two directions, $J_{ij} = J_{ji}$. Therefore since the energy accounted for in the Hamiltonian depends only on the relative alignment of the spins, and its direction (up or down) is conventional, it follows that the inversion of all spins in a given configuration will yield the same energy, and thus that such a model will have (at least) two symmetric degenerate ground states. On the other hand, the connectivity between spins (usually referred to as topology) can be arbitrarily complex and thus they need not be restricted to planar ones, essentially meaning that the edges within the graph can cross each other. On top of that, an external field b_i can be acting on each spin i independently, which effectively accounts for the single-spin terms in the Hamiltonian. The addition of the external field effectively makes spins prefer one orientation over the other, and thus breaks the up-down degeneracy of the ground state.

The Hamiltonian describing this kind of model's physics can hence be written, in its more general form, as

$$H = -\sum_{(ij)} J_{ij} s_i s_j - \sum_i b_i s_i, \tag{2.1}$$

where the first summatory runs over all connected spins, and s_i are the values of the spin variables. The negative sign is used for convention: while a positive or a negative external field makes each spin have a preferred orientation over the other, once the negative sign is taken into account a positive or negative coupling between two spins makes them prefer to stay aligned or unaligned, respectively. These two types of interactions are usually termed ferromagnetic and antiferromagnetic. When, due to conflicting interactions, two spins cannot align in a way that minimizes the energy of the system as given by Eq. (2.1), it is often said that they present spin frustration. Due to all such complexities, spin systems can exhibit tremendously rich behaviours, which have eluded physicists for decades. Analytically challenging even at the mean-field level [39] and numerically hard for non-planar topologies [40], their study continues to rely mostly on numerical simulations.

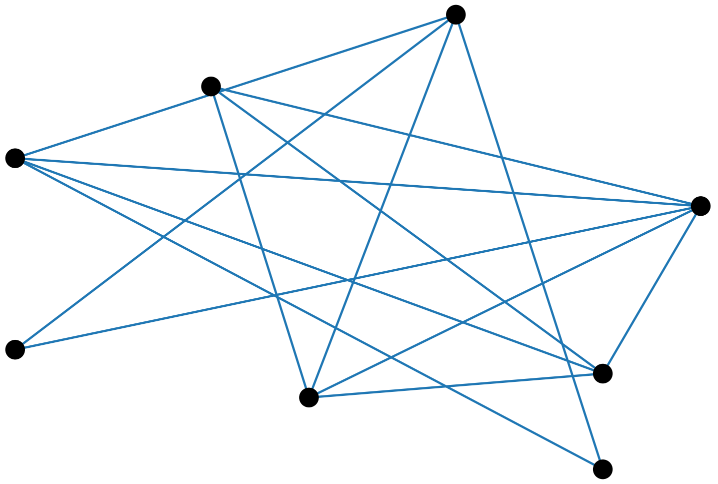

Figure 2.1: **An undirected, non-planar graph.** Spin systems like those described by Hamiltonian Eq. (2.1) can be visualized as graphs with arbitrarily complicated topologies. The nodes represent the spin variables s_i and the edges the couplings J_{ij} between them, while an external magnetic field can be applied on each individual node.

Finally, by assigning a 0 value to the couplings between unconnected spins and to the external fields over spins not presenting linear terms, we can rewrite Eq. (2.1) in matrix notation

$$H = -s^t J s - b^t s, \tag{2.2}$$

with the vector s representing the spin configuration. Note that, since as discussed above, the physicality of the model forces the graph to be undirected, the coupling matrix J must be symmetric. Further, it must also have null elements on its diagonal. A certain joint realization of J and b is usually referred to as disorder.

2.1.1 The Ising model

Among all spin models found across the literature, one of the most prominent and celebrated ones is without a doubt the so called Ising model [41]. It was first proposed by Wilhelm Lenz in 1920 and analytically solved in its one-dimensional version by his at that time student Ernst Ising in his thesis in 1924. We would have to wait until 1944 for Lars Onsager to obtain an analytical solution for the two-dimensional version too, while no solutions for higher dimensions are known to date of the writing of the present thesis. Its relevance comes not only for being one of the first models of its kind, but for being the only one presenting a real phase transition for which an exact solution has been obtained so far. Further, it is also worth noting its applicability to the study of real physical magnetic systems. Concretely, it was first designed to model the spontaneous magnetization arising on ferromagnetic materials such as iron below a certain (finite) critical temperature, based on the idea that the magnetism of the bulk material comes as a consequence of the combined magnetic dipole moments of a huge number of tiny spins within it. It has nevertheless been used to study a plethora of other types of physical systems.

The Ising model describes a lattice of spins with uniform ferromagnetic interactions J, nearest-neighbours only connectivity and an uniform external magnetic field. Each spin is defined as a small magnet that can either point up or down, and that interacts with its neighbours consequently: while two neighbouring spins pointing to the same direction will decrease the total energy of the system a total amount of J, two unaligned ones will make it increase by that exact same quantity. In one dimension the spins' positions form a chain, while in its two- and three-dimensional versions they are placed in quadratic and cubic lattices, respectively. Its Hamiltonian thus reads

$$H = -J \sum_{\langle i,j \rangle} s_i s_j - B \sum_i s_i, \tag{2.3}$$

where the angle brackets $\langle \cdot \rangle$ stand for nearest neighbours. In general, periodic boundary conditions are assumed. Being the external magnetic field B uniform across the whole lattice, its only contribution is defining a preferred direction in the total spin configuration hence breaking the up-down degeneracy. Thus, without loss of generality, it is usually assumed to be zero.

The macroscopic magnetization per spin of a given configuration is computed as the sum of the dipole moments over the whole lattice

$$m = \frac{1}{N} \sum_i s_i \tag{2.4}$$

and the average over configurations of its absolute value serves as an order parameter

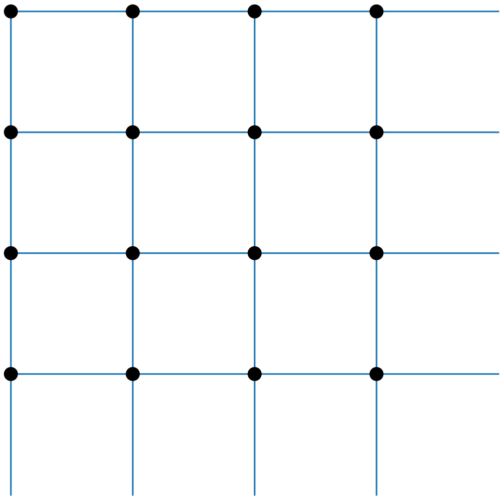

Figure 2.2: **A two dimensional square Ising system.** Each spin (represented by a point) is only coupled to its nearest neighbours. The couplings to the right and to the bottom of the lattice represent the periodic boundary conditions.

of the system, often referred to as μ

$$\mu = \frac{1}{Z} \sum_{\{s\}} |m_s| e^{-\beta E_s}, \tag{2.5}$$

with $\{s\}$ defining the configuration, E_s its corresponding energy, β the inverse temperature and Z the partition function. It reaches non-zero values below the critical temperature T_c (and its maximum one at exactly $T = 0$), where the whole magnet achieves a spontaneous magnetization. Above it, where thermal fluctuations dominate over the Hamiltonian's interactions and therefore the spins behave randomly, the average magnetization is zero. Note that in Eq. (2.5) we set $k_B = 1$. This convention is further adopted throughout the whole thesis.

As given by Onsager's results [42], the critical temperature in two-dimensional systems is

$$T_c = \frac{2}{\log\left(1 + \sqrt{2}\right)}. \tag{2.6}$$

In Fig.2.3 the phase transition of two dimensional lattices of the Ising model of various sizes is shown. Note that, for an increasing number of spins within the system, the region around T_c for which the phase transition occurs gets sharper, eventually being infinitely sharp in the thermodynamic limit.

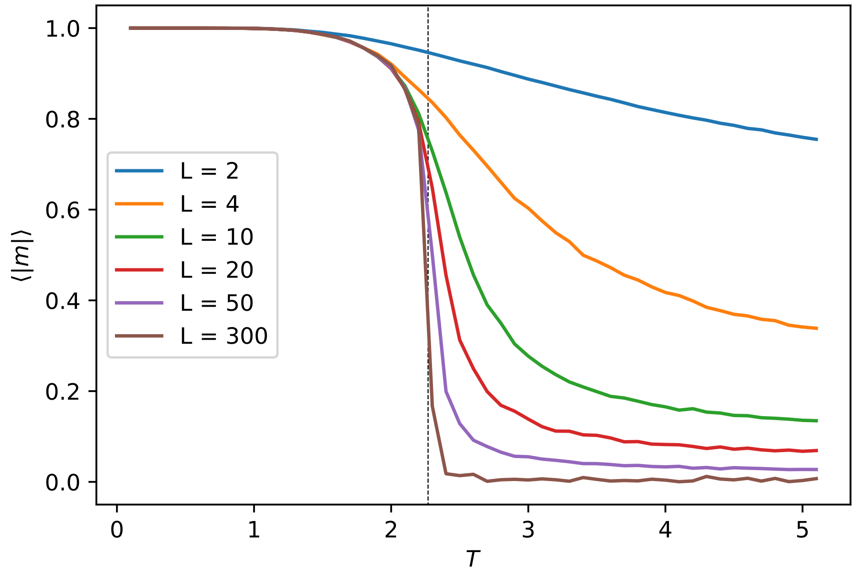

Figure 2.3: **Magnetization per spin in the 2d Ising model.** Using μ as an order parameter, a phase transition can be observed at T_c (marked by a vertical line). The transition gets sharper for bigger systems (L corresponds to the side length of the square lattices, so that the number of spins is $N_s = L^2$).

It is important to clarify that, due to its popularity, the term "Ising lattice" is many times used across the literature to talk about not the original Ising model itself, but about any spin model with spins taking binary values. This notation will also be used sometimes in this thesis.

2.1.2 Spin glass models

While the Ising model distinguishes itself for having allowed physicists to beautifully study phase transitions by essentially proving that statistical physics was indeed able to derive them from microscopic principles, it is also still the only one with an analytical solution for a reason: its (relative) simplicity. The uniformity of the couplings and the low connectivity clearly simplify its study. One can of course think of slightly more complicated models for which, for example, the couplings can have different strengths and more intricate topologies. These more complicated models are usually referred to as spin glasses [43, 44], in analogy to the huge complexity of the structure of normal window glass, in which atoms are not arranged in periodic lattices but form random structures instead. Spin glass behaviour was first experimentally identified in metal alloys such as AuFe [45] and CuMn [46].

Strictly speaking, nevertheless, an intricate connectivity between spins is not a sufficient condition for the system to show spin glass behaviour. Since, as discussed earlier, positive couplings will do nothing but tend to align the connected spins in the same direction, a very densely connected graph with only positive-valued couplings will still tend to present a ferromagnetic configuration. Such system will then not present any frustration, and the only difference with respect to the Ising model will reside in that the Hamiltonian's forces will be overcome by the thermal fluctuations and hence present a phase transition at a (potentially) different temperature. The spin glass behaviour is in fact defined by the spin frustration of the ground state configuration, which requires not only complex enough topologies, but the presence of both ferro- and antiferromagnetic couplings as well.

Spin glasses are characterized by very rugged energy landscapes [47, 48, 49, 50], as schematically depicted in Fig. 2.4. The number of local minima wells increases exponentially with the system's size, and the energy barriers separating them get higher and thus more difficult to be overcome. Moreover, the many energy minima get more and more similar between them, which makes it even harder to distinguish between local ones and the global ground state, which can potentially be degenerate as well. While looking for the ground state, it is then very easy to get stuck in local energy minima within large basins. Finding the ground states of spin glass systems constitutes one of the most challenging and formidable problems of modern statistical physics, to which a huge amount of work is devoted [51, 52, 53, 54, 55, 56].

Another characteristic property of spin glasses is the potential existence of a so called spin glass phase transition. Though fundamentally similar to other phase transitions such as the ferromagnetic-paramagnetic one discussed above, it presents some characteristic differences. To begin with, the spin glass phase transition is not characterized by a change or discontinuity in the specific heat of any type [57], as is typical of other phase transitions. Furthermore, the presence of spin frustration and both ferro- and antiferromagnetic couplings makes all feasible configurations present spins pointing towards any possible direction, which can effectively result in the same value of the total magnetization at any possible temperature. Therefore, the very own properties of spin glasses prevent the magnetization μ of serving as an order parameter for this phase transition.

Contrarily, at the initial stages of the study of spin glass models, Edwards and Anderson [58] proposed that, should such phase transition exist, it could be identified by the vanishing of the now so-called Edwards-Anderson order parameter

$$q_{EA} = \frac{1}{N} \lim_{t \to \infty} \sum_i \langle s_i(0) s_i(t) \rangle_t \tag{2.7}$$

where the inner bracket corresponds to a time average. If the system is in a regime in which the majority of spins do not have any preferred orientation, the two-spin

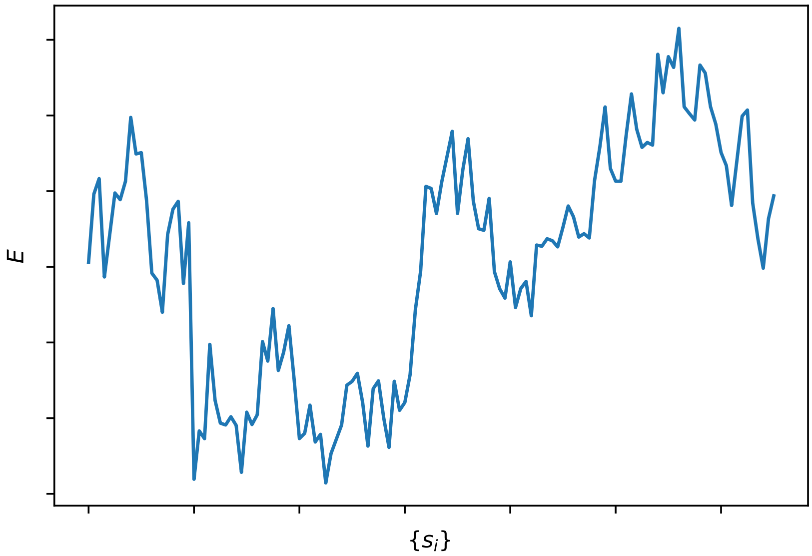

Figure 2.4: **Spin glasses are characterised by rugged energy landscapes.** The number of wells in the energy landscape of systems showing spin glass behaviour exponentially increases with its size, which makes finding its ground state extremely difficult.

products in Eq. (2.7) will vanish on average and yield a null value of the order parameter. Contrarily, if the system is in a regime in which spins do have a preferred orientation, the average will not cancel out and thus $q_{EA} > 0$. Since Eq. (2.7) basically represents an infinite-time autocorrelation function of the spins in a given configuration, q_{EA} can consequently distinguish between the paramagnetic phase (in which spins continuously flip in a random manner), and the spin glass phase (in which an apparent random configuration is frozen in time).

In order to avoid using time averages, Parisi [59, 60, 61] would later further develop this idea by using the powerful replica method, which results in a redefinition of the order parameter with a more profound physical interpretation. Consider different independent replicas of the same system (this is, lattices defined by the same coupling matrix J), labeled by greek indices. The equivalent order parameter is then redefined as

$$q_{\alpha\beta} = \frac{1}{N}\sum_i \langle s_i^{(\alpha)} s_i^{(\beta)} \rangle \tag{2.8}$$

where the average now is taken over all the independent replicas of the system. The order parameter then takes non-vanishing values when the spin overlap between replicas is high, meaning that their configurations resemble each other, while it vanishes when they differ. From this perspective, the vanishing values of the order parameter arise from the breaking of the replica symmetry.

The Edwards-Anderson model

Among all spin glass models, the Edwards-Anderson [58] one is a noteworthy one (again, for its relative simplicity) and has been and still is the focus of innumerable investigations. It consists of spins arranged in a (usually three-dimensional) cubic lattice of side length L, with nearest neighbours interactions, couplings drawn from a normal probability distribution, and no external magnetic field. Its Hamiltonian reads

$$H = -\sum_{\langle i,j\rangle} J_{ij} s_j s_j. \tag{2.9}$$

In other words, it is a modified Ising model in which $J_{ij} = \mathcal{N}(0,1)$. Even if its understanding generally lacks analytical results, extensive numerical efforts have been put to its study. The computational power needed to this end is nevertheless huge, and confident results for three-dimensional lattices of only up to $L = 10$ [51] can be found in the literature, with some others tentatively exploring systems of up to $L = 14$ [62]. Whether the Edwards-Anderson model does indeed undergo a spin glass phase transition or not is still an open question in physics. While it is known and established that its one-dimensional version does not, non-conclusive results suggest that such phase transition does occur in three- and higher-dimensional systems [35, 63, 64, 65, 66, 67, 68]. On the other hand, the case in between, namely $d = 2$, is thought to not present such behaviour [68].

The Sherrington-Kirkpatrick model

The Sherrington-Kirkpatrick (SK) model [69, 70] was proposed in 1975 as an analytically solvable model of a spin-glass, and it can be thought of as a long-range interaction generalization of the Hamiltonian Eq. (2.9). Its physics are governed by

$$H = \frac{-1}{\sqrt{N}} \sum_{i \neq j} J_{ij} s_i s_j, \tag{2.10}$$

with N the number of spins in the system and the couplings J_{ij} again distributed according to a Normal distribution. Te SK model, contrarily to the EA model, considers full connectivity, this is, a graph in which all spins interact with each other in pairs. The $\sqrt{N}$ factor makes the energy intensive with the system size. Furthermore, this model is known to present both ferromagnetic and spin-glass phases from its original conception [69].

While the full connectivity of the SK model eases its mean-field solution, Parisi developed the Replica Symmetry Breaking (RSB) method to thoroughly study this model's properties. His research in this area would led him to be awarded with the 2021 Nobel prize in physics, "for the discovery of the interplay of disorder and fluctuations in physical systems from atomic to planetary scales".

2.2 QUBO formulation of real world problems

We have so far discussed spin systems as what could easily seem to be, to the untrained eye, purely theoretical entelechies, closer to abstract constructs than to the real physical systems they try to be a model of. Actually, nothing could be further from reality: the mathematical models described in previous sections happen to be the foundational pillars on which modern day optimization techniques rely. Optimization problems are present in a huge plethora of industrial applications, and thus solving them efficiently can be of huge advantage. In this section we outline the ideas beyond the important connection between the theoretical study of spin glass systems and their industrially relevant applications [15, 71, 72].

It is first crucial to note that the Ising model applied to non-planar graphs (such as the Edwards-Anderson model in three or more dimensions) is an NP-complete problem. As such, not only it poses a significant challenge in terms of computational complexity, but any other problem within the NP family can be translated into it in polynomial time as well. Concretely, many optimization problems encountered in real-world applications can be rephrased in a Quadratic Unconstrained Binary Optimization (QUBO) form. The essence of QUBO mapping lies in designing a convenient Hamiltonian function of binary variables with quadratic interactions in such a way that its ground state or minimum energy configuration encodes the solution to the original optimization problem. This design involves the careful definition of the binary variables on the one hand and their couplings and local fields on the other one. The binary variables (spins) must be defined in such a way that any given configuration of them represents (*encodes*) a state of the original problem. The couplings and local fields acting on them, accordingly, must be such that the spin system's dynamics reflect the constraints present in the original problem.

The process of finding the optimal solution is then substituted by the one of looking for the ground state of the equivalent spin system, which can be done by the application of standard techniques. Once the ground state is obtained, the corresponding solution to the original problem is obtained by *decoding* its spin configuration.

Mathematically, a QUBO problem is any optimization problem working with binary variables involved in quadratic functions and with linear constraints. In its general form, it can be written as the minimization of the objective function $f(\vec{x})$

$$\min_{\vec{x}} f(\vec{x}) = \min_{\vec{x}} \left[\sum_{i,j} Q_{ij} x_i x_j + \sum_i C_i x_i \right], \tag{2.11}$$

where $\vec{x} = [x_0, x_1, .., x_{N-1}]$ is a vector of binary variables and Q_{ij} and C_i stand for quadratic interactions and linear constraints, respectively. The problem consists then in finding the right collection of values in $\vec{x}$ such that $f(\vec{x})$ reaches its minimum value. If the maximization of the objective function is desired instead, the minimization of its negation, $-f(\vec{x})$, is performed. Note that the objective function $f(\vec{x})$ is formally equivalent to the Ising Hamiltonian (2.1). While, in the study of spin glasses from a physical perspective, the spin variables are usually assigned $s_i = \{\pm 1\}$ values, in the QUBO literature it is more recurrent to use $\sigma_i = \{0, 1\}$ instead. One can easily change from one framework to the other by the simple change of variable

$$s_i = 2\sigma_i - 1. \tag{2.12}$$

QUBO formulations of many relevant problems that arise in fields ranging from protein folding [73, 74] to logistics [75, 76, 77] are given and derived in detail in the classical reference [78], and an example with a real world application in the field of telecommunications is discussed and solved in Appendix A.

2.3 Heuristic optimization methods

In practical optimization problems, the time required to obtain a solution is a critical factor and is often considered as a resource to be optimized along with the primary objective. This is particularly true in fields such as operations research, logistics, engineering design, finance, and many others. Essentially, when dealing with optimization problems, the goal is not just to find a solution that satisfies certain criteria but also to find it within a reasonable amount of time. On top of that, it is usual for many different sub-optimal solutions to differ with the global minimum in an exponentially small amount, and thus for the optimization problem to render into an "optimal stopping" problem in which the goal is to spend the minimum time possible for a good enough solution to be obtained. In real-world applications, a properly balanced compromise between the time-to-solution and the quality of the solution itself arises as an essential need, and therefore the time complexity of the optimization strategy becomes a crucial consideration.

Let us discuss the scalability properties of a real-world example: the Traveling Salesman problem (TSP). Being one of the most paradigmatic examples of NP problems, it consists on finding the shortest route that visits a set of N interconnected places constrained to the fact that each place must be visited exactly once and that the path must be closed, so that the starting and the final point must coincide. The total number of possible paths within the graph representing the problem can be counted as the number of different permutations of a list of N elements and discarding those that follow the same sequence of points but starting at different initial places or going in reverse order (as we are considering the "easy", undirected version of the problem). Thus for such a graph we have a total number of possible n_p paths given by

$$n_p = \frac{1}{2}(N-1)!, \tag{2.13}$$

which yield, for different N, the values given in Table 2.1. Note the rapid explosion of the size of the solution space: already for 64 points the possible number of paths is some orders of magnitude greater than the most generous estimations of the total number of atoms in the universe. The naïve strategy of exploring all possible paths and computing their length in order to compare them and select the shortest among the list is hence clearly unfeasible even for small problems. Still, this problem needs to be solved efficiently in a lot of areas of our everyday life, so alternative strategies to solve it (and many other problems of similar difficulty) have been devised.

N	n_p
4	3
8	2520
16	$6.54 \cdot 10^{11}$
32	$4.11 \cdot 10^{33}$
64	$9.91 \cdot 10^{86}$

Table 2.1: Order of magnitude of the number of possible paths n_p in a Traveling Salesman Problem of different sizes N.

In cases like this, in which finding an absolute optimal solution is computationally unfeasible within a reasonable amount of time, approximation algorithms such as heuristic methods are often employed to trade-off solution quality for computational efficiency. Furthermore, the cost functions of such problems look like the spin-glass energy landscapes discussed in previous sections, in the sense that the many local minima differ by sometimes completely inappreciable amounts. Therefore, such local minima constitute a really good approximation of the global ground state.

2.3.1 Monte Carlo methods

Monte Carlo methods [79, 80] are a collection of tools that allow to solve certain deterministic problems using probabilistic heuristics, by means of sampling a certain probability distribution. Using the same example that, according to one of the inventors of its modern formulation (S. Ulam), led to its envisage, the idea behind its functioning is as follows. Suppose you are playing the solitaire card game and you want to compute the probability of wining. Since only one player takes part in the game, the probability of being successful is entirely determined by the order in which the cards within the deck have been gambled. The game is rather complex and, given that the deck consists of a total of 52 cards, the combinatorial problem of computing that probability is tedious enough so that Ulam himself abandoned the idea of reaching an exact conclusion. On the other hand, he decided to just play the game a large number of times, and then approximate the desired probability by means of a statistical interpretation of the results.

Analogously, Monte Carlo methods are nowadays used in a broad spectrum of problems which happen to be very difficult to solve exactly. Paradigmatic examples are multidimensional integrals in mathematics and the simulation of complex systems with many coupled degrees of freedom in physics.

The idea behind Monte Carlo methods is thus to generate an appropriate set of states of the system according to a certain probability distribution, for which Markov processes [81] are used in most of the cases. The obtained sequence of states is then referred to as Markov Chain, and the method as Markov Chain Monte Carlo (MCMC). A Markov process is characterized by a time-independent matrix of transition probabilities, or Markov matrix, which defines the probabilities of generating a state ν given the fact that it is fed a state μ, $P_{\mu\nu} = P(\mu \to \nu)$. The transition probabilities should be chosen such that the Markov chain visits each of the states of the system with the probability distribution we are interested in sampling, and thus that the expectation value or the estimator of a given observable of the system O can be computed as a time average over the visited states:

$$\langle O \rangle = \frac{\sum_\mu O_\mu p_\mu}{\sum_\mu p_\mu} \approx \frac{1}{N} \sum_n O_{\mu_n}, \tag{2.14}$$

where p_μ is the (non-normalized) probability of the system being in state μ. The trick for Eq. (2.14) to work with a low (compared to the usually huge number of possible states needed to compute the exact expectation in its left hand side) number of samples is clearly the fact that the states appearing on the time average are not chosen uniformly, but rather according to a proper probability distribution that generates the important states more often than the others. This requirement is known as importance sampling. Moreover, it is intuitive to state that, for this scheme to have any sense at all, the time average over the Markov chain should asymptotically reach the exact expectation value in the limit of infinite samples. Thus it is also required for the Markov process to be ergodic, meaning that the probability of visiting any state of the system from any other one is non zero, if a long enough chain is considered. Note that this does not imply that all transition probabilities $P(\mu \to \nu)$ must be non-zero (there can indeed be states that are not directly reachable from some others), but rather that there must at least exist a chain of states that lead from one to any other one. Finally, the last property that the Markov process is required to have is that, when equilibrium is reached, the generated probability distribution is indeed the desired one, which is assured by the detailed balance condition:

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu). \tag{2.15}$$

2.3.2 The Metropolis-Hastings algorithm

Any process following the conditions outlined above will generate a chain of states of the system that will make it possible to sample it according to the desired probability distribution and that will thus allow us to measure its desired observables. Nevertheless, choosing the right transition probabilities and at the same time assuring that the algorithm is as efficient as possible does not necessarily have to be trivial. The Metropolis-Hastings algorithm [82] (usually referred to as Metropolis algorithm, for short) is a recipe for doing so in the case of systems in contact with a thermal bath, which happens to be, in the end, the case of interest of the present thesis.

The algorithm starts by considering a system in a state μ. To go around the difficult question of how to choose the new state ν of the Markov chain, we note that the transition probability can be split in two terms

$$P(\mu \to \nu) = S(\mu \to \nu) A(\mu \to \nu), \tag{2.16}$$

where $S(\mu \to \nu)$ is the probability with which a certain next state is selected, and $A(\mu \to \nu)$ is the probability with which the proposed move is accepted. The Metropolis algorithm considers the simple case in which any new state can be proposed with the same probability at any time, and hence $S(\mu \to \nu) = \Omega^{-1}$ is drawn from a uniform distribution, with Ω the number of states of the system (in Ising-like systems as the ones we are interested in, $\Omega = 2^{N_{\text{spins}}}$). Being uniform, $S(\mu \to \nu) = S(\nu \to \mu)$ for all possible states and therefore the detailed balance condition, Eq. (2.15), reads

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{A(\mu \to \nu)}{A(\nu \to \mu)} = \frac{p_\nu}{p_\mu}. \tag{2.17}$$

Since we are interested in simulating a system in contact with a thermal bath, the probability of it being in a given state is proportional to its Boltzmann weight, $p_\mu \propto \exp\left[-\beta E_\mu\right]$ with β the inverse temperature and E_μ the energy corresponding to the state μ. It then follows that the most efficient acceptance probability that is in accordance with all the above discussion is [79]

$$A(\mu \to \nu) = \begin{cases} e^{-\beta(E_\nu - E_\mu)} & \text{if } E_\nu - E_\mu > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{2.18}$$

The Metropolis algorithm thus accepts all proposed movements that lower the energy of the system, and accepts those that increase it with a probability depending on the temperature and energy difference between the current and the proposed state. Seen from an optimization perspective, this feature essentially allows the system escape from local energy minima with the aim of later on falling towards a more optimal solution. This behaviour is sketched in Fig.2.5. At high temperatures (low values of β) the algorithm then accepts virtually all movements, while it tends to stay in the current minimum at low ones (high β).

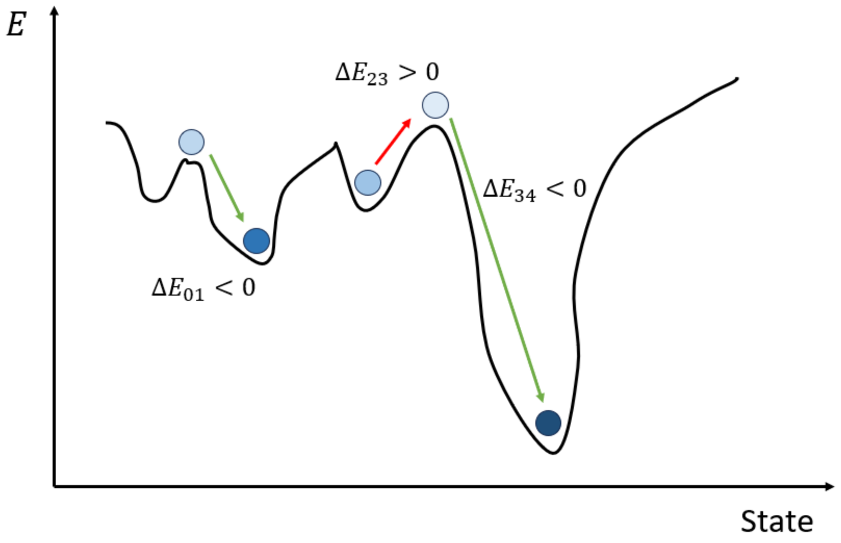

Figure 2.5: **The Metropolis algorithm sometimes accepts new states with a higher energy.** By doing so, the system is allowed to escape from local minima and reach the global ground state.

2.3.3 Simulated Annealing

Simulated Annealing (SA) is the most broadly known (though not the most efficient) Markov Chain Monte Carlo method for optimization purposes. It was first introduced in 1983 by Kirkpatrick [83], and draws inspiration from the annealing process in metallurgy, in which a piece of a certain metal alloy is cooled down from a high temperature to a low one in a carefully enough way such that it crystallizes into a structure without defects and that therefore is as robust as possible. From a physics perspective, this process minimizes the system's Gibbs free energy,

$$F = \langle E \rangle - TS, \tag{2.19}$$

with E the internal energy, T the temperature and S the entropy. If the system is in thermal equilibrium, the states in the canonical ensemble are expected to appear with a probability proportional to its Boltzmann weight

$$P_\mu = \frac{e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}}, \tag{2.20}$$

where $\beta = (k_B T)^{-1}$ is the reduced inverse temperature, k_B is the Boltzmann constant and E_μ is the energy corresponding to the given state. Thus in case the system is properly equilibrated, all states appear with the same probability in the limit of infinite temperature while only the ground state appears, with probability 1, in the limit of zero temperature (or with probability N_{gs}^{-1}, in case the ground state is N_{gs}-degenerate). Accordingly, the exploration of the phase space and thus the equilibration of the system is easier on the former regime and gets more difficult as the temperature is lowered. The objective of the annealing process is then to ensure that the system is in thermal equilibrium at a low temperature, and thus that the ground state can be found. To this end, the process starts in this high-temperature regime in order to guarantee that the system is equilibrated, and then is slowly cooled down. Intuitively, if a system is thermalized at a given temperature and it transitions towards another infinitely close one, the thermal equilibrium should be conserved. Therefore if the system starts at a high enough temperature and the annealing is done adiabatically enough, the ground state will be found with high probability.

Simulated Annealing essentially tries to mimic this procedure. It starts by defining an annealing schedule of decreasing temperatures that the system will sequentially navigate. At each temperature step, the system produces a Markov chain of states with the Metropolis algorithm that effectively samples the corresponding Boltzmann distribution, so that the phase space is properly explored and thus that thermal equilibrium is reached. The Metropolis algorithm thus looks for the minimum energy state and, as discussed previously, allows for less energetically favourable ones to be accepted in order to overcome possible energy barriers.

Note that, if the system presents spin glass behaviour, it will be mandatory for the annealing schedule to span across temperatures both above and below the corresponding spin glass phase transition temperature, to guarantee a good exploration of the energy landscape at the beginning and that the system gets trapped in the ground state towards the end.

2.3.4 Parallellizability of the simulations

We have so far described two regimes at which the simulation of spin systems is of interest and around which the present thesis will orbit, namely equilibrium simulations at the critical temperature and annealing processes for the finding of ground states. Unfortunately, both of them have the property that, in order for proper results to be achieved, long simulation times are required [79]. As will be discussed in following chapters, the former presents diverging correlation times that complicate the measure of statistically independent samples. The later, on the other hand, requires both adiabatic annealing schedules and long thermalization processes to guarantee a proper exploration of the phase space and hence the achievement of thermal equilibrium, such that the global ground state of the system is confidently reached. Further, these times will be dramatically increased for larger systems. The simulation of such physical processes therefore requires huge computational times, and the use of intelligently enough designed algorithms and of specialized hardware accelerators, such as Graphics Processing Units (GPUs) and Field-Programmable Gate Arrays (FPGAs), becomes mandatory.

Luckily, the Ising lattice model with nearest neighbours interactions presented above allows for the utilization of parallel schemes for its simulation, which can effectively boost its performance [84, 85]. To this end, consider any of the spins in the lattice of Fig. 2.2, σ_i, and note that all of the spins located on its immediate diagonal sites (usually referred to as "next-to-nearest neighbours"), $\sigma_{\text{n.n.n. to } i}$, are not connected to it. Contrarily, when proposing an update of spin σ_i from a state μ to a state ν, the energy difference associated to such move does not depend on the state of these next-to-nearest neighbouring spins. From Eq.(2.9) it can be seen that

$$\Delta E(\sigma_i^\mu \to \sigma_i^\nu) = 2\sigma_i^\mu \sum_{j\,\text{n.n. to}\, i} J_j \sigma_j^\mu \neq f(\sigma_{\text{n.n.n. to i}}), \tag{2.21}$$

so that the Metropolis algorithm can be run on two next-to-nearest neighbouring spins independently. A subtle difference in the notation must be stressed: n.n.n. to i refers to "next-to-nearest neighbours to σ_i", and n.n. to i refers to "nearest neighbours to σ_i". Analogously, the same can be done for each of these diagonal spins: none of their next-to-nearest neighbours directly interacts with them and consequently their update does not depend on the state at which the other spins are at a given moment. By extending this idea to the whole lattice, we end up with two sub-lattices containing spins that can be updated independently, in a checkerboard scheme, Fig. 2.6. It is important to note that, for this scheme to work with periodic boundary conditions, the size of the lattice L must be even.

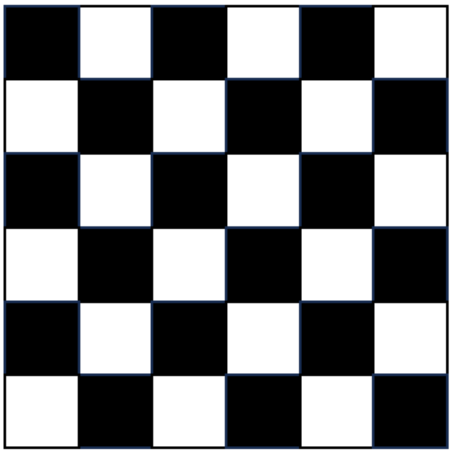

Figure 2.6: **Checkerboard scheme of a two-dimensional spin lattice with nearest neighbours interactions**. The spins pertaining to each of the black/white sub-lattices can be updated independently.

The thermalization of spin lattices is usually carried out for a given number of *Monte Carlo sweeps*, which consist on proposing and accepting/rejecting a total of (usually randomly selected) N spin updates with a Metropolis algorithm, where N is the total number of spins in the system. In this updating scheme, being the two sub-lattices independent, the update proposals of all the spins within the same sub-lattice can be considered in parallel. Consequently, a full Monte Carlo sweep can be carried out with only two steps, corresponding to the updating of each of the two sub-lattices.

It is true that there exists a small fundamental difference when comparing this updating scheme with the standard one. In the later, since the spins proposed for updating are selected at random, it is not guaranteed that all spins will be given the chance to be updated in a given MC sweep, while some others can be proposed more than once. While this one does apparently seem more physically realistic (the order of the updated spins is random and thus not correlated), both schemes satisfy ergodicity and no difference is observed in the obtained results [85]. At the same time, a considerable improvement in performance is achieved with the former parallel version. Furthermore, recall that the generation of (pseudo-) random numbers usually constitutes one of the main bottlenecks of the simulations, which is partially avoided as well in this paradigm.

The nearest-neighbours topology is nevertheless quite simple in the sense that it is a planar graph and thus no hard optimization problems can be embedded in it. This parallelization scheme, though, can be further extended to consider more complex topologies. In the case of the next-to-nearest neighbours connectivity, which is non-planar and NP-complete, one only needs to consider four different sub-lattices instead of two, see Fig.3.1. Analogously, in this case a complete sweep of the lattice can be achieved in four steps.

In this system, each spin is connected to eight different neighbours. It therefore results intuitive to think that for a three-dimensional cubic lattice, in which the spins are only connected to six nearest neighbours, an analogous scheme of at most four colours should be possible. Indeed, a Monte Carlo sweep can be performed in a model with such connectivity with, again, two different sub-lattices: it suffices with considering regular checkerboard sub-lattices as in the two dimensional case, Fig.2.6, but displaced at every alternate layer.

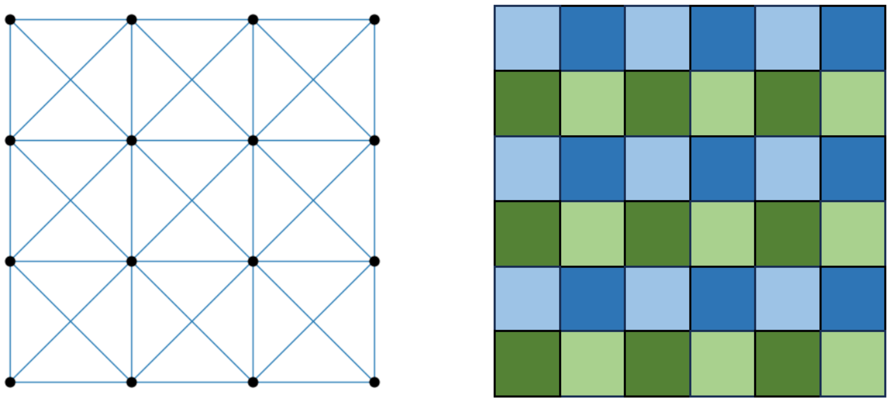

Figure 2.7: **Four-coloured parallelization scheme of a spin lattice with next-to-nearest neighbours connectivity**. The spins pertaining to each of the different coloured sub-lattices can be updated independently.

Chapter 3

Phase transitions of non-planar graphs

In this chapter, based on the results presented in [35], we make use of the Haake-Wilkens-Lewenstein (HLW) approach to spin glasses. This formalism consists on the evaluation of the probability distribution of configurations of two replicas of the system, averaged over quenched disorders. In essence, this is related to the Edwards-Anderson order parameter, q_{EA}, through the definition of an effective system of spins for which the ferromagnetic transition is related to the spin glass one of the original one. This allows us to study the phase transition of two-, three- and four-dimensional EA lattices.

3.1 Introduction

As simple as the general definition of spin-glasses might be, they still represent one of the biggest and more shady topics in modern statistical physics research [43, 68]. As such, many questions related to them remain open and are the focus of a hot debate and, even without the need of going further and considering exotic questions about their behaviour or their quantum version [86, 87, 88], basic concepts such as whether a spin glass phase transition actually exists at all for certain models still pose a tremendous, unresolved challenge. Just as in magnetic order in Ising ferromagnetic metals, the spin-glass phase is defined as a state at which the spins in the system naturally order themselves and remain frozen, unaltered by thermal fluctuations since they are completely overcome by the Hamiltonian's couplings. Contrarily to ferromagnetic systems, though, and due to the more complex distribution of the couplings between spins (which effectively yields to their frustration), this phase is characterized by a configuration that appears to be random. Then since each spin acquires its own orientation, on average half of them point towards one direction and the other half towards the other one, effectively resulting in a zero magnetization for big enough systems. This fact consequently makes μ a useless order parameter to distinguish these two phases. The Edwards-Anderson order parameter was therefore introduced in order to identify this postulated spin-glass phase transition [58].

The efforts devoted to the understanding of this phase transition are nonetheless justified not only from the more fundamental physics part, but from the applied one too. One of the most crucial parts when designing the annealing schedules of heuristic algorithms such as Simulated Annealing, Parallel Tempering and Population Annealing consists of understanding this spin-glass phase transition. Recall that they are designed to anneal the system from (or, in the case of PT, to introduce configurations pertaining to) a high enough temperature in which the exploration of the configuration space is easy but finding the global ground state is difficult (as big thermal fluctuations allow for the easy escape from local minima) towards a low enough temperature, at which the exploration becomes harder but, on the other hand, the probability of laying in the global minimum gets higher. Using annealing schedules with both the initial and the final temperatures within a range at which the system stays in the same phase would consequently not meet this required criteria, and the space exploration would therefore disastrously fail: they definitely need to use temperatures both above and below the critical one. The existence of this phase transition is thus not only the focus of a mere academic curiosity, but rather a question of relevance in environments on which optimization problems are solved by means of these kind of systems. Studying this spin glass problem hence constitutes an important cornerstone for improving currently available optimization techniques. Concretely, most non-planar graphs such as the high-dimensional (3d and beyond) Edwards-Anderson model still lack thorough analytical solutions that allow us to confidentially discriminate whether such phase transition occurs or not. Based on partial results, though, it is believed that such transition does take place on three- and four-dimensional EA models, but not in the 2d case.

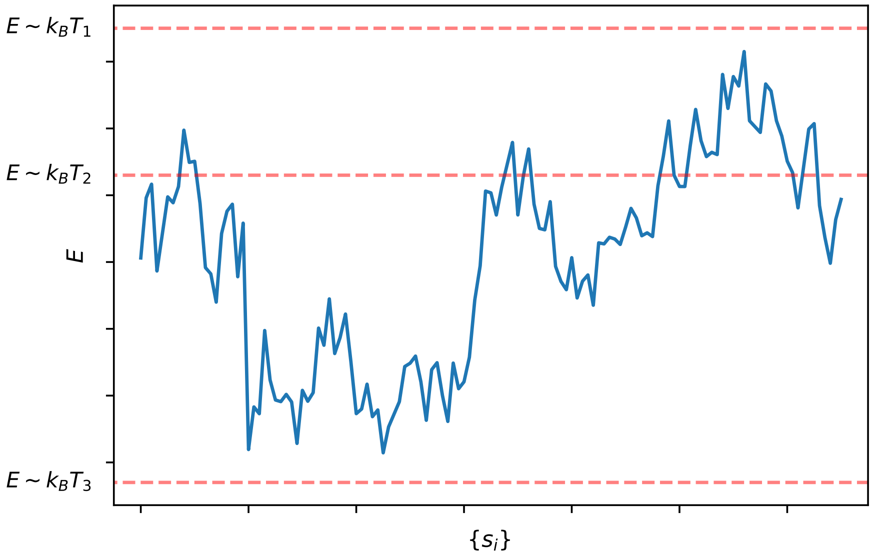

Figure 3.1: **The temperature at which the system is simulated determines the maximum (within thermal fluctuations) energy that can be explored**. In order to explore the whole configurational space, high enough temperatures must be used at the start of the annealing process, while low enough ones are required towards its end to guarantee that the chance of finding the ground state is high.

3.2 The HLW approach

The Edwards-Anderson order parameter for spin glasses, Eq. (2.7), can be understood from different physical perspectives [58, 59, 61, 89], as discussed in previous chapters. An equivalent formulation is given by considering the spin overlap function over two replicas (here labelled α and γ, in order to avoid confusion with the inverse temperature β) of the same system, averaged over both thermal configurations and over the distribution of bonds in the disorder

$$q_{\alpha\gamma} = \big\langle \langle \sigma_i^{(\alpha)} \sigma_i^{(\gamma)} \rangle_T \big\rangle_b . \tag{3.1}$$

The thermal average, $\langle\cdot\rangle_T$, is obtained with the canonical Boltzmann distribution

$$P(\{\sigma\}) = \frac{e^{-\beta H(\{\sigma\})}}{Z(\beta)}, \tag{3.2}$$

while the disorder one, $\langle\cdot\rangle_b$ is computed according to the distribution that the values of the couplings between spins follow in a given model. In our case, since we are considering the EA model, Eq.(2.9), we are interested in Gaussian distributed couplings, and therefore it shall be computed according to

$$P(J) = \frac{1}{\sqrt{2\pi\Delta^2}} \exp\frac{-J^2}{2\Delta^2}. \tag{3.3}$$

Replicas of the same system are defined by the same set of quenched couplings $\{J_{ij}\}$, and must therefore share the same partition function which, for brevity, we denote as $Z(\{J_{ij}\}) \equiv Z(J)$. The thermal average on Eq. (3.1) can thus be computed as

$$q_{\alpha\gamma} = \left\langle \sum_{\{\sigma^{(\alpha)}\}} \sum_{\{\sigma^{(\gamma)}\}} \sigma_i^{(\alpha)}\sigma_i^{(\gamma)} \exp\left[-\beta H(\sigma^{(\alpha)}) - \beta H(\sigma^{(\gamma)}) - 2\ln Z\right] \right\rangle_b, \tag{3.4}$$

where H is the Edwards-Anderson Hamiltonian introduced earlier, Eq. (2.9).

Note that, since the probability distribution of the couplings is even, the sign of J can be absorbed into the spin variables. The following gauge transformation of the couplings can then be considered

$$J_{ij} = J_{ij}\sigma_i\sigma_j, \tag{3.5}$$

which can be used to write the sum of the Hamiltonians for the two replicas in Eq. (3.4) as

$$\sum_{\langle i,j\rangle} J_{ij}\sigma_i^{(\alpha)}\sigma_j^{(\alpha)} + \sum_{\langle i,j\rangle} J_{ij}\sigma_i^{(\gamma)}\sigma_j^{(\gamma)} = \sum_{\langle i,j\rangle} J_{ij} \underbrace{\sigma_i^{(\alpha)}\sigma_j^{(\alpha)}\sigma_i^{(\alpha)}\sigma_j^{(\alpha)}}_{=1} + \sum_{\langle i,j\rangle} J_{ij}\sigma_i^{(\gamma)}\sigma_j^{(\gamma)}.\sigma_i^{(\alpha)}\sigma_j^{(\alpha)} \tag{3.6}$$

For any possible orientation of the spins, $\{\pm 1\}$, the square of a given spin variable of the same replica always yields $\sigma_i^{(\alpha)}\sigma_i^{(\alpha)} = 1$. Further, the spin overlap variables can be defined

$$\tau_i = \sigma_i^{(\alpha)}\sigma_i^{(\gamma)}. \tag{3.7}$$

With all these ingredients, the Edwards-Anderson order parameter reads

$$q_{\alpha\gamma} = \left\langle \sum_{\{\sigma^{(\alpha)}\}} \sum_{\{\sigma^{(\gamma)}\}} \tau_i \exp\left[\beta \sum_{\langle i,j\rangle} J_{ij}(1 + \tau_i\tau_j) - 2\ln Z\right] \right\rangle_b. \tag{3.8}$$

Finally, rearranging some terms and noting that, since both σ and τ are spin variables, the summations can be carried out over any of them indistinguishably, $\sum_{\{\sigma^{(\alpha)}\}}\sum_{\{\sigma^{(\gamma)}\}} \equiv \sum_{\{\sigma\}}\sum_{\{\tau\}}$, one gets

$$q_{\alpha\gamma} = \frac{2^N}{Z^2}\sum_{\{\tau\}}\tau_i\left\langle \exp\left[\beta\sum_{\langle i,j\rangle}J_{ij}(1+\tau_i\tau_j)\right]\right\rangle_b. \tag{3.9}$$

Note that, in this form, the order parameter is reinterpreted as the magnetization of a system with an effective canonical distribution

$$q_{\alpha\gamma} = Z_{\text{eff}}^{-1}\sum_{\{\tau\}}\tau_i e^{-\beta H_{\text{eff}}(\tau)}, \tag{3.10}$$

with the effective probability distribution of the new variables τ

$$P(\tau) = Z_{\text{eff}}^{-1}e^{-\beta H_{\text{eff}}(\tau)} = \left\langle \frac{2^N}{Z^2(J)}\exp\left[\beta\sum_{\langle i,j\rangle}J_{ij}(1+\tau_i\tau_j)\right]\right\rangle_b, \tag{3.11}$$

or, equivalently, the disorder-averaged joint probability distribution of configurations of two replicas, $P(\sigma^{(\alpha)},\sigma^{(\gamma)})$.

The Haake-Lewenstein-Wilkens approach [90] therefore consists on studying the spin glass phase transition of the original spin system with the magnetization of an effective system of variables τ. Correspondingly, we will have:

- Non-zero values of the Edwards-Anderson order parameter in the original system, if the spin-overlap variables τ show magnetic order.

- Disordered spin glass phase, characterised by vanishing q_{EA}, if the effective system shows paramagnetic behaviour.

3.3 Saddle point/steepest descent calculations

For convenience, let us first rescale the couplings with the inverse temperature β and the variance of their Gaussian distribution Δ. To this end we apply the following change of variables to define the new couplings κ_{ij}:

$$J_{ij} = \beta\Delta^2\kappa_{ij}. \tag{3.12}$$

Further, we also define the scaled inverse temperature

$$\tilde{\beta} = \beta\Delta. \tag{3.13}$$

The partition function now is

$$Z(\kappa) = \sum_{\{\sigma\}} \exp\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \kappa_{ij}\sigma_i\sigma_j\right], \tag{3.14}$$

where we have explicitly included the dependency of Z on the couplings used. On the other hand, the bond distribution becomes

$$P(\kappa) = \frac{1}{\sqrt{2\pi\tilde{\beta}^{-2}}} \exp\frac{-\tilde{\beta}^2\kappa^2}{2}. \tag{3.15}$$

And, therefore, the effective probability distribution of the spin-overlap variables τ finally reads

$$P(\tau) = \left\langle \frac{2^N}{Z^2(\kappa)} \exp\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \kappa_{ij}(1+\tau_i\tau_j)\right]\right\rangle_b, \tag{3.16}$$

where the disorder average $\langle\cdot\rangle_b$ is now computed according to the bond distribution Eq. (3.15). The probability distribution is therefore given by the integral

$$P(\tau) = \frac{2^N}{\sqrt{2\pi\tilde{\beta}^{-2}}} \int d\kappa_{ij}\, \underbrace{Z^{-2}(\kappa_{ij}) \exp\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \left\{\kappa_{ij}(1+\tau_i\tau_j) - \frac{\kappa_{ij}^2}{2}\right\}\right]}_{L(\kappa,\tau,\tilde{\beta})} \tag{3.17}$$

which we can approximate with the saddle point/steepest descend (SPSD) method. To this end, we first look for the saddle point of the integrand $L(\kappa,\tau,\tilde{\beta})$, by equating to zero the first derivative with respect to κ_{ij} of its logarithm. The saddle point is found to be given by

$$0 = \tilde{\beta}^2\big[-\kappa_{ij} + 1 + \tau_i\tau_j - 2\langle\sigma_i\sigma_j\rangle_T\big], \tag{3.18}$$

with the thermal average computed according the distribution Eq.(3.2) with the properly rescaled couplings, Eq.(3.12):

$$\langle\sigma_i\sigma_j\rangle_T = \frac{1}{Z}\sum_{\{\sigma\}} \sigma_i\sigma_j \exp\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \kappa_{ij}\sigma_i\sigma_j\right]. \tag{3.19}$$

The probability distribution can be approximated around this point by directly substituting their SPSD values. Further, we can also consider the second order correction by computing the Hessian matrix of the second derivatives of the logarithm of the integrand. This yields

$$\mathcal{H}_{\mu\nu} = -\tilde{\beta}^2\left[\delta_{\mu\nu} + 2\tilde{\beta}^2\left(\langle c_\mu\rangle\langle c_\nu\rangle - \langle c_\mu c_\nu\rangle\right)\right], \tag{3.20}$$

where we have introduced the shortened notation $\langle \sigma_i \sigma_j \rangle = c_\mu$. We further note that the correlation matrix

$$\langle c_\mu c_\nu \rangle - \langle c_\mu \rangle \langle c_\nu \rangle = \langle (c_\mu - \langle c_\mu \rangle)(c_\nu - \langle c_\nu \rangle) \rangle, \tag{3.21}$$

i.e. it is explicitly positively semi-definite. In fact, the Hessian matrix:

$$\hat{\mathcal{H}} < 0, \tag{3.22}$$

so that the logarithm of the integrated function considered is a strictly convex function of many variables. It is therefore expected to have one maximum corresponding to our SPSD solutions. Note also that eigenvalues of the Hessian matrix are all negative and will typically be of order α^4, and they are bounded in modulus from below by $\tilde{\beta}^2$. One should thus expect the SPSD method to become, for $\tilde{\beta} \to \infty$, asymptotically very precise or even exact.

Let us now study the behaviour of the obtained SPSD solutions. Being Ising variables, the spin-overlap variables τ can only take on the values ± 1, and thus so can their product $\tau_i \tau_j = \pm 1$. The saddle point equation, Eq.(3.18), therefore has two distinct regimes depending on the value of the spin overlap $\tau_i \tau_j$:

- $\tau_i \tau_j = 1$. The corresponding couplings κ_{ij} are then restricted by

  $$\kappa_{ij} = 2\left(1 - \langle \sigma_i \sigma_j \rangle_T\right) > 0, \tag{3.23}$$

  hence it represents a ferromagnetic coupling.

- $\tau_i \tau_j = -1$. In this case the couplings read

  $$\kappa_{ij} = -2 \langle \sigma_i \sigma_j \rangle_T, \tag{3.24}$$

  so we should in principle have an antiferromagnetic coupling. Contrarily, an antiferromagnetic coupling between spins located at sites i and j implies, at low temperatures (alternatively high values of $\tilde{\beta}$, for which, as discussed above, these results should be accurate or even exact), a negative correlation between them, $\langle \sigma_i \sigma_j \rangle_T < 0$. Eq. (3.24) therefore yields a contradiction, that could be avoided by setting $\kappa_{ij} = 0$. As we shall see below by solving mean field equations, this is in fact not needed, but it only represents a manifestation of the frustration of the system.

Summarizing, the SPSD equation Eq. (3.18) can be interpreted as describing a system formed by domains of aligned spins coupled with ferromagnetic interactions, $\kappa_{ij} > 0$, characterized by $\tau_i \tau_j = 1$. These islands or domains are then limited by domain walls described by antiferromagnetic couplings, $\kappa_{ij} < 0$, and thus unaligned spins, $\tau_i \tau_j = -1$.

3.4 Thouless argument

We now discuss whether a ferromagnetic phase in τ exists or not and, consequently, a spin glass phase in σ. As first proposed by Thouless [91, 92, 93], for models of electron propagation in the presence of disorder, and subsequently adapted to study Ising models in random magnetic fields [94, 95], and also spin glasses [96], the existence of a spin glass order can be addressed by studying the sensitivity of the system to boundary conditions. To this end, we consider a d-dimensional cylindrical spin lattice of length L and cross section L^{d-1}. Let us define the probability of all spins τ being aligned

$$P_+ = P(\tau_i = 1 \forall i), \tag{3.25}$$

and the probability of the existence of a domain wall, defined as that of all spins to its left pointing towards a certain direction and all spins to its right pointing to the other one:

$$P_- = P(\tau_i = 1 \forall i \in D_l \cap \tau_i = -1 \forall i \in D_r). \tag{3.26}$$

Here, $i \in D_l$ and $i \in D_r$ represent the indices of the spins i to the left and right of the domain wall, respectively. By comparing these two probabilities we can study whether the system acquires ferromagnetic order ($P_+ > P_-$) or not ($P_+ < P_-$). To this end we define

$$\delta = \ln \frac{P_+}{P_-}. \tag{3.27}$$

Recall now that, in the HLW approach, ferromagnetic order in the τ variables implies spin glass order in the original spin system of σ variables. Consequently, considering both the Thouless argument and the HLW perspective of spin systems we can study its spin glass phase. The system of σ spins will lay in:

- a spin glass phase, if $\delta > 0$,
- a disordered phase, if $\delta < 0$.

The existence of a spin glass phase transition can then be inferred by the crossing of the δ parameter from negative to positive values.

3.5 Self-consistent SPSD and local mean field solutions

In order to use the Thouless argument presented above to study the spin glass phase transition of the Edwards-Anderson model, let us consider a d-dimensional hypercubic lattice embedded within a cylinder with L layers of spins and bonds distributed according to a Gaussian law.

Using standard mean field theory, we approximate

$$\sigma_i = \underbrace{\langle\sigma_i\rangle}_{m_i} + \delta\sigma_i, \tag{3.28}$$

where we have defined the mean field (MF) value of the magnetization of a given spin, m_i. It can be computed to yield

$$m_i = \tanh\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \kappa_{ij}\sigma_j\right]. \tag{3.29}$$

Discarding all terms quadratic in the fluctuations, the SPSD equation, Eq. (3.18), becomes

$$\kappa_{ij} \approx 1 + \tau_i\tau_j - 2m_i m_j, \tag{3.30}$$

and thus the MF value of the magnetization reads

$$m_i = \tanh\left[\tilde{\beta}^2 \sum_{\langle i,j\rangle} \left(1 + \tau_i\tau_j - 2m_i m_j\right) m_i\right]. \tag{3.31}$$

Being a hyper-cubic lattice of d dimensions, each spin is coupled to $2d$ neighbours, with only two of them in the cylinder's axis direction. Further, if we assume axial symmetry in the cylinder, all spins belonging to the same layer will have the same value m_i, where now the index i only labels the layer. Note that not only their mean field value will be equal, but their exact one too so that its possible values are restricted to ± 1 and thus their square equals one, $m_i m_i = 1$. On top of that, we are in the ferromagnetic regime and thus $\tau_i\tau_j = 1$. Taking all of this into account, Eq. (3.31) is then interpreted as a chain of layers with magnetization m_i, coupled to the previous and the following layers. Each layer is nonetheless comprised by $2(d-1)$ "sub-layers" in the directions perpendicular to the cylinder's axis, which will each generate the same equation. The magnetization at each axial layer i in the cylinder is then

$$m_i = \tanh F_i(m), \tag{3.32}$$

where we have defined, for brevity

$$F_i(m) = \tilde{\beta}^2\left[2\left(d-1\right)\kappa_{i,i}m_i + \kappa_{i,i-1}m_{i-1} + \kappa_{i,i+1}m_{i+1}\right]. \tag{3.33}$$

By continuity, the boundary conditions are

$$m_0 = m_{L+1} = 0, \tag{3.34}$$

$$\kappa_{0,1} = \kappa_{L,L+1} = 0. \tag{3.35}$$

The values of κ_{ij} are given by the SPSD equations, Eq.(3.30), and thus we again have the two possibilities discussed there, namely $\tau_i\tau_j = \pm 1$. When we compute the magnetization within a ferromagnetic domain, by the discussion of previous sections, we have $\tau_i\tau_j = 1$ and thus

$$\kappa_{i,j} = 2 - 2m_i m_j. \tag{3.36}$$

On the other hand, the magnetization in a domain wall is characterised by $\tau_i\tau_j = -1$, such that

$$\kappa_{i_D,i_D+1} = -2m_{i_D}m_{i_D+1}, \tag{3.37}$$

where i_D labels the layer at which the domain wall is encountered. Recall that we are assuming axial symmetry, and thus the boundaries between domains must be flat hyper-planes that comprise the whole layer. While there is not a strict physical restriction in this regard, we are in essence studying effective, averaged-over-disorder probability distributions of the τ variables, and hence from the construction there are no reasons for this probability distribution to break the translation symmetry.

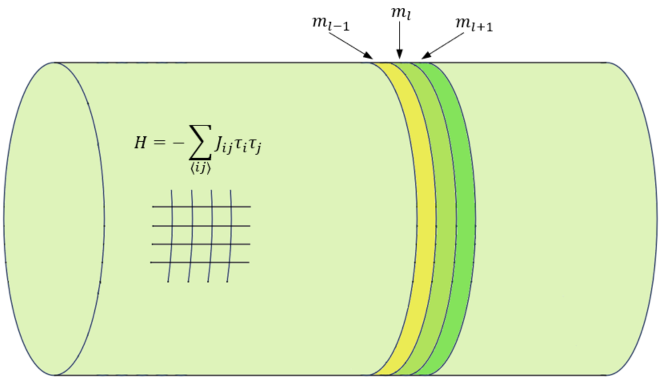

Figure 3.2: **A spin lattice with nearest neighbours interactions in a cylinder**. Due to the cylinder's axial symmetry, all spins belonging to the same layer will have the same magnetization m_i.

We now finally have all the necessary ingredients to address the computation of δ, as given by Eq. (3.27), which will allow us, by the previous' sections discussions, to study the spin glass phase transition in our model. We define the logarithm of the probability distribution

$$H_\pm = \ln P_\pm, \tag{3.38}$$

and call it "energy", in a somewhat abuse of notation, by an analogy to a Boltzmann weight, $P \propto \exp(-\beta E)$. Being an extensive quantity, the energy divided by the volume in all but one dimensions should be, in the leading order, a linear function of the cylinder's length

$$\frac{H_\pm}{L^{d-1}} = A(\tilde{\beta})L + B_\pm(\tilde{\beta}). \tag{3.39}$$

Since the configuration contributing to P_- has a domain wall in the middle of the cylinder and thus a connection between two layers given by a different expression, Eq. (3.37), both energies will have the same bulk contributions $A(\tilde{\beta})$ but different boundary terms, $B_\pm(\tilde{\beta})$. Thus we will have

$$\delta = B_+(\tilde{\beta}) - B_+(\tilde{\beta}). \tag{3.40}$$

Let us now explicitly compute the energies $H_\pm$ from the probability distribution of the τ spin variables, Eq. (3.17), again with the Laplace method. We find

$$H_\pm = L\ln 2 + \tilde{\beta}^2 \sum_{\langle i,j\rangle} \left\{ 2\kappa_{i,j} - \frac{\kappa_{i,j}^2}{2} \right\} - 2\ln Z, \tag{3.41}$$

The partition function is

$$Z = \prod_i 2\cosh\left\{ \tilde{\beta}^2 \sum_{\langle i,j\rangle} \kappa_{ij}\sigma_j \right\} \tag{3.42}$$

with the product over all spins. As discussed previously, all spins within each of the L $(d-1)$-dimensional layers will have the same magnetization. Thus grouping them and taking the logarithm it translates to

$$\ln Z = L^{d-1} \sum_{i=0}^{L-1} \ln\left(2\cosh F_i\right), \tag{3.43}$$

with F_i as defined in Eq. (3.33) and the sum over all layers. Taking into account the different conditions for each energy, one finally finds

$$\begin{aligned} \frac{H_+}{L^{d-1}} &= L\ln 2 + \tilde{\beta}^2 \left\{ (d-1)\sum_{i=1}^{L}\left[2\kappa_{i,i} - \frac{\kappa_{i,i}^2}{2} \right] + \sum_{i=1}^{L-1}\left[2\kappa_{i,i+1} - \frac{\kappa_{i,i+1}^2}{2} \right] \right\} - \\ &- 2\sum_{i=1}^{L} \ln\left[2\cosh F_i(m) \right] \end{aligned} \tag{3.44}$$

for the completely ferromagnetic case, Eq. (3.25), and

$$\begin{aligned} \frac{H_-}{L^{d-1}} &= L\ln 2 + \tilde{\beta}^2 \left\{ (d-1)\sum_{i=1}^{L}\left[2\kappa_{i,i} - \frac{\kappa_{i,i}^2}{2} \right] + \sum_{i=1}^{L/2-1}\left[2\kappa_{i,i+1} - \frac{\kappa_{i,i+1}^2}{2} \right] + \right. \\ &+ \left. \sum_{i=1}^{L/2-1}\left[2\kappa_{i,i+1} - \frac{\kappa_{i,i+1}^2}{2} \right] \right\} - 2\sum_{i=L/2+1}^{L-1} \ln\left[2\cosh F_i(m) \right] \end{aligned} \tag{3.45}$$

for the case with two domains separated by a domain wall, Eq. (3.26). Recall that, in the later case, the condition Eq. (3.37) must be imposed. One could also include in each case the Gaussian fluctuations around the SPSD solution term, by adding

$$\frac{\delta H_{\pm}}{L^{d-1}} = -\frac{1}{2}\ln\left(\det \hat{\mathcal{H}}_{\pm}\right) \tag{3.46}$$

correspondingly. By approximating the eigenvalues by the diagonal elements of the Hessian matrix computed previously, Eq. (3.20), one finds

$$\frac{\delta H_{\pm}}{L^{d-1}} \approx -\frac{1}{2}\sum_{ij}\ln\left[\tilde{\beta}^2 + 2\tilde{\beta}^4(1 - m_i^2 m_j^2)\right], \tag{3.47}$$

where the SPSD solutions must be computed accordingly for each one of the ± cases.

3.5.1 High $\tilde{\beta}$ regime

Let us now focus on the study of the high $\tilde{\beta}$ (low temperature) regime, in which the extensive parameter of the energies, $A(\tilde{\beta})$, can be estimated analytically. The corresponding mean-field (MF) equations in the bulk are

$$\kappa = 2(1 - m^2), \tag{3.48}$$

$$m = \tanh 2\tilde{\beta}^2 d\kappa m. \tag{3.49}$$

It is now useful to define the new variable $x = 2\tilde{\beta}^2 d\kappa$ and combine Eq. (3.48) and Eq. (3.49) into

$$x = 4\tilde{\beta}^2 d \cosh^{-2}\left[x\sqrt{1 - \frac{x}{4\tilde{\beta}^2 d}}\right] \tag{3.50}$$

from which the limit in the low temperature regime can be extracted

$$x = \frac{1}{2}\ln 4\tilde{\beta}^2 d, \tag{3.51}$$

which, through the definition of x, yields

$$\kappa = \frac{\ln 4\tilde{\beta}^2 d}{4\tilde{\beta}^2 d}. \tag{3.52}$$

Thus, as should be expected, κ tends to zero in the limit of large $\tilde{\beta}$ and, using Eq.(3.49), the magnetization is maximum in the bulk, $m = 1$. Finally, this leads to

$$A(\tilde{\beta}) \simeq \ln(2) - \frac{1}{2}\ln(4\alpha^2 d), \tag{3.53}$$

i.e. as expected $\ln(P_{\pm}) = H_{\pm}$ becomes negative at large L (when our analysis makes sense) and at large $\tilde{\beta}$ (when SPSD should work well). Furthermore, $A(\tilde{\beta})$ is seen to diverge with $\tilde{\beta}$, but only in a logarithmic way.

Since in this regime we can then set the local magnetization to $m_i = 1 \forall i$, the expression for the energy Eq.(3.44) simplifies to

$$\frac{H_+}{L^{d-1}} = L\ln(2) - 2\alpha^2 \left[(d-1) \sum_{i=1}^{L} \kappa_{i,i} + \sum_{i=1}^{L-1} \kappa_{i,i+1} \right], \tag{3.54}$$

in which the Gaussian corrections have been neglected, according to the results obtained in previous sections when discussing the eigenvalues of the Hessian matrix, Eq (3.20). Since the numerical analysis in the asymptotic regime is tough, we may and will use this expression there.

3.5.2 Phase transition at moderate $\tilde{\beta}$

The solution of the mean field equations presented above can be seen to change character as the scaled temperature $\tilde{\beta}$ increases. In the high temperature regime (low values of $\tilde{\beta}$), the system is not magnetized and thus $m_i = 0 \forall i$, and below a certain transition temperature $\tilde{\beta}_T$ the system acquires a spontaneous magnetization, $m_i \neq 0$. We can then approximate Eq.(3.32) for temperatures around the critical point as a series expansion for small m_i as

$$m_i \approx \tilde{\beta}^2 \Big[2\,(d-1)\,\kappa_{i,i} m_i + \kappa_{i,i-1} m_{i-1} + \kappa_{i,i+1} m_{i+1} \Big]. \tag{3.55}$$

To first order, around criticality the magnetization of all the layers should be equal, $m_i = m \forall i$, and by Eq. (3.36) we have $\kappa_{ij} = 2$. Taking into account these considerations into Eq. (3.55) and then imposing it to get trivial at $\tilde{\beta}_T$ we finally find the transition temperature to be

$$\tilde{\beta}_T = \frac{1}{2\sqrt{d}}. \tag{3.56}$$

3.5.3 Numerical calculations

After the analytical discussion of the last sections, we shall now address the numerical computations of the derived equations describing the behaviour of the τ spin system on the cylinder. We first compute the magnetization, then discuss the domain wall width separating the two regimes, and finally analyze the crossing of the δ parameter, Eq.(3.27), to discuss about the spin glass phase transition in the original lattice of σ spins.

Magnetization over the cylinder

We now solve numerically the system of equations conformed by Eq. (3.32) and the boundary conditions imposed by the lack of spins in sites $i = 0$ and $i = L+1$, Eq. (3.34) and Eq. (3.35), respectively, and the condition imposed by the existence of a domain wall in the middle of the cylinder, Eq. (3.37). As shown in Fig. 3.3, the systems in all the considered dimensions show qualitatively the same behaviour around their corresponding critical temperatures, as given by Eq. (3.56).

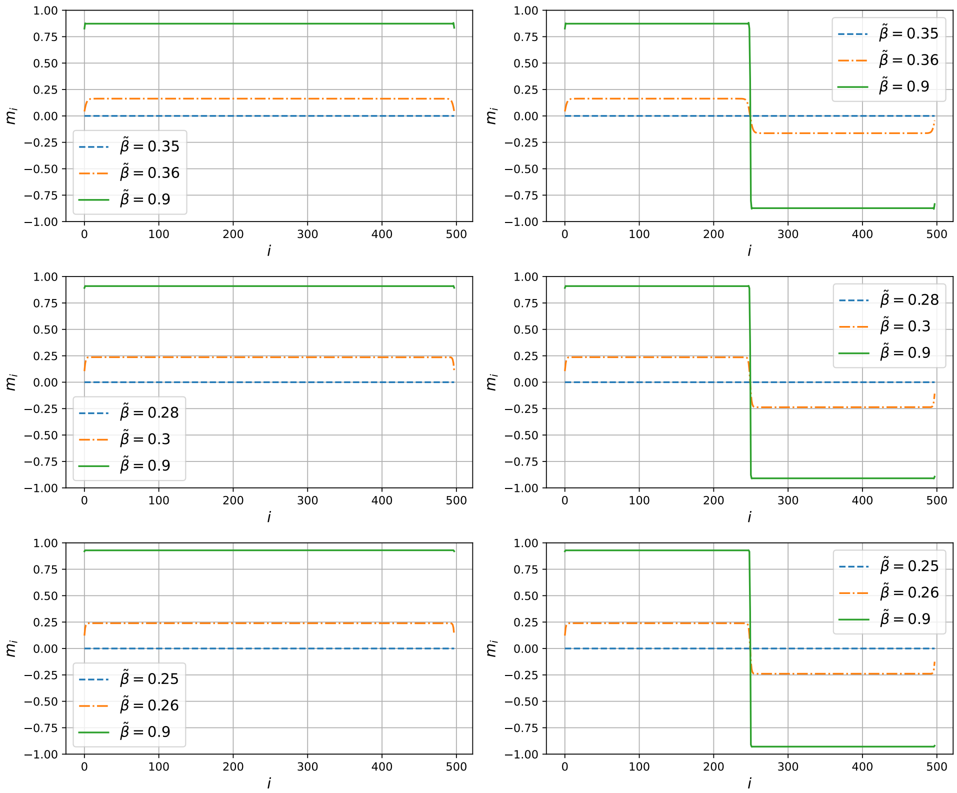

Figure 3.3: **Magnetization over a cylinder with $L = 500$ for different temperatures and dimensions**. Both systems (with and without a domain wall) get magnetized when going from high to low temperatures. Results obtained for temperatures above criticality (blue line), slightly below it (orange line) and far below it (green line), for a system with (right panel) and without (left panel) domain wall. Top, middle and bottom panels correspond to systems in $d = \{2, 3, 4\}$, respectively.

A closer inspection of the domain walls in all dimensions shows that, rather than being a sharp jump, they have in fact a finite width Δ_D. A numerical exploration indicates that, at a fixed temperature, the domain wall reaches an L-independent width limit for large L, as can be seen by the $^{\Delta_D}/_L \propto L^{-1}$ behaviour in the right panel of Fig.3.4. Analogously, for a fixed number of layers L, the domain wall reaches a temperature-independent width as well for low enough temperatures, left panel of Fig.3.4.

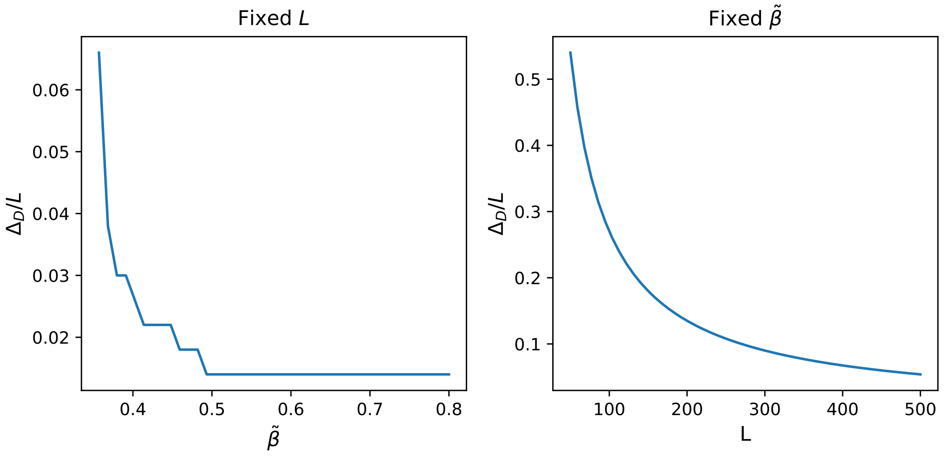

Figure 3.4: **Domain wall width for fixed L (left panel) and fixed $\tilde{\beta}$ (right panel).** Numerical explorations show that, in both regimes, the domain wall asymptotically reaches a fixed width for large enough values of the independent parameter.

Spin glass phase transition

We have already discussed that the energies of the cylinders divided by its volume in all but one dimensions, $H_{\pm}L^{-d+1}$, has to be a linear function of the cylinder's length L, Eq. (3.39). By solving this same system of equations for various values of L and at various temperatures, and fitting the results with a linear regression we can finally compute the functions $A(\tilde{\beta})$, $B_{\pm}(\tilde{\beta})$ and $\delta(\tilde{\beta})$, which effectively tells whether the original spin system undergoes a phase transition between ordered and disordered spin glass phases or not. Furthermore, we also take into account the Gaussian corrections discussed in previous sections. The results show that $A(\tilde{\beta})$,

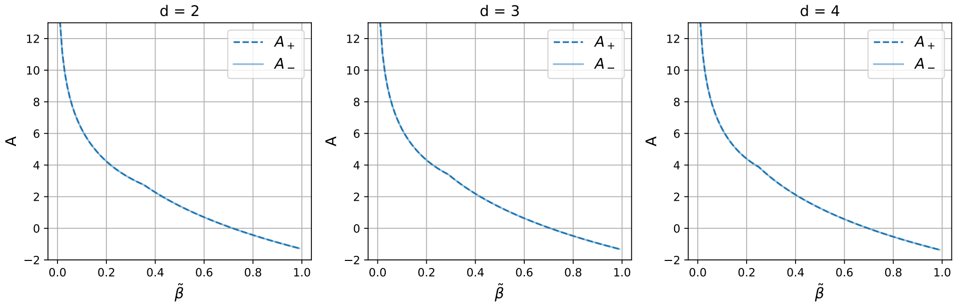

Figure 3.5: $A(\tilde{\beta})$ **for all studied dimensions**. The results show a very similar behaviour of this parameter in all dimensions, both qualitatively and quantitatively.

Fig 3.5, tends to $-\infty$ in a logarithmic way, in accordance to the previous analytic calculations. They behave very similarly both quantitatively and quantitatively in all dimensions. Furthermore, and as expected, the $A(\tilde{\beta})$ functions for the two studied situations (with and without a wall domain) are the same since they represent the bulk energy, which should be only, if at all, slightly affected by the presence of the domain wall. On the other hand, the contributions of the constant term in Eq. (3.39), characterised by $B_{\pm}(\tilde{\beta})$, do diverge between the systems with and without a domain wall, as seen in Fig 3.6. While $B_{+}(\tilde{\beta})$ tends to a positive constant for large $\tilde{\beta}$ (low temperatures), $B_{-}(\tilde{\beta})$ can be seen to tend to infinity for all dimensions. Effectively, this results in δ crossing from positive to negative values in all three cases, Fig. 4.1, thus indicating a spin glass transition at finite temperatures, unfortunately, in all the studied dimensions.

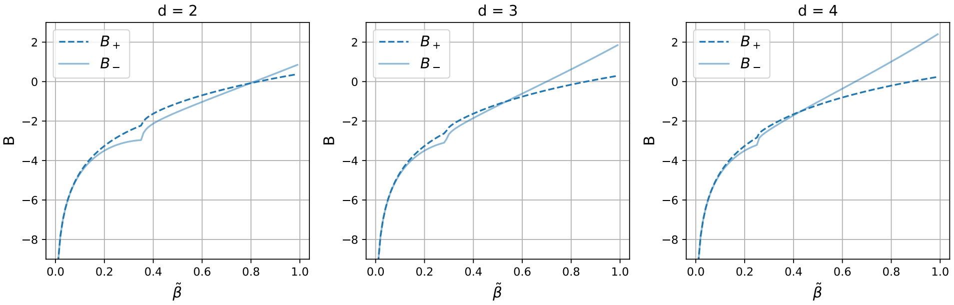

Figure 3.6: $B_{\pm}(\tilde{\beta})$ **for all studied dimensions**. The results show a very similar behaviour of this parameter in all dimensions, both qualitatively and quantitatively. Note that $B_{+}(\tilde{\beta})$ and $B_{-}(\tilde{\beta})$ diverge for large values of $\tilde{\beta}$ (low temperatures).

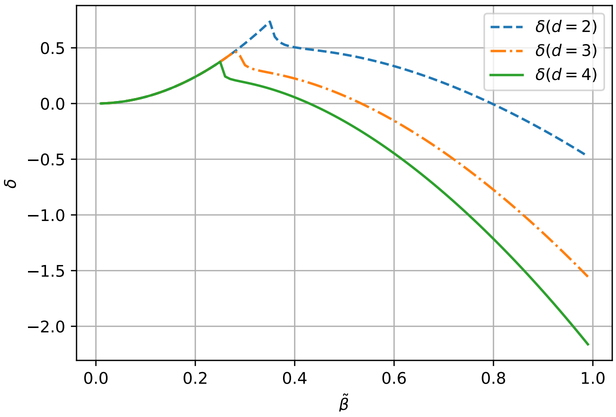

Figure 3.7: $\delta(\tilde{\beta})$ **for all studied dimensions**. The function crossed from positive to negative values at finite temperatures for all dimensions.

As discussed previously, even though this is still an open question, there is broad consensus in that this should be the case in three and more dimensions, but not in two. We therefore hypothesize about the reasons why the used approach could be failing in two dimensions.

First, the energy increase for a domain wall competes with boundary effects more strongly in two dimensions than in the other ones, and thus their local mean field calculations might be too rough, which could be in effect destroying the order. Further, it is possible that the SPSD approximation is not precise enough or even completely incorrect, with many SPSD solutions contributing, when we have only considered one of them. Still, the results obtained for the Hessian matrix suggest that this is not the case as the SPSD solution appears to be the single maximum and is expected to be very precise for large values of $\tilde{\beta}$, as discussed in previous sections.

3.6 Conclusions

The Haake-Lewenstein-Wilkens (HLW) approach is applied to the study of the Edwards-Anderson (EA) Ising spin glass model in its two-, three- and four-dimensional versions. By considering the probability distribution of the spin overlap variables of two replicas of a system averaged over disorder, this formalism allows for the investigation of the behaviour of the standard EA order parameter, q_{EA}. The saddle-point/steepest-descend (SPSD) solutions of the studied system appear to be asymptotically very precise and with a single maximum in the low temperature limit, though a rigorous proof is not explored. Finally, by studying the sensitivity of the SPSD solutions to our system to boundary conditions, the existence of a spin glass phase transition is then inferred. The results indicate that this transition does indeed occur in all the considered dimensions.

The existence of such phase transition in 3 and 4 dimensions agrees with the common belief within the scientific bibliography, but that of two-dimensional systems does not. We consequently identify some reasons why the considered approach may fail in two dimensions. Namely, errors can either arise in the SPSD solutions or in the mean-field (MF) approximation to the boundary effects. The (not rigorous enough) study of the Hessian matrix suggests that the former is unlikely. On the other hand, the competing boundary effects, which are more important in lower-dimensional systems, could be preventing the MF considerations to be precise.

Chapter 4

Correlated RN sequences in MC simulations

In this chapter, based on the results presented in [36], we study how do the correlations present in pseudo random number sequences affect the value of the obtained dynamic critical exponent, z, of two-dimensional Ising models. To this end, we first review some needed theoretical concepts related to the generation of pseudo random numbers, as well as to the physics behind the dynamic critical exponent. We then study some direct effects that using randomness sources of different qualities have in the value of z, to finally address its computation by using completely uncorrelated, true random streams obtained from Quantum Random Number Generators (QRNGs). To conclude, the used (pseudo-) randomness sources are tested by means of standard statistical tests, and the differences between them are compared.

4.1 Introduction

It has been known for decades that some complex deterministic computations can be carried out by intelligently using random numbers, in a process called Monte Carlo simulations [80], and with the often negligible caveat of losing some accuracy. One of the most paradigmatic examples of this type of strategies is the computation of the irrational number π. For it, one first considers a square of unit side length and an inscribed circle and notes that the ratio of the areas between the two geometric figures is $A_{\text{circle}}/A_{\text{square}} = \pi/4$. If one interprets this ratio as the probability of an (x, y)-point drawn from an uniform probability distribution in the $[0, 1]$ interval, $x, y = U(0, 1)$, lying within the circle, then the value of π can be estimated by simply randomly generating a big enough number (N_{total}) of points and counting how many of them (N_{circle}) meet such condition, $\pi \approx \tilde{\pi} = 4N_{\text{circle}}/N_{\text{total}}$.

Still, what one might naively think to be an easy and straightforward task, namely the generation of random numbers, happens to be one of the hardest and less trivial computational problems. From a philosophical perspective, it is actually impossible to generate true random numbers via algorithmic procedures such as those that can be implemented in a computer [97], and any attempt on this regard will, sooner than later, fall into problems such as sequence repetition and detectable correlation. On the one hand, algorithms producing such apparently random sequences rely on deterministic mathematical functions that, given a certain amount of previous states, output a new one, $x_i = f(x_{i-1}, x_{i-2}, ...)$. Therefore, and taking into account that we want our function to output numbers within a certain, finite range and thus the pool of possible states x is limited, it is guaranteed that, for some (hopefully large) iteration i the sequence will start over again and repeat itself. On the other hand, being the output numbers completely deterministic, it is impossible to find a complete way around the intrinsic correlations between them.

This is the reason why one normally refers to this kind of numbers as pseudo random numbers (PRN), and during decades lots of efforts have been put into the improvement of both their generation and their similarity to a real randomness source. Consider again the π's calculation example and suppose that a pseudo random number generator (PRNG) is used to this end. Since the statistical interpretation of the probability is only valid in the limit of large N_{total}, it is clear that, if by some reason the PRNG's algorithm produces numbers that differ from a perfect uniform distribution and, say, tend to slightly accumulate in the center of the square, we will overestimate the value of π when using a large enough amount of PRN. Consider yet another example: assume a PRNG with a long enough period of repetition such that we can in principle approximate π without repeating any sequence, and that outputs numbers with a distribution that, for our purposes, is completely indistinguishable from a true uniform at long enough iterations, but with such correlations between consecutive numbers such that it tends to begin filling the $[0, 1]$ interval from the boundaries to the center. In this case, the number of points laying within the circle will be lower and, even if our PRNG's output was supposed to closely enough resemble a true uniform, the intrinsic time correlations between its numbers will consequently yield an underestimation of π.

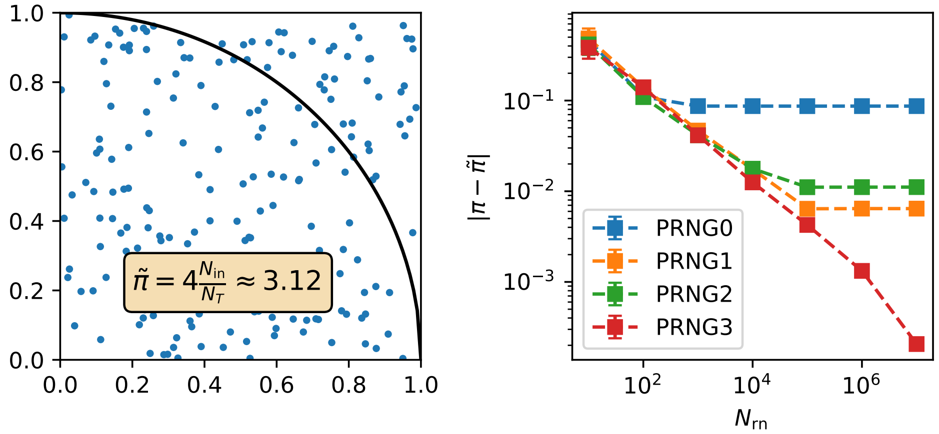

Figure 4.1: **Estimating the value of π with a Monte Carlo method**. Left panel: random numbers are used in numerical calculations such as the calculation of the irrational number π. Right panel: errors in the estimation of π when using RNGs of different qualities, as a function of the amount of random numbers consumed.

So it is clear that the quality of the random numbers is of crucial importance when using them in Monte Carlo (MC) calculations and, while for some applications PRN might be enough, for some others they might not. Modern scientific problems that are tackled with MC strategies often demand vast amounts of random numbers, and therefore one must sometimes completely avoid the problems mentioned above in order to obtain correct and accurate results. For instance, one might need to rely on so called true random number generators (TRNG), which extract real entropy from physical processes such as thermal noise and use it to output numbers. Such generators, though, are normally very slow. Moreover, they often rely on purely classical processes, and thus their apparent true randomness comes from the lack of knowledge of the configuration of a certain system rather than proper and strict randomness. This can eventually have huge implications, specially when using these random numbers in fields such as cryptography. As a consequence, quantum random number generators (QRNG), which randomness can be certified (for example, by means of a Bell test [98]), have been proposed as the ultimate and only possible way to obtain truly unpredictable and uncorrelated randomness. Furthermore, these machines have been implemented in out-of-the-laboratory gadgets in the last decade, and are already starting to be able to generate random bits at fast enough speeds for them to be considered for computing and cryptographic purposes.

A significant challenge for randomness technologies consists of assuring non-predictability of RNG outputs, which is usually done with statistical tests [99, 100, 101]. The National Institute of Standards and Technologies (NIST) has collected and defined a set of tests to measure the quality of PRNGs [102]. These, however, have serious drawbacks, and it has been observed that the simulation of complex physical systems can make some hidden correlations in pseudo random sequences arise and affect their results [32, 33].

Take as an example the dynamic critical exponent z. It is a characteristic constant appearing when studying how do the simulated dynamics of certain systems slow down when approaching the critical point at which a phase transition occurs. This phenomenon, usually referred to as critical slowing down, defines the scaling of the relaxation time with the system size near criticality, and thus effectively measures how does the time needed to extract two independent measures of a given observable of the system diverge near the phase transition. The simulation times at criticality are seen to be huge, and the study of phase transitions via Monte Carlo simulations consequently becomes dramatically difficult: the demand of computational resources and, more explicitly, of random numbers consumed by the algorithm, explode. Being the dynamic critical exponent a feature of phase transitions, it must be computed exactly at the critical temperature and hence suffers from this severe problem. Consequently, it has been seen to be sensitive to the (usually undetectable) long range correlations between the pseudo random numbers used to its computation [33].

4.2 Randomness and Technology

To visualize what randomness is one normally imagines gambling, either playing dice or tossing a coin. For most purposes these are good examples. Strictly speaking, still, these processes are not random at all: their outcome is, in the end, predictable. Given a set of initial conditions, classical mechanics allows us to devise its outcome and thus win the game. They only seem to be random because the lack of knowledge of the initial conditions and the subsequent chaotic behaviour of the equations governing their dynamics prevent this in most real-life situations. Yet randomness is not only present in gambling games, but also exploited, as discussed in previous sections, in accelerating computations. But if the main purpose is to *accelerate* them, it is clear that one has to rely on random numbers that can be generated slightly faster than the speed at which a coin can be tossed. On the practical side, then, *fast* random number generators are needed, and a huge body of technological and scientific effort has been devoted to this end.

4.2.1 Pseudo Random Number Generators

For most practical uses, the main approach to generating randomness consists on lowering one's expectations and conforming with so called pseudo random numbers. Pseudo random numbers are nothing but outputs of a given algorithmic process and, being thus obtained by means of deterministic functions, are completely predictable if the state of the function at a given time is known. Just as tossing a coin, though, they appear to be random if only the stream of output numbers is seen. In its most general form, a pseudo random number generator is a function of the form

$$x_n = f(\vec{x}_{n-1}), \tag{4.1}$$

where the vector $\vec{x}_{n-1}$, potentially depending on many previous outputs of the function (*i.e.* the individual values x_{n-1}, x_{n-1}, etc.), is called the state of the PRNG. Afterwards, a post-processing of the internal state x_n transforms it into the output random number, $r = g(x_n)$.

Ideally, the function $g(f(\vec{x}_{n-1}))$ will appear to output uncorrelated numbers on the short run (where, in reality, short refers to "as long as possible"), while emulating a certain probability distribution on the long one. On a practical side, the function should be able to reach such properties without compromising performance. Consequently, it is mandatory for such functions to be implemented in an efficient manner.

Repetition of sequences and reseeding

Note that, taking into account that we want to be able to code such functions so that they can be used in computer simulations, the state must be comprised of (at least at the bit level) discrete values. Also, any realistic implementation of a probability distribution will inevitably be bound to a certain range of outputs, as those values whose probability of appearance is lower than the machine's precision will be assigned a real probability of exactly zero. Jointly, these two properties that any imaginable PRNG will have, effectively lead to a very undesired conclusion: a PRNG will always have a period of repetition as well. This means that, once a certain amount of pseudo random numbers is output by the function, the stream will reach the same initial state and thus will exactly repeat itself over again.

This repetition of sequences can be nonetheless avoided by *reseeding* the generator. The reseeding operation consists on, at a given frequency (ideally corresponding to a period shorter than the PRNG's period of repetition), externally changing the state with a new random seed. If the seed is truly random, *i.e.* obtained from a true random number generator, then the PRNG will have an infinite period of repetition [33]. Contrarily, if the seed is output by a second PRNG, the first one will then have an enlarged period of repetition, but it will not be infinite and the sequences will repeat themselves over again once the second PRNG reaches its own maximum amount of different output pseudo random numbers.

Linear Congruential Generators

A Linear Congruential Generator (LCG) [103, 104] is a type of algorithm that outputs pseudo random number streams following a recurrence relation such that

$$x_{\mathrm{n}+1} = (a x_{\mathrm{n}} + c) \bmod m, \tag{4.2}$$

where the parameters defining the LCG a, c and m are strictly positive integer constants called multiplier, increment and modulus, respectively (except for the increment, which can be 0 and in which case the LCG is usually referred to as Lehmer RNG). The following bounds must hold

$$0 < m, \tag{4.3}$$

$$0 < a < m, \tag{4.4}$$

$$0 \leq c < m. \tag{4.5}$$

Due to the modulo operation, the output pseudo random numbers may be within the range

$$0 \leq x_{\mathrm{n}} < m, \tag{4.6}$$

and therefore they have a potential period of repetition of at most m. Their easy general formula makes them fast and efficient to implement in computing devices and thus are widely used in computer simulations. Further, as their state is comprised by a single $\bmod m$ number, the memory required for its storage is relatively low; this is helpful specially when multiple independent pseudo random streams are needed. On the other hand, their pseudo randomness quality is seen to be extremely sensitive on their defining parameters, and thus their selection requires much more care. For instance, a bad choice of parameters can yield an LCG with a period of repetition shorter than its modulus.

4.2.2 Quantum Random Number Generators

While numbers output by PRNGs do in principle appear to be random, they are completely predictable and show correlations between them. The former flaw, predictability, results fatal when the sequences are intended to be used in cryptographic environments. The later, on the other hand, can be problematic for sampling probability distributions as in Monte Carlo simulations. Problems in cryptography are obvious since one really relies on the unpredictability of the used keys to avoid a potential attack, while in simulations they can appear to be slightly more subtle.

To easily visualize them, let us recall the previously presented example of the computation of the number π, and consider a real situation in which we use a stream of pseudo random numbers that, while being uniform in the long run, does present correlations in the short one. Say that, for example, the numbers output by such a PRNG are more concentrated towards the beginning of the $[0, 1]$ interval at some stages and towards its end on other ones. In this case, depending on the seed used to initialize the PRNG's state at the beginning of the simulation, our computation will yield a value of π greater or lower than the correct one, as more sampled points will fall within or outside the unit circle.

Based on a huge body of striking experimental evidence, though, quantum mechanics offers a way (actually, the only one) to overcome such problems. An important part of the efforts put into quantum technologies have so far subsequently been directed towards obtaining true randomness by exploiting the inherent unpredictability of quantum processes. Consequently, QRNGs are currently one of the (if not the one) most reliable quantum technologies available in the industry. Further, recent technical progress allows for the construction of faster and more efficient QRNGs, thus allowing, for instance, the first loophole-free observations of the violation of Bell inequalities and the test of non-locality of quantum mechanics [105, 106, 107]. These demonstrations even lead Aspect, Clauser and Zeilinger to be jointly awarded with the Nobel Prize in Physics 2022, "for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science".

Several methods to harness the intrinsic randomness of quantum mechanical systems have been proposed recently. The so called "certified randomness" [108] can be obtained by means of device-independent (DIQIP) randomness protocols, which employ Bell-inequality violation on systems of photons [105, 107], ions [109], nitrogen-vacancy centres [106], neutral atoms [110] and superconducting qubits [111]. Nevertheless, DIQIP approaches are not particularly efficient, as they often rely on measuring properties of few-particle systems [112]. For practical reasons, then, it is useful to measure signals from quantum processes and subsequently extract their randomness with classical devices. These involve devices to observe the timing of nuclear decay, [113] electron shot noise in semiconductors, splitting of photons on beam-splitters, timing of photon arrivals, vacuum fluctuations, laser phase diffusion, amplified spontaneous emission, Raman scattering, atomic spin diffusion, and others. For a review on the topic, see Ref. [112]. Further developing this idea, the fastest quantum random number generators available today make use of macroscopic traces of processes of a quantum origin. For instance, laser phase diffusion (PD) QRNGs [34, 114, 115, 116] base their entire functioning on macroscopic optical signals, Fig.4.2, whose behaviour is understood solely from a quantum-mechanical perspective [117]. Working with well-established technological tools such as lasers allows these QRNGs to achieve much better signal-to-noise ratios than other options as well as faster generation speeds, with the record at the time of writing being

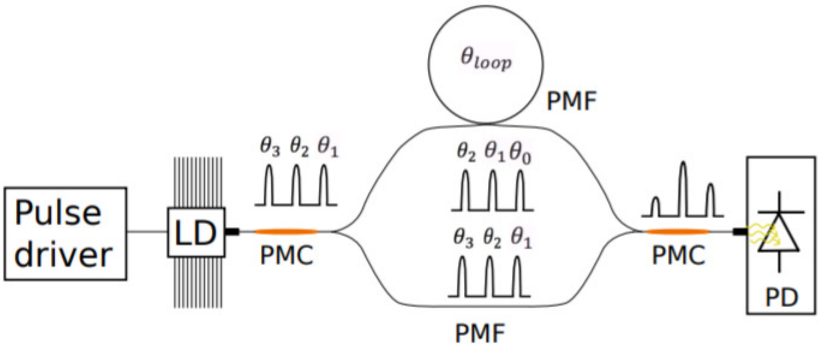

Figure 4.2: **Scheme of Quside's PD-QRNG used in this work.** A laser is modulated above and below threshold, each time generating a pulse with a random phase θ_i. By means of an unbalanced Mach-Zehnder interferometer, each pulse is interfered with a later generated random phase pulse, turning phase fluctuations into intensity fluctuations, which are further converted into random numbers by using conventional photodetectors and electronics. Image taken with permission from [34].

68 Gbps [118].

4.3 The dynamic critical exponent

Understanding the physics behind the dynamic critical exponent z is crucial before attempting to compute it through Monte Carlo simulations and to study how the potential correlations within the random number streams used in such task affect its value. In this section we thus present all the necessary concepts and tools needed to this end.

4.3.1 The characteristic autocorrelation time

In dynamical simulations of physical systems it is usually very insightful to study the time-displaced autocorrelation function of some observable. Concretely in Ising-like spin systems, it is usual to compute this function for the order parameter, *i.e.* the magnetization

$$\chi(t,t') = \langle [m(t) - \langle m \rangle][m(t') - \langle m \rangle] \rangle. \tag{4.7}$$

It essentially tells for how long, on average, the magnetization of our system at a given time t is expected to be correlated with the magnetization at a previous time t'. A value of χ close to one means that most of the spins point to the same direction at both times and thus that the configurations $\{\sigma_i(t)\}$ and $\{\sigma_i(t')\}$ are correlated, while a value of χ close to zero means that are statistically independent. The time-displaced autocorrelation function therefore gives us a measure of the memory effect in the system: it tells for how long a configuration obtained by letting the system evolve (either physically or, in our case of interest, by simulating it with a Markov chain process) keeps any statistically relevant resemblance to the initial state, and thus by studying its characteristic time scales we can gain knowledge of the times needed for our simulations to reach properly thermalized states.

Exponential decay of the time-displaced autocorrelation function

As many physical processes of this kind, it might seem somewhat intuitive to think that the time-displaced autocorrelation function should follow an exponential decay. Let us nevertheless discuss this hypothesis [79, 119].

Let w be the vector of probabilities $w_\mu(t)$, which describe the probability of a given system to be in state μ at time t. Let P be the Markov matrix, with elements $P_{\mu,\nu}$ the transition probabilities from state μ to ν. In matrix notation, the probabilities of laying in a particular configuration μ at two consecutive Monte Carlo steps n and $n+1$ are then related as

$$w(n+1) = Pw(n), \tag{4.8}$$

so that the state at a time step t can be computed by repeatedly applying the Markov transition matrix P. Thus

$$w(t) = P^t w(0), \tag{4.9}$$

where $w(0)$ is the state at which the system is initialized. By writing this initial state in the eigenbasis of the Markov transition matrix, $w(0) = \sum_i a_i v_i$, with $Pv_i = \lambda_i v_i$, we can easily take the exponential of the matrix in the last equation to find

$$w(t) = \sum_i a_i \lambda_i^t v_i. \tag{4.10}$$

In general, the expected value of a given observable of interest can be computed as the sum of the values of the observable at all possible states, weighted by the probability of the system being in such states. In particular, for the magnetization $m(t)$ we find its expected value as

$$m(t) = \sum_\mu w_\mu(t) m_\mu, \tag{4.11}$$

with the sum over all the possible states μ of the system. By defining M as a vector with elements m_μ, in matrix notation we have

$$m(t) = Mw(t), \tag{4.12}$$

and therefore, introducing Eq. (4.10)

$$m(t) = \sum_i M v_i a_i \lambda_i^t. \tag{4.13}$$

For the time-displaced autocorrelation function of the magnetization, Eq. (4.7), we therefore have

$$\chi(t) = \sum_i \sum_j a_i a_j M v_i M v_j \lambda_i^t, \tag{4.14}$$

where the reference time is set to the initial one, $t_0 = 0$, for the sake of simplicity. Note that, consequently, the eigenvalues corresponding to $M(0)$ are all present to the 0th power and so all yield one. We can regroup all constants as

$$b_i = \sum_j a_i a_j M v_i M v_j, \tag{4.15}$$

so the simple equation holds:

$$\chi(t) = \sum_i b_i \lambda_i^t, \tag{4.16}$$

and the autocorrelation of the magnetization is seen to be a linear combination of the eigenvalues of the Markov transition matrix.

Let us now think, for simplicity, of a single spin system that can take values $\sigma_i = \pm 1$. The spin is in contact with a thermal bath but no external magnetic field is present, so that the spin flips randomly at a rate which we denote as $\frac{1}{2\tau}$ (that may depend on the temperature of the bath) but neither state is found with a higher probability than the other. Define the probability of the spin being found in the ± 1 state at time t as $p_\pm(t)$. The master equations for each of this probabilities are:

$$\frac{d}{dt} p_+(t) = -\frac{1}{2\tau} p_+(t) + \frac{1}{2\tau} p_-(t) \tag{4.17}$$

$$\frac{d}{dt} p_-(t) = -\frac{1}{2\tau} p_-(t) + \frac{1}{2\tau} p_+(t), \tag{4.18}$$

which of course are not independent but rather related through the normalization condition at all times, $p_+(t) + p_-(t) = 1 \forall t$. To solve the system of equations we can therefore combine them by defining a more convenient function

$$q(t) = p_+(t) - p_-(t), \tag{4.19}$$

which, recalling that the spin can take only the values $\sigma = \pm 1$, is equal to the expected value of the spin at a given time, $q(t) = \langle \sigma(t) \rangle$. The new function is thus defined by the new master equation

$$\frac{d}{dt} q(t) = -\frac{1}{\tau} q(t) \tag{4.20}$$

which we can easily solve to yield

$$q(t) = q_0 e^{-t/\tau}, \tag{4.21}$$

where q_0 is the value at the initial time, $q_0 = q(0)$.

Since our system has only one spin, the magnetization is directly given by its instantaneous value, $m(t) = \sigma(t)$, and the autocorrelation function is

$$\langle m(t)m(0)\rangle = \sum_{\sigma}\sum_{\sigma'} \sigma p_{\sigma}(t)\sigma' p_{\sigma'}(0), \tag{4.22}$$

with the summations over all possible combinations of $\{\sigma, \sigma'\} = \pm 1$. From the definition of $q(t)$, Eq. (4.19), and its solution, Eq. (4.21), we can write the probabilities of being in a given state as

$$p_{\pm}(t) = \frac{1 \pm q_0 e^{-t/\tau}}{2}. \tag{4.23}$$

Finally, by introducing Eq. (4.23) into Eq. (4.22) we find the autocorrelation function to be

$$\langle m(t)m(0)\rangle = q_0^2 e^{-t/\tau}. \tag{4.24}$$

Comparing the autocorrelation functions found for an arbitrary system, Eq. (4.16), and for a single spin system, Eq.4.24, it makes sense to define the autocorrelation times of each eigenmode as

$$\tau_i = \frac{-1}{\log \lambda_i}, \tag{4.25}$$

and, by introducing them into Eq.4.16, the final autocorrelation reads

$$\chi(t) = \sum_i b_i e^{-t/\tau_i}. \tag{4.26}$$

The system therefore has an autocorrelation time for each of the possible states. At long timescales, nevertheless, only the exponential with the largest value of τ_i will dominate, and so we normally define the autocorrelation time of the system as that corresponding to the smallest eigenvalue, which we denote, for simplicity, by τ:

$$\chi(t) \propto e^{-t/\tau}. \tag{4.27}$$

It is sometimes also referred to as memory time or decorrelation time. It is therefore usual to consider two states of the system to be statistically independent from each other when they are obtained after letting the system evolve for, at least, a time $\Delta t = 2\tau$. To avoid correlations when obtaining dynamical measures from a Monte Carlo simulation it is then important to make sure that they are sampled at a frequency not greater than the imposed by this Δt. Consequently, the number of independent measures n_{ind} that can be drawn from a simulation with a total of N time steps is on the order of

$$n_{\text{ind}} = \frac{N}{2\tau}. \tag{4.28}$$

Measurement of τ in Monte Carlo simulations

While independent measures of any observable of the system can only be obtained between about two full correlation times, in order to compute τ it is obviously mandatory to measure the magnetization at a higher frequency, since it is a dynamical property of shorter timescales. If the magnetization was measured only with a temporal spacing of $\Delta = 2\tau$ or higher, we would see no correlation between them and thus we would only be able to state that the correlation time is lower than the maximum precision achievable by the measurements. A good estimation of the correlation time often requires to measure the magnetization at every time step.

In order to compute $\chi(t)$, we first expand Eq. (4.7) to

$$\chi(t) = \int dt' \left[m(t')m(t'+t) - \langle m \rangle^2 \right]. \tag{4.29}$$

Its discretized version reads

$$\chi_n = \frac{1}{N-n} \sum_{n'=0}^{N-n} m_{n'} m_{n+n'} - \frac{1}{(N-n)^2} \sum_{n'=0}^{N-n} m_n \sum_{n'=0}^{N-n} m_{n+n'}, \tag{4.30}$$

where, in order to obtain a good estimation of the autocorrelation time, the N values of the magnetization are obtained from a simulation spanning a time frame much bigger than the expected value for τ, $T = N\Delta t \gg \tau$. The need for this can be intuitively seen if we pretend to obtain τ by later fitting $\chi(t)$ with an exponential decay: longer simulations will yield smaller fluctuations on the beginning of the temporal range, over which we perform the fit. It can also be analytically seen to be mandatory by simply noting that the multiplicative factors for the sums in Eq. (4.30) diverge if $n \to N$, which will yield huge errors even if small (though intrinsically present) fluctuations appear in the magnetization.

Note that the computational cost of Eq.4.30 goes as n^2. In our specific case, we demand our results to be very accurate and statistically relevant, so both large values of N and a lot of repetitions are required. Therefore, the computation of the autocorrelation function can be quite challenging. In order to avoid this scaling, it is useful to consider the Fourier transform of $\chi(t)$. In the frequency domain, the convolution of $\mu(t) = m(t) - \langle m \rangle$ with itself in Eq. (4.29) is easily computed as the multiplication of the transformed functions $\tilde{\mu}(\omega) = \mathscr{F}\{\mu(t)\}$, and therefore we can obtain the autocorrelation by inverse-transforming its product

$$\chi(t) = \mathscr{F}^{-1}\{\|\tilde{\mu}(\omega)\|^2\}, \tag{4.31}$$

which can be done with a scaling of $O(n \log n)$ by means of a fast Fourier transform

(FFT) algorithm [120]. We can finally obtain the characteristic correlation time of the system by linearly fitting $\log \chi(t)$.

4.3.2 Dependence on the temperature: autocorrelation and scale invariance

The correlation time can be derived, as has been discussed in previous sections, solely from the transition probabilities between the different states of the system. It therefore constitutes an intrinsic physical property of the system. Clearly, nevertheless, the transition probabilities may be different at different temperatures. For instance, one may expect them all to be low, in general, when the system is cooled down to a very low temperature so that the spins basically appear to be in a frozen state. On the other hand, the transition probabilities may be close to one when the temperature is high enough such that the thermal fluctuations dominate over the Hamiltonian's couplings, describing an erratic Markov Chain that produces completely random configurations when $T \to \infty$. The system's dynamics thus depend on its temperature, and so is seen to do the correlation time.

Another parameter quite related to the correlation time τ is the correlation length ξ. It measures the length scale at which spins are seen to be, on average, correlated between each other. In analogy to the discussion of the previous section, the two-point connected spatial correlation function can be seen to behave exponentially as well [121]

$$G_c = \langle s_i s_j \rangle - \langle s^2 \rangle \propto e^{-\Delta_{ij}/\xi} \tag{4.32}$$

with Δ_{ij} the euclidean distance between spins located at sites i and j. Note that G_c effectively measures the fluctuations of the magnetization with respect to its mean value or, in plain words, how far away an island of spins aligned to a given direction is expected to extend for, within a sea of spins aligned to the opposite direction.

Let us now first grasp some intuition on how does the correlation length behave with temperature. At low temperatures most of the spins will be aligned, except for some isolated ones or some small islands of them. On the other hand, at high temperatures the thermal fluctuations make spins behave in an isolated manner as well, with interactions going no further than, again, some small islands. In both limits the characteristic size of the islands, described by the correlation length, is therefore small. Between these two extreme cases, namely at the system's critical temperature T_c, the system undergoes a phase transition and hence understanding the physical intuition requires a bit more work.

Zooming out

If we did a simulation at a temperature near the critical one, we would see that the size of the correlated islands increases. We would actually be able to distinguish islands of all possible sizes, Fig.4.4.

Let's think about what happens if we change the scale of observation of these three scenarios with a pixel analogy [122]. When we zoom out, the scale of observation gets bigger and the resolution of the details is consequently worse. Effectively, if we have a box of 2 pixels per side and we increase our length scale by a factor of two, the old 4-pixel box will become our new unit of pixel. If the pixels can only take binary values, we will then assign to the new bigger one the value of the majority of the old, small ones. In case that no value predominates over the other within the old box, a random value drawn from an uniform distribution is assigned to the new big pixel, thus ensuring no bias towards any concrete value is introduced. Statistically, then, this *decimation* operation (in the language of renormalization theory) of the pixels that our photographic cameras and our very own eyes perform on a daily basis when changing the scale of observation of a given picture, effectively removes the inhomogeneities. Say, for example, that we are looking at an archipelago from a plane. As we increase our altitude, our resolution of the small details (the islands) among the homogeneity (the sea) will decrease until the point at which we are no longer able to see them. Coming back to the spins in our system, in the two extreme cases (both above and below the critical temperature), zooming out will have the effect of decreasing the size (and thus the number, by eliminating the smallest ones) of islands of correlated spins. Since, as discussed earlier, we expect the system to have smaller and smaller correlation lengths as we increase or decrease the temperature further away from the critical temperature, the system at the new amplified scale is physically equivalent to a system at a higher or lower temperature, Fig. 4.3. Effectively, *the change in scale is modifying the physics of the system.*

On the other hand we have experimentally seen that (see Fig.4.4), at the immediate vicinity of the critical temperature, the islands of aligned spins can be of *all possible sizes*: this is the main point. By zooming out we will then loose the resolution of the smaller islands but, due to its fractal geometry, we will still keep seeing islands of all possible sizes. In contraposition to the two extreme cases, then, *the physics at around T_c are not altered by the change of scale operation.* This streaking feature is known as *scale invariance.*

Recall that, as discussed above, noticing the change in scale is physically equivalent to moving away from the critical temperature. Intuitively the change in temperature felt by the system should be greater when the scale invariance assumption starts to fail, *i. e.* when we have zoomed out to a scale at which the system can no longer be considered infinite and thus increasing the scale further effectively starts eliminating islands of aligned spins. It then follows that the bigger the system, the more different scales the islands can have and thus the more times we will be able to zoom out without noticing it. Thus for small systems, which should suffer this kind of finite size effects earlier (they can be zoomed out a smaller number of times before the change in scale manifests itself as a change in temperature), the change in

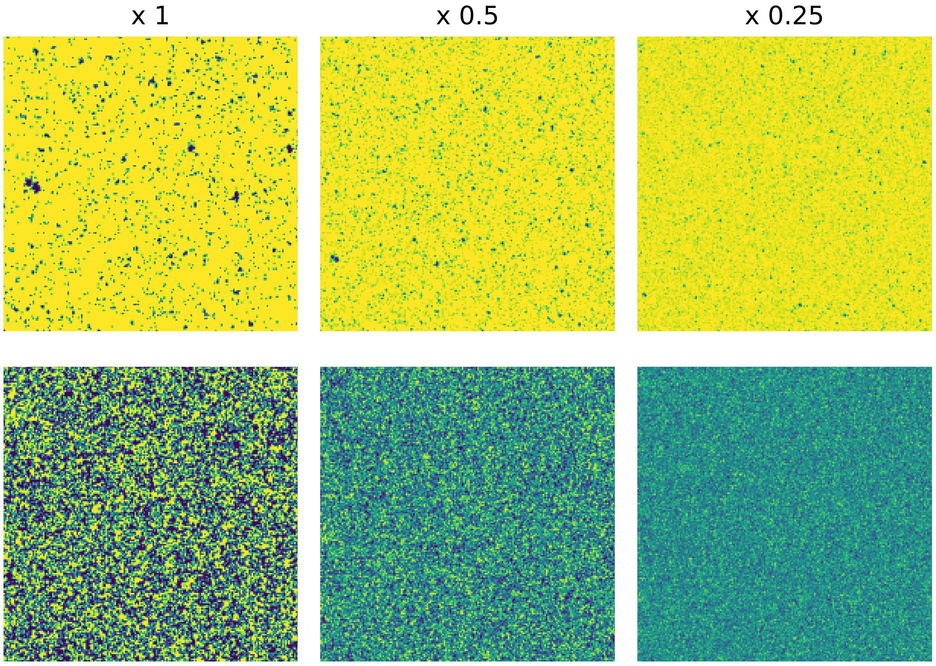

Figure 4.3: **Effects of the change in scale at temperatures away from the critical point.** Since the correlation length ξ is finite at temperatures away from the critical point, a change in scale is seen to effectively reduce the size of the clusters of correlated spins. This is equivalent to consider a system further away from criticality. Top (bottom) panels represent a system at a temperature below (above) the critical one. From left to right, each panel represents a zooming out by a factor of ×0.5 from the previous one, thus physically equivalent to even lower (higher) temperatures. Each colour represents spins pointing to one direction.

temperature should be more dramatic, and therefore the order parameter will have a smoother transition around the critical point. Only in the thermodynamic limit will then the system present a completely sharp transition at T_c, as illustrated in Fig.2.3. Scale invariance is a very characteristic and defining property of phase transitions.

Conclusions: divergence of τ and ξ

By the previous discussion, if at the critical temperature we have autocorrelation lengths happening at all possible scales, we conclude that it must diverge.

$$\lim_{T \to T_c} \xi = \infty. \tag{4.33}$$

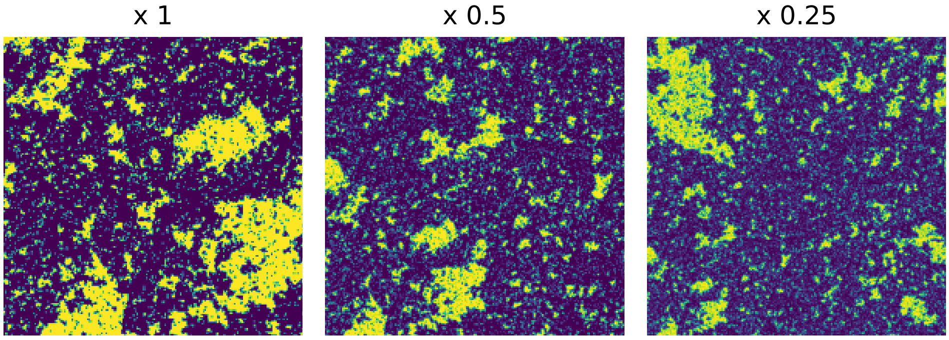

Figure 4.4: **Scale invariance at the critical temperature.** A configuration thermalized at the critical temperature manifests its scale invariance by presenting clusters of aligned spins of all possible sizes, and thus ξ diverges. By zooming out and decimating the image, the system exhibits exactly the same behavior, and therefore the physics are unaltered. From left to right, each panel represents a zooming out by a factor of ×0.5 from the previous one. Each colour represents spins pointing to one direction.

At temperatures around the critical one, the domains of high magnetization in one direction are gradually destroyed by the effect of thermodynamic fluctuations, and converted into zones of high magnetization in the opposite direction: islands of all possible sizes are being created and destroyed at all times. This behaviour is illustrated in fig.4.5. Analogously, so does diverge the correlation time τ.

4.3.3 Universality and critical exponents

Now that the intuition behind the scale invariance of phase transitions has been introduced, and that the divergence of the characteristic time and length has been deduced, we are ready to discuss exactly how do the characteristic time and length diverge when approaching the critical point.

Let us define a scale invariant function as that for which a change in the scale of the argument $x \mapsto \lambda x$ only affects its output in a change in scale (that may depend on λ) too

$$f(\lambda x) = \mu(\lambda) f(x), \tag{4.34}$$

such that, by a simple linear rescaling of the axes of the obtained function, we can recover the initial one. Effectively, this means that the physics of the system described by $f(x)$ remain unaltered under a change in scale. The objective is then to obtain a formal definition of a scale-invariant function by solving Eq. (4.34) for $f(x)$.

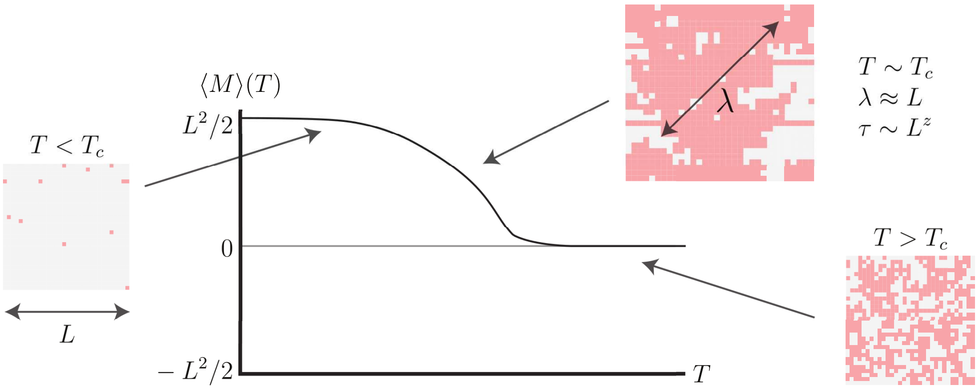

Figure 4.5: **2d Ising model dynamics.** For temperatures below the critical temperature, T_c, the system shows ferromagnetic behavior; for temperatures larger than T_c, the system shows ferromagnetic behavior. At T_c, the system shows a phase transition and the correlation length and time diverge.

By differentiating both sides of Eq. (4.34) we get

$$\kappa f(x) = \frac{df(\lambda x)}{dx} x, \tag{4.35}$$

with $\kappa = \frac{d\mu}{d\lambda}$ a constant. Rearranging terms we find

$$\kappa \frac{dx}{x} = \frac{df(\lambda x)}{f(x)} \tag{4.36}$$

which, by substituting $f(\lambda x)$ as per the definition of scale invariance, Eq. (4.34), can be easily integrated to yield

$$f(x) = ax^b, \tag{4.37}$$

with $a = e^{C/\mu}$, $b = \frac{\kappa}{\mu}$ and C the integration constant. Under the definition given by Eq. (4.34), then, all scale invariant functions must be of the form given by Eq. (4.37), *i.e.* power laws. Both the correlation length and time will therefore diverge as power laws when approaching the critical temperature.

Let us define the reduced temperature

$$t = \frac{T - T_c}{T_c}, \tag{4.38}$$

which effectively measures the distance in temperature from the critical point. Given the scale invariance discussed above, then, the correlation length near the critical temperature must diverge as

$$\xi \propto |t|^{-\nu}, \tag{4.39}$$

where we have used the absolute value of the reduced temperature in order for the expression to be valid both at below and above T_c. Note, however, that there is no restriction in the proportionality constant, which can in general be different at both sides [79]. Analogously, for the correlation time we have

$$\tau \propto |t|^{-\nu z}. \tag{4.40}$$

We have conveniently defined the exponent as a multiplication between ν (the very same exponent appearing in Eq. (4.39)) and a second one, z. This does not pose any restriction on the whole exponent defining the divergence of τ but, as we shall see, will prove helpful.

Many other properties (specific heat, for example) diverge with power laws near the phase transition, each with its corresponding exponent. These exponents are usually referred to as *critical exponents.*

Critical slowing down: the dynamic critical exponent

By combining Eq. (4.39) and Eq. (4.40) we can write a simple relation between the two coefficients:

$$\tau \propto \xi^z, \tag{4.41}$$

where we are left with a single exponent z thanks to the conveniently defined critical exponent of the correlation time. Eq. (4.41) gives us a measure of how rapidly does the correlation time grow as the correlation length diverges near the critical point. This property is usually referred to as *critical slowing down*, and is characterized by the exponent z, which is known as the *dynamic critical exponent.* A higher value of z will indicate that the slowing down of the dynamics of the system when approaching the critical temperature will be more dramatic, and thus that, in order to gain statistically relevant results and in virtue of Eq. (4.28), the simulations at that point will need to run for much longer.

In computer simulations we can nevertheless not work with infinite lattices, and thus this divergence is translated into the largest wavelength approaching the linear size of the system, $\lambda_{max} \approx L$. In a cubic lattice of side length L, for instance, we find

$$\lim_{T \to T_c} \xi = L. \tag{4.42}$$

Note that this happens even when using periodic boundary conditions. The critical slowing down effect is thus described by

$$\tau_{\text{FSS}} \propto L^z, \tag{4.43}$$

where the subscript FFS stands for finite size scaling.

Universality

Note that, so far, we have not restricted the discussion to any specific model. Actually, one of the most striking properties of phase transitions is the fact that they appear to behave in an extremely similar way for systems that have dramatically different dynamics at finite scales. Due to the scale invariance introduced above, their behaviour near the critical point becomes increasingly similar among them as the limit scale is approached. Models that share a scale invariant limit and thus have the same asymptotic behaviour are then said to form a universality class. This asymptotic behaviour is manifested through the aforementioned critical exponents. All models within the same universality class share the same set of critical exponents and they are, for example, exactly the same for two dimensional spin systems of square and triangular lattices, among others. These do not depend either on the strengths of the couplings J. On the contrary, they can depend on the dimensionality of the system.

Universality, as well as the exact power-law scaling functions discussed in previous sections, are properties that can be thoroughly derived by means of renormalization theory. The exact values for the critical exponents can also be tackled theoretically by different techniques (that may yield slightly different values and among which there is not a current consensus within the literature, see below). This, nevertheless, already lays outside the scope of the present thesis. The interested reader is referred to [121, 122].

4.3.4 Values of the dynamic critical exponent in the literature

Over the years, many attempts to give an accurate value of the dynamic critical exponent z have been carried on with theoretical [123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133], experimental [134, 135, 136, 137, 138], and computational [139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162] approaches. With the aim of illustrating how vastly the obtained results vary, we present a long, yet non-exhaustive collection of values found in the literature, for both two- and three- dimensional lattices. Some of the references presented here give various values corresponding to different types of lattices, in an attempt to show the postulated universality of z across models pertaining to the same universality class. We therefore restrict to the so called A class, usually referred to as time-dependent Ginzburg-Landau models [163, 164]. It is the class to which, among many others, the two-dimensional, nearest-neighbour Ising model pertains. We plot in Fig. 4.6 the obtained exponents as a function of the year of its calculation. Surprisingly, even with the expected improvement of the used methods, there is no clear tendency, neither in two-, three-dimension, theoretical or MC calculations, and even the various results obtained in recent years show a wide spreading. The data is gathered in Tables B.1 and B.2 from Appendix B. For completeness, the value

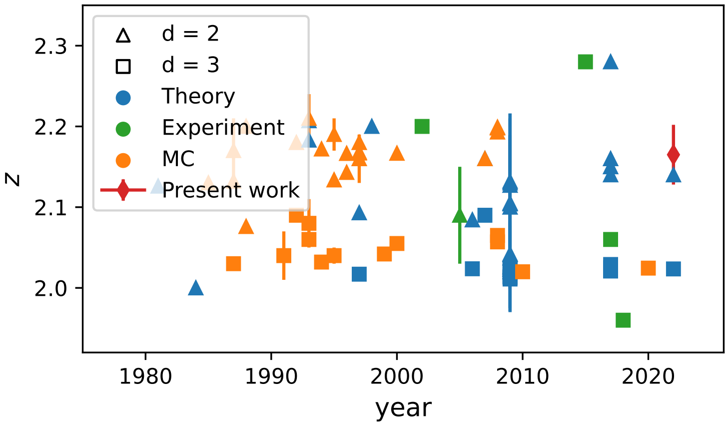

Figure 4.6: **Values of the dynamic critical exponent z in the literature.** Values of z obtained by theoretical (blue), experimental (green) and Monte Carlo (orange) approaches. Triangles and squares label two- and three-dimensional lattices, respectively. The red diamond labels the value obtained in the following sections of the present thesis.

obtained in the following sections of the present work is shown as well.

Let us now take a closer look at the two most numerous families of results, namely those obtained via theoretical studies and via Monte Carlo simulations. The later is also the one we are most interested in as is the one we follow in the present thesis. We plot in Fig. 4.7 the histograms grouping the theoretical predictions and the MC ones. While the values obtained by means of theoretical methods do not clearly show any apparent distribution, the ones obtained with MC calculations can be fitted to a Gaussian distribution with mean $\langle z_{\mathrm{MC}} \rangle = 2.1664$, even if a clear convergence along the years is not observed either.

Given the huge variability of the results found across the literature (specially among theoretical predictions), it is clear that there is not a widely accepted value for z, and we therefore do not have a *correct* value to compare our results against. In the following sections we assess this calculations with Monte Carlo simulations.

4.4 Results

For the calculation of the critical coefficient, we simulate the two dimensional spin model with nearest-neighbour interactions described by Hamiltonian Eq. (2.3) for a time of 1300τ and calculate its magnetization after each MC step. After the discussion of previous sections, we approximate τ by $L^{z_{\mathrm{approx}}}$ with $z_{\mathrm{approx}} = 2$. Note that ξ, the expected theoretical value of the correlation length, equals L, since $\tau \propto \xi^z$ and $\tau_{\mathrm{FSS}}(L) \propto L^z$. From these results, the different values of the correlation time

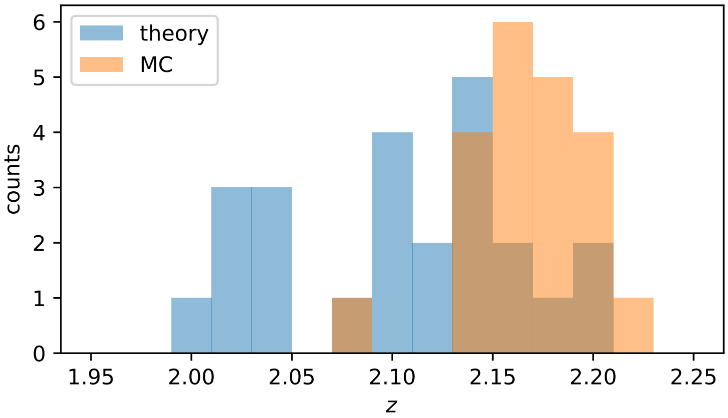

Figure 4.7: **Histograms of the dynamic critical exponent z in the literature.** While the values obtained with theoretical approaches (blue) do not clearly show any apparent distribution, the results of Monte Carlo simulations (orange) do resemble a Gaussian distribution.

$\tau_{\text{FSS}}(L)$ have been obtained by adjusting the correlation to a decreasing exponential. Recall that, according to Eq. (4.26), the time-displaced autocorrelation function can present high non-linearities due to the many correlation times still active at short timescales. Thus in order to avoid, on the one hand, detecting these initial undesired correlation times τ_i, and on the other hand the fluctuations in the tail of the exponential, we adjust the obtained results within a time interval $t \in (0.3\tau_{\text{FSS}}, 1.1\tau_{\text{FSS}})$, following Ref. [33]. We stress that such a choice is only valid as a result of performing very long simulations, spanning a huge amount of correlation times.

Finally, in order to measure the dynamic critical exponent describing the critical slowing down of the algorithm, we compute the autocorrelation time for various lattice sizes and then fit the obtained values, according to Eq.(4.43), with a power law. With the aim of approximating the real value of z, it is desirable for τ_{FSS} to be a good estimation of τ, and thus to simulate as big systems as possible.

4.4.1 Consumption of random numbers

Before attempting the actual calculations, let us first consider the amount of random numbers consumed by such process.

Let α be the fraction of spins that, at a given inverse temperature β, yield a positive change in energy when proposed as a candidate to be flipped, $\Delta E > 0$. Following the Metropolis algorithm, they subsequently request a random number in order to determine whether to be flipped or not, with a probability $p = e^{-\beta \Delta E}$. Consider a thermalization sweep consisting on proposing and accepting/rejecting L^2 spin flips, with L the lattice size. If the spins to be updated are selected at random, the total amount of random numbers consumed in such process is

$$N_{\mathrm{rn}}^{(\mathrm{seq})} = L^2(1+\alpha). \tag{4.44}$$

In a parallel implementation in which the checkerboard scheme is used, all spins are proposed as candidates to be flipped and thus no random numbers are used to select them. On the other hand, in order to optimize the performance and if there is a big enough availability of computing resources, this implementation requests a random number per spin prior to knowing whether it will be used or not in the Metropolis algorithm. Those numbers corresponding to spins whose update proposal yields $\Delta E < 0$ (and thus are directly flipped, without the need of any random number) are discarded. The randomness consumption of such a parallel-scheme sweep is then

$$N_{\mathrm{rn}}^{\mathrm{par}} = L^2, \tag{4.45}$$

which equals the lowest possible sequential value. By avoiding the discard of such $L^2\,(1-\alpha)$ unused random numbers an optimal consumption of

$$N_{\mathrm{rn}}^{\mathrm{opt}} = \alpha L^2 \tag{4.46}$$

would be achieved.

As discussed in previous sections, the simulations are run for a total tome of $1300\tau_{\mathrm{FSS}}$ and, in order to gain statistically relevant results, each computation is repeated N_{it} times. In Table 4.1 an estimation of the total random bits consumed for each lattice size is given.

L	MC updates/ simulation	N_{it}	Random bits/ update	Total bits required (GB)
16	533192	100	512	3.2
24	1285292	100	1152	17.2
32	2399489	100	2048	57.2
48	5784114	100	4608	310.3
64	10798263	100	8192	1029.8
96	26029866	100	18432	5585.4
128	48594709	100	32768	18537.4
256	218687559	100	131072	333690.7
512	984145175	10	524288	600674.5
			TOTAL	959905.8

Table 4.1: Estimated randomness consumption.

Given the huge amounts of random numbers consumed to its computation, the value of the critical dynamic exponent obtained via Monte Carlo simulations is known to be sensitive to the long-term correlations present within the used pseudo-random streams.

	Modulus	Multiplier	Increment
PRNG0	$2^{32}-1$	16807	0
PRNG1	$2^{25}-39$	12836191	0
PRNG2	$2^{23}-15$	422527	0
PRNG3	$2^{17}-1$	43165	0

Table 4.2: Parameters used in the Linear Congruential Generators used.

4.4.2 Detecting correlations in PRNG

As stated above, the dynamic exponent z is known to be sensitive to the correlations of the random numbers employed in its computation. Therefore, in order to study the effects that such correlations have in our results, we make use of different PRNGs to simulate our system.

Fundamental tests with PRNGs

Before introducing the calculation with the QRNGs, we address the computation of the dynamic exponent with four different Linear Congruential Generators (LCGs), all of them presumably showing low correlations [103]. The modulus, parameter and increment used for each LCG are resumed in Table 4.2.

With each PRNG and for each lattice size (ranging from $L = 4$ to $L = 512$), we compute the dynamic exponent z with a Monte Carlo simulation. For each point in the simulation, we run a total number of $N_{it} \geq 500$ iterations in order to extract statistically relevant results (the number of iterations is restricted to $N_{it} \sim 10$ for $L = 512$ lattices, due to long computation times). We fit the obtained curves with exponential laws in order to obtain τ for different lattice sizes, and then plot them in a logarithmic scale as a function of the lattice size to obtain the dynamic exponent z. We then compare them with the theoretical estimate of z obtained by studying the stochastic matrices governing the physics of our system in the classical Ref. [157], that is, $z = 2.1667 \pm 0.0005$, since it is also the most similar one to the mean value found with MC approaches in the literature, as seen in previous sections. We summarize our results in Table 4.3, in which we show the obtained dynamic exponents z, as well as their respective errors relative to the theoretical estimate, ϵ_r (we take as reference $z_{\text{ref}} = 2.1667$ and approximate ϵ_r in the fourth decimal).

	z	ϵ_r
PRNG0	2.1087	0.0268
PRNG1	2.1159	0.0234
PRNG2	2.1047	0.0286
PRNG3	2.1162	0.0233

Table 4.3: Dynamic exponents z and relative errors with respect to theoretical estimate ϵ_r, obtained for each PRNG.

All tested PRNG's yield a reasonable approximation of the dynamic exponent, as can be seen in Table 4.3, but they differ in their predicted value in the second significant decimal.

Importantly, the statistical variance of the results poses a considerable caveat that must be taken into account when interpreting them. For any given PRNG, once the amount of random numbers consumed by the algorithm exceeds its period of repetition (modulus in Table 4.2), the sequence repeats itself, introducing a large amount of correlation at large time scales. We observe that these extra correlations affect the obtained dynamic exponent by exponentially enlarging the variance between iterations of the same simulation while keeping its mean value constant. To quantify these correlations, we define the normalized variance as σ^2/μ^2, where σ is the variance of the results and μ their mean value. For this case, the normalized variance stays more or less constant (around a value of $\frac{\sigma^2}{\mu^2} \approx 0.01$), until the lattice reaches a size large enough so that its simulation requires more random numbers than the generator's period of repetition. Beyond this point, the variance starts to increase exponentially with lattice size, as shown in Fig. 4.8. Furthermore, as we keep increasing L far beyond the variance's explosion point, the obtained magnetization autocorrelation function $\chi(t)$ does no longer resemble an exponential decay, but instead starts showing a noisy behavior. Therefore one cannot strictly speak of, nor extract, a reliable value for the dynamic exponent.

Effects of reseeding

The repetition of sequences in the pseudo random streams has dramatic effects in the obtained results. As discussed previously, nevertheless, the reseeding of a PRNG yields theoretically infinite periods of repetition, so this behaviour could in principle be avoided. We now explore this option by reseeding PRNG3 of the previous section. Concretely, in our reseeding we substitute the entire state of the LCG by a true random number obtained from the computer's physical entropy. We reseed it every $\kappa(m-1)$ pseudo-random numbers, where m is the modulus parameter of the LCG, and we first consider a low frequency reseeding with $\kappa = \{1, 2, 4, 8\}$. Note that, in this situation, we still have repeating sequences for all values of κ except for the first one, $\kappa = 1$, for which we barely avoid it by exploiting the full period of repetition of the LCG. As shown in Fig. 4.9, for all cases the variance explosion observed before is avoided. In this case, the normalized variance does not grow monotonically, but instead reaches a plateau whose value appears to be proportional to the number of repetitions introduced in the pseudo-random sequences before the reseeding is carried out (i.e., the amount of extra correlations introduced in our algorithm, compared to the infinite-period PRNG).

After these conclusions and in light of the fact that the case $\kappa = 1$, (*i.e.* an example of an infinite-period pseudo-random number generator), exhibits the lowest

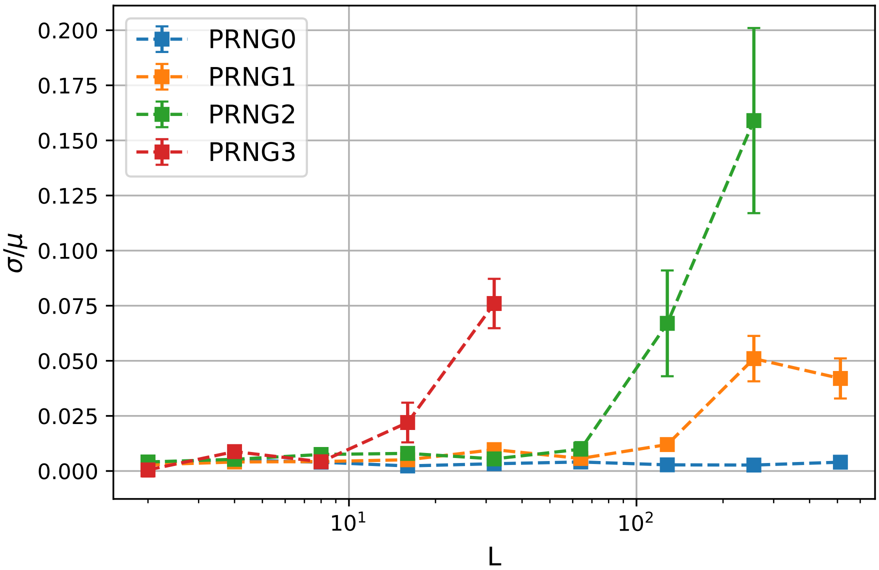

Figure 4.8: **Reduced variance of z.** The statistical variance of the dynamic exponent z normalized to its mean value is severely degraded when sequences of PRNG are repeated in its computations. Results obtained from $N_{\mathrm{it}} \geq 100$ iterations for all points except for $L = 512$, for which $N_{\mathrm{it}} = 30$, as a function of lattice size, for different PRNG's.

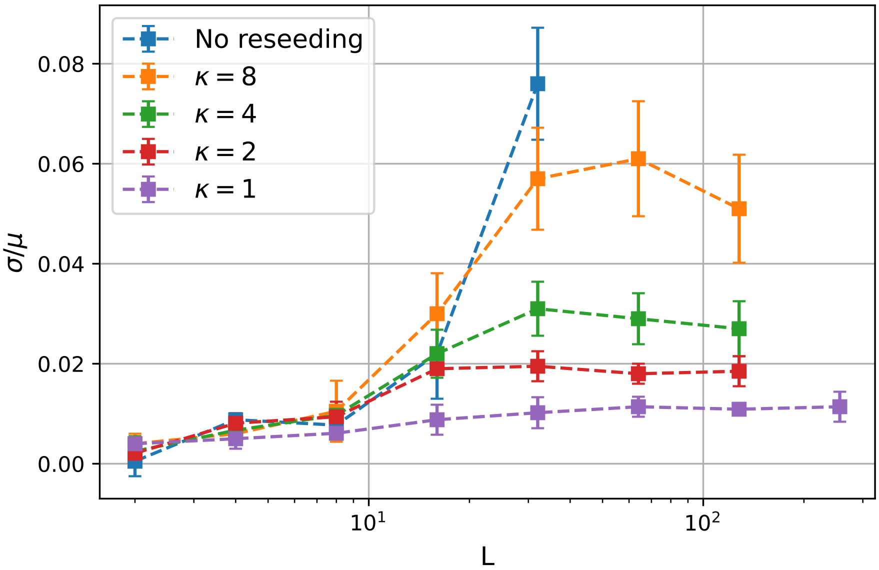

Figure 4.9: **Normalized variance with low-frequency reseeding of PRNG3 as a function of lattice size.** Low-frequency physical reseeding of the generator every $(m-1)$, $2\,(m-1)$, $4\,(m-1)$, $8\,(m-1)$ random numbers is able to avoid the variance explosion. PRNG3 without reseeding is shown for comparison.

normalized variance of all presented cases, it then feels somewhat natural to ask oneself the question of how would a high-frequency reseeding with $\kappa < 1$ affect the obtained results. This scenario minimizes the correlations appearing in the pseudo-random sequences, approaching a true RNG as κ gets smaller. By doing so (and thus paying the high computational cost of such a reseeding), we observe that the variance can still be pushed further down, see Fig. 4.10. This fact allows us to conclude that the use of PRNG's, even those showing low correlations and having a theoretically infinite period, can indeed affect the quality of our results in terms of variance for the problem at hand. Moreover, as we show here, once κ is fixed, the variance of the obtained z converges to a plateau and barely changes with L. Hence, performing longer simulations will not improve the approximation. We tested the PRNGs using the NIST statistical test suite and observed that, while those having repeated sequences do not pass it, we are able to decrease the number of failed tests within the suite until eventually making it pass all of them for sufficiently high reseeding frequencies. PRNG0 already passed the tests without reseeding, while PRNG1 and PRNG2 only needed reseeding with $\kappa = 1$. On the other hand, PRNG3 needed $\kappa = \frac{1}{16}$ as it was usually failing, out of the more than 200 different tests within the suite, only one of them after being reseeded with $\kappa = 1$ (the FFT test). Thus, given that all of them pass the NIST tests and we can still see improvement

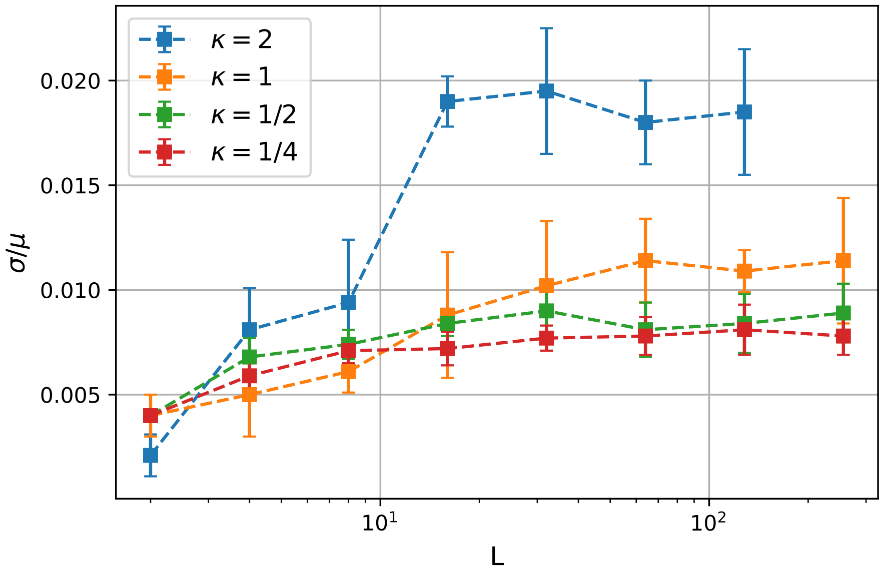

Figure 4.10: **Normalized variance with high-frequency reseeding of PRNG3 as a function of lattice size.** High-frequency physical reseeding of the generator every $2\,(m-1)$, $(m-1)$, $\frac{(m-1)}{2}$ and $\frac{(m-1)}{4}$ random numbers improves the variance obtained by lower frequency reseedings. Beyond $\kappa = \frac{(m-1)}{4}$ the error bars do not allow us to distinguish the results for higher reseeding frequencies.

when further reducing κ, we can conclude that this protocol is able to distinguish between two pseudo-random sequences containing different amounts of randomness amplification, this is, that have used more or less extensively the PRNG's algorithm (so that their next outcome can be predicted with a higher probability), and to therefore identify if one is closer to a true RNG than the other one.

Interestingly, the values that the obtained reduced variances reach at the plateaus show a linear dependence on the reseeding period κ, Fig. 4.11. By means of a linear fitting, this behaviour can be extrapolated to obtain the reduced variance that a TRNG would yield, this is, for the case of $\kappa = 0$. The obtained value is $\frac{\sigma^2}{\mu^2}(\kappa = 0) \approx 0.004$, which coincides with the initial points of the curves of Figs. 4.9 and 4.10, corresponding to simulations of very small latices for which correlations between pseudo random numbers are still non-detectable.

Besides the effect on the variance of the estimation of the dynamic exponent, the reseeding of the PRNG shows an effect on its mean value as well, yielding a more accurate result as higher reseeding frequencies are considered, see Table 4.4. We note that, for every case with reseeding, and for every frequency κ implemented, the relative error to the theoretical value is reduced to about half the one obtained with the different PRNG's used in Table 4.3, which were not reseeded. Nevertheless, we

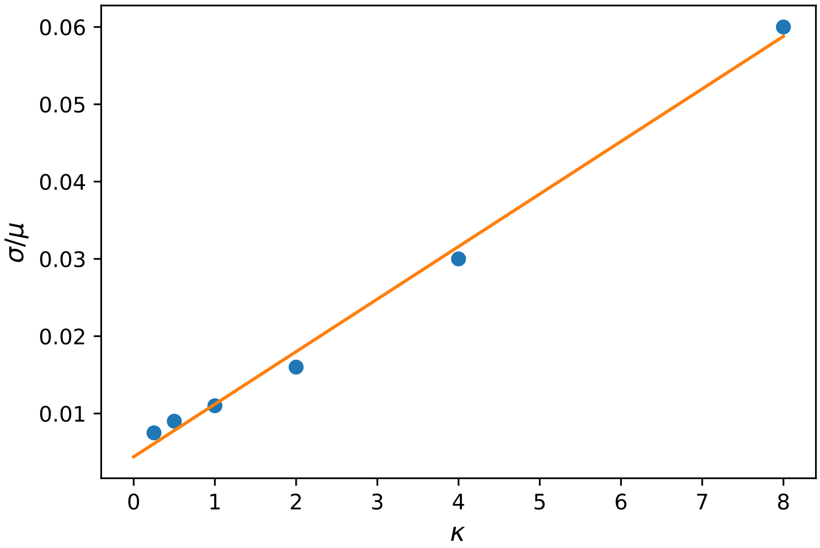

Figure 4.11: **Fit of the reduced variance with κ.** Reduced variance for different values of the reseeding period κ of PRNG3 (dots) and linear fit (line).

cannot reduce this error further. We emphasize that reducing κ reduces the variance, bringing the result closer to those of a TRNG, but at a larger computational cost for smaller and smaller κ. The results here point out that in the limit of very small κ, one should obtain results close to those obtained with a TRNG. Nonetheless, there is no practical way of testing this due to the large computational cost.

	z	ϵ_r
$\kappa = 2$	2.1815	0.0068
$\kappa = 1$	2.1477	0.0088
$\kappa = {}^1/_2$	2.1482	0.0085
$\kappa = {}^1/_4$	2.1441	0.0104

Table 4.4: Dynamic exponents z and relative errors with respect to theoretical estimate ϵ_r, for different reseeding frequencies of PRNG3.

4.4.3 Computing z with a QRNG

Finally, we address the computation of the dynamic critical exponent with Monte Carlo simulations employing streams of completely uncorrelated numbers, as obtained from QRNGs.

As in previous sections, we performed simulations for different lattice sizes. In Fig. 4.12 we show the time-delayed correlation of the order parameter (magnetization), $\chi(t)$, as a function of time. We see the expected exponential decay described by Eq. (4.27). By fitting the previous equation, we extract the average value of the exponent τ for each L. We performed 100 simulations for L up to 256 and 15 simulations for $L = 512$ due to the large computational cost in this last case (see discussion below). It is very apparent from last panel in Fig. 4.12 that the results for $L = 512$ are more noisy that those obtained for smaller lattice sizes, probably due to the smaller number of simulations.

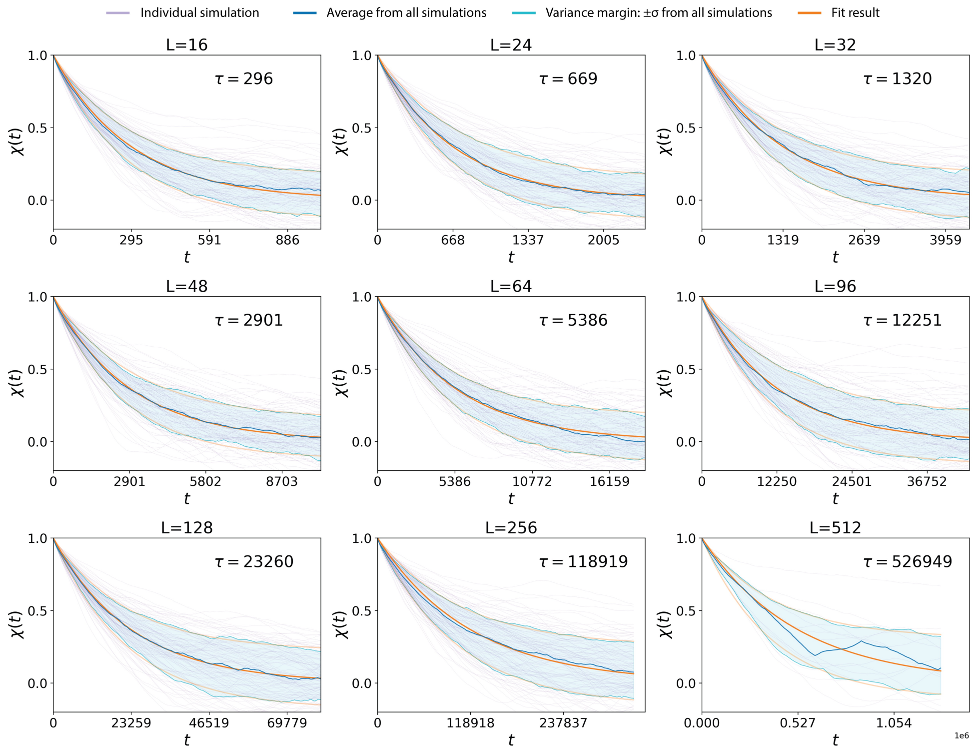

Figure 4.12: **Determination of the relaxation time as a function of lattice size.** Inset τ values are the decay times associated to the orange curves, which correspond to the averages obtained by the fitting of the multiple repetitions simulated at each side length. The individual simulations (purple), as well as the variance interval for all of them (cyan) are also shown in the graphs.

In Fig. 4.13 we represent the average correlation time $\tau(L)$ obtained from these results. Performing a linear fitting of the obtained curve, we find a value of the critical exponent $z = 2.165 \pm 0.039$. We emphasize that, from here, there is a clear strategy to improve this value: performing multiple simulations for bigger lattices would add more points to the fit, resulting in a better approximation of this coefficient. However, unlike the case of pseudo-generators, obtaining correlation times for larger cell sizes is limited by the vast need for random numbers required by the simulation. Note that, for each step of the simulation, $N \propto L^2$ random numbers are required. As the simulations are run for $1300\tau_k = 1300L^z$ steps, we require on the order of $1300L^{z+2}$ random numbers. Assuming that each of these numbers has 32-bit precision, we face a massive consumption of about $41600 \cdot L^{z+2}$ randomly-distributed bits. For small cell sizes, these requirements are innocuous; however, the $\sim L^4$ exponential growth in demand for random numbers is prohibitive. In the case of Quside's QRNG apparatus used in our simulations, which reaches quantum random number generation rates of 400 Mbps, we spend around 12 hours for each simulation of $L = 256$. Doubling the size of the network (L =512) requires twenty days for each simulation; by doubling it again ($L = 1024$), we estimate a simulation time of almost six months per simulation. To avoid this computational bottleneck, and speeding-up the simulation of the $L = 512$ case, we used an amplification of the QRNG's random numbers. This amplification consists on the implementation of a PRNG on the FPGA, which is reseeded as fast as the QRNG provides new seeds. This decision introduces some correlations that are not present for smaller lattices. This fact, along the very low number of repetitions, could potentially be the reason of the more noisy behaviour observed in the last panel of Fig.4.12, as discussed in the PRNG section. Nevertheless, due to the small size of the statistical sample at hand, we cannot conclude which one of them is the predominant reason without additional simulations.

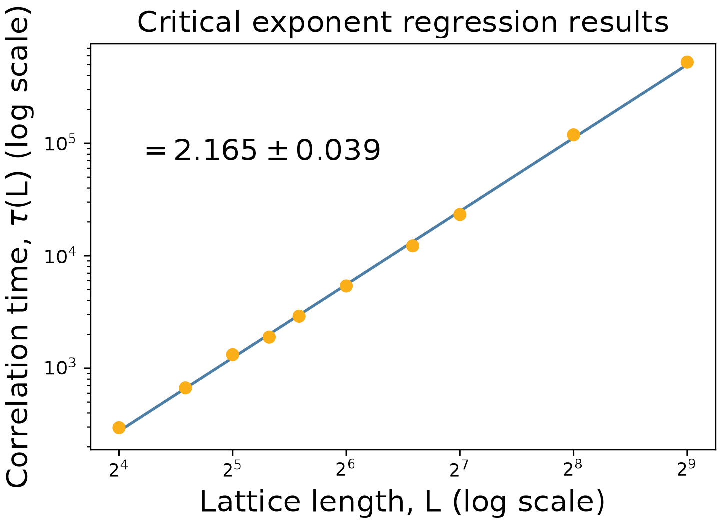

Figure 4.13: **Critical exponent calculation with a Quantum Random Numbers source.** Orange: estimated relaxation times (in log10 scale) as a function of the lattice size (in log2 scale); blue, the linear fitting $\log_2(\tau) = z \cdot \log_2(L) + b$ associated. The slope of the fitting $z = 2.165 \pm 0.039$ corresponds to the critical exponent, in agreement with theoretical results.

4.4.4 Offset in the linear fit of the dynamic exponent

As discussed in previous sections, the relaxation time of the system is expected to have a power-law relation with the lattice size L. In order to obtain the dynamic exponent z, one usually takes logarithms to both sides of Eq. (4.43) by naively reducing it to an equality $\tau = L^z$ and then makes a linear fit $\log(\tau) = z\log(L)$. As a final remark, we finish this chapter by strictly considering the proportional sign in Eq. (4.43) by stating that $\tau = \tau_0 L^z$, therefore allowing the linear fit to have an offset, $\log(\tau) = z\log(L) + \log(\tau_0)$.

We summarize our findings in Table 4.5, in which we compare the offset obtained by the fitting of the QRNG data against those obtained by the PRNG data. There are two main things worth noting about the obtained results. First, we observe that all of them yield a non-zero, negative offset. And, secondly, we note that those PRNG using reseeding (thus having an infinite period, and therefore being closer to a TRNG) yield values closer to the one obtained by the QRNG.

	$log(\tau_0)$
QRNG	-0.362
PRNG3 $k = \frac{1}{2}$	-0.1701
PRNG3 $k = 2$	-0.2454
PRNG2	-0.0906
PRNG0	-0.1007
PRNG1	-0.1183

Table 4.5: Results obtained for $log(\tau_0)$ by using different RNG. These values correspond to the fittings yielding the dynamic exponents shown in previous sections.

4.4.5 Comparison with other standard randomness tests

This chapter finally concludes by taking a look at how the used PRNGs and QRNG perform under the National Institute of Standard and Technology (NIST) Statistical Test Suite for Random and Pseudo Random Number Generators NIST SP800-22 [102].

In Fig. 4.14 we show the violin plots of the NIST tests' results obtained with all the considered PRNGs with high frequency reseeding and the QRNG, to illustrate how these are distributed. The ordinate axis labels the percentage of passed tests, while a wider figure indicates that a bigger proportion of tests obtain results within that ordinate value when using that particular generator. We observe that PRNG0 and PRNG3 obtain clearly worse results than the other ones, as their violin plots' tails extend all the way below the expected failure rate $\alpha = 0.01$. In contrast, PRNG1 and PRNG2 show a distribution that is localized around the expected failure rate. Moreover, this distribution strikingly resembles that of the QRNG. The bad results obtained by PRNG3 are easily explained by the fact that, even with a high frequency

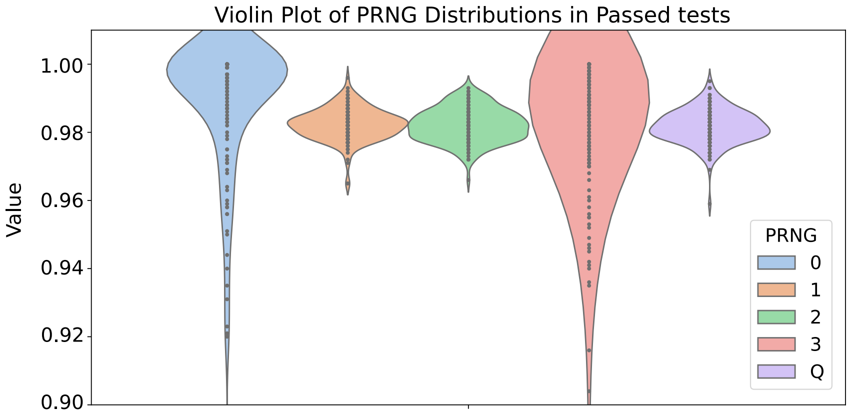

Figure 4.14: **Violin plots of the randomness tests passed by all the PRNGs and the QRNG.** While PRNG0 and PRNG3 obtain clearly worse results than the other ones, PRNG1 and PRNG2 are indistinguishable from the QRNG.

reseeding, its period of repetition is short enough for the NIST tests to catch the inner correlations. Contrarily, the explanation for the bad results of PRNG0, which has the longest period of repetition among all of them, is more subtle and can indicate that its inner correlations are more apparent to the NIST tests than those present in other generators. While those inner correlations are not seen to affect the computation of the dynamic exponent as those present in PRNG1, PRNG2 and PRNG3 (see Fig.4.8), the correlation shown by these first two, which indeed affect the variance in z, seem to pass unseen under the NIST SP800-22 tests. We thus conclude that there are certain types of correlations that can be detected with standard methods but not with the present one, and vice-versa, so that both strategies can be jointly considered in order to improve the set of randomness tests. Further, it is important to stress on the fact that the computation of the dynamic exponent is affected by the correlations present in some generators that are, from the NIST SP800-22 STS perspective, indistinguishable from a QRNG.

4.5 Conclusions

In this chapter, we have computed the dynamic critical exponent, z, of the two-dimensional Ising model by means of Monte Carlo simulations. In them, (pseudo-) random number generators of various qualities have been used, ranging from poor PRNGs showing lots of correlations and repeating sequences, to high-frequency re-seeded PRNGs and even a completely and truly random QRNG.

To begin with, an exhaustive review of the literature on the value of z is carried on, not showing any convergence to a commonly accepted one. Since its computation is known to be affected by the quality of the random numbers used in the MC method, the value obtained with a QRNG given here is expected to be more precise. Further, this value is seen to closely resemble the mean of the Gaussian-distributed results obtained with Monte Carlo approaches in the literature.

The ill-behaved PRNGs that output repeating sequences of pseudo-random numbers are seen to have a dramatic impact on the computation of z, specifically on its variance. A high-frequency reseeding version of them, nevertheless, is able to go around this problem. Furthermore, we have seen that the higher the frequency of the reseeding (and thus the closer to a true randomness source), the better the results. An extrapolation of the error in the limit of the highest possible frequency (a TRNG) is explored and seen to coincide with the error obtained for very small lattices, whose simulation does not require a high enough amount of random numbers for the correlations to arise and get noticed. This detected error in the variance paves the way to show the superiority of QRNGs for computing purposes: once QRNGs achieve fast enough generation throughput and similar tests can be carried on with them, they might obtain a lower error.

Finally, the comparison of the used PRNGs and the QRNG with the more standard tests commonly used to detect hidden correlations among pseudo-random sequences show that some of the PRNGs are indistinguishable from the QRNG. On the other hand, some of those performing worst under such standard tests are also some of the ones obtaining smaller errors in the variance. The new detected error is therefore seen to be complementary to the standard tests, as none of the two approaches can clearly, at this point and until further investigation with QRNGs, unambiguously point out which are the best ones.

Chapter 5

Defect-driven Population Annealing

In this chapter, based on the results presented in [37], a quantum-inspired modification of Population Annealing is presented. In analogy to the Toric Code, single bond flips allow for the creation and movement of topological defects within the lattice and, when these defects collide and are therefore annihilated, the outcome is equivalent to a non-local move. Further, in two dimensional lattices these non-local moves are equivalent to cluster updates such as the ones implemented in the Wolff algorithm.

5.1 Introduction

At the beginning of the second decade of this century, the first analog quantum annealing machines were commercialized by the Canadian company D-Wave Systems [165, 166]. These were specifically designed to solve NP complete problems with a discrete solution space and with a spin-glass-like energy landscape, characterized by a huge number of energy minima. Besides the exponential growth of the Hilbert space in which the quantum states representing the solutions live, such machines exploit yet another quantum property that allows them to efficiently navigate the solution space in search for the ground state of the problem (in which its solution is encoded): quantum tunneling [167, 168]. Essentially, quantum tunneling is the capacity of quantum wave functions to pass through a high potential barrier, even when the classical description prohibits it due to the particle not having a high enough energy. In the quantum annealing paradigm of solving optimization problems, quantum tunneling therefore allows the state to escape from local energy minima, and pass through potential barriers to reach the global ground state without the need of the thermal excitations used by classical annealing in metallurgy and simulated by the Metropolis algorithm. Following the discussion of Fig.2.5, the state can be transferred directly to the deepest well without first having to accept a less energetically favourable configuration to surmount it, see Fig.5.1.

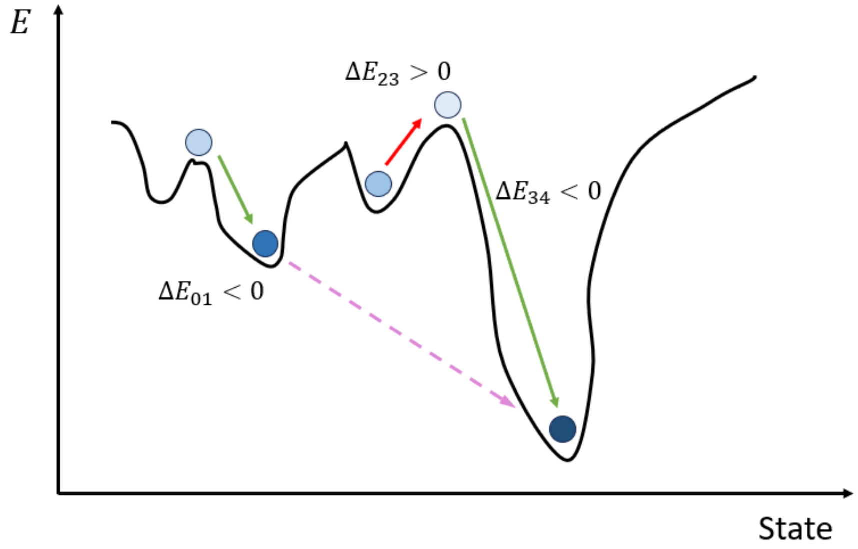

Figure 5.1: **Quantum Annealing exploits quantum tunneling to reach the global ground state.** The question is: is it possible to simulate processes with similar properties in classical algorithms to avoid the systems getting trapped in local minima?

Quantum annealing, nevertheless, is still far from being a plausible solution to the optimization problems faced today. Even if it has been demonstrated to be able to effectively solve (relatively) small-scale problems, only classical computation can nowadays achieve the number of reliable bits needed to tackle problems of an industrially-relevant size [24]. These facts have led to the design of classical algorithms that draw inspiration of quantum processes in order to improve the capacity of the simulations of avoiding getting trapped within local minima.

In particular, the usually used thermal-based classical algorithms (Population Annealing and Parallel tempering [169, 170], which are in essence equivalently efficient [51]) for solving spin glass systems sometimes encounter hard disorder instances for which their capacity of thermalizing the simulated Hamiltonians is severely damaged [62]. At each temperature step, the system does not reach a proper thermal equilibrium and thus the adiabatic assumption over which the annealing process relies to find the minimum energy state can no longer be applied with enough confidence.

5.2 Population Annealing

Population Annealing [171, 172] is a generalized ensemble extension of SA in which a family of a total of R_0 replicas of the same system are independently simulated in parallel. PA starts considering all R_0 replicas at infinite temperature, at which equilibration is easy but encountering the global ground state is difficult, and anneals the population towards a low temperature at which equilibration is difficult but the probability of finding the ground state is higher. We denote the annealing schedule as $\Xi_\beta = [\beta_0, ..., \beta_{N_T}]$, with β the inverse temperature and $\beta_i < \beta_{i+1}\ \forall i$, where i labels each resampling and equilibrating step.

Moreover, PA utilizes a resampling step that, taking advantage of the population of replicas, considerably speeds up the thermalization process. To this end, each replica is eliminated or replicated according to its ensemble weight, so that the newly produced population of replicas is kept near statistical equilibrium with respect to the desired distribution. Specifically in PA, the ensemble distribution targeted by the resampling step is the Gibbs distribution. Therefore, when resampling from an inverse temperature β to β', the statistical normalized weight for each replica is computed as

$$\tau_j(\beta, \beta') = \frac{w_j}{R_\beta^{-1} \sum_r w_r} = R_\beta \frac{e^{-(\beta'-\beta)E_j}}{\sum_r e^{-(\beta'-\beta)E_r}}, \tag{5.1}$$

where the index r runs over all replicas within the population, of size R_β. The normalization factor is needed to keep the population size at the new temperature β' close to the previous one, β. Note that the ensemble weight for each replica, $w(\beta, \beta')$, is proportional to its corresponding energy, and thus lower energy states are more prompt to be replicated while those far away from the system's minima will tend to be eliminated. This process can be somewhat related to the selection part of genetic algorithms, in which replicas with lower energy (this is, better fitness if the goal is to find the ground state of the system) are set to have several offspring, and is known to enhance the efficiency of the algorithm compared to SA [51]. The resampling step nevertheless introduces correlations between replicas in the new produced ensemble, and therefore an equilibration of the population must be performed. Just as in SA, each replica then undergoes an MCMC method for a number of sweeps N_{sweeps}. MCMC methods are know to equilibrate slowly when a system is annealed from one temperature to the other in the critical region or when dealing with rugged energy landscapes [79], and thus highly adiabatic processes are usually required.

Essentially, then, PA combines SA's thermal annealing with a population of independent replicas of the system under study. Since, for thermally equilibrated and sufficiently large replicas population, resampling allows to change population temperature in a single step while the population remains approximately equilibrated, thermal annealing is then significantly sped up by replicas population resampling procedure done according to the Gibbs distribution.

5.2.1 Resampling protocol

In order to reproduce the desired ensemble distribution at each resampling step, the expected number of descendants of each replica within the population must be given by its corresponding statistical weight, Eq. (5.1)

$$\langle n_j \rangle = \tau_j(\beta, \beta'). \tag{5.2}$$

Several options can be considered to this end [173]. Specifically, instead of a protocol that keeps the population size fixed to R_0, we choose this number of descendants n_j to be computed as

$$n_j = \begin{cases} \lceil \tau_j \rceil & \text{with probability } P_{\text{ceil}} = \tau_j - \lfloor \tau_j \rfloor \\ \lfloor \tau_j \rfloor & \text{with probability } P_{\text{floor}} = 1 - P_{\text{ceil}}. \end{cases} \tag{5.3}$$

This choice of probability minimizes the variance of n_j and lets the population's size vary around a mean value R_0 with a fluctuation of $\sqrt{R_0}$. This widely used choice reduces the correlation between replicas in the descendant population [62], and therefore minimizes the amount of numerical power that must be invested in MCMC updates in order to decorrelate them.

5.2.2 Free energy estimation

One of the main features of Population Annealing is the fact that it is able to give an estimate of the free energy of the system at no additional computational cost. To see it, first note that the ratio between partition functions at two different temperatures,

β and β', can be written as

$$\begin{aligned} \frac{Z(\beta')}{Z(\beta)} &= \frac{1}{Z(\beta)} \sum_\gamma e^{-\beta' E_\gamma} = \\ &= \frac{1}{Z(\beta)} \sum_\gamma e^{-\beta E_\gamma} e^{-(\beta'-\beta)E_\gamma} = \\ &= \langle e^{-(\beta'-\beta)E_\gamma} \rangle_\beta \end{aligned} \tag{5.4}$$

where the index γ runs through all spin configurations. In Population Annealing, resampling the replica population from β to β' with the Gibbs distribution produces, for each replica r, an expected number of descendants given by Eq. (5.1). To this end, the normalization factor $Q(\beta, \beta')$ must be computed

$$Q(\beta, \beta') = \frac{1}{R_\beta} \sum_r e^{-(\beta'-\beta)E_r}. \tag{5.5}$$

If the population of replicas is a proper representation of a sufficiently significant portion of the configuration space, the average in the right hand side of Eq. (5.4) can then be estimated as a population average over the replicas

$$\langle e^{-(\beta'-\beta)E_\gamma} \rangle_\beta \approx Q(\beta, \beta'). \tag{5.6}$$

From the thermodynamic relation between the free energy and the partition function of a given system we see that

$$F = -T \ln Z = -T \ln \frac{Z(\beta)}{Z(\beta')} - T \ln Z(\beta'). \tag{5.7}$$

If the partition function at a given temperature β' is known, then by introducing Eq. (5.6) in Eq. (5.7) and rearranging the terms one finds an estimator for the free energy, $\tilde{F}$, at temperature β

$$-\beta \tilde{F} = -\ln Z(\beta') + \ln Q(\beta, \beta'). \tag{5.8}$$

The partition function, nevertheless, is not an easily accessible function to compute at any desired temperature. On the other hand, considering an infinite initial temperature ($\beta_0 = 0$) and thus an equiprobable distribution of configurations, the partition function in the first term of the right hand side of Eq. (5.8) can be calculated exactly as $Z = \Omega^{-1}$, with Ω the total number of possible states in the system. In Ising-like systems with a total number of spins N_s, spins take two different values and thus $\Omega = 2^{N_s}$. Therefore

$$-\beta \tilde{F} = N_s \ln 2 + \ln Q(\beta, \beta'). \tag{5.9}$$

Finally, by recursively decreasing temperatures one finds the estimator for the free energy at any given β_k to be

$$-\beta_k \tilde{F} = N_s \ln 2 + \sum_{i=0}^{k-1} \ln Q(\beta_i, \beta_{i+1}), \tag{5.10}$$

assuming that the simulation was started at an infinite temperature.

Let us stress on the fact that Population Annealing already had to compute the normalization factors Q at each temperature to perform the resampling step, so the estimation of the free energy comes at no additional cost.

A small numerical consideration for big lattices

Being the total energy of the system extensive, for big enough lattices the computation of the exponentials appearing in the Boltzmann weights needed for the number of descendants in the resampling step, Eq. (5.1), becomes numerically challenging and must be addressed carefully [174]. In order to proceed, we introduce an offset E_{off} that decreases the exponents:

$$\tilde{w}_r(\beta, \beta') = w_r(\beta, \beta') e^{(\beta-\beta')E_{\text{off}}} = e^{-(\beta-\beta')(E_r - E_{\text{off}})}, \tag{5.11}$$

where the energy offset is the same for all replicas and chosen as the minimum energy among them,

$$E_{\text{off}} = \min_r E_r, \tag{5.12}$$

so that all exponents are guaranteed to be negative and therefore numerical overflow is avoided. Also note that, since all replicas use the same offset, the additional exponential factor gets canceled out when computing the corresponding expected number of descendants, and therefore Eq.5.1 remains unchanged. On the other hand, the free energy estimator must be modified in order to take into account the introduced offsets:

$$-\beta \tilde{F} = N_s \ln 2 + \sum_{i=0}^{k-1} (\beta_{i+1} - \beta_i) E_{\text{off}}^{(i)} + \sum_{i=0}^{k-1} \ln \tilde{Q}(\beta_i, \beta_{i+1}). \tag{5.13}$$

Here, $\tilde{Q}$ now refers to the normalization constant directly given by the population annealing algorithm, so with the energy offsets already taken into account.

5.2.3 Confidence on the solution found

Due to their heuristic nature, algorithms such as PA or SA may not reach the global, *true* ground state of the studied system. On the contrary, the only strategy to doubtlessly find it consists on exploring the whole configuration space, a task that becomes intractable for even small lattice sizes due to its exponential growth. On the other hand, one has to conform with a measure of the likelihood of the solution found being the global minimum, or as close as possible to it. Specifically for PA, this likelihood can be assessed by measuring two different parameters on the final population of the algorithm's run, namely the probability of being in the ground state, and the family entropy.

Probability of being in the ground state

The first considered parameter is the probability of being in the ground state at a given temperature, g_0, which is given by its Boltzmann weight

$$g_0(E_0, F; \beta) = \frac{2e^{-\beta E_0}}{Z(\beta)}. \tag{5.14}$$

Here E_0 is the ground state energy of the system, $Z(\beta)$ is the partition function at inverse temperature β and the factor 2 comes from the up-down degeneracy of the ground state. Since the partition function is difficult to compute directly but, on the other hand, as previously discussed, PA yields an accurate estimation of the free energy, it feels natural to introduce the thermodynamic relation $F = -\beta^{-1} \ln Z$ into Eq 5.14 and thus work with the more accessible definition

$$g_0(E_0, F; \beta) = 2e^{\beta(F-E_0)}, \tag{5.15}$$

which will prove useful later on. Being N_0 a variable counting how many times the minimum energy state found during the whole process appears among the population of a total of R_f replicas in the final stage of the simulation, Population Annealing can easily estimate this probability using a statistical interpretation. Thus, in the limit of a large population size R_f, we have an independent measure of g_0:

$$g_0(N_0) = \frac{N_0}{R_f}. \tag{5.16}$$

Let us define the ground state and free energies per spin, $e_0 = {}^{E_0}/_{N}$ and $f = {}^{F}/_{N}$ with N the number of spins, respectively. Since the free energy is always lower than the internal energy of the system (and, in particular, lower than the ground state energy), the term $f - e_0$ converges to a negative value that, in the thermodynamic limit, should be independent of the disorder instance. Equation Eq. (5.15) therefore yields a probability of being in the ground state that decreases exponentially with the number of particles of the system. Consequently, in order for the estimation of g_0 from the statistical interpretation to be accurate enough, this has to be compensated by an exponential increase in the number of replicas used in the simulation with the lattice size.

Family entropy

The second considered parameter is the effective number of surviving families N_{eff}, which measures the number of replicas that have been able to find the lowest energy state *independently*. Let us first define a family of replicas as those group of replicas at a given point in the simulation that descend from the same replica in the initial population. Note that, at the start of the annealing process, each family is therefore constituted by one single replica. At each resampling step in PA some families of replicas will increase or decrease its size, while some might even be completely eliminated with no chance to appear again on later stages. Let ν_j be the size of family j relative to the total population size or, in other words, the fraction of replicas present in the current population that descend from replica j in the initial population. The family entropy is hence defined as

$$S_f = -\sum_j \nu_j \log \nu_j \tag{5.17}$$

with the index j running over all families, and the effective number of surviving families can be computed from it as

$$N_{\mathrm{eff}} = \exp\left[S_f\right]. \tag{5.18}$$

Family entropy is a measure of the equilibration or thermalization of the sample [62].

From Eq. (5.1) follows that reducing the temperature differences $\Delta\beta = \beta' - \beta$ by increasing the number of temperature steps in the annealing schedule, implies a lower dependency between the energy of a given replica and the number of descendants that it will have in the resampling step. In the limit of very small steps, $\Delta\beta \to 0$, this expected number of descendants will actually approach 1 and therefore the family entropy at the end of the annealing process will be maximum, as $\nu_j = \frac{1}{R} \forall j$ and thus $N_{\mathrm{eff}} = R$. This is of course in agreement with the intuition that, if we take smaller temperature steps, then it is easier to equilibrate the replicas given that the previous population was already in equilibrium. Since the initial population is simulated at an infinite temperature, $\beta = 0$, and is therefore guaranteed to be in equilibrium if the states are initialized randomly, an induction argument suggests that so will be the final population in such limit. Furthermore, in accordance with the adiabatic theorem, the whole annealing process is slower and therefore the probability of finding the ground state is higher, eventually reaching 1 for small enough $\Delta\beta$. A better thermalization is therefore exhibited by a higher family entropy.

Concluding, if the number of replicas used is high enough and the population is properly thermalized, the chances of finding the global ground state of the system increase. Therefore our confidence on achieving such goal is higher when the measured g_0 and N_{eff} at the end of the simulation are higher too [62, 175].

5.3 Bond representation and TC topological defects

In this section we present the spin glass model we will work with throughout the whole chapter, which we refer to as Random Field Wall (RFW) model. As we shall see, it is an adequate three-dimensional extension of the two-dimensional Edwards-Anderson model that allows for the introduction of Toric Code topological defects [176].

5.3.1 The 2d Edwards-Anderson model

For simplicity, consider first a standard two-dimensional square lattice with only nearest-neighbours interactions and with periodic boundary conditions. It is described by the Ising Hamiltonian

$$H = \sum_{\langle i,j \rangle} J_{ij} s_i^z s_j^z, \tag{5.19}$$

where $\langle i, j\rangle$ indicates that the indices run over the nearest neighbours, in this case four per spin. When the couplings J are distributed according to a Gaussian probability function, this Hamiltonian corresponds to the 2d Edwards-Anderson model. Let us now shift to the bond picture by introducing the change of variable $\sigma_b^z = s_i^z s_j^z$. The Hamiltonian now reads

$$H = \sum_{b} J_b \sigma_b^z. \tag{5.20}$$

In the bond picture of a spin lattice the new binary variables σ_b represent the state of the bond between two spins, which depends on their relative alignment instead of that of the spins themselves. The bond is either said to be up (the spins connected by it have the same value, thus $\sigma_b = 1$), or down (the spins are not aligned and $\sigma_b = -1$). Note that therefore a bond configuration can be translated into a spin one with up to a decision of the value of an initial spin, which corresponds to the up-down degeneracy of the spin picture.

However, in contrast to the spin representation, not all $\{\sigma_b^z\}_b$ configurations represent a physical state, and Hamiltonian Eq.5.20 is not that of a collection of free particles. Rather, the Hamiltonian must be understood along with the constraints associated to the set of feasible configurations. To see this, take an elementary closed path within a given lattice, which we henceforth refer to as plaquette. Arbitrarily define the value of a first spin and then those of the neighbouring ones. It can be seen that, in order to be able to unequivocally assign a value to all spins composing the given plaquette, the product of all the bonds between them must be one, see Fig.5.2. To account for this we define the plaquette operator

$$B_{\square} = \prod_{b \in \square} \sigma_b^z, \tag{5.21}$$

which is equivalent to TC's B operator. The notation $b \in \square$ indicates the indices of the bonds pertaining to a given plaquette.

By the definition of the bond variables, if the plaquette operator applied over a given site yields $B_\square = -1$ one of its surrounding spins must be aligned along two different directions at the same time and thus can non unequivocally be assigned neither of them, see 5.2. In such case we say that the site enclosed within the plaquette contains a defect, as it is not possible to translate it back to a physical configuration of spins. When considering periodic boundary conditions, nevertheless, assuring that all plaquettes within a lattice yield $B_\square = 1$ is a necessary but not sufficient condition for the physicality of its configuration, since the plaquettes conformed by the outer bonds must also be taken into account. In this scenario one can see that, further than the plaquette defects, one has to define as well defects over all the straight lines that cross the lattice, see Fig.5.3. We define the line defect operator as:

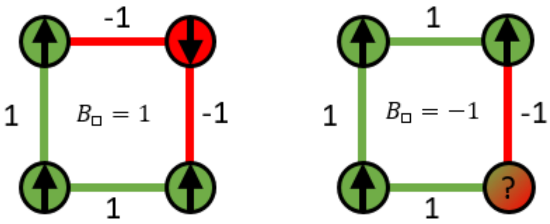

Figure 5.2: **Physical states in the bond representation are characterized by $B_\square = 1$.** When a plaquette yields $B_\square = -1$, spins located on its vertices cannot be unequivocally determined and therefore the configuration does not represent a physical state of the Ising model. We thus say it contains a defect.

$$B_| = \prod_{b\in|} \sigma_b^z, \tag{5.22}$$

where the notation $b \in |$ indicates the indices of the bonds pertaining to a given line. Consequently, a line is said to contain a defect whenever there is an odd number of non aligned bonds in it. When a line is defected, there is at least one spin in it that can not be unequivocally determined and, analogously to the plaquette discussion above, the configuration of bonds cannot be translated to a spin configuration.

Summarizing, in order for a lattice in the bond representation to have a proper translation into spin variables and thus to represent a physical state of the Ising system, it must fulfill the following conditions:

- All elementary plaquettes yield 1: $\sum_\square B_\square = N_p$.
- All straight lines across the lattice (both horizontal and vertical) yield 1: $\sum_| B_| = N_L$.

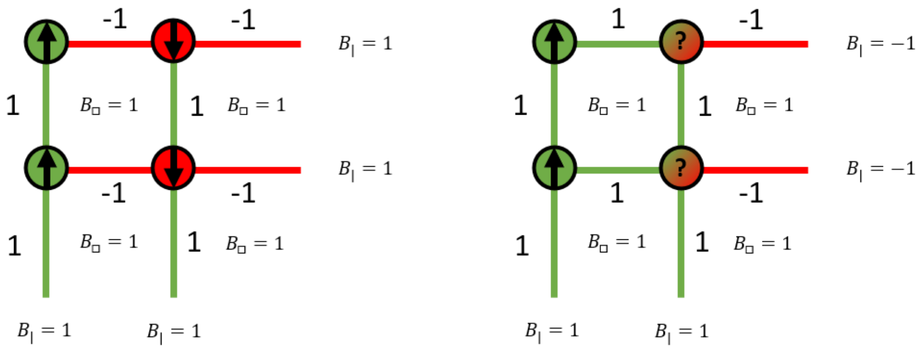

Figure 5.3: **If periodic boundary conditions are assumed, physical states in the bond representation are characterized by $B_| = 1$.** When a line yields $B_| = -1$, spins located on it cannot be unequivocally determined and therefore the configuration does not represent a physical state of the Ising model. We thus say it contains a defect.

Any configuration that does not fulfill either of these two conditions has topological defects and then it is not properly described by Hamiltonian Eq.5.19. In what follows we refer to such configurations as non-physical. For the case at hand, being L the side length of the 2d lattice, the total number of plaquettes is $N_p = L^2$ and the total number of lines is $N_L = 2L$.

5.3.2 Defects' dynamics in 2d: cluster updates

Consider a collection of neighbouring spins within a given lattice, which we will refer to as cluster. In the leftmost panel of Fig.5.4 a quadratic lattice in its spin picture is shown, with green and red dots indicating spins pointing to up/down directions and black lines indicating the nearest-neighbours coupling topology. For simplicity, consider a ferromagnetic model such that the system's energy is minimized when all its spins are aligned. To this end, the cluster of red spins, indicated by a surrounding blue dashed line, is proposed to be flipped. Let us now switch to the bond representation of this same configuration, and identify those aligned spins with a green bond and those not aligned with a red one, center panel of Fig.5.4. Note that, since all the spins within the cluster are aligned, they are connected by means of green bonds, and so happens with all the spins outside the defined cluster as well (a bigger cluster containing all the spins not included in the considered one could have been defined alternatively). Finally, in the rightmost panel of Fig.5.4 we can see the effect of accepting the proposed move. On the one hand, from the spin representation's perspective, accepting the proposed cluster move means that all spins within it are flipped. On the other hand, from the bond representation's perspective, only those bonds that were coupling the spins defining the frontier of the cluster to those outside it are flipped.

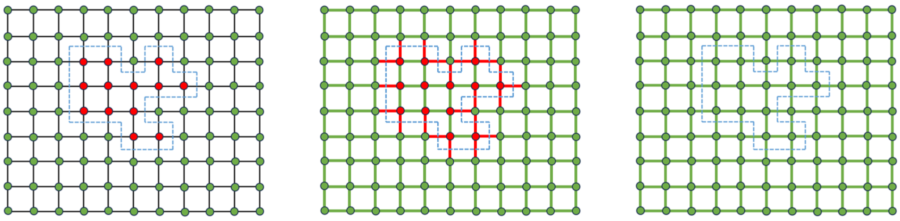

Figure 5.4: **Cluster update in a 2d quadratic spin lattice.** In the bond representation of a spin lattice, when a cluster of contiguous spins is flipped at a time, only the bonds connecting its frontier spins with those outside it change its value.

Let us now take a look at the flipping of a given cluster. To this end, first recall the discussed model described by Hamiltonian Eq.5.20, in which only certain configurations were allowed in order to preserve its physical meaning. Note that, if we do not restrict the system to physical states and instead consider an expanded configuration space, we can introduce topological defects by performing single bond flips, by means of the operator σ_b^x. When applied to a bond separating two plaquettes with $B_\Box = 1$, operator σ_b^x creates a defect in each of them such that afterwards $B_\Box = -1$, Fig.5.5 left panel. Contrarily, if the bond to which σ_b^x is applied separates two plaquettes that already hold defects, these will annihilate themselves. On the other hand, when applied to bonds separating a defected and a non-defected plaquettes it effectively moves the defect from one to the other, Fig.5.5 center panel. Repeatedly flipping single bonds therefore creates, moves and annihilates topological defects within the lattice. Eventually, when the path followed by a pair of defects happens to close itself, it defines a line through which all bonds have been flipped, Fig.5.5 right panel. This is equivalent to the cluster updates discussed previously, and thus the spins enclosed within the closed path are effectively flipped as a cluster.

Note that, from this perspective, a single spin flip can now be seen as flipping all bonds between it and its neighbours, equivalently described by a single spin cluster, by means of creating a pair of defects, moving them around a given spin and finally annihilating them, Eq.5.23, which is equivalent to TC's star operator, A_s.

$$s_i^x = \prod_{b \in +_i} \sigma_b^x, \tag{5.23}$$

with $b \in +$ the indices of the bonds surrounding spin s_i. In other words, while standard Metropolis MC simulations performs spin dynamics by applying the TC's A operator, we now seek to enhance the thermalization of the system by means of TC's B operator. First, however, we must look for a proper 3d model in which such dynamics can be performed.

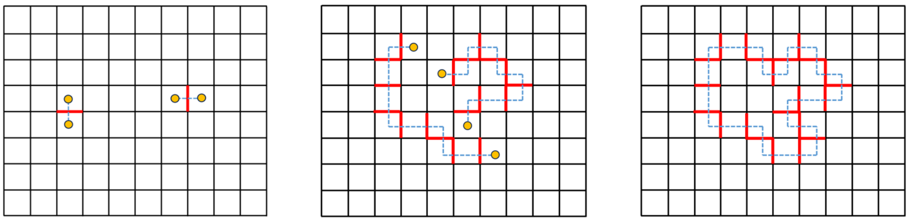

Figure 5.5: **The dynamics of the defects in 2d lattices induce cluster moves.** By creating, moving and annihilating defects in a 2d lattice, the spins enclosed within the closed path followed by the defects are updated as a cluster.

5.3.3 A proper three-dimensional generalization

The extensively studied cubic lattice, this is, the straightforward three-dimensional generalization of the square lattice discussed above, does not allow for the definition of equivalent topological defects, since the bonds between spins do not conform an enclosed cube. Thought of from the perspective of the defect's dynamics, a bond connecting two spins in a 3d space does not separate two single sites within which a defect can be defined, and thus its flip does not move a defect from one to the other.

On the other hand, we can shift to an appropriate 3d extension of such a 2d Ising model that allows for the introduction of topological defects in bond representation by considering the Kramers-Wannier dual [177] to the 3d uniform Ising model, which is the 3d Ising gauge theory model [178]. In it, spins are located on lattice edges and are subjected to plaquette interactions on each cube face. The latter model can be represented as a "wall model" where the walls are the new binary variables taking values of cube face plaquettes, the plaquette interaction J is now a field acting on walls, and walls are not independent variables but are subjected to the constrain that in each cube the product of all 6 planes must equal 1. The given configuration of walls determines energy and represents all 2^N gauge-equivalent states (where N is the number of vertices). Consequently, in this representation entropy is greatly reduced.

To visualize the 3d wall model, one may consider an extrusion onto the new dimension of all the elements described so far. This means that we add an extra dimension to all of them, and therefore degrees of freedom change as

- Spins (points) become edge spins (lines).
- Bonds (lines) become walls (planes).

And with respect to operators (constrains)

- Plaquettes (which may contain a defect) become cubes (which may contain a defect).

- Lines (which may contain a defect) become planes (which may contain a defect).

As we previously had a total of 4 spins connected by bonds (each bond connecting two single spins) and forming a plaquette surrounded by 4 bonds that could have a defect in it, we now have 12 edge spins connected by walls (each wall connecting four edges) and all of them forming a cube surrounded by 6 walls that can hold a defect as well; see Fig.5.6. In this case, the change of variable between edge spin and wall models is:

$$\eta_w^z = e_i^z e_j^z e_s^z e_t^z \tag{5.24}$$

where $e_{\{i,j,s,t\}}$ are the edge spins surrounding wall η_w.

Figure 5.6: **The 3d Random Field Wall (RFW) model (right) is a generalization of the 2d bond model (left).** By extruding an elementary plaquette and all its elements onto an extra dimension, we obtain a cube surrounded by edge spins which are 4-to-4 connected by a wall. The site enclosed inside the cube can hold a defect.

The Hamiltonian of the 3d Ising gauge model in the wall representation, or RFW, finally reads

$$H = \sum_w J_w \eta_w^z. \tag{5.25}$$

However, and analogously to Eq. (5.20), this is not a free particle model. Again, it has to be understood along with the constraints associated to the set of feasible configurations.

As with the spin flip in the 2d case discussed above, Eq. (5.23), from this perspective the edge-spin flip operator, Eq. (5.26), now consists of updating all walls surrounding a given edge-spin, which can also be seen as creating a pair of defects and moving them around a given vertex until they collide and annihilate again, Fig. 5.7:

$$e^x_\mu = \prod_{w\in +_\mu} \eta^x_w. \tag{5.26}$$

It is important to note here that, in the 3d case, the dynamics of the defects do not directly yield, as did happen previously, equivalent cluster updates. Being a cluster now a volume containing a set of spins, the boundary separating it from the rest of the lattice is a an enclosed surface. Therefore, while it is not strictly impossible, it indeed is quite difficult that the path followed by the defects, constituted by stripes of updated walls, fills the whole surface. Still, the importance of the dynamics of the defects will now fall on the fact that they yield non-local updates.

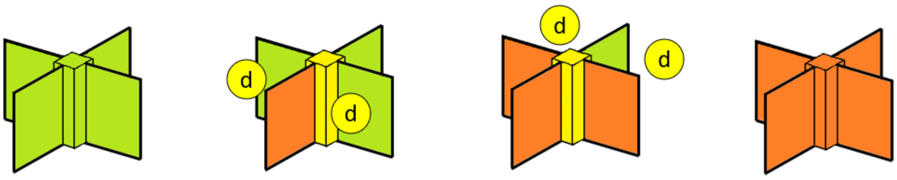

Figure 5.7: **Graphical representation of an edge spin flip as the flipping of four walls.** Green and orange colors label ±1 values for the walls and the edge spin, while yellow means that there are defects in the adjacent cubes and therefore the edge cannot be unequivocally determined. It is important to stress that, even if only one edge spin is shown for clarity, each of the walls is surrounded by (is coupling) four edge spins and thus the shown defects must be seen as pertaining to a whole cube (for instance, they could move to the cube on its top instead).

To determine whether a cube or a plane hold defects and thus if a given configuration of a RFW lattice can be unequivocally translated into a physical edge-spin one or not, we generalize the plaquette and line operators, Eq.. (5.21) and Eq.. (5.22), to cube and plane ones as

$$B_c = \prod_{w\in c} \eta^z_w, \tag{5.27}$$

$$B_p = \prod_{w\in p} \eta^z_w, \tag{5.28}$$

respectively, with $w \in c$ the indices of the walls pertaining to cube c and $w \in p$ the indices of the walls pertaining to plane p. The physicality condition of the wall configuration, and hence the fulfillment of constraints for Eq. (5.25), is now ensured when

- The cube operator applied over all elementary sites yields 1: $\sum_c B_c = N_c$.

- The plane operator applied over all planes across the lattice yields 1: $\sum_p B_p = N_p$.

Now the total number of cubes and of planes within the lattice are $N_c = L^3$ and $N_p = 3L$, respectively, and we can thus count the total number of defects that a given configuration contains as

$$N^{(D)} = \left[\left(N_c - \sum_c B_c \right) + \left(N_p - \sum_p B_p \right) \right] . \tag{5.29}$$

To conclude, let us recall that, being the RFW model a 3d generalization of the 2d bond model, one should be able to recover the later from the former. To this end, first consider a wall lattice with a single layer of cubes on the, say, $(x, y)-$plane, and then set an infinite-valued local field to all edge-spins that have its edge pointing in either x or y directions, effectively freezing their motion. Doing so we restrict the edge-spin flips to only those whose edge's direction is perpendicular to the plane of the lattice, and thus the resulting model is equivalent to the desired 2d bond one with an additive constant that does not change the physics of the Hamiltonian.

5.3.4 Hardness of the RFW Lattice model

All spin glass models are intrinsically hard to solve, since their energy landscape is characterized, by definition, by a huge number of wells separated by hard to overcome barriers. Different topologies yield, nonetheless, different hardnesses, and even within the same model there can be disorder instances that are harder to solve than other ones. As a final discussion of the presented RFW model, it is therefore interesting to study its hardness by comparing it against that of more common models. Particularly, we find it illustrative to contrast it with the 3d Edwards-Anderson model, being it a 3d, nearest-neighbours interacting spin lattice as well. Extensive research has been conducted for this later model, with reliable results reported for sizes of up to $L = 10$ [51], or equivalently $N_s = 1000$ spins. Bigger lattice sizes have also been explored in the literature, but with a lower degree of confidence on the quality of the thermalization achieved [62].

As discussed earlier, the RFW model does not correspond to the standard 3d-EA model but allows for the effective dynamics of topological defects. On the one hand, since it is well known that multi-spin interacting models (with *multi* referring to more than two) can in general be very difficult to sample and optimize, the RFW model, Eq. (5.25), which is equivalent to 3d random Ising gauge model with four-spin plaquette interactions, can indeed expected to be difficult. On the other hand, the form of the interactions depends on the representation (bond or spins) and, for the case at hand, one could therefore expect the 2d Edwards-Anderson model, Eq. (5.20), which in the bond representation looks like a free-particle model with four-particle constrains, to be hard as well. Contrarily, it is known that the later lacks hard instances and moreover, considering that the 2d EA model in bond representation can be seen as a special case of a 3d RFW lattice, the multi-spin nature of the constrains in the wall representation of the RFW model is not an obvious argument that it is more difficult to simulate than the already hard 3d EA model. On top of that, the RFW is not a well studied system and thus the literature results about it are very scarce.

In order to asses the hardness of the model and to compare it against that of the 3d EA one, in what follows we present results of the standard PA simulations of both models, organized in four different lines of argumentation.

Free energy against ground state energy

Let us first take a look at the disorder-averaged free energy as a function of the temperature. This thermodynamic quantity can be theoretically computed as:

$$F = \langle E \rangle - TS, \tag{5.30}$$

where $\langle E \rangle$, F and S are the expected value of the energy, the free energy and the entropy of the system at temperature $T = \beta^{-1}$, respectively. Since at sufficiently low temperatures $\langle E \rangle \approx E_0$ being E_0 the energy of the global ground state of the system, from Eq. (5.30) it is straightforward to see that, for decreasing temperatures, the free energy and the ground-state energy will coincide and thus their ratio will approach 1. Alternatively, a ratio close to 1 serves as a hint that the global ground state has been found and that the model has been properly thermalized at low temperatures. The hardness of two different systems can then be compared by studying how fast this ratio approaches unity with decreasing temperatures, as lower values of the ${}^{E_0}/_{F}$ ratio will indicate a higher presence of entropy S, and therefore a higher difficulty to explore a significant and representative portion of its energy landscape.

When studying the disorder averaged ratio of this fraction for different sizes of EA and RFW lattices with a Population Annealing algorithm utilizing the same set of parameters and the same amount of computational resources one gets the results plotted in Fig.5.8. There, the solid lines represent the averaged quantity and the error bars the standard deviation over different disorder realizations. Their rather small sizes indicate that there is little variability among instances, while the mean value clearly indicates that the RFW model yields quite lower values than the EA one for all of them. Concretely, and in accordance with most of the literature exploring numerical simulations of EA lattices, an inverse temperature of $\beta = 5$ seems to be more than enough to confidently state that PA has found the global ground state for both lattice sizes. On the other hand, when annealing RFW systems towards $\beta = 20$ (and therefore using four times as much computation time, since the used annealing schedule is the same for both models and utilizes equally spaced inverse

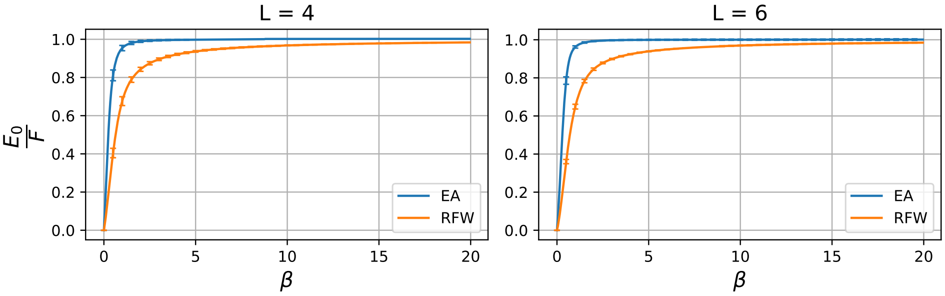

Figure 5.8: **RFW models are computationally harder than their EA counterparts.** Comparison of the ratio between the minimum energy found and the estimated free energy, at a given β during a simulation, for EA and RFW models. Solid lines indicate the average among 100 different disorder realizations and error bars indicate the standard deviation of such disorders over their average. Lattice sizes of $L = 4$ (top panel) and $L = 6$ (bottom panel).

temperatures), PA still does not yield convincing values of the ${}^{E_0}/_{F}$ ratio.

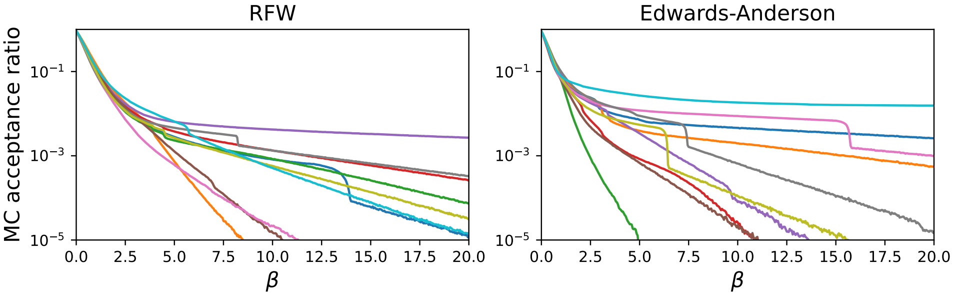

Figure 5.9: **RFW and EA models have similar MC acceptance ratios when solved using the PA algorithm.** Population Annealing's acceptance ratios of MC updates for different disorders of RFW lattices (top panel) and Edwards-Anderson lattices (bottom panel), as a function of inverse temperature β. Each line is the result obtained for a different disorder.

The acceptance ratio of Metropolis steps complements these results and sheds some more light into the differential hardness of the two models. A brief examination of Fig.5.9, in which this parameter is plotted for various disorder instances of both RFW and EA models, quickly indicates that, for both of them, at $\beta = 5$ the majority of proposed spin updates are rejected. This fact is actually more noteworthy for the RFW lattices. Furthermore, we observe that the minimum energy state found during the whole simulation is, in general, the exact same one that has already been encountered at stages previous to $\beta = 5$, see Fig.5.10. This means that, from then on, the only factor improving the results (increasing g_0, but potentially decreasing the family entropy) is the resampling of the population, but not a further exploration of the system's energy landscape. Combining the results of Figs.5.8, 5.9 and 5.10 thus yields the conclusion that the increase in $^{E_0}/_{F}$ with high values of β observed for the RFW lattice model can only be due to the decrease of the TS term in Eq.5.30, which means that this model has much more entropy at lower temperatures. Since, as discussed previously, the wall representation removed the entropy related to the gauge freedom of the underlying RFW model, this entropy only indicates that the phase space of the model at hand is more difficult to explore. We hypothesize that this larger entropy might be related to the non-ordered nature of the "confining phase" of the RFW model versus the spin-glass order in the 3d EA one.

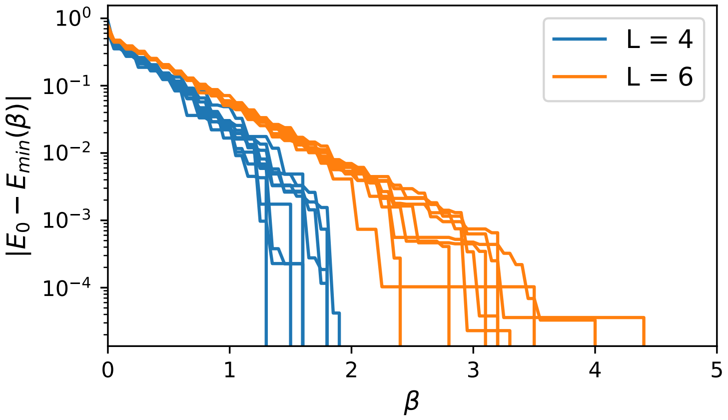

Figure 5.10: **Annealing towards $\beta > 5$ does not yield lower energy states.** For $L = 4$ and $L = 6$ RFW lattices, $\beta = 5$ is a sufficiently low temperature such that no states with lower energy are found beyond this point.

Reduced-frustration disorders

Another way in which the comparably greater hardness of the RFW model emerges is by looking at disorder instances with reduced frustration. To this end, let us consider disorders with coupling strengths drawn from modified Gaussian distributions. In particular, we consider distributions with a varying ratio of negative bonds, by taking their absolute value and then changing the sign of a randomly selected portion of them. Following the intuition behind the embedding process used to solve densely connected QUBO problems into more restrictive topologies, since a positive-valued coupling will tend to align the two spins it connects, a portion of the lattice in which its spins are all coupled by positive bonds will act as a single, big spin at low enough temperatures. An unaligned spin within such cluster would always be clearly energetically unfavourable, and thus the number of potential ground state configurations is effectively reduced. Therefore the lower the ratio of negative bonds that the disorder instance has, the lesser the amount of frustrated spins it can potentially have, and thus the easier it will be to find its minimum energy configuration, eventually reaching the trivial limit of a ferromagnetic disorder. In Fig.5.11 the metrics obtained by a Population Annealing algorithm when solving a large collection of different disorder instances of RFW lattices with a reduced ratio of negative bonds, r_{nb} are shown. For a fair comparison, the same set of parameters was used in all the simulations. In it one can easily appreciate the shifting towards higher values of the histograms of both metrics with the decreasing values of r_{nb}, and how they compare against the metrics that regular EA lattices obtain.

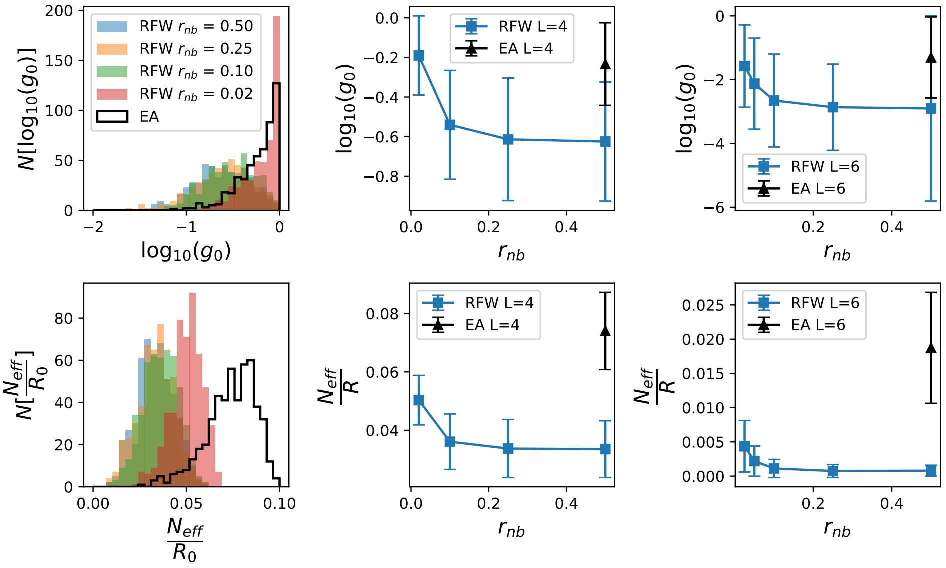

Figure 5.11: **The hardness of RFW models against EA clearly emerges when varying the frustration level of the solved lattices.** Left panels: histograms of g_0 (top) and ${}^{N_{eff}}/_{R_0}$ (bottom) obtained for $L = 4$ RFW lattices of various ratio of negative bonds, r_{nb}, along with the ones obtained for Edwards-Anderson lattices. Center panels: mean values and standard deviations of such histograms, as a function of r_{nb}. Right panels: mean values with standard deviations of the same histograms but for $L = 6$ lattices. All simulations are obtained with PA with the same set of parameters.

We highlight two main consequences from the results shown in Fig.5.11. On the one hand, we confirm that, as discussed earlier, higher values of g_0 and the family entropy are correlated with higher confidence in the solution found being the global ground state and, therefore, can be used as a metric to this end. On the other hand, and as a direct consequence of this first fact, it shows how strikingly harder the RFW model is compared to typical Edwards-Anderson instances with uncorrelated disorders, obtaining comparable values of g_0 only when restricting the frustration of disorders to a ratio of negative bonds of about 2%, while still getting much worse values for the family entropy. To improve it, besides reducing the frustration, greater computational resources should be invested by increasing the adiabaticity of the annealing process. Looking at the same curves for $L = 6$, we observe pretty similar results, bottom panels of Fig.5.11. Although the g_0 obtained for the EA model is approximately equal to the one obtained for RFW with about 2% negative bonds, the family entropy is still far from it even for such low-frustration cases. It is also worth noting that, for $L = 6$, this difference is even greater than for $L = 4$.

Independent measures of g_0 in PA

As previously discussed, PA yields independent measures for $g_0(N_0)$, estimated from the final population of replicas, and for F, estimating it in an iterative process during the simulation. Using Eq. (5.15) with the estimated free energy, the algorithm thus yields two independent estimations of the probability of being in the ground state. Therefore in the limit of infinite replicas these two measures should coincide if the system is really in thermal equilibrium

$$\lim_{R\to\infty} g_0(N_0) = g_0(E_0, F; \beta). \tag{5.31}$$

Equivalently, given a certain amount of computational resources, a larger discrepancy between these two measures of g_0 indicates a poorer thermalization and thus a landscape that is more difficult to explore.

In Fig.5.12, the histograms comparing these two measures for several disorder instances of both the RFW and EA models of $L = 4$ and $L = 6$ are shown. While the measures of g_0 for EA lattices are mostly similar in both considered sizes, RFW lattices yield much poorer results, which are, in all cases, not even comparable by orders of magnitude. Further, the difference between $g_0(N_0)$ and $g_0(E_0, F)$ appears to increase for larger RFW lattices but not for EA ones, hinting that the increase in hardness with size (which, as previously discussed, already demands an exponential growth of the number of replicas by itself) is more dramatic in the former model.

Bigger lattices

Further, we implement $L = 8$ RFW lattices with Gaussian disorder ($r_{\mathrm{nb}} = 0.5$), and attempt to solve them with standard PA using a large amount of computational resources. For equally large EA lattices, the parameter set $(R_0, N_T, N_{\mathrm{sweeps}}) = (5 \cdot 10^5, 200, 10)$ is reported to be *more-than-adequate resources* in [51], where R_0 is the initial population size, N_T is the number of temperatures in the annealing schedule and N_{sweeps} is the number of sweeps carried on per temperature (each sweep consists of attempting N_{spins} MC updates on the lattice). For comparison,

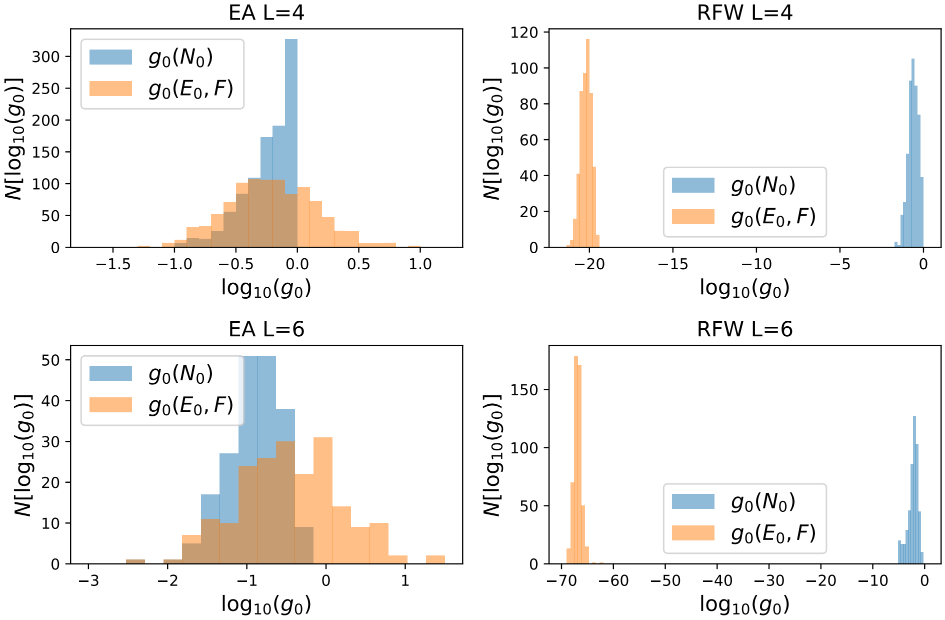

Figure 5.12: **RFW models are computationally harder than their EA counterparts.** Histograms of the values of g_0 obtained with the two discussed independent measures, for various disorder instances of EA (left panels) and WRF (right panels), of sizes $L = 4$ (top panels) and $L = 6$ (bottom panels).

in our simulations we used $(R_0, N_T, N_{\text{sweeps}}) = (2.6 \cdot 10^5, 500, 20)$ and repeatedly solved 20 different disorder realizations in order to see how many times the algorithm converged to the same minimum energy state. Among all of them, the convergence ratios between independent runs varied from a minimum of 0% (in 11 out of the 20 studied disorders) to a maximum of 12% (in only three of them).

5.4 Defect-driven Population Annealing

Let us now combine PA's features with the non-local moves discussed above, in a modification that we refer to as Defect-driven Population Annealing (DPA). To this aim, we relax the constraints imposed on the possible configurations of the system by including the constraints directly into the Hamiltonian through a Lagrange-multiplier-like approach. This effectively expands its configuration space by the addition of non-physical states with topological defects. We nonetheless want this feature to be a possibility only during the early stages of the simulation, as towards the end we require the system to encode the solution to a real problem and therefore to have some physical meaning.

The terms resembling Lagrange multipliers are used to control the presence of defects, and are introduced through an additional field parameter κ into the Hamiltonian, which has the effect of assigning an energy penalty to them. Recalling that at any time the number of defects present within the lattice can be counted with the help of the cube and plane operators, Eqs. (5.27) and (5.28), respectively, the final Hamiltonian reads:

$$H = \sum_w J_w \eta_w^z + \kappa \left[\left(N_c - \sum_c B_c \right) + \left(N_p - \sum_p B_p \right) \right]. \tag{5.32}$$

By annealing κ from 0 to a high value the system acquires a lot of defects at the beginning of the evolution, when they contribute with a minimal energy cost, and gets rid of the non-physical states towards its end, when their presence implies huge energy penalties. The method therefore consists on annealing a population of replicas following a certain schedule on the two parameter space, $\Xi_{\beta,\kappa} = [(\beta_0, \kappa_0), ..., (\beta_{\mathrm{N_T}}, \kappa_{\mathrm{N_T}})]$, with $\beta_i \leq \beta_{i+1}$ and $\kappa_i \leq \kappa_{i+1}$ $\forall i$. As in regular PA, at each point in the schedule a resampling step is carried on the population and then the replicas are set to evolve with an MCMC procedure in order to ensure thermal equilibrium. Note that now the energy of the system is a function of the varying parameter κ:

$$E(\kappa) = \sum_w J_w \eta_w + \kappa N^{(D)}, \tag{5.33}$$

with the number of defects $N^{(D)}$ defined by Eq. (5.29). Since at each resampling step $(\beta, \kappa) \rightarrow (\beta' \kappa')$ both parameters can potentially be updated, the normalized weights are now

$$\tau_j(\beta, \beta', \kappa, \kappa') = R_0 \frac{e^{-[\beta' E_j(\kappa') - \beta E_j(\kappa)]}}{\sum_r e^{-[\beta' E_r(\kappa') - \beta E_r(\kappa)]}}. \tag{5.34}$$

Regarding the MCMC evolution, at each iteration a spin and a wall updates are proposed, e^x_μ (Eq. (5.26)) and η^x_w, respectively, and accepted with the corresponding probability.

5.4.1 Constant-fraction entropy-loss annealing

As previously discussed, a bad thermalization of a given system in Population Annealing is characterized by a low family entropy at the end of the annealing process, which is reduced every time a resampling procedure is carried on the population. In order to maximally preserve it, one then could think that the best strategy consists on loosing a constant portion of surviving families at each step, thus following an exponential decay. On the contrary, when utilizing the standard linear-in-β annealing schedule, widely used in the literature regarding simulations of this type of systems with PA [51, 62, 175, 179], one finds that the effective number of surviving families follows a rather sharper decay, see Fig.5.13. However, the difference between hard and easy disorder cases is manifested in subtle changes in the decay line and, importantly, its end value. Since the DPA algorithm anneals the system in a two-parameter space, (β, κ), it allows for a more flexible exploration. Concretely, in order to maximize the value of the family entropy during the whole simulation and thus its final value as well, DPA adapts the annealing process such that, at each resampling step, a constant fraction of the number of surviving families is lost. This way the family entropy follows an exponential decay towards an objective value, which can be tuned externally. Ultimately, this allows the user to (try to) push the obtained final family entropy to a desired value, which can be higher than the one obtained by PA.

To reach this objective value DPA implements an adaptive step procedure for both annealing schedules, Ξ_β and Ξ_κ. In it, before each resampling step $(\beta, \kappa) \to (\beta', \kappa')$, the optimal values of β' and κ' are computed such that the portion of families lost in the resampling is the closest possible to the one needed in order for it to follow the specified exponential decay

$$\frac{N_{\text{eff}}(t_n)}{R_0} = \left(\frac{N_{\text{eff}}^{\text{desired}}}{R_0}\right)^{\frac{t_n}{N_T}}, \tag{5.35}$$

where t_n is the simulation step, $N_{\text{eff}}(t_n)$ is the number of surviving families at that time step, and $N_{\text{eff}}^{\text{desired}}$ is the objective number of surviving families at the end of the simulation.

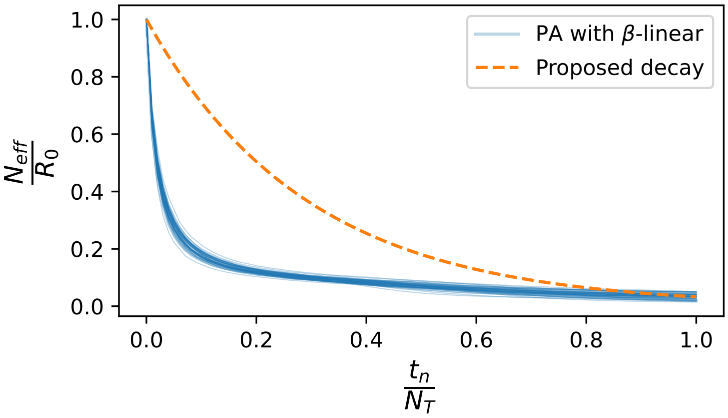

Figure 5.13: **Constant-fraction entropy-loss annealing.** Comparison between the decay in family entropy obtained in PA simulations of $L = 4$ RFW lattices with a β-linear annealing schedule (blue lines) and the objective, implemented decay (orange line). Each blue line is the obtained decay for a different disorder. Although final family entropy may seem very similar in this scale, its fluctuations in PA simulations describe differences between hard and easy disorder cases. The implemented annealing schedule tries to loose a constant portion of families at each simulation step, while the classic one loses a larger amount at the beginning.

5.4.2 Remaining defects

As per any spin glass model, the energy landscape of the Wall Lattice is rough and has an enormous number of difficult to overcome energy barriers. In the example at hand, nevertheless, this fact can have some additional consequences, as the configuration space has been enlarged by the introduction of the already discussed Toric Code topological defects.

For low values of the field parameter κ, the defects start to appear and proliferate, and since their number is not hardly penalized by the second term in the Hamiltonian, Eq. (5.32), the system is allowed to explore non-physical configurations with energies that can be potentially lower than that corresponding to the physical ground state. Then, even if a later high value of κ penalizes the existence of defects within the lattice, it does not guarantee their total elimination but, on the other hand, it is indeed possible that an adiabatic annealing through the (β, κ)-space yields configurations in which some defects have got stuck within deep energy wells. If this happens at low temperatures and no other adjacent sites have defects, an increase in κ will not be able to move and eliminate them by means of regular Metropolis dynamics. As discussed earlier, these non-physical states do not have an unequivocal translation back to spin variables, and as we are interested in real life QUBO problems they are therefore not an acceptable outcome of the simulation.

In order to address this issue and get rid of the defects that may end up in the final stages of the simulations, we make use of two different strategies. On the one hand, we apply a kinetic Monte Carlo algorithm, namely an N-fold way algorithm, that speeds up their dynamics. On the other hand, we also include an attractive potential energy term between defects that makes them get closer to each other and thus collide and annihilate.

Review of the N-fold way algorithm

At very low temperatures, the acceptance probability of any move approaches zero and therefore the Metropolis algorithm becomes very inefficient at accepting new states. This is of course in accordance with the idea that, at that point, the current state should be very similar to the global minimum and thus that no sub-optimal configurations should be needed in order to overcome energy traps.

The N-fold way algorithm, first proposed by Bortz, Kalos and Lewovitz [180], is a kinetic Monte Carlo strategy that speeds up the low temperature dynamics of the simulated system by realizing about the fact that there is in reality a small number of different classes of possible updates. Consider first a one dimensional Ising spin system with ferromagnetic couplings, no external magnetic fields and periodic boundary conditions. As in the Metropolis algorithm, pick a spin to be updated at random. Whatever the configuration of the system at a given time of the simulation is, there is only three possible updating energies depending on the state of the two neighbouring spins with respect to the picked one: either none, one or two spins are initially aligned with it, and thus the flip of the picked spin will change the energy of the system with $\Delta E_c = \{-2J, 0, +2J\}$, respectively. The N-fold way algorithm then proceeds as:

- Classify all spins according to their corresponding updating energy
- assign to each of the identified classes a probability $P_c = e^{-\beta \Delta E_c}$
- Select a class with probability $P_c = \frac{P_c}{\sum_c P_c}$
- randomly update a spin within the selected class

Note that, since the acceptance probability is now always 1, the evolution of the system is faster. Just as in the Metropolis algorithm, those spins whose flip is more energetically favourable will have a higher probability of being updated, and the difference resides in how the time a spin remains static is computed. In reality, we have now in some sense computed the expected value of the time that a spin pertaining to any of the different classes will remain in its current state at temperature β, and thus one has to consider time steps of length $\tau = P_c$ in order to track the thermal dynamics of the system.

N-fold way in the RFW model and DPA

When dealing with more complex systems (such as those having higher dimensionality or heterogeneous connectivity and coupling strengths), the couplings between spins will not be the same and therefore the number of different classes will increase. Then the computations that must be carried on in order to define the probability for each spin to be selected accordingly grows, and hence it is possible that the kinetic speed up does not pay off when considering the computational cost. In the case of Gaussian distributed disorders as those considered in the Edwards-Anderson or RFW models, the strengths of the couplings are drawn from an infinite set of possible values. One therefore ends up having a different class for each single spin, which is the reason why a plain Metropolis algorithm is normally used in spin glass systems, while the N-fold way strategy is discarded.

Nevertheless, as we have discussed before, when dealing with the Wall Lattice model we sometimes encounter some disorders for which the defects do not get completely annihilated by sole Metropolis dynamics. In such cases it is mandatory to reactivate the defects' mobility in order to get rid of them. To this end we use a slightly modified N-fold way algorithm that only considers the possibility to update those walls surrounding the already present defects. This way, and even if moving any of the present defects is energetically non-favourable and thus their movement would be discarded if the creation of new defects was considered, we force them to move while keeping the number of different classes (and therefore the computational cost) relatively low.

Attractive potential

Forcing the movement of the remaining defects is still not a synonym of completely getting rid of them. In low dimensional (1d, 2d) systems the annihilation of all of them can eventually happen for relatively acceptable time scales, but this is by no means the case in three dimensions. Contrarily, in three- and higher- dimensional systems (which are the ones presenting hard instances and thus those we are interested in) it is much more probable that, in big enough sizes, forcing the defects to move will only make them travel through the lattice with a ridiculously little chance of encountering a partner with which to collide against and to get annihilated with.

To avoid this free particle behaviour of the defects and to reward their convergence and thus annihilation, an attractive potential between them is considered. This can be done by adding a gravitational term into the Hamiltonian describing the RFW system, Eq. (5.32):

$$H_G = G \sum_{i=1}^{N_{\text{def}}} \sum_{j=0}^{i-1} \frac{1}{D_{ij}^{\alpha}}, \tag{5.36}$$

where G is a negative constant that can be tuned, N_{def} is the number of defects present in the lattice, D_{ij} is the euclidean distance between defects at sites i and j, and α can be used to make the interaction potential shorter or longer ranged.

It is important to note that this term has its owns caveats and that has to be used wisely and with precaution. First of all, one should note that a larger number of particles within a given gravitational system implies a bigger potential energy. This can be directly translated into the non-desired fact that, besides the accretion of the already existing defects, the creation of additional ones, and not their annihilation, will be instigated by this term. It can indeed be seen that, if the attractive potential is turn on during early stages of the simulation at which the existence of defects is not heavily enough penalized yet, for low values of the ratio $\frac{\kappa}{G}$ the system undergoes a phase transition in which the lattice gets filled up with defects until their number reaches the maximum allowed by the system, $N_c^{(D)} = N_c$. On the other hand, if this potential term effectively penalizes their annihilation, they will not end up colliding even if they get side by side.

As a consequence it is important that the attractive potential term is only turned on under two circumstances:

- The single bond-flip operator, μ_w^x, works under the modified n-fold way dynamics described earlier, such that only those walls surrounding defected sites are allowed to move and then no new defects are created.

- The ratio between the field parameter and the gravitational constant, $\frac{\kappa}{G}$, is set to a high enough value such that the contribution of the two last terms of the Hamiltonian (the one penalizing the existence of defects and the one getting them close to each other) to the total energy of a proposed movement effectively foments the annihilation of defects.

5.4.3 Free energy estimation in DPA

Let us now give some hints on the derivation of the recipe for the free energy estimation in DPA. When resampling from (β, κ) to (β', κ'), the expected number of descendants in DPA is given by Eq.5.34, with the normalization factor

$$Q(\beta, \beta', \kappa, \kappa') = \frac{1}{R_{\beta,\kappa}} \sum_r e^{-\beta' E_r(\kappa') + \beta E_r(\kappa)}, \tag{5.37}$$

and the energy at a given κ being:

$$E_r(\kappa) = E_r^w + \kappa N_r^{def}, \tag{5.38}$$

where we have defined $E_r^w \equiv \sum_i J_i w_i$, the energy due to the wall configuration, and N_r^{def} is the number of defects within the lattice.

As we did in PA, now the ratio between partition functions is:

$$\begin{aligned}\frac{Z(\beta',\kappa')}{Z(\beta,\kappa)} &= \frac{1}{Z(\beta,\kappa)}\sum_{\gamma} e^{-\beta' E_\gamma(\kappa')} = \\ &= \frac{1}{Z(\beta,\kappa)}\sum_{\gamma} e^{-\beta E_\gamma(\kappa)} e^{-\beta' E_\gamma(\kappa')+\beta E_\gamma(\kappa)} = \\ &= \langle e^{-\beta' E(\kappa')+\beta E(\kappa)}\rangle_{\beta,\kappa}, \end{aligned} \tag{5.39}$$

so that we approximate:

$$\langle e^{-\beta' E(\kappa')+\beta E(\kappa)}\rangle \approx Q(\beta,\beta',\kappa,\kappa'). \tag{5.40}$$

And everything else follows as in PA.

Note that, by allowing single bond (wall) flips and thus the introduction of topological defects, we are now effectively dealing with a different physical system. This implies that a different value of the free energy should be expected in the (β,κ) region of the parameter space in which the expected number of defects present in the lattices is non zero.

On the other hand, the total number of possible configurations is greatly increased. In order for the ensemble mean approximation in Eq. (5.40) to hold, a sufficiently significant portion of the configuration space must be represented by the replica population. Consequently, a much greater population size is required as well if good estimations are desired.

In conclusion, PA should always be preferred over DPA in terms of free energy estimation.

5.4.4 Outline of the algorithm

To conclude, in this section we present a pseudo-code of the Defect-driven Population Annealing method, see Alg. 1.

As in regular PA, the algorithm starts by randomly initializing the configurations of each replica in the population. The annealing schedules, Ξ_β and Ξ_κ, last a total of N_β steps, and then for each annealing step the algorithm estimates the next point in the schedules, (β',κ') such that the family entropy obtained from the resampling follows the proposed exponential decay. After resampling, each replica r is thermalized with a total number of sweeps N_{sweeps} at (β',κ'). To this end, a number of N_{spins} spin updates and wall updates are proposed and accepted with a Metropolis algorithm. In case the simulation is reaching its end ($i_\beta \geq i_\beta^{(\text{critic})}$), and a given lattice still has sites containing topological defects, the dynamics are accelerated with an NFoldWay algorithm instead of using a regular Metropolis one, such that no more defects can be created but only moved around the lattice (with the additional gravitational term, Eq.(5.36)) so that their probability of closing a path and annihilating

Algorithm 1 Defect-driven Population Annealing

```
for each replica r do
    r ← initialize
for $\mathrm{i}_\beta$ in $\mathrm{N}_\beta$ do
    $\mathrm{N}_{\mathrm{eff}}^{(\mathrm{desired})} \leftarrow \mathrm{ProposedDecay}[\mathrm{i}_\beta]$
    Estimate $(\beta', \kappa')$ such that $\mathrm{N}_{\mathrm{eff}}[\mathrm{i}_\beta] \approx \mathrm{N}_{\mathrm{eff}}^{(\mathrm{desired})}$
    $\mathrm{Resample}(\beta, \beta', \kappa, \kappa')$
    for each replica r do
        for sweep in $\mathrm{N}_{\mathrm{sweeps}}$ do
            for n in $\mathrm{N}_{\mathrm{spins}}$ do
                $\mathrm{s} \leftarrow \mathrm{U}(0, \mathrm{N}_{\mathrm{spins}})$
                $\mathrm{Metropolis}(\mathrm{r}, \mathrm{s}, \beta', \kappa')$
                if $\mathrm{i}_\beta < \mathrm{i}_\beta^{(\mathrm{critic})}$ then
                    $\mathrm{w} \leftarrow \mathrm{U}(0, \mathrm{N}_{\mathrm{walls}})$
                    Metropolis(r, w, $\beta'$, $\kappa'$)
                else
                    if $\mathrm{N}_{\mathrm{defects}}(\mathrm{r}) \neq 0$ then
                        $\mathrm{NFoldWay}(\mathrm{r}, \beta', \kappa')$
                    else
                        $\mathrm{w} \leftarrow \mathrm{U}(0, \mathrm{N}_{\mathrm{walls}})$
                        Metropolis(r, w, $\beta'$, $\kappa'$)
    $\beta \leftarrow \beta'$
    $\kappa \leftarrow \kappa'$
```

increases. After all replicas are thermalized, the next annealing steps starts until the schedules reach its maximum length N_β.

5.5 Comparison between PA and DPA

An important factor to be taken into account when comparing different algorithms performing the same task is to make sure that an equivalent amount of computational effort is invested in all of them. In the case at hand, we can define the computational work W as the total number of MCMC updates attempted during a simulation, namely

$$W = N_{\text{sweeps}} \sum_{i=0}^{N_T} R_i \approx N_{\text{sweeps}} N_T R_0. \tag{5.41}$$

A sweep on PA consists on proposing N_{spins} flips on randomly chosen spins, such that on average all spins within the lattice are given the chance to be updated at each sweep. N_T is the number of temperatures in the annealing schedule and R_i is the population size at annealing step i. The last approximation in Eq.5.41 follows from the fact that, in our concrete implementation, a resampling protocol that minimizes the fluctuations in population size is used. The computational work due to the resampling steps is neglected. One of the most computationally-intensive workloads in Monte Carlo simulations is the generation of (pseudo-)random numbers. Thus we note that Eq. (5.41) can be split into two main contributions: the generation of random numbers and the rest of the computations needed to be carried on in the updating scheme.

Recall that every time PA proposes one spin to be updated, DPA proposes one spin and one wall for such task, therefore potentially requiring twice as many random numbers if the same number of sweeps is used in both algorithms. On the other hand, as discussed earlier, a spin flip is equivalent to four wall flips and thus a DPA sweep needs to do five fourths as many computations than a PA one. In order to invest an equivalent computational work in both algorithms, Eq. (5.41) can therefore be considered from two different perspectives:

- **Scenario I.** The generation of random numbers is the bottleneck of the process. The computational work can therefore be counted as the amount of random numbers consumed by the algorithm, and a fair comparison between PA and DPA is given if $N_{\text{sweeps}}^{(\text{PA})} = 2N_{\text{sweeps}}^{(\text{DPA})}$.
- **Scenario II.** The computations needed to process the flips are the bottleneck of the whole process. The computational work can therefore be counted as the amount of walls flipped, and a fair comparison between PA and DPA is given if $N_{\text{sweeps}}^{(\text{PA})} = \frac{5}{4}N_{\text{sweeps}}^{(\text{DPA})}$.

While the most usually encountered scenario corresponds to the generation of random numbers being the bottleneck, this is not always the case. In fact, fast random numbers generators are a hot topic in current technology development. In what follows we therefore study the comparative performance of PA and DPA under both assumptions. Concretely, we address and compare both scenarios using $N_{\text{sweeps}}^{(\text{PA})} = 10$ and $N_{\text{sweeps}}^{(\text{DPA})} = 5$ for the first and $N_{\text{sweeps}}^{(\text{PA})} = 10$ and $N_{\text{sweeps}}^{(\text{DPA})} = 8$ for the second.

5.5.1 $L = 4$ lattices

We first apply the DPA method to the study of $L = 4$ RFW lattices and take a look at the histograms of the parameters g_0 and N_{eff} obtained when solving several disorders, comparing the solutions with those that a normal PA algorithm yields. The top panels of Fig. 5.14 show the obtained histograms when considering that the computational bottleneck in the simulations is the generation of the random numbers (first discussed scenario), while the bottom panels show those same histograms when a fast enough random number generator is assumed (second scenario). In each row, the leftmost panels show the histograms for g_0 and the center ones the histograms for N_{eff}. Finally, the rightmost panels relate, for each solved disorder, the values of the two discussed metrics. In these last panels the disorders that PA finds the hardest ones to solve are also marked with red crosses, as well as how these same disorders score when solved by DPA, with black crosses. The hard disorders are identified as those obtaining the lowest values of family entropy at the end of a PA simulation. In a somewhat arbitrary way, we set the limit of hard disorders on PA not being able to obtain a final effective number of surviving families of $N_{\mathrm{eff}} = 0.015 R_0$, which is equivalent to about 2% of the total amount of considered instances.

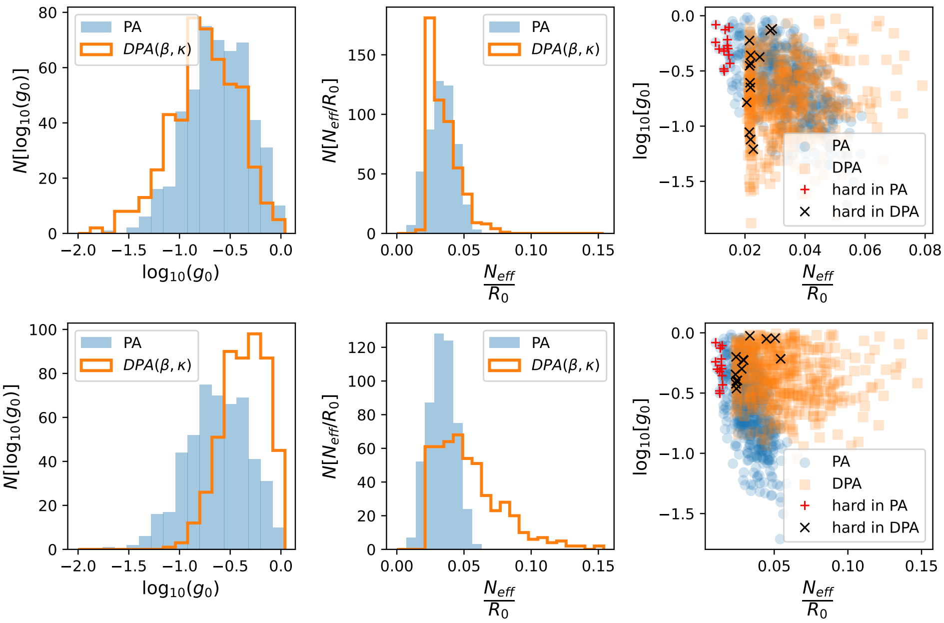

Figure 5.14: **DPA outperforms PA for L=4 lattices when MCMC process simulations are the computational bottleneck.** Histograms of g_0 and N_{eff}/R obtained by solving, by both PA and DPA, a total of $N_{\mathrm{dis}} = 500$ different disorders of $L = 4$ RFW lattices, with bonds randomly distributed according to a normal distribution. Simulations using the same amount of random numbers (top panels) and of MCMC updates (bottom panels). Crosses in the rightmost panels mark these instances PA finds the hardest (red), and how these same instances score in DPA (black).

In the first scenario (Fig. 5.14, top panels) we observe a certain trade-off between parameters (Fig. 5.14, top left and top center panels), resulting in an equivalent performance of the two methods. The fact that the adaptive steps procedure over the two control parameter space (β, κ) is able to properly drive the evolution of the system and set an objective cutoff family entropy is nevertheless noticeable (Fig. 5.14, top center panel). Also, the fact that the the family entropy is no worse than for PA is relevant, taking into account that, in this scenario, the total amount of spin updates is lower and thus one might expect a worse thermalization. With the cut-off imposed on family entropy, scores on this parameter are increased even for hard cases, but they seem to lay, generally, in the lower range for DPA as well (Fig. 5.14, top right panel).

On the other hand, in the second scenario (Fig. 5.14, bottom panels) we see a substantial improvement with the DPA method, as the histograms for both measured parameters seem to be shifted towards bigger values and thus imply better metrics than those obtained with standard PA (Fig. 5.14, bottom left and bottom center panels). As can be seen in the bottom right panel, even some of the hard cases' metrics are improved, as they manage to escape from the lower range of family entropy without diminishing its g_0 score.

In Fig. 5.15 we take a closer look at the results obtained by the two methods, by comparing the measured parameters obtained by both for each of the solved disorder realizations. A brighter color indicates a higher density of points and the red straight line is plotted to indicate the region in which both methods yield the same results for a given parameter. We also marked those disorder instances classically labeled as hard for PA with red crosses. Again, the top and bottom panels correspond to the first and second scenarios, respectively.

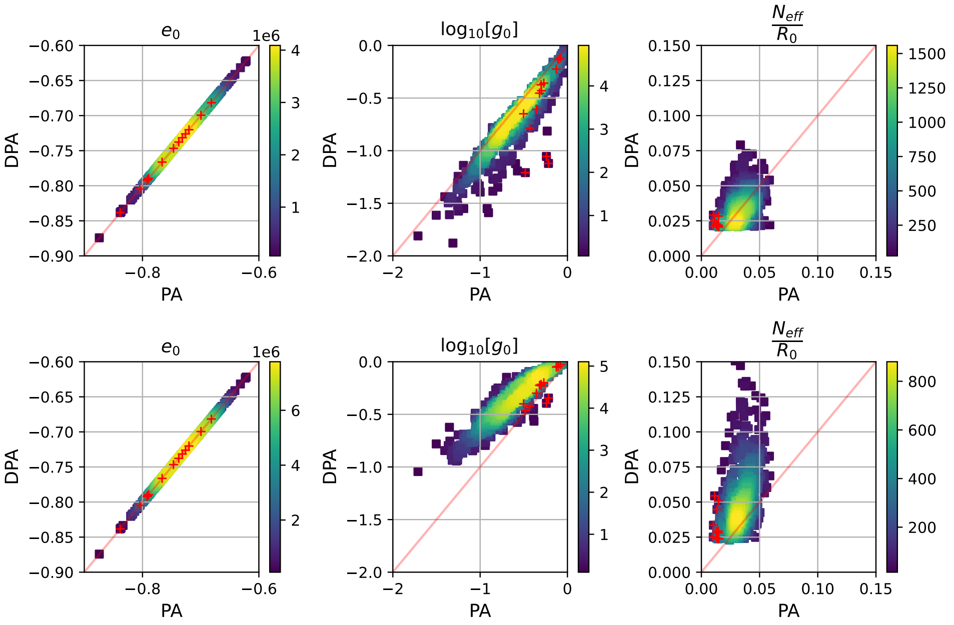

Figure 5.15: **DPA provides on average higher family entropies than PA when solving $L = 4$ lattices.** Comparison of results obtained by PA and DPA for $L = 4$ RFW lattices for the three measured parameters. From left to right: ground state energy, g_0 and effective number of surviving families. Simulations using the same amount of random numbers (top) and of MCMC updates (bottom). Red crosses mark these instances that PA finds the hardest.

Lastly, let us focus on those cases of disorder realizations that PA finds the hardest to solve. As previously discussed (recall Eq. (5.1)), a possible way to overcome the difficulty of equilibration and improving the final family entropy obtained by PA is to reduce the differences $\Delta\beta = \beta' - \beta$ by increasing the number of temperature steps within the same temperature range, thus effectively realizing a more adiabatic simulation. Following this idea, to further contrast both methods, we now take a collection of disorders that PA finds difficult to equilibrate and for which a very similar g_0 is obtained for PA and DPA, so that they lay close to the $g_0^{PA} = g_0^{DPA}$ red line in the center panels of Fig. 5.15, and repeatedly solve them several times with an increasing number of temperature steps in the PA annealing schedule (*i.e.* more adiabatically). For the cases at hand, we find that for PA to obtain a family entropy similar to that obtained by DPA, we would have to invest approximately between 4 and 16 times as much computational power depending on the disorder instance, Fig. 5.16. Furthermore, one should also note here that, contrary to the addition of more replicas, that extra computational effort would not be parallelized in PA, since the annealing schedule must always be followed sequentially.

5.5.2 $L = 6$ lattices

We now apply the same study to $L = 6$ RFW lattices. As for the $L = 4$ case, we solve several different Gaussian disorders with both PA and DPA considering the two scenarios discussed above, and plot the obtained histograms in Fig. 5.17. For the first scenario, (Fig. 5.17, top panels), we again see a certain trade-off between g_0 and N_{eff}. Nevertheless, the gain in family entropy is now more noticeable than for smaller lattices, as now PA does encounter some hard instances for which it obtains a single surviving family, preventing adequate thermalization. On the contrary, the entropy-preserving adaptive steps in DPA are capable of driving the population toward nonzero values, thus ensuring proper thermalization (Fig. 5.17 top center panel). When considering the computational work limited by the number of MC updates, the g_0 obtained with DPA is comparable to that obtained with PA (Fig. 5.17 top left panel), while the gain in family entropy remains the same (Fig. 5.17 top left panel). Again, those hard instances of PA are properly thermalized under DPA. On top of that, comparing the results obtained for each individual disorder between the two scenarios (Fig. 5.17 left panels), we see that, while for PA the hard instances are more or less evenly distributed along the range of g_0 values, they are shifted towards bigger values, especially for the second scenario.

Looking at the right panels of Fig. 5.18 we again confirm that both methods find the same ground-state energy for all disorder realizations. Speaking of g_0, the hardest instances are more or less uniformly distributed around their mean in both scenarios (Fig. 5.18 center panels). They are slightly shifted towards PA when restricted by the generation of random numbers and centered between the two methods when restricted by the amount of MC updates.

We finally study in Fig. 5.19, as in the case $L = 4$, a collection of instances that

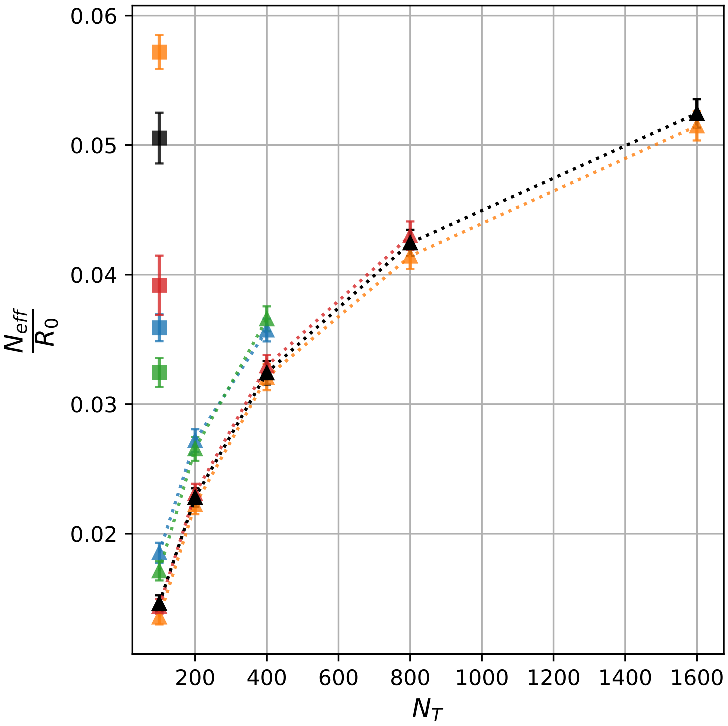

Figure 5.16: **Hard cases on PA require more adiabatic annealing processes to reach an equivalent family entropy to DPA.** Study of the amount of computational power that would have to be invested to solve $L = 4$ RFW lattices with PA, in order to obtain a similar final family entropy as with DPA, for hard instance for which $g_0^{PA} \approx g_0^{DPA}$. Error bars are standard deviations over $N_{\text{rep}} = 100$ repetitions of the same process at each value of N_T. Square markers are obtained with DPA and triangle ones with PA using an increasing number of temperature steps. Each color indicates a different disorder.

PA finds hard and that obtain a similar value of g_0 with both methods. Again, we use an increasingly adiabatic process in PA to see how much more computational power would be necessary to get results comparable to those obtained with PA. In this case, DPA achieves an equivalent performance to PA with about between 2 and 5 more computational investment, depending on the instance. We note a more linear behaviour of the family entropy with N_T than in Fig. 5.19, probably because the considered values of N_T are not big enough. This indicates that $L = 6$ is on the limit of our numerical capabilities.

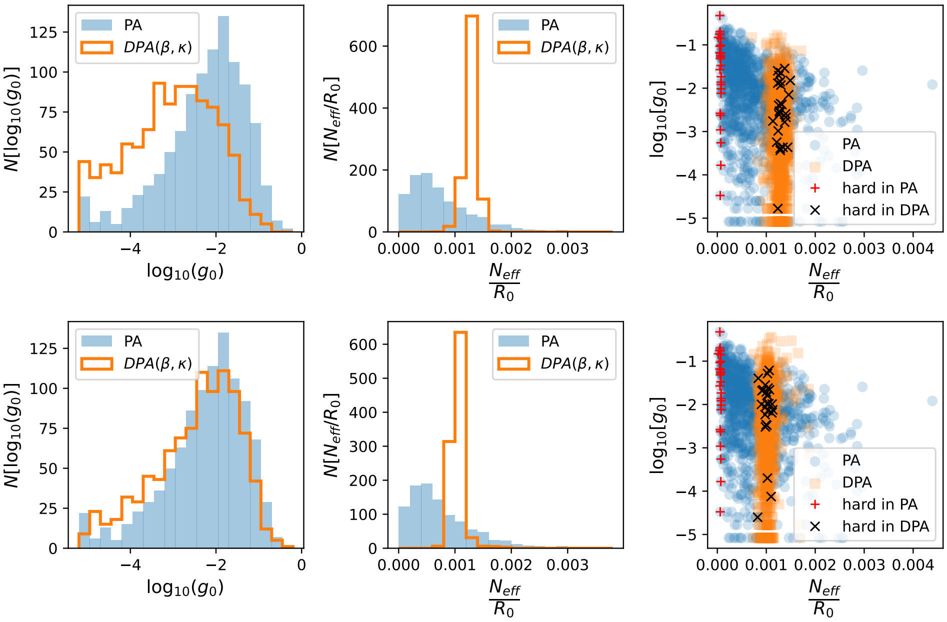

Figure 5.17: **DPA enables thermalization of $L = 6$ lattices, even when these are hard for the PA algorithm.** Histograms of g_0 and $N_{\rm eff}/R$ obtained by solving, by both PA and DPA, a total of $N_{\rm dis} = 1000$ different disorders of $L = 6$ RFW lattices, with bonds randomly distributed according to a normal distribution. Simulations using the same amount of random numbers (left) and of MCMC updates (right). Crosses in the bottom panels mark these instances PA finds the hardest (red), and how these same instances score in DPA (orange).

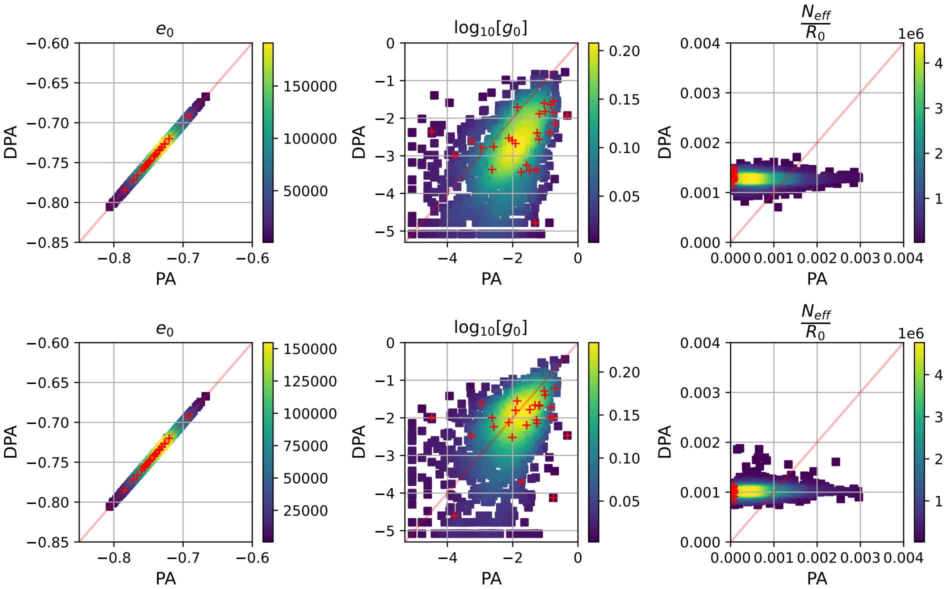

Figure 5.18: **DPA provides on average higher family entropies than PA when solving $L = 6$ lattices.** Comparison of results obtained by PA and DPA for $L = 6$ RFW lattices for the three measured parameters. From top to bottom: ground state energy, g_0 and effective number of surviving families. Simulations using the same amount of numbers (left) and of MCMC updates (right). Black crosses mark these instances that PA finds the hardest.

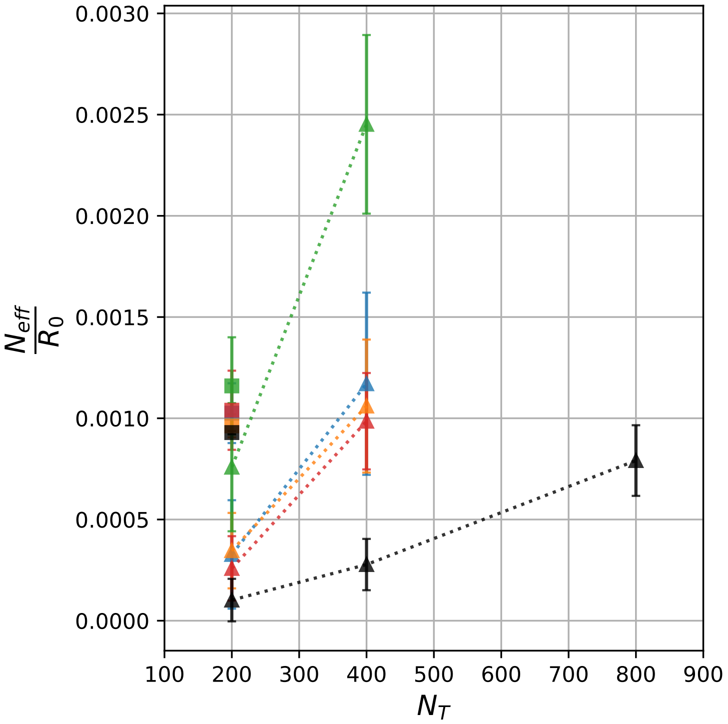

Figure 5.19: **Hard cases on PA require more adiabatic annealing processes to reach an equivaent family entropy to DPA.** Study of the amount of computational power that would have to be invested to solve $L = 6$ RFW lattices with PA, in order to obtain a similar final family entropy as with DPA, for hard instances for which $g_0^{PA} \approx g_0^{DPA}$. Error bars are standard deviations over $N_{rep} = 100$ repetitions of the same process at each value of N_T. Square markers are obtained with DPA and triangle ones with PA using an increasing number of temperature steps. Each color indicates a different disorder.

5.5.3 $L = 8$ lattices

In order to study hard EA disorders on big lattice sizes for which a proper thermalization is not guaranteed, one classically relies on running many times the same instance independently. We apply this methodology to $L = 8$ RFW lattices with Gaussian disorder to test how both algorithms perform. Concretely, we study 20 different disorder instances and run each of them 50 times with each algorithm, considering the computational work limited by the amount of MC updates (second discussed scenario in previous sections). As the systems are not thermalized, we only focus here on the minimum energies per spin found during the simulations, e_{min}, and not on the previously discussed metrics. Contrarily to what should be expected from a thermalized system, none of the algorithms clearly converges to the same energy among different runs for a given disorder. In fact, when studying it with PA, for 11 out of the 20 studied cases the same minimum energy was not found in any of the runs. For 4 of them the minimum energy was found twice, for 2 of them four times and only for the 3 easiest ones it was repeatedly found six times. We note that, under this metric, the ratio of disorders labelled as hard is greatly increased in comparison to smaller lattice sizes. In Fig. 5.20 we show, for one of the three easy studied instances (top panel) and for one of the 11 hard ones (bottom panel), the minimum energies found among independent runs by each algorithm.

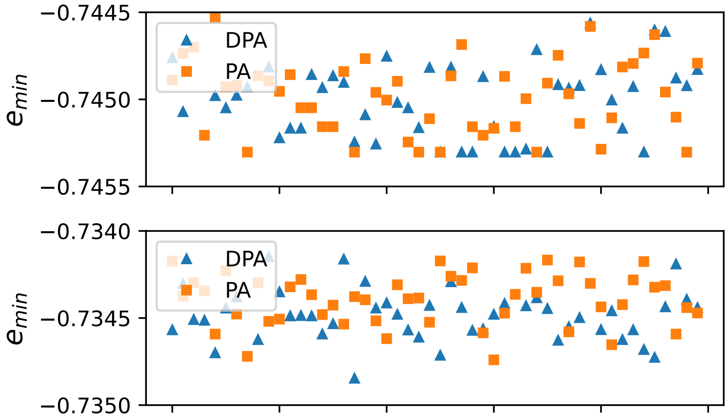

Figure 5.20: **Both algorithms have a low convergence ratio to the same minimum energy state among independent runs.** Upper panel: example of an easy case, for which both algorithms find the same solution after running it several times (both PA and DPA show a convergence ratio of 12%). Bottom panel: example of a hard case, for which both algorithms completely fail to converge towards any minimum energy state, and for which DPA finds the best solution between the two.

On the one hand, we observe that the spread of the minimum energies found among several runs of the same disorder tends to be larger with DPA than with PA. To illustrate it we compute, for each disorder, the median absolute deviation of the minimum energies achieved over the runs. In Fig.5.21 the box plots of these medians obtained with both algorithms are shown. Indicated by a smaller box, clearly all the medians obtained by PA are closer to zero than those obtained by DPA. This noticeably larger spread is probably caused by the larger configuration space due to the introduction of the topological defects. On the other hand, when comparing the minimum energy found with both algorithms, we observe that DPA generally performs better, yielding lower energies (see Fig. 5.22, the "tie" cases contain easy disorders). The fact that DPA generally finds states with lower energies when solving hard instances for which thermalization is poor (or, equivalently, for which no convergence to the same solution is achieved) indicates that the proposed non-local moves

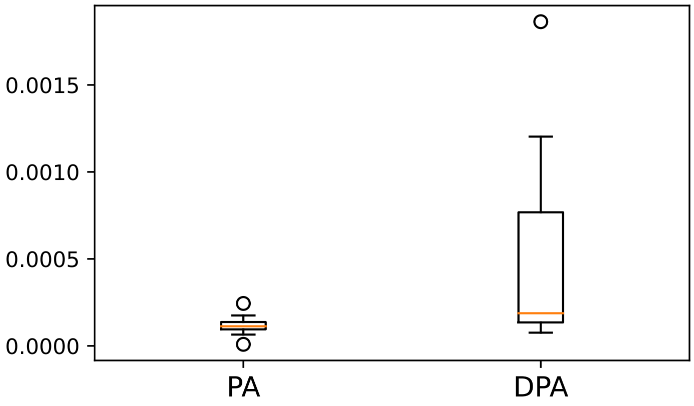

Figure 5.21: **DPA yields a larger spread of minimum energies.** Box plots of the median absolute deviations of minimum energies achieved by PA and DPA over many runs of the 20 studied $L = 8$ RFW disorder instances.

are indeed useful to explore energy landscape in hard disorder instances, effectively allowing the system escape local minima.

5.5.4 Constant-fraction entropy-loss annealing in PA

To further test the effectivity of the newly introduced non-local moves, we implement the family constant-fraction entropy-loss annealing in PA. We use the same parameters used in the simulations of previous sections and find the results shown in Fig. 5.23 for $L = 4$ (top panels) and $L = 6$ (bottom panels). While the adaptive steps are again able to generally impose a cut-off value on the final family entropy, the obtained g_0 is quite worse than in regular PA. For $L = 6$ lattices the cut-off is actually not achieved for some disorders, some of which are even seen as easy (obtaining a high value of N_{eff}) by PA. This fact suggests that different annealing schedules may make PA find hard or easy different instances. More noticeably, the constant-fraction entropy-loss annealing makes PA fail to find the same ground state than when using the standard annealing schedule in some $L = 6$ cases. We therefore conclude that the constant-fraction entropy-loss annealing is not effective by itself when restricted to a one dimensional annealing space, and thus the non-local moves must be the reason why DPA is able to improve thermalization of hard instances, as seen in previous sections.

5.6 Conclusions

In this chapter a quantum-inspired modification of Population Annealing has been presented and tested in the exploration of rugged energy landscapes of 3d random pla-

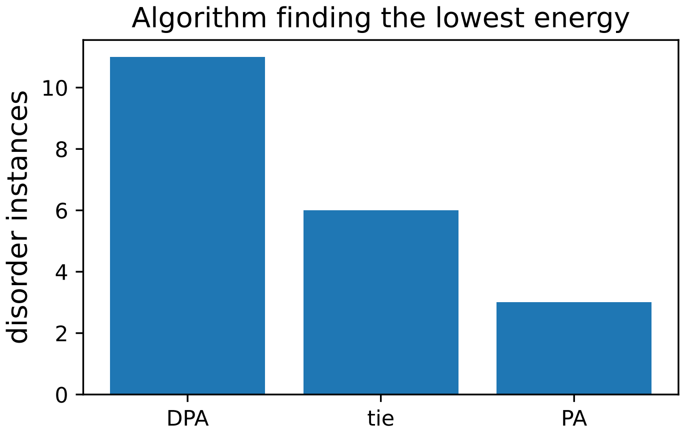

Figure 5.22: **DPA finds lower energy states.** When studying non-thermalized RFW $L = 8$ lattices, DPA is generally able to find lower energies than PA using the same computational resources (in the same amount of MCMC updates scenario). Results obtained after solving 20 different disorder instances, 50 times each with each algorithm.

quette gauge models. The new method, which is inspired on the Toric Code, consists on performing non-local updates by creating, moving and annihilating topological defects within the lattices. Moreover, in 2d Ising square lattices these new moves are equivalent to non-local cluster updates. The defects' dynamics are controlled by an extra field parameter in the Hamiltonian, κ, through which the algorithm anneals the systems along with temperature. A constant-fraction entropy-loss adaptive annealing procedure has been devised to effectively navigate the (β, κ)-space in order to drive the replica population towards an objective value, thus ensuring a better thermalization.

The advantages of the new method, which we refer to as Defect-driven Population Annealing (DPA), are:

- The non-local moves allow the system to overcome high-energy barriers and thus to escape local minima.
- The defects' dynamics are still compatible with massively parallel implementations.
- The annealing in the two-parameter space (β, κ) offers more flexibility, and the starting temperature can now be arbitrary.

On the other hand, the disadvantages are:

- the phase space is substantially enlarged at the early stages of the simulation, and thus the estimation of the free energy is poorer.

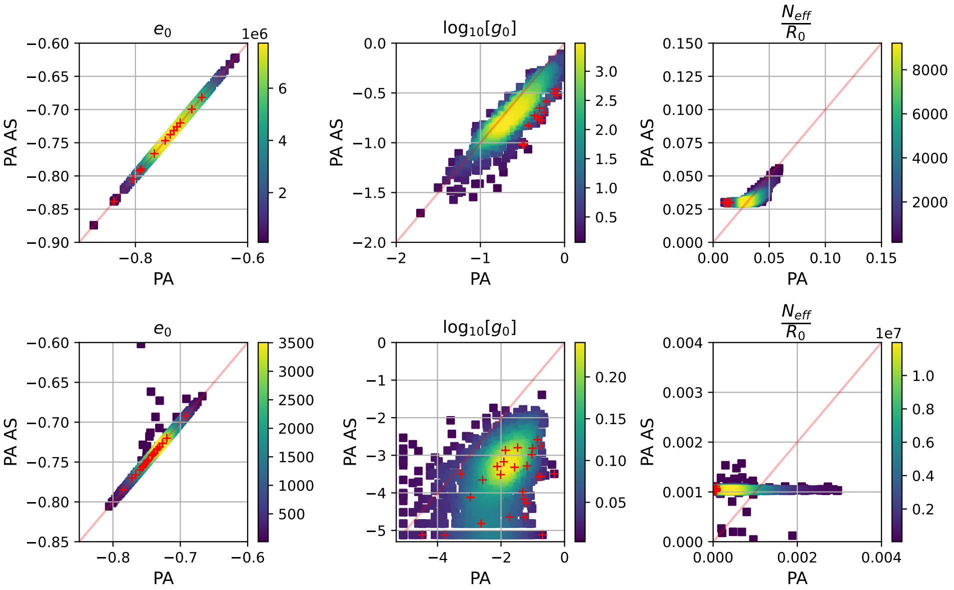

Figure 5.23: **The entropy-preserving steps are not optimal in standard PA.** For $L = 4$ (top panels) and $L = 6$ (bottom panels) RFW lattices the adaptive steps procedure is not effective when the parameter space is reduced to only temperature. Red crosses mark the disorders that normal PA finds the hardest to solve.

The application of the developed method to small RFW lattices shows an improved thermalization quality, which enhances the confidence on the solution found being the global minimum. Moreover, when studying hard cases for which, with Population Annealing, thermalization is not properly achieved nor the global ground state is reached, the proposed non-local moves still show an advantage in overcoming energy barriers and thus yield lower-energy states. It is also worth noting that the results are greatly improved when fast enough random number generators are used, such that the Monte Carlo updates themselves constitute the real bottleneck of the simulations and thus a fair measure of the computational work invested in them.

To conclude, it is important to mention that specialized annealing schedules have been recently proposed and successfully applied in Population Annealing [172, 175, 181]. The merging of these more optimal schemes with the exploration of the two-dimensional annealing parameter space leaves room for improvement. This, which is left as an outlook of the present work, could potentially further increase the spin-glass solving capabilities of DPA, therefore allowing for the exploration of bigger lattices.

Conclusions

The present thesis explores various ways to improve the Monte Carlo-based strategies for solving optimization problems, which consist in simulating complex physical systems such as spin glasses. To this end, three main lines of research have been exploited, each of them intended to improve the understanding and functionality of different parts of such optimization strategies.

The first one is intended to study the spin glass phase transition of the Edwards-Anderson (EA) model in two, three, and four dimensions, by means of the Haake-Lewenstein-Wilkens (HLW) approach. The Saddle-point/steepest-descend (SPSD) solutions to the system are found to be very precise, and the study of their sensitivity to boundary conditions indicates a phase transition in all the considered dimensions. While this is in accordance with the commonly accepted belief in 3d and 4d, it is not in the 2d case, in which no phase transition is expected. Some reasons that may lead to this unfortunate result are identified and left as an outlook.

The second line of research focuses on studying how the hidden correlations present in the pseudo-random numbers used in Monte Carlo simulations can affect the quality of the obtained solutions. The study reveals that poor pseudo-random number generators can have a severe impact on the calculation of the dynamic critical exponent of the two-dimensional Ising model, z. Further, we show that a high-frequency reseeding of the PRNGs improves the results, thus emphasizing the superiority of true randomness. On the other hand, we also apply standard tests for detecting correlations and find that some of the used PRNGs are, from their perspective, indistinguishable from a QRNG. Consequently, the calculation of z can detect some correlations that standard tests cannot, which highlights the need for further investigation with QRNGs in order to determine which is the best choice. Moreover, as z is known to be sensitive to the quality of the randomness used and the value assigned to it on the scientific literature varies widely, we contribute as well by computing it with a QRNG.

To conclude the thesis, we propose a novel type of quantum-inspired non-local moves that effectively help Population Annealing (PA) navigate the rugged energy landscape of 3d random plaquette gauge models, and which in 2d square lattices are equivalent to cluster updates. We show that this method improves the thermalization of small lattices and thus increases the confidence on having found the global minimum, while its use in harder disorder instances yields lower energy states than standard PA. As an outlook, we propose to investigate the merging of this new method with more advanced annealing schedules in PA.

The output of this thesis can be therefore summarized, from a fundamental point of view, in an improvement of the understanding of the processes used to find the minimum energy configurations of spin glass systems. And, besides the main, obvious field of spin glasses, this has potential implications in a broad range of other research areas. Concretely, many fields related to Machine Learning make use of models that can be understood, from a physics point of view, as spin glass systems. The difficulty to train Deep Boltzmann Machines (which are particularly useful for unsupervised learning), for example, is a direct consequence of the spin frustration arising from their intricate topology, and can consequently benefit greatly from understanding spin glass' physics and the improvement of optimization methods.

On the other hand, examples of direct applicability are numerous as well, and can be found in many areas of industrial relevance (for an example, see appendix A or reference [24]). Further, as was discussed in the introduction section, optimization problems arise not only in industrial environments, but, more importantly, in many areas of the society as a whole's concerns as well. These range from the distribution of resources such as water or vaccines to the optimization of public transport routes and, quoting the introduction's words, *we need the best solutions, and we need them to be obtained as fast as possible.* The results of the present thesis will, hopefully, pave the way to this ultimate goal.

Appendix A

A real optimization problem

In this appendix we briefly describe an example of a real QUBO problem. We first discuss its mapping to an Ising system and then solve it by means of an heuristic algorithm, with the aim of demonstrating the whole process of optimization.

A.1 The Maximum Covering Problem

The Maximum Covering Problem (MCP) is one of the paradigmatic examples of optimization problems for which a QUBO formulation is used for addressing real-world problems of relevance. In an abstract, mathematical formulation, it reads:

Let $V = \{u_j\}$ *and* $T = \{\alpha_i\}$ *be two sets of covered and covering elements, respectively. Find the minimum number of elements of* T *such that the maximum number of elements of* V *are covered.*

Additionally, the two following restrictions are usually considered in real-world applications:

- Only a maximum number T of covering elements can be used
- Each covered element is weighted by an importance factor

This problem can be seen to be NP-hard, and therefore no known polynomial-time algorithm is able to solve it optimally in the general case. By using its QUBO formulation, then, it is usual to address it by means of classical algorithms such as Simulated Annealing.

The MCP finds its relevance in many applications. For example, it is used in the telecommunication industry when planning the deployment of a new telecommunication system over a city. In such case, the new antennas can only be placed at certain sites, and one must find the minimum number of antennas that will provide service to the maximum area of the city. Furthermore, a given a maximum number of antennas is restricted by the current budget and each neighbourhood has an importance weighted by the amount of people living in it. The idea is, as has been previously discussed, to formulate this problem with an equation of the form of Eq. (2.11).

In order to formulate it in its QUBO form, let us first define the binary (Ising) variables in which the states of the problem can be encoded.

$$y_j = \begin{cases} 1 & \text{if element } u_j \text{ is covered} \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

$$g_{jk} = \begin{cases} 1 & \text{if element } u_j \text{ is covered exactly } k \text{ times} \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

$$x_i = \begin{cases} 1 & \text{if element } \alpha_i \text{ is active} \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

$$z_t = \begin{cases} 1 & \text{if there is a total of } t \text{ active covers} \\ 0 & \text{otherwise} \end{cases} \tag{A.4}$$

with $t \in [0, T]$, $i \in [0, N)$, $k \in [1, N)$, and $j \in [0, n)$, where n is the number of areas to be covered, N is the total number of covering elements and T is the maximum number of them that we are allowed to use. The problem is completed by the definition of the weight vector d_j differentiating the priority of covering each one of the elements u_j, and the covering matrix a, whith elements a_{ij} defined as

$$a_{ij} = \begin{cases} 1 & \text{if the element } \alpha_i \text{ can cover the element } u_j \\ 0 & \text{otherwise} \end{cases} \tag{A.5}$$

With the binary variables defined above, y_j, g_{jk}, x_i and z_t, Eqs. (A.1) to (A.4), the Hamiltonian to be minimized is [Lucas]

$$\begin{aligned} H &= -A\sum_j d_j y_j + B\sum_j \left[\sum_k g_{jk} - y_j\right]^2 + B\sum_j \left[\sum_k k g_{jk} - \sum_i a_{ij} x_i\right]^2 + \\ &+ C\left[\sum_t z_t - 1\right]^2 + C\left[\sum_t t z_t - \sum_i x_i\right]^2 . \end{aligned} \tag{A.6}$$

where the first term represents the core of the minimization problem (maximize the number of covered elements), the two following ones represent the minimization of the number of covering elements used, and the last two ones represent the constraint on the maximum number of covering elements. The proportionality constants A, B and C weight the importance of fulfilling each of the constraints in front of the minimization problem itself, with the condition that $0 < B, C < A$ in order to make sure that the ground state of the Hamiltonian does not violate them.

Expanding Eq. (A.6) the simplified expression of the Ising Hamiltonian reads

$$\begin{aligned} H = &= \sum J_{jkl} g_{jk} g_{jl} + \sum J_{jjk} y_j g_{jk} + \sum J_{ih} x_i x_h + \\ &+ \sum J_{jki} g_{jk} x_i + \sum J_{tp} z_t z_p + \sum J_{ti} z_t x_i + \\ &+ \sum b_{jk} g_{jk} + \sum b_j y_j + \sum b_i x_i + \sum b_t z_t, \end{aligned} \tag{A.7}$$

with the coupling elements computed as

$$J_{jkl} = (1 - \delta_{kl}) \frac{B}{4} (1 + kl), \tag{A.8}$$

$$J_{jjk} = -\frac{B}{2}, \tag{A.9}$$

$$J_{ih} = \frac{1}{4} \left[C + B \sum_{j=0}^{n} a_{ij} a_{hj} \right], \tag{A.10}$$

$$J_{jki} = -\frac{B}{2} k a_{ij}, \tag{A.11}$$

$$J_{tp} = \frac{C}{4} (1 + tp), \tag{A.12}$$

$$J_{ti} = -\frac{C}{2} t, \tag{A.13}$$

and the individual external fields as

$$b_{jk} = \frac{B}{2} \left[N - 2 + k \frac{N(N-1)}{2} - k \sum_{i=0}^{N} a_{ij} \right], \tag{A.14}$$

$$b_j = \frac{1}{2} \left[B(2 - N) - A d_j \right], \tag{A.15}$$

$$b_i = -C \frac{T(T+1)}{4} + \frac{B}{2} \left[CN + \sum_{j=0}^{n} a_i j \left(1 - \frac{N(N-1)}{2} + \sum_{h=0}^{N} (1 - \delta_{ih}) a_{hj} \right) \right], \tag{A.16}$$

$$b_t = \frac{C}{2} \left[T - 1 + t \left(\frac{T(T+1)}{2} - N \right) \right]. \tag{A.17}$$

A.2 A toy model example

Let us consider a real (small-scale) example now. Suppose a set of antennas α_i must be deployed to cover the telecommunication needs of a set of neighbourhoods u_j within a city. The map of the city along with the possible emplacements of the antennas and their corresponding covering range are shown in Fig.A.1. Each square represents a neighbourhood and each circle the covering range that an antenna would have if installed at position α_i. According to the definition above, Eq. (A.5), such

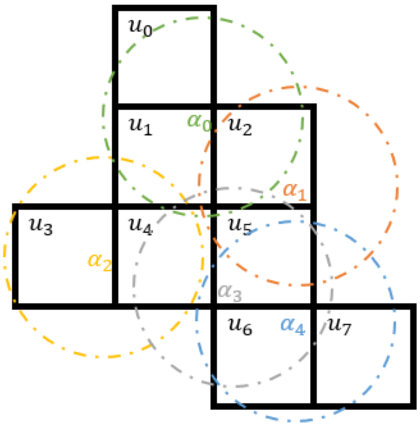

Figure A.1: **Map of the city and the covering range of the antennas**. Squares represent different neighbourhoods and circles the covering range of each possible emplacement of an antenna.

configuration can be described by the covering matrix

$$a = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.18}$$

and by a certain weight vector, d_j, which we consider homogeneous for simplicity as it does not alter the discussion of the example. We also set a maximum number of used antennas to $T = 3$, which is also seen to not alter the solution because it is enough to cover all the city. The graph describing all the interactions between the spin variables is shown in Fig. (A.2), in which each dot represent each of the Ising variables defined to encode the problem's states, y_j, g_{jk}, x_i and z_t, Eqs. (A.1) to (A.4), and the lines between them represent the interactions between them as given by Eq.(A.7). Furthermore, each line has a different colour depending on the strength of the coupling, Eqs.(A.8) to (A.13), as well as each dot has a different colour depending on the strength of the external individual field, Eqs. (A.14) to (A.17).

Finding the ground state of this problem by means of a simulated annealing algorithm finally yields the optimal distribution of active antennas, Fig.A.3.

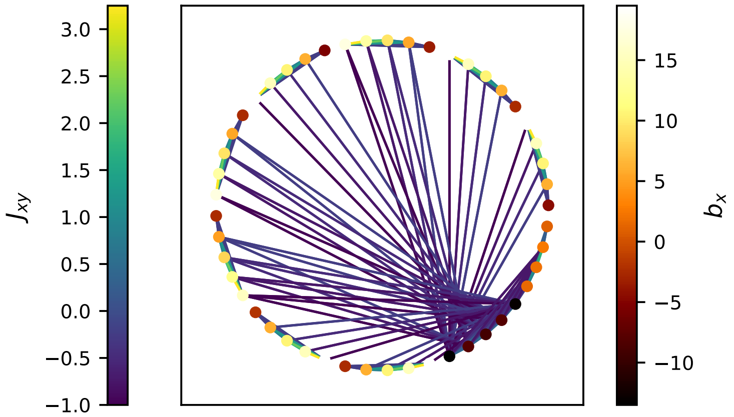

Figure A.2: **Graph of the QUBO formulation of the problem**. Each dot represents a spin variable (y_j, g_{jk}, x_i and z_t), and lines between them represent their connections as given by Eq. (A.7). Colours of the dots and the lines label the external individual fields' and couplings' strengths.

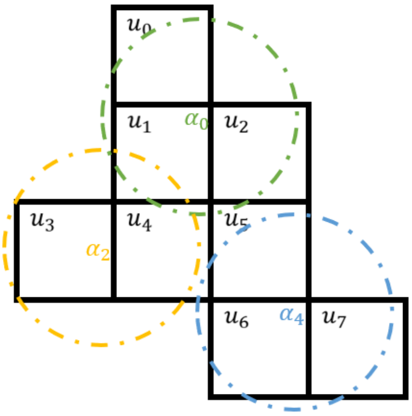

Figure A.3: **Optimal distribution of the antennas**. The whole city is covered if antennas are placed in positions α_0, α_2 and α_4.

Appendix B

Values of the dynamic critical exponent

In Tables B.1 and B.2 we gather all calculated exponents from the literature in two and three dimensions, to the best of our knowledge, for theoretical, MC as well as experimental approaches. This is the data plotted in Fig. 4.6.

Year	**Reference**	**Method**	z
1981	Bausch	Theory	2.126
1984	Domany	Theory	2
1985	Williams	MC	2.13(3)
1987	Ito	MC	2.132 ± 0.008
1987	Tang	MC	2.17 ± 0.04
1988	Ito	MC	2.2
1988	Mori	MC	2.076 ± 0.005
1992	Stauffer	MC	2.18
1993	Dammann	Theory	2.183 ± 0.005
1993	Wang	Theory	2.207 ± 0.008
1993	Muenkel	MC	2.21 ± 0.03
1995	Grassberger	MC	2.172 ± 0.006
1995	Gropengiesser	MC	2.18 ± 0.02
1995	Li	MC	2.1337(41)
1996	Li	MC	2.143(5)
1996	Nightingale	MC	2.1665(12)
1997	Soares	MC	2.16 ± 0.03
1997	Wang	MC	2.168 ± 0.005
1997	Wang	MC	2.180 ± 0.009, TP
1997	Wang	MC	2.167 ± 0.008, hc
1998	Prudnikov	Theory	2.093
1998	Wang	Theory	2.2
2000	Nightingale	MC	2.1667 ± 0.0005
2005	Dunlavy	Experiment	2.09 ± 0.06
2006	Krinitsyin	Theory	2.0842 ± 0.0039
2007	Canet	Theory	2.16(1)
2007	Lei	MC	2.16
2007	Murase	MC	2.193(5)
2007	Murase	MC	2.198(4), hc
2007	Murase	MC	2.199(3), TP
2009	Nalimov	Theory	2.020 ± 0.045
2009	Nalimov	Theory	2.023 ± 0.053
2009	Nalimov	Theory	2.026 ± 0.055
2009	Nalimov	Theory	2.100 ± 0.089
2009	Nalimov	Theory	2.105 ± 0.084
2009	Nalimov	Theory	2.104 ± 0.080
2009	Nalimov	Theory	2.127 ± 0.089
2009	Nalimov	Theory	2.132 ± 0.084
2009	Nalimov	Theory	2.130 ± 0.080
2009	Nalimov	Theory	$2.037^{+0.033}_{-0.0}$
2009	Nalimov	Theory	$2.041^{+0.040}_{-0.0}$
2009	Nalimov	Theory	$2.042^{+0.041}_{-0.0}$
2017	Duclut	Theory	2.28
2017	Duclut	Theory	2.16
2017	Duclut	Theory	2.15
2017	Duclut	Theory	2.14
2022	Adzhemyan	Theory	2.14(2)

Table B.1: All calculated exponents from the literature in two dimensions, to the best of our knowledge.

Year	**Reference**	**Method**	z
1987	Wansleben	MC	2.03 ± 0.04
1991	Wansleben	MC	2.04 ± 0.03
1992	Stauffer	MC	2.09
1993	Ito	MC	2.06(2)
1993	Muenkel	MC	2.08 ± 0.03
1994	Grassberger	MC	2.032 ± 0.004
1995	Gropengiesser	MC	2.04 ± 0.01
1997	Prudnikov	Theory	2.017
1999	Jaster	MC	2.042(6)
2000	Ito	MC	2.055(10)
2002	Livet	Experiment	2.2
2006	Krinitsyin	Theory	2.0237 ± 0.0055
2007	Canet	Theory	2.09(4)
2007	Murase	MC	2.065(25), bcc
2007	Murase	MC	2.057(25), fcc
2009	Nalimov	Theory	2.011 ± 0.012
2009	Nalimov	Theory	2.013 ± 0.012
2009	Nalimov	Theory	2.014 ± 0.011
2009	Nalimov	Theory	2.021 ± 0.006
2009	Nalimov	Theory	2.022 ± 0.005
2009	Nalimov	Theory	2.022 ± 0.005
2009	Nalimov	Theory	2.023 ± 0.006
2009	Nalimov	Theory	2.024 ± 0.005
2009	Nalimov	Theory	2.024 ± 0.005
2009	Nalimov	Theory	$2.013^{+0.011}_{-0.0}$
2009	Nalimov	Theory	$2.014^{+0.011}_{-0.0}$
2009	Nalimov	Theory	$2.014^{+0.011}_{-0.0}$
2010	Collura	MC	2.020(8)
2015	Livet	Experiment	2.28
2015	Niermann	Experiment	2.06
2017	Duclut	Theory	2.029
2017	Duclut	Theory	2.024
2017	Duclut	Theory	2.023
2017	Duclut	Theory	2.025
2017	Duclut	Theory	2.021
2017	Duclut	Theory	2.021
2018	Livet	Experiment	1.96(11)
2020	Hasenbusch	MC	2.0245(15)
2022	Adzheyman	Theory	2.0235(8)

Table B.2: All calculated exponents from the literature in three dimensions, to the best of our knowledge.

Bibliography

[1] R.K. Arora. *Optimization: algorithms and applications.* Chapman and Hall, CRC, 2015.

[2] P. Bangert. *Optimization for Indsutrial Problems.* Springer Berlin Heidelberg, 2012.

[3] G.R. McGhee. *Convergent Evolution: Limited Forms Most Beautiful.* Vienna Series in Theoretical Biology, 2011.

[4] D.P. Bebber, J. Hynes, P.R. Darrah, B. Lynne, and M.D. Fricker. Biological solutions to transport network design. *Proceedings of the Royal Society B: Biological Sciences*, 2007.

[5] T. Nakagaki, H. Yamada, and M. Hara. Smart network solutions in an amoeboid organism. *Biophysical Chemistry*, 2004.

[6] A. Tero, S. Takagi, T. Saigusa, K. Ito, D.P. Bebber, M.D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki. Rules for biologically inspired adaptive network design. *Science*, 2010.

[7] S. Nikkhah, K. Jalilpoor, E. Kianmehr, and G.B. Gharehpetian. Optimal wind turbine allocation and network reconfiguration for enhancing resiliency of system after major faults caused by natural disaster considering uncertainty. *IET Renewable Power Generation*, 2018.

[8] A.F. Nematollahi, A. Rahiminejad, B. Vahidi, H. Askarian, and A. Safaei. A new evolutionary-analytical two-step optimization method for optimal wind turbine allocation considering maximum capacity. *Journal of Renewable Sustainable Energy*, 2018.

[9] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information.* Combridge University Press, 2010.

[10] F. Arute, K. Arya, R. Babbush, D. Bacon, J.C. Bardin, R. Barends, R. Biswas, S. Boixo, F.G.S. L. Brandao, D.A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen,

A. Fowler, C. Gidney, M. Giustina, . Graff, K. Guerin, S. Habegger, M.P. Harrigan, M.J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T.S. Humble, S.V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P.V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J.R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M.Y. Niu, E. Ostby, A. Petukhov, J.C. Platt, C. Quintana, E.G. Rieffel, P. Roushan, N.C. Rubin, D. Sank, K.J. Satzinger, V. Smelyanskiy, K.J. Sung, M.D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z.J. Yao, P. Yeh, A. Zalcman, H. Neven, and J.M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 2019.

[11] E. Farhi, J. Goldstone, S. Gutmann, J. M. Lapan, A. Lundgren, and D. Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 2001.

[12] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum computation by adiabatic evolution. *arXiv:quant-ph*, 2000.

[13] J. Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2018.

[14] T. Albash and D.A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 2018.

[15] F. Glover, G. Kochenberger, R. Hennig, and Y. Du. Quantum bridge analytics i: a tutorial on formulating and using qubo models. *Annals of Operation Research*, 2022.

[16] K. Pudenz, T. Albash, and D.A. Lidar. Error-corrected quantum annealing with hundreds of qubits. *Nature communications*, 2014.

[17] T. Albash and D.A. Lidar. Decoherence in adiabatic quantum computation. *Phys. Rev. A*, 2015.

[18] K.C. Young, M. Sarovar, and R. Blume-Kohout. Error suppression and error correction in adiabatic quantum computation: Techniques and challenges. *Physical Review X*, 2013.

[19] A. Auffèves. Quantum technologies need a quantum energy initiative. *PRX Quantum*, 2022.

[20] D. Jaschke and S. Montangero. Is quantum computing green? an estimate for an energy-efficiency quantum advantage. *Quantum Science and Technology*, jan 2023.

[21] F. Meier and H. Yamasaki. Energy-consumption advantage of quantum computation. *arXiv:quant-ph*, 2023.

[22] 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. *Physical Review A*, 2016.

[23] V.S. Denchev, S. Boixo, S.V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven. What is the computational value of finite-range tunneling? *Physical Review X*, 2016.

[24] M.J.A. Schuetz, J.K. Brubaker, H. Montagu, Y. van Dijk, J. Klepsch, P. Ross, A. Luckow, M.G.C. Resende, and H.G. Katzgraber. Optimization of robot-trajectory planning with nature-inspired and hybrid quantum algorithms. *Physical Review Applied*, 2022.

[25] M. Mitchell. *An introduction to genetic algorithms.* the MIT Press, 1998.

[26] A. Lambora, K. Gupta, and K. Chopra. Genetic algorithm- a literature review. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019.

[27] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun. Graph neural networks: A review of methods and applications. *AI open*, 2020.

[28] Y. Yu, X. Si, C. Hu, and J. Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 2019.

[29] T. Hao, X. Huang, C. Jia, and C. Peng. A quantum-inspired tensor network algorithm for constrained combinatorial optimization problems. *Frontiers in Physics*, 2022.

[30] K. Han and J. Kim. Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *IEEE transactions on evolutionary computation*, 2002.

[31] D. Beloborodov, A.E. Ulanov, J.N. Foerster, S. Whiteson, and A. Lvovsky. Reinforcement learning enhanced quantum-inspired algorithm for combinatorial optimization. *Machine Learning: Science and Technology*, 2020.

[32] T.H. Click, A. Liu, and G.A. Kaminski. Quality of random number generators significantly affects results of monte carlo simulations for organic and biological systems. *Journal of computational chemistry*, 2011.

[33] Y. Lin and F. Wang. Linear relaxation in large two-dimensional ising models. *Physical Review E*, 2016.

[34] C. Abellán, W. Amaya, M. Jofre, M. Curty, A. Acín, J. Capmany, V. Pruneri, and M.W. Mitchell. Ultra-fast quantum randomness generation by accelerated phase diffusion in a pulsed laser diode. *Optics express*, 2014.

[35] M. Lewenstein, D. Cirauqui, M.Á. García-March, G. Guigó i Corominas, P. Grzybowski, J.R.M. Saavedra, M. Wilkens, and J. Wehr. Haake-lewenstein-wilkens approach to spin-glasses revisited. *Journal of Physics A: Mathematical and Theoretical*, 2022.

[36] D. Cirauqui, M.Á. García-March, G. Guigó i Corominas, T. Graß, P.R. Grzybowski, G. Muñoz-Gil, J.R.M. Saavedra, and M. Lewenstein. Quantum random number generators: Benchmarking and challenges. *arXiv preprint arXiv:2206.05328*, 2022.

[37] D. Cirauqui, M.Á. García-March, J.R.M. Saavedra, M. Lewenstein, and P.R. Grzybowski. Population annealing with topological defect driven nonlocal updates for spin systems with quenched disorder. *Physical Review B*, 2024.

[38] S. Rychkov. Lectures on the random field ising model: From parisi-sourlas supersymmetry to dimensional reduction. *Lectures on the Random Field Ising Model*, 2023.

[39] G. Parisi. Mean field theory of spin glasses: statics and dynamics, 2007.

[40] F. Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 1982.

[41] J.P. Casquilho and P.I.C. Teixeira. *Introduction to Statistical Physics*. Cambridge University Press, 2014.

[42] L. Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, 1944.

[43] M. Mézard, G. Parisi, and M.A. Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company, 1987.

[44] K. Binder and A.P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern physics*, 1986.

[45] V. Cannella and J.A. Mydosh. Magnetic ordering in gold-iron alloys. *Physical Review B*, 1972.

[46] C.A.M. Mulder, A.J. van Duyneveldt, and J.A. Mydosh. Susceptibility of the CuMn spin-glass: Frequency and field dependences. *Physical Review B*, 1981.

[47] Q. Zhou and W.H. Wong. Energy landscape of a spin-glass model: exploration and characterization. *Physical Review E*, 2009.

[48] L. Berthier and G. Biroli. Theoretical perspective on the glass transition and amorphous materials. *Reviews of modern physics*, 2011.

[49] A. Cavagna. Supercooled liquids for pedestrians. *Physics Reports*, 2009.

[50] I. Pemartín, V. Martin-Mayor, G. Parisi, and J.J. Ruiz-Lorenzo. Numerical study of barriers and valleys in the free-energy landscape of spin glasses. *Journal of Physics A: Mathematical and Theoretical*, 2019.

[51] 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. *Physical Review E*, 2015.

[52] B. Heim, T.F. Rønnow, S.V. Isakov, and M. Troyer. Quantum versus classical annealing of ising spin glasses. *Science*, 2015.

[53] J.J. Moreno, H.G. Katzgraber, and A.K. Hartmann. Finding low-temperature states with parallel tempering, simulated annealing and simple monte carlo. *International Journal of Modern Physics C*, 2003.

[54] F. Barahona. Ground-state magnetization of ising spin glasses. *Physical Review B*, 1994.

[55] M. Laguna and P. Laguna. Applying tabu search to the two-dimensional ising spin glass. *International Journal of Modern Physics C*, 1995.

[56] R. Yaacoby, N. Schaar, L. Kellerhals, O. Raz, D. Hermelin, and R. Pugatch. Comparison between a quantum annealer and a classical approximation algorithm for computing the ground state of an ising spin glass. *Physical Review E*, 2022.

[57] C.A. Amey. *Popoulation Annealing: analysis, optimization and application to glassy systems.* PhD thesis, University of Massachusstes Amhersts, 2021.

[58] S. F. Edwards and P. W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 1975.

[59] G. Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 1979.

[60] G. Parisi. A sequence of approximated solutions to the s-k model for spin glasses. *Journal of Physics A: Mathematical and General*, 1980.

[61] G. Parisi. Order parameter for spin-glasses. *Physical Review Letters*, 1983.

[62] W. Wang, J. Machta, and H.G. Katzgraber. Population annealing: Theory and application in spin glasses. *Physical Review E*, 2015.

[63] N. Kawashima and A. P. Young. Phase transition in the three-dimensional $\pm j$ ising spin glass. *Physical Review B*, 1996.

[64] B.A. Berg and W. Janke. Multioverlap simulations of the 3d edwards-anderson ising spin glass. *Physical Review Letters*, 1998.

[65] E. Marinari, G. Parisi, and J.J. Ruiz-Lorenzo. *Numerical simulations of spin glass systems.*

[66] E. Marinari, G. Parisi, F. Ricci-Tersenghi, J.J. Ruiz-Lorenzo, and F. Zuliani. Replica symmetry breaking in short-range spin glasses: Theoretical foundations and numerical evidences. *Journal of Statistical Physics*, 2000.

[67] G. Parisi, F. Ricci-Tersenghi, and J.J. Ruiz-Lorenzo. Equilibrium and off-equilibrium simulations of the 4d gaussian spin glass. *Journal of Physics A: Mathematical and General*, 1996.

[68] D.L. Stein and C.M. Newman. *Spin glasses and complexity.* Princeton University Press, 2013.

[69] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 1975.

[70] D. Panchenko. *The sherrington-kirkpatrick model.* Springer Science, 2013.

[71] M. Zaman, K. Tanahashi, and S. Tanaka. Pyqubo: Python library for mapping combinatorial optimization problems to qubo form. *IEEE Transactions on Computers*, 2022.

[72] A. Lucas. Hard combinatorial problems and minor embeddings on lattice graphs. *Quantum Information Processing*, 2019.

[73] A. Irbäck, L. Knuthson, S. Mohanty, and C. Peterson. Folding lattice proteins with quantum annealing. *Physical Review Research*, 2022.

[74] N.M. de Oliveira, R.M. de Abreu Silva, and W. de Oliveira. Qubo formulation for the contact map overlap problem. *International Journal of Quantum Information*, 2018.

[75] F. Hernández, K. Díaz, M. Forets, and R. Sotelo. Application of quantum optimization techniques (qubo method) to cargo logistics on ships and airplanes. In *2020 IEEE Congreso Bienal de Argentina (ARGENCON).* IEEE, 2020.

[76] F. Phillipson and I. Chiscop. Multimodal container planning: A qubo formulation and implementation on a quantum annealer. In *International Conference on Computational Science.* Springer, 2021.

[77] J.F. Ariño Sales and R.A. Palacios Araos. Adiabatic quantum computing for logistic transport optimization. *arXiv*, 2023.

[78] A. Lucas. Ising formulations of many np problems. *Frontiers in physics*, 2014.

[79] M.E.J. Newman and G.T. Barkema. *Monte Carlo methods in statistical physics.* Clarendon Press, 1999.

[80] J.E. Gubernatis. *The Monte Carlo Method in the Physical Sciences.* American Institute of Physics, 2003.

[81] James R Norris. *Markov chains.* Cambridge university press, 1998.

[82] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 1953.

[83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 1983.

[84] M. Weigel. Monte carlo methods for massively parallel computers. In *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory.* World Scientific, 2018.

[85] M. Weigel. Performance potential for simulating spin models on gpu. *Journal of Computational Physics*, 2012.

[86] S. Sachdev. *Quantum Phase Transitions.* Cambridge University Press, 2 edition, 2011.

[87] T. Graß, D. Raventós, B. Juliá-Díaz, C. Gogolin, and M. Lewenstein. Quantum annealing for the number-partitioning problem using a tunable spin glass of ions. *Nature communications*, 2016.

[88] V. Ahufinger, L. Sanchez-Palencia, A. Kantian, A. Sanpera, and M. Lewenstein. Disordered ultracold atomic gases in optical lattices: A case study of fermi-bose mixtures. *Physical Review A*, 2005.

[89] H. Sompolinsky. Time-dependent order parameters in spin-glasses. *Phys. Rev. Lett.*, 1981.

[90] F. Haake, M. Lewenstein, and M. Wilkens. Relation of random and competing nonrandom couplings for spin-glasses. *Physical review letters*, 1985.

[91] J.T. Edwards and D.J. Thouless. Numerical studies of localization in disordered systems. *Journal of Physics C: Solid State Physics*, 1972.

[92] D.C. Licciardello and D.J. Thouless. Conductivity and mobility edges for two-dimensional disordered systems. *Journal of Physics C: Solid State Physics*, 1975.

[93] E. Abrahams, P.W. Anderson, D.C. Licciardello, and T.V. Ramakrishnan. Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Physical Review Letters*, 1979.

[94] Y. Imry and S. Ma. Random-field instability of the ordered state of continuous symmetry. *Physical Review Letters*, 1975.

[95] T.C. Proctor, D.A. Garanin, and E.M. Chudnovsky. Random fields, topology, and the imry-ma argument. *Physical Review Letters*, 2014.

[96] J.R. Banavar and M. Cieplak. Nature of ordering in spin-glasses. *Physical Review Letters*, 1982.

[97] M. Bera, A. Acín, M. Kuś, M.W. Mitchell, and M. Lewenstein. Randomness in quantum mechanics: philosophy, physics and technology. *Reports on Progress in Physics*, 2017.

[98] M.A. Rowe, D. Kielpinski, V. Meyer, C.A. Sackett, W.M. Itano, C. Monroe, and D.J. Wineland. Experimental violation of a bell's inequality with efficient detection. *Nature*, 2001.

[99] L.E. Bassham, A.L. Rukhin, J. Soto, J.R. Nechvatal, M.E. Smid, S.D. Leigh, M. Levenson, M. Vangel, N.A. Heckert, and D.L. Banks. *A statistical test suite for random and pseudorandom number generators for cryptographic applications*. US Department of Commerce, Technology Administration, National Institute of ..., 2001.

[100] G. Marsaglia and W. Tsang. Some difficult-to-pass tests of randomness. *Journal of Statistical Software*, 2002.

[101] D. Eddelbuettel and R.G. Brown. Rdieharder: An r interface to the dieharder suite of random number generator tests. *Initial Version as of May*, 2007.

[102] National Institute of Standards and Technology. Security requirements for cryptographic modules. Technical Report Federal Information Processing Standards Publications (FIPS PUBS) 140-2, Change Notice 2 December 03, 2002, U.S. Department of Commerce, Washington, D.C., 2001.

[103] P. L'ecuyer. Tables of linear congruential generators of different sizes and good lattice structure. *Mathematics of Computation*, 1999.

[104] E. Faure, E. Fedorov, I. Myronets, and S. Sysoienko. Method for generating pseudorandom sequence of permutations based on linear congruential generator. In *CMIS*, 2022.

[105] L.K. Shalm, E. Meyer-Scott, B.G. Christensen, P. Bierhorst, M.A. Wayne, M.J. Stevens, T. Gerrits, S. Glancy, D.R. Hamel, M.S. Allman, K.J. Coakley, S.D. Dyer, C. Hodge, A.E. Lita, V.B. Verma, C. Lambrocco, E. Tortorici, A.L. Migdall, Y. Zhang, D.R. Kumor, W.H. Farr, F. Marsili, M.D. Shaw, J.A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M.W. Mitchell, P.G. Kwiat,

J.C. Bienfang, R.P. Mirin, E. Knill, and S.W. Nam. Strong loophole-free test of local realism. *Physical Review Letters*, 2015.

[106] B. Hensen, H. Bernien, A. Reiserer A.E. Dréau, M.S. Blok N. Kalb, R.F.L. Vermeulen J. Ruitenberg, R.N. Schouten, W. Amaya C. Abellán, V. Pruneri, M.W. Mitchell, M. Markham, D.J. Twitchen, D. Elkouss, S. Wehner, T.H. Taminiau, and R. Hanson. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 2015.

[107] M. Giustina, M.A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J. Larsson, C. Abellán, W. Amaya, V. Pruneri, M.W. Mitchell, J. Beyer, T. Gerrits, A.E. Lita, L.K. Shalm, S.W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger. Significant-loophole-free test of bell's theorem with entangled photons. *Physical Review Letters*, 2015.

[108] A. Acín and L. Masanes. Certified randomness in quantum physics. *Nature*, 2017.

[109] S. Pironio, A. Acín, S. Massar, et al. Random numbers certified by bell's theorem. *Nature*, 2010.

[110] W. Rosenfeld, M. Weber, J. Volz, F. Henkel, M. Krug, A. Cabello, M. Zukowski, and H. Weinfurter. Towards a loophole-free test of bell's inequality with entangled pairs of neutral atoms. *Advanced science letters*, 2009.

[111] M. Jerger, Y. Reshitnyk, A. Potocnik M. Oppliger, M. Mondal, A. Wallraff, S. Wehner K. Goodenough, K. Juliusson, N.K. Langford, and A. Fedorov. Contextuality without nonlocality in a superconducting quantum system. *Nature Communications*, 2016.

[112] M. Herrero-Collantes and J.C. Garcia-Escartin. Quantum random number generators. *Review of Modern Physics*, 2017.

[113] M. Isida and H. Ikeda. Random number generator. *Annals of the Institute of Statistical Mathematics*, 1956.

[114] F. Xu, B. Qi, X. Ma, H. Xu, H. Zheng, and H. Lo. Ultrafast quantum random number generation based on quantum phase fluctuations. *Optics express*, 2012.

[115] M. Jofre, M. Curty, F. Steinlechner, G. Anzolin, J. Pérez-Torres, M.W. Mitchell, and V. Pruneri. True random numbers from amplified quantum vacuum. *Optics express*, 2011.

[116] Z.L. Yuan, M. Lucamarini, J.F. Dynes, B. Fröhlich, A. Plews, and A.J. Shields. Robust random number generation using steady-state emission of gain-switched laser diodes. *Applied Physics Letters*, 2014.

[117] A.L. Schawlow and C.H. Townes. Infrared and optical masers. *Physical Review*, 1958.

[118] Y. Nie, L. Huang, Y. Liu, F. Payne, J. Zhang, and J.W. Pan. The generation of 68 gbps quantum random number by measuring laser phase fluctuations. *Review of Scientific Instruments*, 2015.

[119] R.J. Glauber. Time-dependent statistics of the ising model. *Journal of mathematical physics*, 1963.

[120] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 1965.

[121] J.A. Bonachela Fajardo. *Universality in Self-Organized Criticality.* PhD thesis, Universidad de Granada, 2008.

[122] A. Lesne and M. Laguës. *Scale invariance: From phase transitions to turbulence.* Springer Science & Business Media, 2011.

[123] R. Bausch, V. Dohm, H. K. Janssen, and R.K.P. Zia. Critical dynamics of an interface in $1 + \epsilon$ dimensions. *Physical Review Letters*, 1981.

[124] E. Domany. Exact results for two- and three-dimensional ising and potts models. *Phys. Rev. Lett.*, 1984.

[125] J. Wang. Critical slowing down of the two-dimensional kinetic ising model with glauber dynamics. *Physical Review B*, 1993.

[126] B. Dammann and J.D. Reger. Dynamical critical exponent of the two-dimensional ising model. *Europhysics Letters (EPL)*, 1993.

[127] J.S. Wang and C.K. Gan. Nonequilibrium relaxation of the two-dimensional ising model: Series-expansion and monte carlo studies. *Physical Review E*, 1998.

[128] V.V. Prudnikov, S.V. Belim, E.V. Osintsev, and A.A. Fedorenko. Critical dynamics of disordered magnets in the three-loop approximation. *Physics of the Solid State*, 1998.

[129] A.S. Krinitsyn, V.V. Prudnikov, and P.V. Prudnikov. Calculations of the dynamical critical exponent using the asymptotic series summation method. *Theoretical and mathematical physics*, 2006.

[130] L. Canet and H. Chaté. A non-perturbative approach to critical dynamics. *Journal of Physics A: Mathematical and Theoretical*, 2007.

[131] M.Y. Nalimov, V.A. Sergeev, and L. Sladkoff. Borel resummation of the -expansion of the dynamical exponent z in model a of the ϕ 4 (o (n)) theory. *Theoretical and Mathematical Physics*, 2009.

[132] C. Duclut and B. Delamotte. Frequency regulators for the nonperturbative renormalization group: A general study and the model a as a benchmark. *Physical Review E*, 2017.

[133] L.T. Adzhemyan, D.A. Evdokimov, M. Hnatic, E.V. Ivanova, M.V. Kompaniets, A. Kudlis, and D.V. Zakharov. The dynamic critical exponent z for 2d and 3d ising models from five-loop expansion. *Physics Letters A*, 2022.

[134] F. Livet, F. Bley, J.P. Simon, R. Caudron, J. Mainville, M. Sutton, and D. Lebolloch. Statics and kinetics of the ordering transition in the auagzn 2 alloy. *Physical Review B*, 2002.

[135] M.J. Dunlavy and D. Venus. Critical slowing down in the two-dimensional ising model measured using ferromagnetic ultrathin films. *Physical Review B*, 2005.

[136] F. Livet, M. Fèvre, G. Beutier, and M. Sutton. Ordering fluctuation dynamics in auagzn$_2$. *Physical Review B*, 2015.

[137] D. Niermann, C.P. Grams, P. Becker, L. Bohatý, H. Schenck, and J. Hemberger. Critical slowing down near the multiferroic phase transition in mnwo$_4$. *Physical Review Letters*, 2015.

[138] F. Livet, M. Fèvre, G. Beutier, F. Zontone, Y. Chushkin, and M. Sutton. Measuring the dynamical critical exponent of an ordering alloy using x-ray photon correlation spectroscopy. *Physical Review B*, 2018.

[139] J. K. Williams. Monte carlo estimate of the dynamical critical exponent of the 2d kinetic ising model. *Journal of Physics A: Mathematical and General*, 1985.

[140] S. Tang and D.P. Landau. Monte carlo study of dynamic universality in two-dimensional potts models. *Physical Review B*, 1987.

[141] N. Ito, M. Taiji, and M. Suzuki. Accurate estimation of the dynamical critical exponent of the two-dimensional kinetic ising model based on the ising machine m-tis. *Journal of the Physical Society of Japan*, 1987.

[142] S. Wansleben and D.P. Landau. Dynamical critical exponent of the 3d ising model. *Journal of Applied Physics*, 1987.

[143] N. Ito, M. Taiji, and M. Suzuki. Critical dynamics of the ising model with ising machine. *Le Journal de Physique Colloques*, 1988.

[144] M. Mori and Y. Tsuda. Vectorized monte carlo simulation of large ising models near the critical point. *Physical Review B*, 1988.

[145] S. Wansleben and D.P. Landau. Monte carlo investigation of critical dynamics in the three-dimensional ising model. *Physical Review B*, 1991.

[146] D. Stauffer. Simulation of time exponent in 2d glauber kinetic ising model. *Physica A: Statistical Mechanics and its Applications*, 1992.

[147] C. Münkel, D.W. Heermann, J. Adler, M. Gofman, and D. Stauffer. The dynamical critical exponent of the two-, three- and five-dimensional kinetic ising model. *Physica A: Statistical Mechanics and its Applications*, 1993.

[148] N. Ito. Non-equilibrium relaxation and interface energy of the ising model. *Physica A: Statistical Mechanics and its Applications*, 1993.

[149] Z.B. Li, L. Schülke, and B. Zheng. Dynamic monte carlo measurement of critical exponents. *Physical review letters*, 1995.

[150] P. Grassberger. Damage spreading and critical exponents for "model a" ising dynamics. *Physica A: Statistical Mechanics and its Applications*, 1995.

[151] U. Gropengiesser. Damage spreading and critical exponents for 'model a'ising dynamics. *Physica A: Statistical Mechanics and its Applications*, 1995.

[152] Z. Li, L. Schülke, and B. Zheng. Finite-size scaling and critical exponents in critical relaxation. *Physical Review E*, 1996.

[153] M.P. Nightingale and H.W.J. Blöte. Dynamic exponent of the two-dimensional ising model and monte carlo computation of the subdominant eigenvalue of the stochastic matrix. *Physical review letters*, 1996.

[154] F.G. Wang and C.K. Hu. Universality in dynamic critical phenomena. *Physical Review E*, 1997.

[155] M. S. Soares, J. Kamphorst Leal da Silva, and F.C. SáBarreto. Numerical method to evaluate the dynamical critical exponent. *Physical Review B*, 1997.

[156] A. Jaster. Computer simulations of the two-dimensional melting transition using hard disks. *Physical Review E*, 1999.

[157] M.P. Nightingale and H W.J. Blöte. Monte carlo computation of correlation times of independent relaxation modes at criticality. *Physical Review B*, 2000.

[158] N. Ito, K. Hukushima, K. Ogawa, and Y. Ozeki. Nonequilibrium relaxation of fluctuations of physical quantities. *Journal of the Physical Society of Japan*, 2000.

[159] X. Lei, J. Zheng, and X.Y. Zhao. Monte carlo simulations for two-dimensional ising system far from equilibrium. *Chinese Science Bulletin*, 2007.

[160] Y. Murase and N. Ito. Dynamic critical exponents of three-dimensional ising models and two-dimensional three-states potts models. *Journal of the Physical Society of Japan*, 2007.

[161] M. Collura. Off-equilibrium relaxational dynamics with an improved ising hamiltonian. *Journal of Statistical Mechanics: Theory and Experiment*, 2010.

[162] M. Hasenbusch. Monte carlo study of a generalized icosahedral model on the simple cubic lattice. *Physical Review B*, 2020.

[163] K. Binder. Time-dependent ginzburg-landau theory of nonequilibrium relaxation. *Physical Review B*, 1973.

[164] P.C. Hohenberg and A.P. Krekhov. An introduction to the ginzburg–landau theory of phase transitions and nonequilibrium patterns. *Physics Reports*, 2015.

[165] S. Boixo, T. Albash, F.M. Spedalieri, N. Chancellor, and D.A. Lidar. Experimental signature of programmable quantum annealing. *Nature communications*, 2013.

[166] N.G. Dickson, M.H. Amin M.W. Johnson, R. Harris, F. Altomare, A.J. Berkley, P. Bunyk, J. Cai, E.M. Chapple, P. Chavez, F. Cioata, T. Cirip, P. deBuen, M. Drew-Brook, C. Enderud, S. Gildert, F. Hamze, J.P. Hilton, E. Hoskinson, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Mahon T. Lanting, R. Neufeld, T. Oh, I. Perminov, C. Petroff, A. Przybysz, C. Rich, A. Tcaciuc P. Spear, M.C. Thom, E. Tolkacheva, J. Wang S. Uchaikin, A.B. Wilson, Z. Merali, and G Rose. Thermally assisted quantum annealing of a 16-qubit problem. *Nature communications*, 2013.

[167] Y.W. Koh and H. Nishimori. Quantum and classical annealing in a continuous space with multiple local minima. *Physical Review A*, 2022.

[168] Y. Liu, W.J. Su, and T. Li. On quantum speedups for nonconvex optimization via quantum tunneling walks. *Quantum*, 2023.

[169] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 1996.

[170] H.G. Katzgraber, S. Trebst, D.A. Huse, and M. Troyer. Feedback-optimized parallel tempering monte carlo. *Journal of Statistical Mechanics: Theory and Experiment*, 2006.

[171] J. Machta. Population annealing with weighted averages: A monte carlo method for rough free-energy landscapes. *Physical Review E*, 2010.

[172] C. Amey and J. Machta. Analysis and optimization of population annealing. *Physical Review E*, 2018.

[173] D. Gessert, M. Weigel, and W. Janke. Resampling schemes in population annealing—numerical results. In *Journal of Physics: Conference Series*. IOP Publishing, 2022.

[174] P. Blanchard, D. J Higham, and N.J. Higham. Accurately computing the log-sum-exp and softmax functions. *IMA Journal of Numerical Analysis*, 2021.

[175] A. Barzegar, C. Pattison, W. Wang, and H.G. Katzgraber. Optimization of population annealing monte carlo for large-scale spin-glass simulations. *Physical Review E*, 2018.

[176] A.Y. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 2003.

[177] H.A. Kramers and G.H. Wannier. Statistics of the two-dimensional ferromagnet. part i. *Physical Review*, 1941.

[178] F.J. Wegner. Duality in Generalized Ising Models and Phase Transitions Without Local Order Parameters. *Journal of Mathematical Physics*, 1971.

[179] P.L. Ebert, D. Gessert, and M. Weigel. Weighted averages in population annealing: Analysis and general framework. *Physical Review E*, 2022.

[180] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for monte carlo simulation of ising spin systems. *Journal of Computational Physics*, 1975.

[181] L.Y. Barash, M. Weigel, L.N. Shchur, and W. Janke. Exploring first-order phase transitions with population annealing. *The European Physical Journal Special Topics*, 2017.