ABSTRACT

Title of dissertation: DUALITY METHODS IN NETWORKS, COMPUTER SCIENCE MODELS, AND DISORDERED CONDENSED MATTER SYSTEMS

Joe Mitchell, Doctor of Philosophy, 2014

Dissertation directed by: Dr. V. M. Galitski
Department of Physics

In this thesis, I explore lattice independent duality and systems to which it can be applied. I first demonstrate classical duality on models in an external field, including the Ising, Potts, and $x-y$ models, showing in particular how this modifies duality to be lattice independent and applicable to networks. I then present a novel application of duality on the boolean satsifiability problem, one of the most important problems in computational complexity, through mapping to a low temperature Ising model. This establishes the equivalence between boolean satisfiability and a problem of enumerating the positive solutions to a Diophantine system of equations. I continue by combining duality with a prominent tool for models on networks, belief propagation, deriving a new message passing procedure, dual belief propagation. In the final part of my thesis, I shift to propose and examine a semiclassical model, the two-component Coulomb glass model, which can explain the giant magnetoresistance peak present in disordered films near a superconductor-insulator transition as the effect of competition between single particle and localized pair transport. I

numerically analyze the density of states and transport properties of this model.

DUALITY METHODS IN NETWORKS, COMPUTER SCIENCE MODELS, AND DISORDERED CONDENSED MATTER SYSTEMS

by

Joseph Dan Mitchell

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Dr. V. Galitski, Chair/Advisor
Dr. J. Sau
Dr. W. Gasarch
Dr. M. Girvan
Dr. V. Yakovenko

Acknowledgments

Many people have helped me over the past few years, making my graduate work possible and enjoyable. I would like to thank Victor Galitski for his support and endless creativity as my advisor. Despite his wide range of interests and exciting work, he was always helpful and interested if I wanted to step outside of that range. His grasp of the big picture was invaluable when I was stuck in the details, and his amenable manner allowed me to work when I needed but always left an open door for problems or exciting successes. I will cherish my time spent learning from and working with him while I envy his ability to immediately strike to the heart of the most random projects and ideas.

Second, my collaborators deserve credit for much of the progress I made through my graduate career. I worked intensely with Dr. Markus Müller of the International Center for Theoretical Physics along with colleague Anirban Gangopadhyay and my advisor on the two-component Coulomb glass project from Chapter 5. His knowledge of the subject and his ability to find every kink in the project and paper made everything flow smoothly, and he has been helpful in many ways I was not expecting, including pointing out key references for the SAT duality project in Chapter 3. Dr. Benjamin Hsu eased many of the difficulties of the SAT project, especially early with his knowledge of a field neither Victor nor I had much experience with. I would like to also give special thanks to people who have aided the work in this thesis with short and dense conversations, including Tigran Sedrakyan, Boris Shklovskii, Brian Skinner, and Zohar Nussinov.

I would like to thank Victor's group, my coworkers, for their support and friendship, and also for adding to my questions with a keen understanding of the physics and adding to my understanding with keen questions. I am especially grateful to Justin Wilson and Anirban Gangopadhyay. Justin always had some tidbit of info stored up, and usually a correct tidbit. His fondness for a beautiful equation or derivation made me appreciate refining a piece of work until it was intuitive as well as correct. I shared ideas and current work with Anirban until I would be more interested in his problems than my own. He would listen when I needed to talk something out, question when I wasn't explaining, and crack jokes when I needed to relax. I will remember quite fondly the hours we spent programming side by side, working through dozens of separate programs for the Coulomb glass project.

Too many friends have helped me in the past five years to list, but I need to thank Jeff Grover, Zach Smith, Sergei Pershoguba, Guilherme Miranda, Caitlin Thomas, Steve Ziemak, Trystan Koch, Matt Wise, and especially Steve Lynam, Jon Duggan, and Alex Knoll. My parents, brother, and sisters provide a welcome relief when all else is chaos, giving affection and encouragement in spades. Finally, my girlfriend Rachael DeNale simply makes my life easier in too many ways to count. I could not do without her love and support.

Thank you, everyone!

Table of Contents

List of Figures vi

1 Introduction 1

2 Lattice independent duality 6
2.1 Kramers-Wannier duality 7
2.2 Ising Duality in an external field 11
2.3 Vector Potts duality 17
2.4 $x - y$ duality 20
2.5 Conclusion 22

3 Boolean satisfiability 24
3.1 The SAT problem 25
3.2 Duality for the SAT problem through the Ising model 27
3.3 Simplification to a Diophantine system of equations 29
3.4 Examples of SAT duality 32
3.5 Other satisfiability problems 35
3.6 Conclusion 37

4 Message passing under duality 40
4.1 Normal and dual belief propagation for the Ising model 41
4.2 Examples: Ising model transition temperature approximations 44
4.3 Normal and dual belief propagation for the vector Potts model 46
4.4 Conclusion 50

5 Two-component Coulomb glass in insulators with a local attraction 52
5.1 Motivation for two-component Coulomb glass 55
5.2 Model 58
5.3 Single site density of states 60
5.3.1 Definitions 60
5.3.2 Spatially Uniform Interaction - Anomalous Coulomb gap . . . 62
5.3.3 Numerical simulations 68
5.3.4 Spatially Disordered Interaction 70

5.4 Resistor network mapping 71
5.4.1 Choice of parameters 71
5.4.2 Resistor-network construction for first order processes 75
5.4.3 Extension to include second order processes 78
5.4.4 Simplified algorithms to include pair breaking/formation . . . 80
5.5 Transport 82
5.5.1 Results 84
5.6 Conclusion 85

6 Conclusions and future work 89

Bibliography 98

List of Figures

2.1 Behavior of spin, link, and dual spin variables under Kramers-Wannier duality transformation on a periodic square lattice. 9
2.2 Behavior of dual coupling $\tilde{J}$ as a function of original coupling J. . . . 10

3.1 Schematic of the SAT duality process 30
3.2 A SAT instance, $\mathcal{F} = (x_1 \vee x_2 \vee x_3) \wedge (x_3 \vee x_4 \vee x_5)$. Each circle represents a site, the location of a spin associated with a boolean variable. Each square represents a link, the location of an interaction (constraint) between spins. In the dual picture, the spins are exchanged with interactions. 30

5.1 Illustration of the two-component model: The energy landscape is due to the combination of on-site disorder and Coulomb interactions. The arrows indicate typical hopping processes relevant for the complex low T transport in the two-component Coulomb glass. 55
5.2 DOS for different uniform interaction U. $U = 0$ is a critical point at which both $\rho_{1,2}$ have a linear pseudogap. The slope of the single particle DOS ρ_1 is suppressed to $\alpha/4e^4$. For net repulsion, $U > 0$, ρ_1 has the canonical slope α/e^4 at lowest energy, followed by a hump at the scale $U/2$, crossing over to the critical slope, while pairs are gapped up to $E = U$. For $U < 0$, single electrons have a hard gap $|U|/2$, while pairs are pseudogapped with slope $\alpha/16e^4$. Note: For these plots, the chemical potential was explicitly zeroed when averaging the DOS over the various initial occupancy-distributions . . 63
5.3 Breakup of the single-particle DOS $\rho_1(\epsilon) = \rho_1^{(0)}(\epsilon) + \rho_1^{(1)}(\epsilon) + \rho_1^{(2)}(\epsilon)$ for repulsive $U = 0.7$, split according to the site occupancies, as described by Eqns. 5.12, 5.13 and 5.14. Note that the contribution to $\rho_1(\epsilon)$ from singly-occupied sites ends at $\epsilon = \pm U$, exactly where the pair-DOS $\rho_2(\epsilon)$ begins. Inset: The dip in $\rho_1(\epsilon)$ corresponds to $\rho_1^{(1)}(\epsilon)$ going to zero at $\epsilon = \pm U$, as emphasized by the dashed lines. 66
5.4 Comparison of the DOS's for constant (left) and random (right) U, with strong scatter $\Delta U = 2$ (units of e^2/a). The sharp gaps and humps are smoothed out by disorder, but the overall trend of increase/decrease of the low energy DOS remain intact. 70

5.5 One activated resistance, $R(i, n_i, j, n_j; k, n_k, l, n_l)$, corresponds to the above two hops, as the particles are indistinguishable from each other 78
5.6 Top panel: Peak in the resistance upon tuning $\overline{U}$ at $T = 0.04$ vs. the average interaction $\overline{U}$ (with $\xi_2 = 4\xi_1$); and the fraction of pair hops in the percolating cluster (with resistances within 30% of the percolating resistance). The resistance peaks when roughly half of the critical resistors are pair and single moves, resp. Lower panel: fraction of paired electrons in a typical metastable state. This fraction smoothly decreases across the "mixed regime", since the bulk of such pairs is inactive in transport. 88

Chapter 1

Introduction

Duality is a subject that has shown up time and again in several fields of physics. Kramers and Wannier introduced duality in the square lattice Ising model [1] to derive the exact critical temperature three years before the model's exact solution by Onsager in 1944 [2]. Duality has since spread to many different models: the standard Potts, vector Potts, $x - y$, heisenberg, quantum Ising, and toric code models, just to name a few. Duality has been applied in general dimensions, for discrete and continuous variables on a number of algebras.

The idea is that auxiliary variables can be introduced into a hamiltonian wherever the elementary variables interact, the elementary variables may be summed out to leave constraints for the auxiliary variables, and these constraints may be satisfied by defining the auxiliary variables in terms of a set of variables at "plaquettes" of some sort. This transformation exactly relates the partition function of the original model to the partition function of the dual model. Often, the transformation maps the original couplings in such a way that large coupling maps to small and small to large, or maps low temperature to high and high to low.

Recently, Cobanera, Ortiz, and Nussinov have developed a technique for approaching dualities even more generally [3, 4, 5]. This bond-algebraic approach looks closely at the interactions of the model and uses them to derive dualities directly, paying particular attention to the link between gauge symmetries and dualities. Not only does this allow duality transformations to be studied more systematically, but it treats quantum and classical dualities far more closely than previous studies. Also, much of the work in the bond-algebraic approach can be done locally, even while including such things as the nonlocal Jordan-Wigner transformation with the discussion of duality.

However, there are still many models that could benefit from this technique. In particular, much of the study on duality has focused on dual models on the plaquettes of the original model's lattice, or similar features related to the embedding of the graph. In two dimensions with two-body interactions, the dual variables reside on the actual plaquettes of the lattice, while in three or higher dimensions or with multi-body interactions the dual variables may reside at some other lattice-dependent locations. With traditional dualities, models with long range interactions, complicated lattice structure, or especially a model on a network (with no lattice structure) could not be studied.

This removes many models from the consideration of duality. There is much interesting work for glassy systems on random networks or with long range interactions [6]. Many compelling computer science problems occur on random networks, such as the boolean satisfiability problem (SAT) [7, 8], the graph coloring problem [9], and the travelling salesman problem [10]. Evolution and artificial intelli-

gence often necessitate large complex networks for stimulating results [11, 12].

Fortunately, duality can be applied to these network problems. Simply including a nonzero external field along with the various interactions in the model is enough to make duality a lattice-independent transformation in many cases. The extra field means that extra auxiliary variables are introduced in the transformation and when constraints need be satisfied, these extra variables can automatically satisfy them, leaving the dual variables to be the auxiliary variables introduced at the site of the original interactions. The new dual interactions reside at the site of the original variables. Duality just switches the vertices and edges of the network, leaving the structure intact.

In this thesis, I will expand upon this idea, applying duality to several new models. I will first show duality in an external field in an illustrative and practical model, the Ising model, just as used originally by Kramers and Wannier. I will proceed to describe an application of duality to the SAT problem of computer science, mapping it to a completely different computer science problem. Then I will show how duality may be incorporated with message passing algorithms to give a new way to approximate on large tree-like networks. I will conclude with a discussion of a useful model for thin disordered films on the insulating side of the superconductor-insulator transition, a model that could benefit from a duality analysis in the vein of Fisher and Lee [13].

In Chapter 2, I examine the idea of duality in an external field. I first review the original Kramers-Wannier duality for the square lattice Ising model. Then I show the differences accompanying the introduction of an external field, including

the lattice-independence of the transformation. I go on to produce similar results for the vector-Potts and $x - y$ models, showcasing the generality of the phenomena and giving a foundation for further work on these models.

In Chapter 3, I apply Ising duality in an external field to study the SAT problem. I review this problem and briefly discuss its far-reaching importance and intriguing properties. I then show how it may be mapped onto a limit of many-body interacting Ising model in an external field which in turn may be mapped by duality to another Ising model. I continue by simplifying this model through the appropriate limiting behavior to solving an underconstrained system of Diophantine equations and summing over the positive solutions. I present examples verifying the SAT duality and showcasing important features of the system to be solved and the accompanying summation. Finally, I quickly apply duality to related satisfiability problems, Not All Equal SAT and 1-in-3 SAT.

In Chapter 4, I combine duality and a popular message passing approximation technique, belief propagation. I review the derivation and usage of belief propagation as a method to calculate thermodynamic quantities and its restriction to only apply to networks with no small loops, intrinsically latticeless. Then I apply belief propagation to the dual Ising model to discover another message passing procedure. I apply this procedure to various simple lattices and compare transition temperatures approximated with normal belief propagation approximation and the new message passing algorithm with the exact temperatures. Finally, I combine duality and belief propagation for the vector Potts model, extending the Ising results.

The content of Chapters 3 and 4 is based on an extension of work done in

collaboration with Benjamin Hsu and Victor Galitski that is under review for publication [14].

In the final part of my thesis, I shift from duality to discuss a simple model for disordered thin films on the insulating side of a superconducting-insulator transition, the two-component Coulomb glass model. This is a primarily classical model and is able to capture various important features of these films, including a magnetoresistance peak, by studying the coexistence of electrons and localized pairs with thermally induced variable-range hopping and a Coulomb interaction. There is some possibility of applying a duality to a generalization of the model, to study the fermions and localized bosons that coexist and possibly to relate the physics of superconductor and insulator. However, this is not presented in the thesis, in favor of a numerical analysis of the model.

In Chapter 5, I introduce the two-component Coulomb glass model and examine it at length. I first describe the model and various parameters of interest. I then discuss the density of states for the model and physics accessible from its study. I give an account of the resistor network used to study transport across the material. Finally, I discuss the transport results themselves, which include a reproduction of the magnetoresistance peak seen in experiment. The content of this chapter is based on published work done in collaboration with Anirban Gangopadhyay, Victor Galitski and Markus Müller [15].

Chapter 2

Lattice independent duality

Kramers and Wannier showed the power of duality applied to statistical mechanics in 1941, when they derived a relation between the 2D square lattice Ising model at high and at low temperature and through this relation determined the exact critical temperature [1]. Since then, dualities in statistical mechanics have been discovered and studied in an incredible range of models, both classical and quantum [16, 17, 5]. They have been used in a number of different applications, from describing difficult models in terms of their simpler dual models [18] to improving computational efficiency by equating normal and dual expansions to partially solve problems [19]. Furthermore, the recently invented bond algebraic dualities [3, 4, 5, 20] have not only grouped many of these classical and quantum dualities under a general framework, but also present methods to discover new dualities, perhaps even helping to deal with the notoriously difficult Non-Abelian models.

In this chapter, I will focus on duality when including an external field. The duality of models with an external field is interesting by itself [21, 22, 23], but duality also takes a particularly simple form with the inclusion of such a field.

To showcase this, I examine the Ising [24], vector Potts [23, 21, 25, 20], and $x-y$ [26, 27, 28] models, to give examples with both discrete and continuous degrees of freedom. In Section 2.1, I review the original Kramers-Wannier duality on the Ising model. Sections 2.2, 2.3, and 2.4 cover the dualities of the Ising, Potts, and $x-y$ models in an external field, respectively. I conclude in Section 2.5.

2.1 Kramers-Wannier duality

Assume a periodic square lattice with Ising spins $\sigma_s = \pm 1$ on its vertices, or *sites*. These spins interact across the edges, or *links*, of the lattice with negative (ferromagnetic) energy J. Here and for the rest of the dissertation, I absorb the temperature into the energy parameters, $\beta J \to J$.

$$\mathcal{H} = \sum_{\langle s,s' \rangle} J\sigma_s\sigma_{s'} \tag{2.1}$$

The partition function is

$$\mathcal{Z}(J) = \sum_{\{\sigma_s=\pm 1\}} \exp\left\{ \sum_{\langle s,s' \rangle} J\sigma_s\sigma_{s'} \right\} \tag{2.2}$$

Kramers-Wannier duality relies on a simple relation, an expansion of the exponentials.

$$\mathrm{e}^{J\sigma} = \cosh J + \sigma \sinh J = \sum_{k=0,1} C_k(J)\sigma^k \tag{2.3}$$

This is a discrete Fourier transformation, and introducing it for every exponential means that the original variables can be summed out and replaced with constraints on the new link variables. Let l be the links of the lattice, let $l \in \partial s$ be the links with interactions that involve σ_s, and let $s \in \partial l$ be the sites involved in the interaction

at link l. Note that $\sum_{\langle s,s'\rangle} J\sigma_s\sigma_{s'} = \sum_l J\prod_{s\in\partial l}\sigma_s$ and that $\prod_{s\in\partial l}\sigma_s = \pm 1$.

$$\begin{aligned}\mathcal{Z} &= \sum_{\{\sigma_s=\pm1,k_l=0,1\}}\left(\prod_l C_{k_l}(J)(\prod_{s\in\partial l}\sigma_s)^{k_l}\right) \\ &= \sum_{\{k_l=0,1\}}\left(\prod_l C_{k_l}(J)\right)\left(\prod_s\sum_{\sigma_s}\sigma_s^{\sum_{l\in\partial s}k_l}\right) \\ &= \sum_{\{k_l=0,1\}}\left(\prod_l C_{k_l}(J)\right)\left(\prod_s 2\delta_2\left(\sum_l k_l\right)\right)\end{aligned} \tag{2.4}$$

These steps will be outlined in detail in the next section. The constraint $\delta_2(n)$ is 1 if n is even and 0 if n is odd. The constraints may be satisfied by defining the link variables in terms of new spins introduced on the plaquettes of the lattice. Let d be the plaquettes, let $l \in \tilde{\partial}d$ be the links bordering plaquette d, and let $d \in \tilde{\partial}l$ be the two plaquettes with link l as a border. Introduce a spin $\nu_d = \pm 1$ at each of these plaquettes, so that $k_l = \frac{1}{2}(1 - \prod_{d\in\tilde{\partial}l}\nu_d)$. This definition satisfies all constraints:

$$\sum_{l\in\partial s} k_l = 2 - \tfrac{1}{2}\nu_1\nu_2 - \tfrac{1}{2}\nu_2\nu_3 - \tfrac{1}{2}\nu_3\nu_4 - \tfrac{1}{2}\nu_4\nu_1 \underset{\text{mod } 2}{\equiv} 0 \tag{2.5}$$

It can be shown that the definition is also necessary up to a global sign change in $\{\nu\}$, i.e. any set of $\{k\}$ that satisfies the constraints can be represented with the $\{\nu\}$. With this, the constraints vanish and the partition function becomes:

$$\mathcal{Z} = 2^N\sum_{\{\nu_d\}}\prod_l C_{\frac{1}{2}(1-\prod_{d\in\partial l}\nu_d)}(J) \tag{2.6}$$

Simplifying the $C(J)$ will turn this into a recognizable dual partition function.

$$C_{\frac{1}{2}(1-\nu)}(J) = (\cosh J\sinh J)^{1/2}\tanh^{-1/2\nu}J = (\cosh J\sinh J)^{1/2}\mathrm{e}^{\tilde{J}\nu}, \tag{2.7}$$

$$\tilde{J} = -\tfrac{1}{2}\ln\tanh J \tag{2.8}$$

$$\mathcal{Z} = \mathcal{Z}_0\sum_{\nu_d}\exp\left\{\sum_{\langle d,d'\rangle}\tilde{J}\nu_d\nu_{d'}\right\}, \qquad \mathcal{Z}_0 = (\sinh 2J)^N \tag{2.9}$$

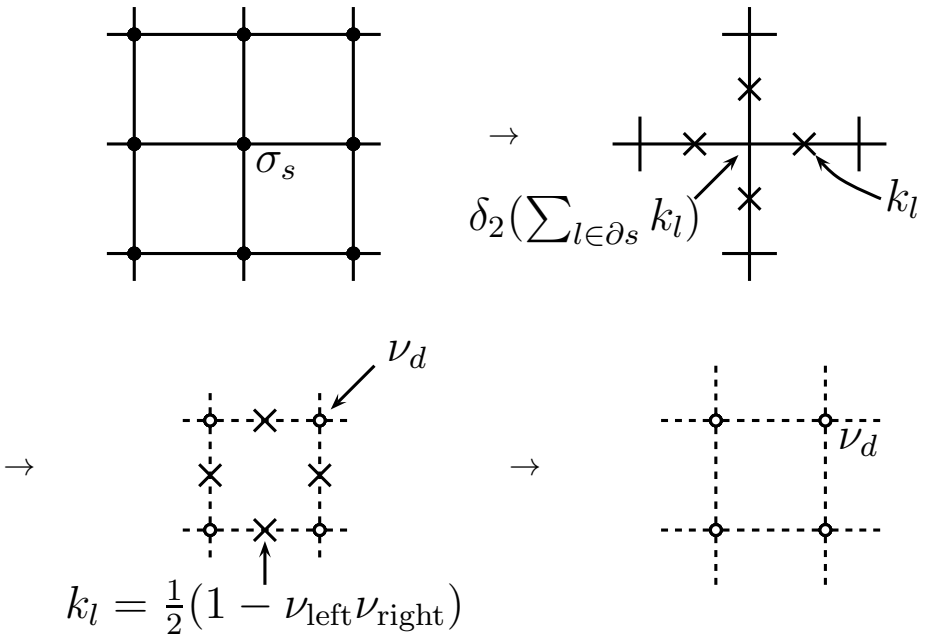

Figure 2.1: Behavior of spin, link, and dual spin variables under Kramers-Wannier duality transformation on a periodic square lattice.

In this way, the plaquettes of the original lattice become the vertices of the dual lattice. The graphical dual of the square lattice is itself, so the square lattice Ising model is self-dual.

$$\mathcal{Z}_{\text{Ising square lattice}}(J) = (\sinh 2J)^N \, \mathcal{Z}_{\text{Ising square lattice}}(\tilde{J}) \tag{2.10}$$

Note that the dual coupling constant, $\tilde{J}$, is a positive, monotonically decreasing function of the original coupling, seen in Fig. 2.2.

The transition temperature of the square lattice Ising model may now be simply calculated. At the ferromagnetic phase transition, the partition function should diverge. Since no part of Eq. (2.10) can diverge at finite J except for the partition function, the partition functions must diverge simultaneously on the left and right sides, and so $\mathcal{Z}_{\text{Ising square lattice}}$ must diverge at both J_C and $\tilde{J}_C$. Assuming

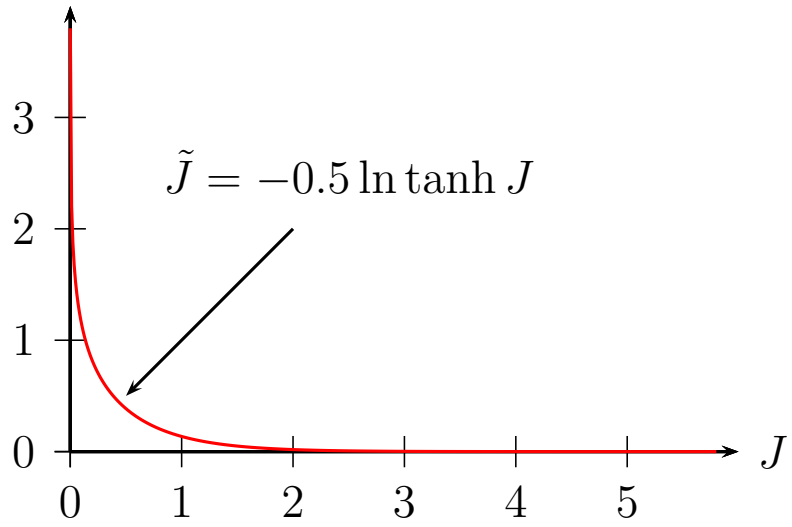

Figure 2.2: Behavior of dual coupling $\tilde{J}$ as a function of original coupling J.

there is only one transition at finite positive J, this means the transition occurs when $J_C = \tilde{J}_C$.

$$J_C = \tilde{J}_C = -\tfrac{1}{2} \ln \tanh J_C = \tfrac{1}{2} \ln(1 + \sqrt{2}) \tag{2.11}$$

This is just one example of the utility of duality. Kramers and Wannier derived this in their original work, and similar arguments may be made for other lattices, e.g. triangular. However, these others are not self dual and instead establish the relation $J_{C,\text{ original lattice}} = \tilde{J}_{C,\text{ dual lattice}}$, e.g. $J_{C,\text{ triangular}} = \tilde{J}_{C,\text{ hexagonal}}$.

Dualities of this sort have been generalized to a broad range of models [16, 17], including the vector Potts and $x - y$ models. Vector q-state Potts model:

$$\mathcal{H} = \sum_{\langle s,s' \rangle} J \cos\left(\frac{2\pi}{q}(n_s - n_{s'})\right) \qquad n_s = 0, 1, \ldots q - 1 \tag{2.12}$$

$x - y$ model:

$$\mathcal{H} = \sum_{\langle s,s' \rangle} J \cos(\theta_s - \theta_{s'}) \qquad \theta_s \in (-\pi, \pi] \tag{2.13}$$

I will specifically address these dualities in Sections 2.3 and 2.4.

Furthermore, the recently invented bond algebraic dualities [3, 4, 5, 20] have not only grouped many of these classical dualities and their quantum counterparts under a general framework, but also present methods to discover new dualities.

2.2 Ising Duality in an external field

The boolean satisfiability duality I derive in Chapter 3 is essentially the classic Kramers-Wannier duality for a specific Ising model with an external field. This additional field changes the constraints introduced so that the variables and interactions switch under duality, rather than the variables switching with the plaquettes.

There are multiple benefits to studying the model with an external field. First, many models that are related to the Ising model require such a field. Setting the field to zero will reproduce the normal Kramers-Wannier duality, although some care needs to be taken to appropriately remove the field, as seen later in this section. Finally, unlike normal duality, the duality produced with such a field is actually a lattice-independent transformation. This allows us to apply duality not only for models with complicated lattice structure, but also for models with multi-spin interactions and with networks of spins, with no lattice structure required.

I start by introducing some definitions. There is a network of sites, s, where the free variables reside, in this case classical $\mathbb{Z}_2$ spins $\sigma_s = \pm 1$. I introduce a external

field term in the hamiltonian and let this field be spatially inhomogeneous: $\mathcal{H}_h = -\sum_s h_s \sigma_s$. I introduce an interaction along the links, l, of the lattice between two neighboring sites, $s = l1$ and $s = l2$, and again can let it be spatially inhomogeneous: $\mathcal{H}_J = -\sum_l J_l \sigma_{l1} \sigma_{l2}$. The 2D square lattice, for example, has spins arranged in a square grid, with links between nearest neighbor spins in the $\hat{x}$ and $\hat{y}$ directions.

The sites and links may be thought of as vertices and edges of the graph describing the structure of variables and interactions. This graph need not have a specified embedding on a surface since the duality I discuss is not structure-changing, but the faces, or plaquettes, of the graph are important in the limit of zero external field. Later, I will extend the model to have multi-spin interactions; the graph will become a hypergraph, and the links will be hyperedges.

The partition function for the system:

$$\mathcal{Z} = \sum_{\{\sigma_s\}} \exp\left[\sum_s h_s \sigma_s + \sum_l J_l \sigma_{l1} \sigma_{l2}\right]. \tag{2.14}$$

Now I can perform the steps of duality, most closely resembling the derivation in Savit's review [16]. Now that there is the extra external field term, I must simply introduce two types of summation variables, $j_s \in \{0, 1\}$ and $k_l \in \{0, 1\}$.

$$\mathcal{Z} = \sum_{\{\sigma_s\}} \sum_{\{j_s, k_l\}} \left(\prod_s C_{j_s}(h_s)(\sigma_s)^{j_s}\right) \left(\prod_l C_{k_l}(J)(\sigma_{l1}\sigma_{l2})^{k_l}\right) \tag{2.15}$$

$$C_0(x) = \cosh x, \qquad C_1(x) = \sinh x. \tag{2.16}$$

As before, ∂s is the set of links which connect site s to other sites, and ∂l is the set of sites which have variables involved in the interaction at link l, in the Ising model

just two sites $s = l1, l2$. Now I pull all σ_s for each site into its own product:

$$\mathcal{Z} = \sum_{\{\sigma_s, j_s, k_l\}} \left(\prod_s C_{j_s}(h_s) \right) \left(\prod_l C_{k_l}(J_l) \right) \left(\prod_s \sigma_s^{j_s + \sum_{l \in \partial s} k_l} \right). \tag{2.17}$$

$\sum_{l \in \partial s} k_l$ is the sum of k_l over all links around site s. With this, the spins can each be separately summed out to leave δ_2, the Kronecker delta function, mod 2, since the sum is 0 if $j_s + \sum_{l \in \partial s} k_l$ is odd and 2 if even.

$$\mathcal{Z} = \sum_{\{j_s, k_l\}} \left(\prod_s C_{j_s}(h_s) \right) \left(\prod_l C_{k_l}(J_l) \right) \left(\prod_s 2\delta_2(j_s + \sum_{l \in \partial s} k_l) \right). \tag{2.18}$$

Normally, this is where a subset of the link variables, k_l, is chosen to automatically satisfy the delta function, and this subset reveals new elementary dual variables associated with the plaquettes of the original lattice. However, now I can satisfy every delta function by setting site variable j_s. By picking $j_s = -\sum_{l \in \partial s} k, \text{ mod } 2$, the partition function is expressed with no constraints for the full set of k_l.

See that setting h_s to zero means a summand is only nonzero if $j_s = 0$, so j_s is no longer free to be chosen to satisfy the delta function. This freedom is only allowed by nonzero external field.

With the expression of k in terms of spinlike variables, $k_l = \frac{1}{2}(1 - \nu_l)$, $\nu_l = \pm 1$, it can be shown that $j_s = \frac{1}{2}(1 - \prod_{l \in \partial s} \nu_l)$, expressed with the product of ν_l on the links around site s. Now the functions C can be simplified by noting that

$$C_{k_l}(J_l) = (\tfrac{1}{2} \sinh 2J_l)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2} \ln \tanh J_l \cdot \nu_l \right], \tag{2.19}$$

$$C_{j_s}(h_s) = (\tfrac{1}{2} \sinh 2h_s)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2} \ln \tanh h_s \cdot \prod_{l \in \partial s} \nu_l \right]. \tag{2.20}$$

Finally, I can reexpress the partition function in terms of dual variables and dual

couplings.

$$\mathcal{Z} = \left(\prod_s 2\sinh 2h_s\right)^{1/2} \left(\prod_l \tfrac{1}{2}\sinh 2J_l\right)^{1/2} \sum_{\{\nu_l\}} \exp\left[\sum_l \tilde{J}_l \nu_l + \sum_s \tilde{h}_s \prod_{l\in\partial s} \nu_l\right], \tag{2.21}$$

$$\tilde{h}_s = -\tfrac{1}{2}\ln\tanh h_s, \qquad \tilde{J}_l = -\tfrac{1}{2}\ln\tanh J_l.$$

Consider a 1D ring Ising model. The links are just between adjacent sites, so $\prod_{l\in\partial s}\nu_l$ will be the product of adjacent spins, giving us the Ising model again. With an external field included, the Ising model is self dual in 1D instead of the usual 2D square lattice [5]. The dual forms for both external field and interaction parameters are identical to the Kramers-Wannier form, but they have switched roles so that the interaction acts as a dual external field on the new link spins and the external field acts as a dual interaction between adjacent link spins, across where the original sites existed.

Consider the 2D square lattice. The only thing that changes is the type of interaction, $\prod_{l\in\partial s}\nu_l$; the dual spins are still on the links, the external field and interaction switch roles, and the dual interaction operates across where the original sites existed. Now instead of the Ising model with external field, the dual model looks like Ising ring exchange on a checkerboard lattice with external field.

Now the structure of the duality transformation is apparent. With an external field term included, it is the variables and the interactions that are dual to one another. The sites and links switch. This means that unlike traditional duality, the structure of the network of variables and interactions persists. While without an external field, the dual model for an triangular lattice Ising model is a hexagonal

lattice Ising model, with an external field the dual model is again triangular, but with the variables on the edges of the original lattice while the interactions occur across the vertices.

At no point in the derivation did I use the fact that interactions took place between specifically two variables, i.e. that $\prod_{s\in\partial l}\sigma_s = \sigma_{l1}\sigma_{l2}$. It is simple to moderately generalize the classical spin model.

$$\mathcal{Z} = \sum_{\sigma_s} \exp\left[\sum_s h_s\sigma_s + \sum_l J_l \prod_{s\in\partial l}\sigma_s\right]. \tag{2.22}$$

Under the duality transformation, every link becomes the site for a dual variable, and every site becomes the link for a dual interaction. The dual to this partition function is exactly that seen in Eq. (2.21), where the links l now can involve any number of spins.

Other possibly interesting dualities can be quickly deduced from this duality relation; for example, 2D square lattice Ising model with external field and four variable interactions across the plaquettes instead of the normal two variable interactions is self dual. Other interesting Ising spin models fall into this category [22].

I will now examine how this duality reduces to, for example, normal Kramers-Wannier duality on the 2D square lattice. The Ising model with nonzero h_s gives under duality a checkerboard ring exchange model. See what happens if we take $h_s \to 0$:

$$(2\sinh 2h_s)^{\frac{1}{2}} \mathrm{e}^{\tilde{h}_s \prod_{l\in\partial s}\nu_l} \to \sqrt{4h_s (h_s)^{-\prod_{l\in\partial s}\nu_l}} \to \begin{cases} 2, & \text{for } \prod_{l\in\partial s}\nu_l = 1; \\ 0, & \text{for } \pi_{l\in\partial s}\nu_l = -1. \end{cases} \tag{2.23}$$

So the only sets of ν_l that contribute are those who have $\prod_{l\in\partial s}\nu_l = 1$ for all sites.

But this gives the rest of the constraint that the normal Ising model acquires under duality, and the link spins must be expressed by a product of new spins in adjacent plaquettes, μ_d,

$$\mathcal{Z} = 2^{(2N-L)/2} \left(\prod_l \sinh 2J_l \right) \sum_{\{\mu_d\}} \exp \left[\sum_l \tilde{J}_l \mu_{l1} \mu_{l2} \right]. \tag{2.24}$$

This is exactly Kramers-Wannier duality. It should be pointed out that this splits the normal duality into two separate ideas. The celebrated mapping between low and high temperatures or large and small couplings results from the switch between variables and interactions, while the structural change switching the variable locations to the plaquettes of the original lattice is caused by zero external field, preserving the $\mathbb{Z}_2$ symmetry.

Finally, I would like to put focus on negative h_s or J_l, since from the definitions in Eq. (2.21) these would give complex and multivalued dual couplings. Even naively, none of the derivation breaks down with such couplings, and complex couplings in an Ising model could be managed and are of interest [29]. An advantage in this case is that the dual couplings take a very specific form.

A careful rederivation with the possibility for $h_s < 0$ and $J_l < 0$ shows that the dual model has real and single-valued Boltzmann terms, though they are no longer necessarily non-negative:

$$\begin{aligned} \mathcal{Z} = &\left(\prod_s 2 \sinh 2|h_s| \right)^{\frac{1}{2}} \left(\prod_l \tfrac{1}{2} \sinh 2|J_l| \right)^{\frac{1}{2}} \\ &\times \sum_{\{\nu_l\}} \left(\prod_s \mathrm{sgn} h_s^{\frac{1-\prod_{l \in \partial s} \nu_l}{2}} \right) \left(\prod_l \mathrm{sgn} J_l^{\frac{1-\nu_l}{2}} \right) \exp \left[\sum_l |\tilde{J}_l| \nu_l + \sum_s |\tilde{h_s}| \prod_{l \in \partial s} \nu_l \right], \end{aligned} \tag{2.25}$$

Note that the new dual couplings themselves, $|\tilde{h_s}|$ and $|\tilde{J}_l|$, are all non-negative.

In a sense, duality trades frustration from negative couplings for signed Boltzmann terms.

2.3 Vector Potts duality

Duality has been studied at length for the standard Potts model and for the vector Potts model considered here [23, 21, 25, 17, 16, 20]. I present the following duality in an external field primarily as a generalization of the Ising duality (since the $q = 2$ vector Potts model is the Ising model) and to showcase the persistence of the ideas of Section 2.2.

Consider the partition function for the vector Potts model in an external field. I will continue to use a general lattice with sites, s, and links, l, as for the Ising duality, although in this case only two-variable interactions are considered.

$$\mathcal{Z} = \sum_{\{n_s\}} \exp\left\{ \sum_s h_s \cos\left[\frac{2\pi}{q}(n_s - p_s)\right] + \sum_l J_l \cos\left[\frac{2\pi}{q}(n_{l1} - n_{l2})\right] \right\} \tag{2.26}$$

The variables are $n_s = 0, 1, \ldots q-1$, and there is an inhomogeneous, (two-dimensional) directional external field applied, with strength h_s and direction $\phi_s = \frac{2\pi p_s}{q}$. The usual method for discovering duality is to Fourier expand the Boltzmann terms. This expansion is aided by the generalized hyperbolic functions.

$$\tfrac{1}{q} \sum_{n=0,1,\ldots q-1} \mathrm{e}^{-\frac{2\pi i m n}{q}} \mathrm{e}^{J\mathrm{e}^{\frac{2\pi i n}{q}}} = F_{q,m}(J) = \sum_{k=0}^{\infty} \frac{J^{qk+m}}{(qk+m)!} \tag{2.27}$$

Generalized hyperbolic functions such as $F_{q,m}(J)$ ($q \in \mathbb{N}$, $m \in \mathbb{Z}$, $J \in \mathbb{C}$) were first recorded by Riccati with the introduction of the hyperbolic functions in 1957. They are natural generalizations of the exponential ($F_{1,0}(J) = \mathrm{e}^x$) and hyperbolic

Periodic in m:	$F_{q,m+q}(J) = F_{q,m}(J)$
Sum to exponential:	$\sum_{m=0}^{q-1} F_{q,m}(J) = \mathrm{e}^{J}$
Derivative:	$\frac{\mathrm{d}}{\mathrm{d}J} F_{q,m}(J) = F_{q,m-1}(J)$
Addition formula:	$\sum_{n=0}^{q-1} F_{q,n}(x) F_{q,m-n}(y) = F_{q,m}(x+y)$
Phase:	$F_{q,m}(\mathrm{e}^{\frac{2\pi \mathrm{i} n}{q}} J) = \mathrm{e}^{\frac{2\pi \mathrm{i} n m}{q}} F_{q,m}(J)$
Pythagorean identity:	$\prod_{n=0}^{q-1} \sum_{m=0}^{q-1} \mathrm{e}^{\frac{2\pi \mathrm{i} m n}{q}} F_{q,m}(J) = 1$

Table 2.1: Properties of the generalized hyperbolic functions

($F_{2,0} = \cosh J$, $F_{2,1} = \sinh J$) functions and have several nice properties, as seen in Table 2.1. Notably, these functions define the circulant matrix entries for the solutions to the ODE $x^{(q)} = \omega^q x$ at time $T = J/\omega$. Duality for the vector Potts model may be derived with the help of these functions. A discrete Fourier expansion of the Boltzmann terms in the partition function and Parseval's theorem gives

$$\frac{1}{q} \sum_{n=0}^{q-1} \mathrm{e}^{-\frac{2\pi \mathrm{i} m n}{q}} \mathrm{e}^{J \cos \frac{2\pi n}{q}} = \sum_{k=0}^{q-1} F_{q,k}(J/2) F_{q,m+k}(J/2) =: G_{q,m}(J) \tag{2.28}$$

$$\mathcal{Z} = \sum_{\{n_s, m_s, m_l\}} \left(\prod_s G_{q,m_s}(h_s) \right) \left(\prod_l G_{q,m_l}(J_l) \right) \mathrm{e}^{\sum_s \frac{2\pi \mathrm{i} m_s (n_s - p_s)}{q} + \sum_l \frac{2\pi \mathrm{i} m_l (n_{l1} - n_{l2})}{q}} \tag{2.29}$$

$$\begin{aligned} &= \sum_{\{m_l\}} \mathrm{e}^{\sum_s \frac{-2\pi \mathrm{i} \Delta_s m p_s}{q}} \left(\prod_s G_{q,-\Delta_s m}(h_s) \right) \left(\prod_l G_{q,m_l}(J_l) \right) \\ &= \sum_{\{m_l\}} \left(\prod_s \exp\left(\frac{2\pi \mathrm{i} \Delta_s m p_s}{q} \right) \right) \exp\left(\sum_s \ln G_{q,-\Delta_s m}(h_s) + \sum_l \ln G_{q,m_l}(J_l) \right) \end{aligned}$$

Here, $\Delta_s m$ is a sum of dual variables m_l over links connected to site s, but with opposite signed m_l for sites on each side of the link.

This model is not typically self dual on any lattice. Of course, $q = 2$ returns

this analysis to the Ising model which is self dual on a line, and $q = 3, 4$ are similarly self dual on a line. There is also a more general model, the Z_N model, which has self dual special cases, e.g. the Villain model.

The lattice independent structure of the duality can nonetheless be seen in this model. The interactions in the original model become external field-like interactions in the dual model, $\ln G_{q,m_l}(J_l)$, while the original external fields become interactions involving all links the site is connected to, $\ln G_{q,\Delta_s m}(h_s)$. Also note the alternating sign seen in the Ising model when $h_s < 0$ reappears here with the product of cosines. If the original external field does not bias for $n_s = 0$ (with $p_s = 0$ and $h_s > 0$), then the cosine will complicate the dual model, reflecting the complicating frustration of the original model.

In addition, the low temperature limit $h_s \to \infty$, $J_l \to \infty$ turns the dual model into a Potts-like model, just as in the case without external field [16, 5].

$$G_{q,m}(J \gg 1) \approx \mathrm{e}^{J} + 2\cos\frac{2\pi m}{q}\mathrm{e}^{J\cos\frac{2\pi}{q}} \approx \exp\left(J + 2\cos\tfrac{2\pi m}{q}\mathrm{e}^{J(\cos\frac{2\pi}{q}-1)}\right) \tag{2.30}$$

$$\begin{aligned}\mathcal{Z} &\approx \sum_{\{m_l\}}\left(\prod_s \exp\frac{2\pi\mathrm{i}\Delta_s m p_s}{q}\right)\exp\left[\sum_s\left(h_s + 2\mathrm{e}^{h_s(\cos\frac{2\pi}{q}-1)}\cos\frac{2\pi\Delta_s m}{q}\right)\right. \\ &\qquad \left. + \sum_l\left(J_l + 2\mathrm{e}^{J_l(\cos\frac{2\pi}{q}-1)}\cos\frac{2\pi m_l}{q}\right)\right] \qquad (2.31)\end{aligned}$$

$$\begin{aligned}&= \mathrm{e}^{\sum_s h_s + \sum_l J_l}\sum_{\{m_l\}}\left(\prod_s \exp\tfrac{2\pi\mathrm{i}\Delta_s m p_s}{q}\right)\exp\left(\sum_s \tilde{h}_s\cos\tfrac{2\pi\Delta_s m}{q} + \sum_l \tilde{J}_l\cos\tfrac{2\pi m_l}{q}\right) \\ &\qquad \tilde{h}_s = 2\mathrm{e}^{-h_s(1-\cos\frac{2\pi}{q})}, \qquad \tilde{J}_l = 2\mathrm{e}^{-J_l(1-\cos\frac{2\pi}{q})} \qquad (2.32)\end{aligned}$$

In 1D, this is the Potts model with exponentially small interaction parameters $\tilde{h}_s$ and $\tilde{J}_l$. In 2D, the interactions will involve four variables, but is otherwise similar to the Potts model.

2.4 $x - y$ duality

I have shown the form this duality takes in the Ising and vector Potts models, and now I will look at a model with continuous variables, the $x - y$ model.

Note that this model can be obtained as a limit of the Potts model with $q \to \infty$. I will demonstrate that much of the general reasoning above applies for continuous variables. The derivation is very similar to the previous section, with a different expansion.

$$\mathcal{H} = -\sum_s h_s \cos(\theta_s - \phi_s) - \sum_l J_l \cos(\theta_{l1} - \theta_{l2}). \tag{2.33}$$

The elementary variables are now $\theta_s \in (-\pi, \pi]$ at every site s. ϕ_s is an inhomogeneous direction for the external field, similar to the p_s in the vector Potts model. The partition function is

$$\mathcal{Z} = \left(\prod_s \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta_s}{2\pi} \right) \mathrm{e}^{\sum_s h_s \cos(\theta_s - \phi_s) + \sum_l J_l \cos(\theta_{l1} - \theta_{l2})}. \tag{2.34}$$

Expanding in Fourier components gives modified Bessel functions,

$$\mathrm{e}^{J \cos\tau} = \sum_{-\infty < \phi < \infty} \mathrm{e}^{\mathrm{i}\phi\tau} I_\phi(J), \tag{2.35}$$

$$\mathcal{Z} = \sum_{\{\psi_s, \phi_l\}} \left(\prod_s I_{\psi_s}(h_s) \right) \left(\prod_l I_{\phi_l}(J_l) \right) \left(\prod_s \int \frac{\mathrm{d}\theta_s}{2\pi} \mathrm{e}^{\mathrm{i}\theta_s(\psi_s + \Delta_s\phi)} \right), \tag{2.36}$$

$$\mathcal{Z} = \sum_{\{\phi_l \in \mathbb{Z}\}} \exp\left[\sum_l \ln I_{\phi_l}(J_l) + \sum_s \ln I_{-\Delta_s\phi}(h_s) \right]. \tag{2.37}$$

This is the dual model. Again, $\Delta_s\phi$ is a sum of ϕ_l with opposite signs for sites on each side of the link. The dual variables exist on the links and the dual interactions, $\ln I_{-\Delta_s\phi}(h_s)$, occur across the original sites. However, in this case the form as

expressed is unfamiliar and unintuitive. In order to go further, restrict to the low temperature regime. Specifically, the temperature is much lower than both the external field and interaction parameters, $h_s >> 1$ and $J_l >> 1$. Note that this means after this point I cannot set the external field to zero without ruining the approximation, so I cannot look for the normal dual to the $2D$ XY model.

With this assumption [18], $\ln I_n(x) \approx \ln I_0(x) - \frac{1}{2x}n^2$. Let $C = (\prod_s I_0(h_s))(\prod_l I_0(J_l))$, so that

$$\mathcal{Z} = C \sum_{\{\phi_l \in \mathbb{Z}\}} \exp\left[-\sum_l \frac{1}{2J_l}\phi_l^2 - \sum_s \frac{1}{2h_s}(\Delta_s \phi)^2 \right]. \tag{2.38}$$

At this point, I can introduce a set of integers, m_l, to make ϕ_l continuous with the Poisson summation formula

$$\mathcal{Z} = C \sum_{\{m_l \in \mathbb{Z}\}} \left(\prod_l \int \frac{\mathrm{d}\phi_l}{2\pi} \right) \exp\left[-\sum_l \frac{1}{2J_l}\phi_l^2 - \sum_s \frac{1}{2h_s}(\Delta_s \phi)^2 + 2\pi \mathrm{i} m_l \phi_l \right]. \tag{2.39}$$

Assuming h_s and J_l are spatially constant and using a square lattice for concreteness, the ϕ_l may be integrated away with a Fourier transformation:

$$\mathcal{Z} = \mathcal{Z}_0 \sum_{\{m_q \in \mathbb{Z}\}} \exp\left[\int \mathrm{d}^2 q \frac{1}{4\Gamma(q)} |m_q|^2 \right], \tag{2.40}$$

$$\mathcal{Z}_0 = \left(\prod_q \sqrt{\frac{1}{4\pi\Gamma(q)}} \right) \left(\prod_s I_0(h_s) \right) \left(\prod_l I_0(J_l) \right), \tag{2.41}$$

$$\Gamma(q) = \tfrac{1}{2J} - \tfrac{2}{h}\left(\sin\tfrac{q_x}{2} + \sin\tfrac{q_y}{2} \right)^2. \tag{2.42}$$

Finally, I should return to the real space representation, however the exact form in real space is difficult to acquire, and I do not show this here, as this is enough to see that the calculation is little more difficult than that for the XY model without external field. Full details of the vortex unbinding at high temperature using a similar duality relation are given by Fertig [28, 30].

Once again see that the exact duality can be interpreted as switching the sites with the links, the elementary variables with the interactions, along with switching strong coupling with weak coupling, as seen in Eq. (2.37). The dual model is as geometry independent as the Ising model. Switching from the discrete variables of the Ising model to continuous variables has not changed the crucial simplification and generality afforded by including an external field coupling into the model.

2.5 Conclusion

This chapter has shown that when extending the usual classical models by including a coupling to an external field, dualities may still be derived. In fact, including the field shows that normal duality factors into two separate notions. First, duality possibly changes the nature of the variables and of the interactions, it switches the locations of the variables with that of the interactions, and it often maps large coupling to small coupling and vice versa. Second, if there is no external field applied then there is an additional constraint on the system that changes the structure of the network and calls for reduced dual variables and dual interactions on a new network. On a 2D planar lattice, for example, the reduced dual variables sit on the plaquettes of the original lattice while the reduced dual interactions take place across the original links.

This reiterates a point brought up by Cobanera et al. [5]: symmetry and duality are often inaccurately related. Without the symmetry afforded by zero external field, duality can still be performed. This symmetry may allow a connection

between more disparate models since it changes the lattice structure, for example the connection between the triangular and hexagonal lattices or face-centered cubic and diamond lattices. However, duality without enforcing the symmetry allows the freedom to deal with more complicated lattices or no lattice at all. This will be important applying Ising duality to the SAT problem in the next chapter, as most study of this problem is for random or at least non-lattice networks of variables and interactions.

In addition, there are tools that could be combined with duality that are only present on networks with no small loops or no lattice structure. In particular, message passing algorithms are heavily used in these kinds of glassy, frustrated models but rely on the assumption of few small loops in the network. I will discuss combining duality and message passing in Chapter 4.

Chapter 3

Boolean satisfiability

Computationally intractable problems are ubiquitous and occur in many areas of the natural and computational sciences, with problems as diverse as spin glasses, optimization problems, and cryptography [31, 32, 6, 33]. Generally speaking, they are problems whose solutions require a time that grows faster than polynomial in the input size N of the problem. Understanding what sorts of problems have solutions that require only polynomial time is an important issue that has many practical implications [34].

However, proving that a problem is indeed computationally intractable is by itself a difficult task. One tool utilized by theoretical computer scientists is mapping between problems with a known computational complexity. The relative difficulty of two problems may be established by mapping one problem into another. In this chapter, I use for the first time Kramers-Wannier duality to analyze the famous boolean satisfiability problem. I discover a new and exact dual formulation of the counting problem, known as #SAT, which maps onto an under-constrained system of Diophantine equations.

I use the classic Kramers-Wannier duality for an Ising model in an external field, discussed in detail in Chapter 2. In addition to more traditional statistical mechanical models, the unnecessity of a defined lattice allows a study of the SAT problem. This problem considers a set of boolean clauses and asks if they are simultaneously satisfiable by some instance of the boolean variables in the clauses. It is extremely important in computational complexity theory and computer science in general, not least because its complexity is at a significant position in the theory of complexity classes and because it is closely connected to the famous question P$\stackrel{?}{=}$NP.

First, I review the SAT problem and its importance in Section 3.1. In Sections 3.2 and 3.3, I use the duality developed in Chapter 2 to derive a SAT duality, showing the equivalence between the SAT counting problem and enumerating the positive solutions to a linear Diophantine system of equations. Section 3.4 discusses this duality and displays some concrete examples. Section 3.5 deals with related SAT problems, and in Section 3.6 I summarize and also discuss future work. [1]

3.1 The SAT problem

Let $\{x_s = T, F\}$ be a set of N boolean variables. Define a "clause" as the logical OR ($\vee$) of some combination of these variables, possibly with negation, e.g. $(x_1 \vee \bar{x}_2 \vee x_5)$. A SAT problem asks whether some set of M clauses of the N boolean variables may be simultaneously satisfied by some assignment of the variables. SAT

[1] The content of this chapter is based on work done in collaboration with Benjamin Hsu and Victor Galitski [14].

problems that have the same number k of variables in each clause are known as k-SAT problems.

This problem is in the non-deterministic polynomial time computational complexity class (NP), meaning it is verifiable, but not necessarily solvable, in polynomial time. It is in fact an NP-complete problem, meaning that in addition to being in NP itself, it is also as difficult as all other problems in NP [7, 8]. A polynomial time solution for this problem would imply a polynomial time solution for all NP problems, proving that all problems verifiable in polynomial time are also solvable in polynomial time, or P=NP. On the other hand, a proof that there is no polynomial time algorithm to solve any SAT problem would prove that P$\neq$NP. The SAT problem is actually the first proven example of an NP-complete problem, and many other NP-complete problems were proven so by reducing the SAT problem to them.

It is straightforward to examine the SAT problem from a statistical mechanics standpoint. Consider that the boolean variables are equivalent to Ising variables and that the clauses can be expressed as hamiltonian contributions that are zero if satisfied and positive otherwise, represented as a combination of multispin interactions. Many methods used to study sets of random SAT problems are also used to study disordered Ising models, including the replica method, belief propogation, and survey propogation [35, 36, 6].

Random k-SAT problems with a large number of variables and clauses have a variety of interesting behavior and have been a topic of study in statistical mechanics for some time. There are a number of transitions related to the organization of satisfactory solutions in the space of possible solutions, dependent on the ratio of

number of clauses to number of variables, $\alpha_c = M/N$. There is a transition, α_c below which nearly all random instances are satisfiable and above which nearly all are not. For $k > 3$, there is an intermediate state below α_c with a drastically increased median running time for many of the best known SAT solvers to decide if the instance is satisfiable. This exponential slow down is presumed closely related to a spin glass transition in an equivalent disordered Ising model [37, 38]. One of the primary motivations for studying duality in SAT problems is because of the close connection between phase transitions and duality.

To be specific, I am concerned with calculating the number of assignments of the boolean variables that make the set of clauses true. This problem is known as sharp-SAT, or #SAT. While the "counting problem", #SAT, is closely related to the "decision problem", SAT, they are distinct. SAT is in the NP-complete class of problems; it is in NP and is as difficult as any other problem in NP. #SAT is in the #P-complete class, where #P roughly is the class of counting problems associated with decision problems that are in NP. Because Kramers-Wannier duality is a relation between the partition functions of two problems and not, for example, between low energy states, this duality analysis will be able to investigate counting problems but not decision problems without modification.

3.2 Duality for the SAT problem through the Ising model

A SAT instance can be mapped onto a spin model. Define the Ising spin variable as $\sigma_s = (-1)^{x_s}$ where $x_s = 0, 1$ is the boolean variable in a given clause,

and define c_s^μ to be -1 if x_s is involved in clause μ with a negation, 1 if it is involved without a negation, and 0 if it is not involved in the clause. One can define an energy cost for a violated clause as proportional to a large number $X_\mu > 0$ with the hamiltonian

$$\mathcal{H} = \sum_\mu X_\mu \left[\prod_s (1 - c_s^\mu \sigma_s) \right]. \tag{3.1}$$

It is easy to see that $\mathcal{H} \geq 0$. The decision problem determines if this hamiltonian has any ground states with zero energy, while the counting problem calculates how many states have zero energy. The latter can be compactly described by calculating the associated partition function, $\mathcal{Z} = \sum_{\{\sigma_s\}} \mathrm{e}^{-\mathcal{H}}$, in the limit as each $X_\mu \to \infty$.

This hamiltonian can be rearranged to suggest its nature as an Ising spin glass:

$$\mathcal{H} = \mathcal{H}_0 - \sum_s h_s \sigma_s - \sum_l J_l \prod_{s \in \partial l} \sigma_s. \tag{3.2}$$

$$\mathcal{H}_0 = \sum_\mu X_\mu, \qquad h_s = \sum_\mu X_\mu c_s^\mu, \qquad J_l = -\sum_\mu X_\mu \prod_{s \in \partial l} (-c_s^\mu). \tag{3.3}$$

Note that each of $\mathcal{H}_0$, h_s, and J_l are linearly proportional to the large constants X_μ. This is the model of a network of Ising spins with multispin interactions and a site-dependent magnetic fielde, e.g. there are two- and three-body interactions in the case of 3-SAT.

The duality for this Ising spin glass was discussed in Chapter 2.

$$\begin{aligned} \mathcal{Z} = &\left(\prod_s 2 \sinh 2|h_s| \right)^{\frac{1}{2}} \left(\prod_l \tfrac{1}{2} \sinh 2|J_l| \right)^{\frac{1}{2}} \\ &\times \sum_{\{\nu_l\}} \left(\prod_s \mathrm{sgn} h_s^{\frac{1 - \prod_{l \in \partial s} \nu_l}{2}} \right) \left(\prod_l \mathrm{sgn} J_l^{\frac{1-\nu_l}{2}} \right) \exp \left[\sum_l |\tilde{J}_l| \nu_l + \sum_s |\tilde{h}_s| \prod_{l \in \partial s} \nu_l \right], \end{aligned} \tag{3.4}$$

Now I will simplify using the specifics of the SAT model. The limit $X_\mu \to \infty$ must be taken prudently, Taylor expanding the partition function in terms of e^{X_μ}. With

careful work, the partition function becomes:

$$\mathcal{Z} = \sum_{k_s, m_l \in \mathbb{Z}_+} \mathrm{e}^{-\mathcal{S}_{\{k,m\}}} (-1)^{\sum_s k_s + \sum_l m_l} \times z_{\{k,m\}} \tag{3.5}$$

$$z_{\{k,m\}} = \sum_{\{\alpha_s\}} \left(\prod_s D_{k_s}^{\mathrm{sgn} h_s \alpha_s} \right) \left(\prod_l D_{m_l}^{\mathrm{sgn} J_l \prod_{s \in \partial l} \alpha_s} \right), \tag{3.6}$$

$$D_k^\alpha = \tfrac{1}{2}(1+\alpha) - \tfrac{1}{2}(1-\alpha)(1-\delta_{k,0}), \tag{3.7}$$

$$\mathcal{S}_{\{k,m\}} = \sum_\mu X_\mu \Bigg[1 - \sum_s (1-2k_s) \mathrm{sgn} h_s c_s^\mu - \sum_l (1-2m_l) \mathrm{sgn} J_l \prod_{s \in \partial l} c_s^\mu \Bigg]. \tag{3.8}$$

3.3 Simplification to a Diophantine system of equations

This does not seem a particularly enlightening form of the partition function, or of the #SAT problem. We have introduced extra summations over the natural numbers for each site and link. However, there are two points that should be considered.

First, consider the X_μ independent part of the partition function, Eq. (3.6). Given a set of $\{k_s, m_l\}$, this calculation is simpler than it appears, and the α_s summation can be performed in time polynomial in N. Effectively, this is an unimportant calculation when compared to the full #SAT.

The second point to consider is that the limit $X_\mu \to \infty$ will greatly restrict the summation over k_s and m_l. Since each X_μ is independent, I can set them so that X_μ is much greater than $X_{\mu+1}$. Given this, $\mathrm{sgn} h_s = c_s^\mu$ in the *first* clause μ where c_s^μ is nonzero, and $\mathrm{sgn} J_l = -\prod_{s \in \partial l} (-c_s^\mu)$ in the first clause where the product is nonzero.

SAT instance $\{\bigvee_{s\in\mu} x_s,\ \text{clause } \mu\}$

$\downarrow$ Large energy X_μ for unsatisfied clauses

Ising spin glass $\mathcal{H} = \sum_s h_s\sigma_s + \sum_l J_l \prod_{s\in\partial l}\sigma_s$

$\downarrow$ Kramers-Wannier duality

Dual Ising spin glass $\tilde{\mathcal{H}} = \sum_s \tilde{H}_s \prod_{l\in\partial s}\nu_l + \sum_l \tilde{T}_l\nu_l$

$\downarrow X_\mu \to \infty$

Diophantine system $A^\mu = \sum_s B_s^\mu k_s + \sum_l C_l^\mu m_l$

Figure 3.1: Schematic of the SAT duality process

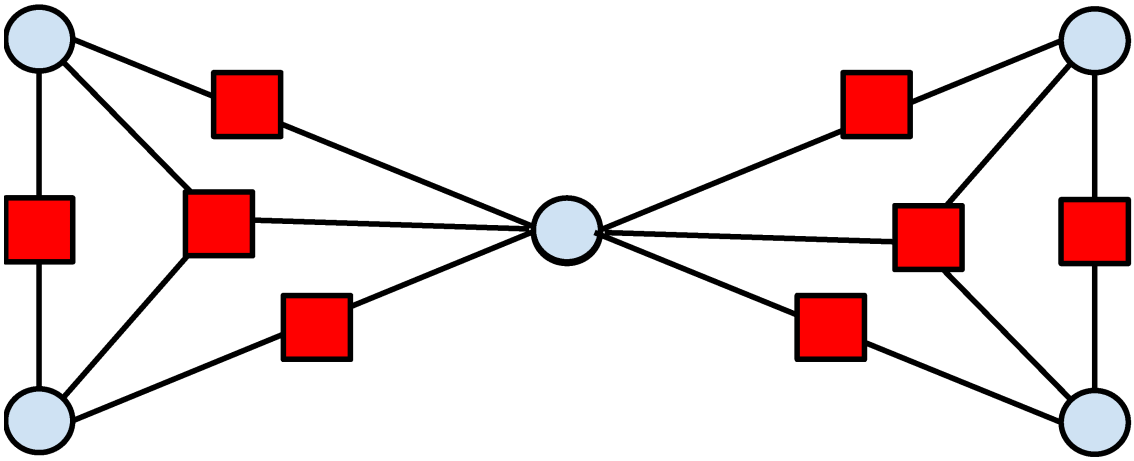

Figure 3.2: A SAT instance, $\mathcal{F} = (x_1 \vee x_2 \vee x_3) \wedge (x_3 \vee x_4 \vee x_5)$. Each circle represents a site, the location of a spin associated with a boolean variable. Each square represents a link, the location of an interaction (constraint) between spins. In the dual picture, the spins are exchanged with interactions.

Now consider $S_{\{k,m\}}$, a sum of terms proportional to the X_μ. If this sum is positive, then the contribution to the partition function vanishes. If the sum is negative, then the sums must arrange to cancel this term or else the partition function would diverge. The only relevant sets of $\{k, m\}$ in the summation are those that give $\mathcal{S}_{\{k,m\}} = 0$.

Each X_μ is controlled separately, so I can set each term of $\mathcal{S}$ to zero separately:

$$A^\mu = \sum_s B_s^\mu k_s + \sum_l C_l^\mu m_l, \tag{3.9}$$

$$B_s^\mu = \mathrm{sgn} h_s c_s^\mu, \qquad C_l^\mu = -\mathrm{sgn} J_l \prod_{s\in\partial l}(-c_s^\mu), \tag{3.10}$$

$$A^\mu = -\tfrac{1}{2}\left(1 - \sum_s B_s^\mu - \sum_l C_l^\mu\right).$$

This is a system of linear Diophantine equations. Finding the nonnegative k_s and m_l that satisfy these equations gives exactly the $\{k, m\}$ that should be summed over in the partition function, Eq. (3.5):

$$\mathcal{Z} = \sum_{k_s, m_l \text{ s.t. } \mathbf{A}=\mathbf{B}k+\mathbf{C}m} (-1)^{\sum_s k_s + \sum_l m_l} \times z_{\{k,m\}} \tag{3.11}$$

#SAT is equivalent to solving Eq. (3.9) for $\{k, m\}$ and summing over these solutions in Eq. (3.11). The complexity of calculating the number of solutions to a boolean satisfaction instance has been transformed into finding the nonnegative solutions of this (likely) underconstrained system of equations and summing an integer function over these solutions. A problem in #SAT reduces to listing the solutions to this system of integer equations.

It would be useful to simplify these equations to also find an equivalent for the decision problem, i.e. whether $\mathcal{Z} = 0$ or not. However, this is no trivial task.

A given solution to Eq. (3.9) could give a positive or negative contribution to the partition function. Knowing a portion of the solutions may give little information, as the rest may cancel the contribution to the partition function from the portion known, and so it will be difficult to say whether an instance is satisfiable or not.

3.4 Examples of SAT duality

Now I will show simple examples of this duality. First, look at a single clause of three variables:

$$\mathcal{F} = (x_1 \vee x_2 \vee x_3). \tag{3.12}$$

This equation is satisfiable, and has $\mathcal{Z} = 7$ solutions that satisfy $\mathcal{F} = T$, all possible configurations except $x_1 = x_2 = x_3 = F$. I will use Eqs. (3.5) and (3.9) to calculate this number.

First, the sites are $s = 1, 2, 3$ and the links are $l = 12, 13, 23, 123$. Seeing that $c_s^\mu = 1$, the signs are $\mathrm{sgn}h_s = \mathrm{sgn}J_{123} = 1$ and for all $l \neq 123$, $\mathrm{sgn}J_l = -1$. This means the Diophantine system of equations is just

$$k_1 + k_2 + k_3 + m_{12} + m_{13} + m_{23} + m_{123} = 3. \tag{3.13}$$

The partition function is

$$\mathcal{Z} = \sum_{\{k_s, m_l \text{ s.t. } \mathcal{S}_{\{k,m\}}=0\}} (-1)^{\sum_s k_s + \sum_l m_l} z_{\{k,m\}} \tag{3.14}$$

Using Eq. (3.6), $z_{\{k,m\}}$ can be calculated with relative simplicity. There are eight configurations of $\{\delta_{k_s,0}, \delta_{m_l,0}\}$ that leave z nonzero. One of these requires all seven

of the k and m to be nonzero, while the others require exactly three nonzero and four zero. These each contribute 1 to the partition function, leaving us with $\mathcal{Z} = 7$.

Now look at an unsatisfiable case:

$$\begin{aligned}\mathcal{F} =&(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \overline{x}_3) \\ &\wedge (x_1 \vee \overline{x}_2 \vee x_3) \wedge (x_1 \vee \overline{x}_2 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_2 \vee x_3) \\ &\wedge (\overline{x}_1 \vee x_2 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee \overline{x}_2 \vee x_3) \wedge (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3)\end{aligned} \tag{3.15}$$

First, note that the calculation of $z_{\{k,m\}}$ is exactly the same. When calculating a new #SAT case, $z_{\{k,m\}}$ need not be recalculated unless the sites or links are changed. A simple benefit from this is that it makes it easier to analyze all bit flips of a SAT instance. However, the $\{k, m\}$ that give $\mathcal{S} = 0$ must be recalculated each time. In matrix form, this Diophantine system is

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ m_{12} \\ m_{13} \\ m_{23} \\ m_{123} \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \tag{3.16}$$

This system is not linearly independent, and upon row reduction it is obvious that it is inconsistent. This immediately gives $\mathcal{Z} = 0$. This is the first of three ways in which a SAT instance can reveal its unsatisfiability: there are no solutions to the

Diophantine system.

This is not the only possibility that leads to unsatisfiability. The Diophantine system could have only negative solutions. Consider the instance

$$\begin{aligned}\mathcal{F} =&(x_1 \vee x_2) \wedge (\overline{x}_2 \vee \overline{x}_3) \\ &\wedge (x_1 \vee \overline{x}_2) \wedge (x_2 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_3)\end{aligned} \tag{3.17}$$

The sites are $s = 1, 2, 3$ and the links are $l = 12, 13, 23$. $\mathrm{sgn}h_1 = 1$, $\mathrm{sgn}h_2 = 1$, $\mathrm{sgn}h_3 = -1$, $\mathrm{sgn}J_{12} = -1$, $\mathrm{sgn}J_{13} = 1$, and $\mathrm{sgn}J_{23} = -1$. The Diophantine system is

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ -1 & 0 & -1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ m_{12} \\ m_{13} \\ m_{23} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ -1 \end{pmatrix} \tag{3.18}$$

After row reduction, there are five equations:

$$\begin{aligned} &k_1 = 0, \qquad k_2 - m_{23} = 0, \qquad k_3 = 0, \\ &m_{12} + m_{23} = 1, \qquad m_{13} = -1. \end{aligned} \tag{3.19}$$

There are solutions to this system, e.g. $k_1 = k_2 = k_3 = m_{23} = 0$, $m_{12} = 1$, and $m_{13} = -1$. However, there are no completely nonnegative solutions to this system, as evidenced by the final equation, $m_{13} = -1$. This demonstrates the second way a SAT instance can be shown unsatisfiable: the system is consistent but admits no nonnegative solutions.

The third way is for there to be some set of positive solutions to the Diophantine system, but for the weights of the solutions in the dual partition function to all cancel out. While possible, it seems unlikely that these cases do not exist, as there is no obvious reason $z_{\{k,m\}}$ cannot arrange to cancel out in the final summation. I did not find such a case in the limited number of instances investigated.

3.5 Other satisfiability problems

The SAT problem as described in Sections 3.1-3.4 is in what is known as conjunctive normal form (CNF), a conjunction of clauses that are themselves a disjunction of boolean variables. Any boolean satisfiability problem may be represented in a CNF formula.

However, this is not always the most efficient way of representing such problems, and there are several different representations or particular cases that are commonly studied. I will now show how SAT duality changes for two of these: Not All Equal satisfiability (NAE-SAT) and 1-in-3 SAT.

NAE-SAT consists of a conjunction of clauses that contain of list of negated or unnegated boolean variables that can not all be equal, neither all be true or all be false. This is simply represented in CNF by adding a pair of CNF clauses for every NAE clause, a disjunction of the variables and a disjunction of the logically negated variables.

$$\begin{aligned} &\mathrm{NAE}(x_1, \overline{x}_2, x_3) \wedge \mathrm{NAE}(x_4, x_5, \overline{x}_1) \wedge \ldots \\ &\quad = (x_1 \vee \overline{x}_2 \vee x_3) \wedge (\overline{x}_1 \vee x_2 \vee \overline{x}_3) \wedge (x_4 \vee x_5 \vee \overline{x}_1) \wedge (\overline{x}_4 \vee \overline{x}_5 \vee x_1) \wedge \ldots \end{aligned} \tag{3.20}$$

After applying duality to the CNF problem, the partition function returns to duality Eqs. (3.9)-(3.11). However, it can be seen from Eq. (3.10) that if clauses μ_1 and μ_2 are a NAE pair, then $B_s^{\mu_2} = -B_s^{\mu_1}$ and $C_l^{\mu_2} = (-1)^{|\partial l|} C_l^{\mu_1}$, where $|\partial l|$ is the number of sites a link connects. In other words, for unpaired clauses μ the Diophantine matrix equation splits into two equations, one for the sites and odd-sized links and one for the even-sized links.

$$\begin{aligned} \tfrac{1}{2}\left(\sum_s B_s^\mu + \sum_{l \text{ s.t. } |\partial l| \text{ odd}} C_l^\mu\right) &= A_1^\mu = \sum_s B_s^\mu k_s + \sum_{l \text{ s.t. } |\partial l| \text{ odd}} C_l^\mu m_l \\ -\tfrac{1}{2}\left(1 - \sum_{l \text{ s.t. } |\partial l| \text{ even}} C_l^\mu\right) &= A_2^\mu = \sum_{l \text{ s.t. } |\partial l| \text{ even}} C_l^\mu m_l \end{aligned} \tag{3.21}$$

Naively, this looks easier to solve, and in fact planar NAE-SAT is in the polynomial time complexity class P, unlike normal 3-SAT and other SAT variants [39]. However, NAE-SAT is still NP-complete in the general case.

1-in-3 SAT consists of clauses in which strictly one of the three boolean variables are true. This could be treated in the same fashion as NAE-SAT, by converting to CNF and simplifying the Diophantine equation. It is simpler, though, to just express the restictions directly in the hamiltonian.

$$\mathcal{H} = \sum_\mu X_\mu \left[3 + \sigma_1 + \sigma_2 + \sigma_3 + \sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_2\sigma_3 - \sigma_1\sigma_2\sigma_3\right] \tag{3.22}$$

It is easy to see that Eq. (3.22) gives terms that are only zero if exactly one of σ_1, σ_2, and σ_3 is $+1$. Propagating this change through duality reproduces Eqs. (3.9)-(3.11) except that $\mathrm{sgn}h_s \to -\mathrm{sgn}h_s$ and so $B_s^\mu \to -B_s^\mu$ and

$$A^\mu \to -\tfrac{1}{2}\left(3 - \sum_s B_s^\mu - \sum_l C_l^\mu\right). \tag{3.23}$$

3.6 Conclusion

I have presented an exact relation between problems in #SAT and solving a linear Diophantine system of equations, given from a modified version of the established Kramers-Wannier duality. This relation serves both as a novel avenue for study of #SAT and a glimpse at the power statistical mechanics dualities can bring to computer science.

It is important to ask how difficult it is to find the solutions to Eq. (3.9). If restricted to interesing instances, with a large number of variables and clauses roughly of the same order $N \sim M >> 1$, and assuming the clauses contain a small number of variables, then the matrix size is quadratic in input size N. There is much research addressing algorithms for reducing a Diophantine system to useable forms, in particular the Hermite and Smith normal forms [40, 41]. The computational time is polynomial in N.

However, acquiring the nonnegative solutions to the system is not a polynomial task. Indeed, the number of nonnegative solutions need not be polynomial in N. For example, a collection of M clauses with 3 variables where no clauses share variables will have 7^M solutions. Simply listing the solutions would take exponential time.

Of course, there are instances where $\mathcal{Z}$ is easier to calculate without using duality. However, this does not preclude situations where it will be easier to use the dual equations, notably when the number of solutions to the Diophantine system is small or zero. This situation could coincide to some extent with a small or zero number of solutions to the original SAT problem, the region of greatest interest in

boolean satisfiability research.

Beyond this, the particular form of the matrix (the fact that it is a $(-1, 0, 1)$ matrix and that the link columns are completely determined from the site columns) could lead to great simplification in algorithms for finding the nonnegative solutions. Also, by categorizing which solutions will give negative values, the equations could be simplified to describe the decision problem that is of more interest, SAT itself.

Additionally, it would be revealing to examine the dual model with more traditional complexity techniques, in particular the FKT algorithm [42, 43] from Fisher, Temperley, and Kasteleyn and holographic reduction [44, 45] from Valiant for planar instances. The planar requirement means that no clause can have more than three variables, as the graph for just a single clause of four variables or higher could not be embedded in the plane. However, planar 3-SAT is a nontrivial foundation for further work.

Other computer science problems could also have interesting dual problems, even other constraint satisfaction problems as seen with NAE-SAT or 1-in-3-SAT. Many of these problems, e.g. graph coloring [46], are easily stated as some limit of a statistical mechanics model, which likely has a simple duality relation to another model. The $x - y$ and vector Potts models discussed in Chapter 2 could be used to connect with more computational problems, or any of the classical models with well studied, simple duality relations [16, 17].

In fact, many NP problems even have Ising formulations, so much of the framework from this chapter may be employed without change [47]. With the large number of duality relations and the ease of relating a computer science problem

to the limit of a physical model, the study of duality could prove fruitful. This also goes for combining duality with the belief propagation method, as discussed in Chapter 4.

The recent work with bond algebraic dualities [3, 4, 5, 20] opens up the number of models to study even further. With this new way to categorize and explore dualities, even quantum complexity classes could be accessible for study with duality [48, 49].

Chapter 4

Message passing under duality

Belief propagation [35, 36, 6] is used to calculate marginals in large networks of various models, notably the SAT and Ising models. The idea is this. There are variables and interactions, and a set of probabilities called "messages" are defined between them. These messages pass information to the messages around them iteratively, using the Boltzmann probabilities at the interactions in the updates. After enough time passes, and assuming the model and network have certain properties, the messages will converge to some fixed point or set of fixed points. These fixed point messages can be used to calculate the marginals and any other desired statistical mechanical properties.

There are benefits to using message passing algorithms to calculate thermodynamic quantities. The ability to apply this method to random networks by assuming the convergence is well behaved and that the loops in the network are large has seen much use, particularly for boolean satisfiability problems. The main benefit is that as opposed to computing the partition function directly, belief propagation is designed to sum and multiply terms in a computationally efficient order, saving

operations.

It is straightforward to see why belief propagation and duality could work well together. Taking as example the Ising model with multispin interactions and an external field, duality does not change the structure of the network of sites and links but merely switches the variables with the interactions and redefines the energy parameters. Naively, "dual belief propagation" should work in the kinds of networks where belief propagation does but give completely different messages, since the Boltzmann probabilities will change and the update equations will switch site and link. Most importantly, if belief propagation is inaccurate for one reason or another, dual belief propagation can provide a second approximation both to narrow the window of the exact result and to gauge the accuracy of the first approximation.

In Section 4.1, I closely follow the derivation Mézard and Montanari's book on the subject [6] and give expressions for both normal and dual belief propagation for an Ising model in an external field. Section 4.2 follows that with examples of belief propagation and dual belief propagation, including approximations for the phase transition of the Ising model on different lattices. In Section 4.3, I extend the theory to the Potts model, and Section 4.4 has concluding remarks. [1]

4.1 Normal and dual belief propagation for the Ising model

First, define messages passing from site s to link l as $\nu^{(t)}_{s\to l}(\sigma_s)$ and from link l to site s as $\hat{\nu}^{(t)}_{l\to s}(\sigma_s)$. These are functions of the variable on the connected site s

[1] The content of this chapter is based partially on work done in collaboration with Benjamin Hsu and Victor Galitski [14].

and are normalized so that $\nu^{(t)}_{s\to l}(1) + \nu^{(t)}_{s\to l}(-1) = 1$ and $\hat{\nu}^{(t)}_{l\to s}(1) + \hat{\nu}^{(t)}_{l\to s}(-1) = 1$.

Belief propagation consists of defining a starting message configuration and then updating the messages. The messages update by defining the message $\nu_{s\to l}$ flowing out of site s into link l in terms of the messages $\hat{\nu}$ flowing into site s from links other than l and defining message $\hat{\nu}_{l\to s}$ flowing into site s from link l from the messages ν flowing into link l from sites other than s. After some time and with certain assumptions, the messages should update to a fixed point configuration, and this configuration can be used to calculate important thermodynamic properties.

To a large extent, the initial configuration is unimportant and the message update equations drive the physics. For the Ising model with hamiltonian $\mathcal{H} = \sum_s h_s\sigma_s + \sum_l J_l \prod_{s\in\partial l}\sigma_s$, they are (up to normalization):

$$\nu^{(t+1)}_{s\to l}(\sigma_s) \cong \mathrm{e}^{h_s\sigma_s} \prod_{l'\in\partial s\setminus l} \hat{\nu}^{(t)}_{l'\to s}(\sigma_s), \tag{4.1}$$

$$\hat{\nu}^{(t)}_{l\to s}(\sigma_s) \cong \sum_{\sigma_{\partial l\setminus s}} \mathrm{e}^{J_l \prod_{s\in\partial l}\sigma_s} \prod_{s'\in\partial l\setminus s} \nu^{(t)}_{s'\to l}(\sigma_{s'}). \tag{4.2}$$

Since these are simple exponentials, it will be beneficial to use the log-likelihood of these messages, $\nu^{(t)}_{s\to l}(\sigma_s) \cong \mathrm{e}^{g^{(t)}_{s\to l}\sigma_s}$ and $\hat{\nu}^{(t)}_{l\to s}(\sigma_s) \cong \mathrm{e}^{\hat{g}^{(t)}_{l\to s}\sigma_s}$. With these, the belief propagation update equations are:

$$g^{(t+1)}_{s\to l} = h_s + \sum_{l'\in\partial s\setminus l} \hat{g}^{(t)}_{l'\to s}, \tag{4.3}$$

$$\tanh \hat{g}^{(t)}_{l\to s} = \tanh J_l \prod_{s'\in\partial l\setminus s} \tanh {g_{s'\to l}}^{(t)}. \tag{4.4}$$

After iteration, the partition function may be calculated with the limiting values of

the messages, $g_{s\to l}$ and $\hat{g}_{l\to s}$. Explicitly,

$$\mathcal{Z}[\underline{\mathrm{g}},\underline{\hat{\mathrm{g}}}] = \left(\prod_s \prod_{l\in\partial s} 2\cosh(g_{s\to l}+\hat{g}_{l\to s})\right)^{-1} \tag{4.5}$$
$$\times \left(\prod_s 2\cosh(h_s+\sum_{l\in\partial s}\hat{g}_{l\to s})\right)$$
$$\times \left(\prod_l \left[\cosh J_l \prod_{s\in\partial l}\cosh g_{s\to l} + \sinh J_l \prod_{s\in\partial l}\sinh g_{s\to l}\right]\right).$$

One of the more important restrictions determining when this method works is tree-like networks. Accurate results are not expected when there are small loops in the network. For example, this would not work well for lattices. Belief propogation gives exact results in a finite amount of time for a tree network, asymptotically exact results for a network with a single loop, and worse for more loops. The loops in networks for random SAT instances are large enough that belief propogation is feasible, though it requires some refinement.

Under duality, the partition function acquires a constant multiplier, switches the sites and links, and redefines the energy constants. Immediately, the dual belief propagation update equations and the resulting partition function can be calculated:

$$\tilde{h}_s = -\tfrac{1}{2}\ln\tanh h_s, \qquad \tilde{J}_l = -\tfrac{1}{2}\ln\tanh J_l, \tag{4.6}$$

$$\tilde{g}^{(t+1)}_{l\to s} = \tilde{J}_l + \sum_{s'\in\partial l\setminus s} \hat{\tilde{g}}^{(t)}_{s'\to l}, \tag{4.7}$$

$$\tanh\hat{\tilde{g}}^{(t)}_{s\to l} = \left(\tanh\tilde{h}_s \prod_{l'\in\partial s\setminus l}\tanh\tilde{g}^{(t)}_{l'\to s}\right), \tag{4.8}$$

$$\tilde{\mathcal{Z}}[\underline{\tilde{\mathrm{g}}},\underline{\hat{\tilde{\mathrm{g}}}}] = \left(\prod_l\prod_{s\in\partial l} 2\cosh(\tilde{g}_{l\to s}+\hat{\tilde{g}}_{s\to l})\right)^{-1}\left(\prod_l 2\cosh(\tilde{J}_l+\sum_{s\in\partial l}\hat{\tilde{g}}_{s\to l})\right) \tag{4.9}$$
$$\times\left(\prod_s\left[\cosh\tilde{h}_s\prod_{l\in\partial s}\cosh\tilde{g}_{l\to s}+\sinh\tilde{h}_s\prod_{l\in\partial s}\sinh\tilde{g}_{l\to s}\right]\right),$$

$$\mathcal{Z}_0 = \left(\prod_s 2\sinh 2h_s\right)\left(\prod_l \tfrac{1}{2}\sinh 2J_l\right) \tag{4.10}$$

$$\mathcal{Z} = \mathcal{Z}_0 \cdot \tilde{\mathcal{Z}}[\tilde{\mathrm{g}}, \hat{\tilde{\mathrm{g}}}]. \tag{4.11}$$

There is a crucial issue with this dual belief propagation. As mentioned in the Chapter 3, a negative h_s or J_l will give a negative Boltzmann probability. This means that the messages are no longer probabilities, no longer constrained between 0 and 1. Not only could diverging messages pose a problem, but also a probabilistic analysis is often used for these problems rather than the messages themselves, so it could cause complication in application. Because of this and the small loops induced by representing a SAT problem with an Ising model, I will not be applying dual belief propagation to the dual SAT model from Chapter 3.

4.2 Examples: Ising model transition temperature approximations

As an example, I can approximate the well known transition temperatures for the Ising model on different lattices (without external field) by searching for when the dual update equations change from having a single fixed point to having more than one, just as Mézard and Montanari do with the normal update equations [6]. This is an inaccurate approximation, since this is a lattice and has many small loops. But I can compare the success of the normal belief propagation approximation with that of the dual approximation.

Begin with the simplest Ising hamiltonian. To start, assume a square lattice.

$$\mathcal{H} = -J\sum_l \sigma_{l1}\sigma_{l2} \tag{4.12}$$

The belief propagation update equations:

$$g_{s\to l}^{(t+1)} = \sum_{l'\in\partial s\setminus l} \hat{g}_{l'\to s}^{(t)}, \tag{4.13}$$

$$\hat{g}_{l\to s}^{(t)} = \operatorname{atanh}\left(\tanh J \prod_{s'\in\partial l\setminus s} \tanh {g_{s'\to l}}^{(t)}\right). \tag{4.14}$$

Assume that the equations have settled to a fixed point: $g_{s\to l}^{(t)} = g_{s\to l}$, $\hat{g}_{l\to s}^{(t)} = \hat{g}_{l\to s}$. To make things analytically calculable, assume that the fixed point messages have no spatial dependence. The fixed point update equations give

$$g = 3\hat{g}, \qquad \hat{g} = \operatorname{atanh}(\tanh J \tanh g) \qquad \Rightarrow \qquad \tanh\hat{g} = \tanh J \tanh 3\hat{g} \tag{4.15}$$

It is important to ask when the number of possible fixed point solutions changes, as this signals a phase transition. For small J, the only solution to the above equation is $g = 0$, and for large J there are three possible solutions. The crossover occurs when

$$0 = \frac{\mathrm{d}}{\mathrm{d}\hat{g}}(\tanh\hat{g} - \tanh J \tanh 3\hat{g})\,|_{\hat{g}=0} \Rightarrow J_C = \operatorname{atanh}\tfrac{1}{3} = \tfrac{1}{2}\ln 2 \tag{4.16}$$

With a few steps and sweeping assumptions, belief propagation estimates the transition temperature of the $2D$ square lattice Ising model to be $T_C^{\mathrm{BP}} = (\frac{1}{2}\ln 2)^{-1} \approx 2.89$. Compare this with the exact transition temperature $T_C^{\mathrm{exact}} = (\frac{1}{2}\ln(1+\sqrt{2}))^{-1} \approx 2.27$.

Now I will apply dual belief propagation to the same model. The dual model is the same but with dual coupling $\tilde{J} = -\frac{1}{2}\ln\tanh J$, so it is quick to show that

$$J_C = \operatorname{atanh}\mathrm{e}^{-\ln 2} = \operatorname{atanh}\tfrac{1}{2} = \tfrac{1}{2}\ln 3 \tag{4.17}$$

These results are promising. The exact transition temperature is $T_c^{\mathrm{exact}} = (\frac{1}{2}\ln(1+\sqrt{2}))^{-1} \approx 2.27$. The approximation from normal belief propagation is $T_c^{BP} =$

Lattice	Exact T_c	BP T_c	Dual BP T_c
Square	2.27	2.89	1.82
Triangular	3.67	4.93	2.88
Honeycomb	1.52	1.82	1.24
Cubic	4.51	4.93	4.22
Diamond	2.70	2.88	1.84

Table 4.1: A comparison of exact transition temperature with crude approximations using belief propagation and dual belief propagation. While both are inaccurate due to the small loops in the lattices, the approximations bound the exact result and have similar performance.

$(\frac{1}{2}\ln 2)^{-1} \approx 2.89$. The approximation from dual belief propagation is $T_c^{DBP} = (\frac{1}{2}\ln 3)^{-1} \approx 1.82$. The approximations neatly bound the exact result, with nearly as much difference from one bound as the other.

The same crude approximations can be made for various other lattices, including triangular, honeycomb, cubic, and diamond. In these cases, the models are not self dual, but the steps can be performed nonetheless. The belief propagation and dual belief propagation approximations bound the exact result and have similar performance on most lattices, as seen in table 4.2

4.3 Normal and dual belief propagation for the vector Potts model

I will now extend the belief propagation treatment to apply to the vector Potts model. The structure of Eqs. (4.1), (4.2) is applicable beyond the Ising model. Given

variables $n_s = 0, 1, \ldots, q-1$ and partition function $\mathcal{Z} = \sum_{\{n_s\}} \left(\prod_s \phi_{n_s}(h_s)\right) \left(\prod_l \psi_{n_{\partial l}}(J_l)\right)$, the belief propagation update equations are written as:

$$\nu_{s\to l}^{(t+1)}(n_s) \cong \phi_{n_s}(h_s) \prod_{l' \in \partial s \backslash l} \hat{\nu}_{l'\to s}^{(t)}(n_s), \tag{4.18}$$

$$\hat{\nu}_{l\to s}^{(t)}(n_s) \cong \sum_{n_{\partial l \backslash s}} \psi_{n_{\partial l}}(J_l) \prod_{s' \in \partial l \backslash s} \nu_{s'\to l}^{(t)}(n_{s'}). \tag{4.19}$$

As for the Ising model, these messages are normalized with $\sum_n \nu_{s\to l}^{(t)}(n) = \sum_n \hat{\nu}_{l\to s}^{(t)}(n) = 1$. This means that, as for the Ising model, there are more update equations than degrees of freedom. In Section 4.1, dealing with log-likelihood variables satisfied the normalization constraint. The Potts model has a similar simplifying change of variables.

There should be $q-1$ free site-to-link messages and $q-1$ free link-to-site messages. A discrete Fourier expansion of the messages, expressed in terms of the $G_{q,m}$ functions defined in Chapter 2 in terms of generalized hyperbolic functions, gives the simplest update equations.

$$\begin{aligned} \nu_{s\to l}^{(t)}(n) &= \sum_m \mathrm{e}^{2\pi \mathrm{i} mn/q}\, G_{q,m}(g_{s\to l,m}^{(t)}) \\ \hat{\nu}_{l\to s}^{(t)}(n) &= \sum_m \mathrm{e}^{2\pi \mathrm{i} mn/q}\, G_{q,m}(\hat{g}_{l\to s,m}^{(t)}) \end{aligned} \tag{4.20}$$

This expression simplifies things because $G_{q,m}$ is the Fourier transform of ϕ_{n_s} and $\psi_{n_{\partial l}}$. Setting $g_{s\to l,0}^{(t)} = 0$ enforces normalization and leaves the correct number of independent messages.

For simplicity, I will not consider a site-dependent direction for the external field ($p_s = 0$), and I will also modify the interaction to be the sum of neighboring angles instead of difference. Both of these changes can be rectified if desired, and

the latter is convenient not only for simplicity but also for studying n-body Potts interactions. The hamiltonian:

$$\mathcal{H} = \sum_s h_s \cos\left(\frac{2\pi n_s}{q}\right) + \sum_l J_l \cos\left(\frac{2\pi \sum_{s\in\partial l} n_s}{q}\right) \tag{4.21}$$

Now to derive the update equations:

$$\begin{aligned} G_{q,m_s}(g^{(t+1)}_{s\to l,m_s}) &= \frac{1}{q}\sum_{n_s} \mathrm{e}^{-2\pi \mathrm{i} m_s n_s/q}\ \nu^{(t+1)}_{s\to l}(n_s) \\ &\cong \frac{1}{q}\sum_{n_s} \mathrm{e}^{-2\pi \mathrm{i} m_s n_s/q}\ \mathrm{e}^{h_s \cos 2\pi n_s/q} \prod_{l'\in\partial s\setminus l} \hat{\nu}^{(t)}_{l'\to s}(n_s) \\ &= \frac{1}{q}\sum_{n_s}\sum_{\{m_{\partial s\setminus l}\}} \mathrm{e}^{-2\pi \mathrm{i} m_s n_s/q}\ \mathrm{e}^{h_s \cos 2\pi n_s/q} \prod_{l'\in\partial s\setminus l} \mathrm{e}^{2\pi \mathrm{i} n_s m_{l'}/q} G_{q,m_{l'}}(\hat{g}^{(t)}_{l'\to s,m_{l'}}) \\ &= \sum_{\{m_{\partial s\setminus l}\}} G_{q,(m_s-\sum_{l'\in\partial s\setminus l} m_{l'})}(h_s) \prod_{l'\in\partial s\setminus l} G_{q,m_{l'}}(\hat{g}^{(t)}_{l'\to s,m_{l'}}) \end{aligned} \tag{4.22}$$

$$\begin{aligned} G_{q,m_s}(\hat{g}^{(t)}_{l\to s,m_s}) &= \frac{1}{q}\sum_{n_s} \mathrm{e}^{-2\pi \mathrm{i} m_s n_s/q}\ \hat{\nu}^{(t)}_{l\to s}(n_s) \\ &\cong \frac{1}{q}\sum_{n_s} \mathrm{e}^{-2\pi \mathrm{i} m_s n_s/q} \sum_{n_{\partial l\setminus s}} \mathrm{e}^{J_l \cos 2\pi(\sum_{s'\in\partial l} n_{s'})/q} \prod_{s'\in\partial l\setminus s} \nu^{(t)}_{s'\to l}(n_{s'}) \\ &= G_{q,m_s}(J_l) \prod_{s'\in\partial l\setminus s}\sum_{n_{s'}} \mathrm{e}^{2\pi \mathrm{i} m_s n_{s'}/q} \nu^{(t)}_{s'\to l}(n_{s'}) \\ &= G_{q,m_s}(J_l) \prod_{s'\in\partial l\setminus s} q\ G_{q,m_s}(g^{(t)}_{s\to l,m_s}) \end{aligned} \tag{4.23}$$

Potts model belief propagation update equations:

$$G_{q,m_s}(g^{(t+1)}_{s\to l,m_s}) = \sum_{\{m_{\partial s\setminus l}\}} \frac{G_{q,(m_s-\sum_{l'\in\partial s\setminus l} m_{l'})}(h_s)}{G_{q,(\sum_{l'\in\partial s\setminus l} m_{l'})}(h_s)} \prod_{l'\in\partial s\setminus l} \frac{G_{q,m_{l'}}(\hat{g}^{(t)}_{l'\to s,m_{l'}})}{\sum_m G_{q,m}(\hat{g}^{(t)}_{l'\to s,m})} \tag{4.24}$$

$$G_{q,m_s}(\hat{g}^{(t)}_{l\to s,m_s}) = \frac{G_{q,m_s}(J_l)}{G_{q,0}(J_l)} \prod_{s'\in\partial l\setminus s} G_{q,m_s}(g^{(t)}_{s\to l,m_s}) \tag{4.25}$$

This result can reproduce Ising belief propagation for $q = 2$. The partition function for the Potts model may be expressed in terms of these messages, as in Section 4.1.

Similar steps can produce dual belief propagation update equations. The main difference between the dual and normal models is the form of $\phi_{n_s}(h_s)$ and of $\psi_{n_{\partial l}}(J_l)$ and the switched roles of sites and links. First, express the messages in the form of the discrete Fourier transforms of ϕ_{n_s} and $\psi_{n_{\partial l}}$.

$$\tilde{\nu}^{(t)}_{l\to s}(n) = \sum_m \mathrm{e}^{2\pi \mathrm{i} mn/q}\ \mathrm{e}^{\tilde{g}^{(t)}_{l\to s,m}\cos\frac{2\pi m}{q}} \tag{4.26}$$

$$\hat{\tilde{\nu}}^{(t)}_{s\to l}(n) = \sum_m \mathrm{e}^{2\pi \mathrm{i} mn/q}\ \mathrm{e}^{\hat{\tilde{g}}^{(t)}_{s\to l,m}\cos\frac{2\pi m}{q}}$$

Again, set $\tilde{g}^{(t)}_{l\to s,0} = \hat{\tilde{g}}^{(t)}_{s\to l,0} = 0$ for normalization. The update equations:

$$\begin{aligned}
\mathrm{e}^{\tilde{g}^{(t+1)}_{l\to s,m_l}\cos\frac{2\pi m_l}{q}} &= \frac{1}{q}\sum_{n_l} \mathrm{e}^{-2\pi \mathrm{i} m_l n_l/q}\ \tilde{\nu}^{(t+1)}_{l\to s}(n_l) \qquad (4.27)\\
&\cong \frac{1}{q}\sum_{n_l} \mathrm{e}^{-2\pi \mathrm{i} m_l n_l/q}\ G_{q,n_l}(J_l) \prod_{s'\in\partial l\setminus s} \hat{\tilde{\nu}}^{(t)}_{s'\to l}(n_l)\\
&= \frac{1}{q}\sum_{n_l}\sum_{\{m_{\partial l\setminus s}\}} \mathrm{e}^{-2\pi \mathrm{i} m_l n_l/q}\ G_{q,n_l}(J_l) \prod_{s'\in\partial l\setminus s} \mathrm{e}^{2\pi \mathrm{i} n_l m_{s'}/q}\ \mathrm{e}^{\hat{\tilde{g}}^{(t)}_{s'\to l,m_{s'}}\cos\frac{2\pi m_{s'}}{q}}\\
&= \sum_{\{m_{\partial l\setminus s}\}} \exp\left(h_s \cos\frac{2\pi(m_s - \sum_{s'\in\partial l\setminus s} m_{s'})}{q} + \sum_{s'\in\partial l\setminus s} \hat{\tilde{g}}^{(t)}_{s'\to l,m_{s'}} \right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{e}^{\hat{\tilde{g}}^{(t)}_{s\to l,m_l}\cos\frac{2\pi m_l}{q}} &= \frac{1}{q}\sum_{n_l} \mathrm{e}^{-2\pi \mathrm{i} m_l n_l/q}\ \hat{\tilde{\nu}}^{(t)}_{s\to l}(n_l) \qquad (4.28)\\
&\cong \frac{1}{q}\sum_{n_l} \mathrm{e}^{-2\pi \mathrm{i} m_l n_l/q} \sum_{n_{\partial s\setminus l}} G_{q,\sum_{l'\in\partial s} n_{l'}}(h_s) \prod_{l'\in\partial s\setminus l} \tilde{\nu}^{(t)}_{l'\to s}(n_{l'})\\
&= \mathrm{e}^{h_s\cos\frac{2\pi m_l}{q}} \prod_{l'\in\partial s\setminus l}\sum_{n_{l'}} \mathrm{e}^{2\pi \mathrm{i} m_l n_{l'}/q} \tilde{\nu}^{(t)}_{l'\to s}(n_{l'})\\
&= \mathrm{e}^{h_s\cos\frac{2\pi m_l}{q}} \prod_{l'\in\partial s\setminus l} q\ \mathrm{e}^{\tilde{g}^{(t)}_{l'\to s,m_l}\cos\frac{2\pi m_l}{q}}
\end{aligned}$$

Potts model dual belief propagation update equations:

$$\mathrm{e}^{\tilde{g}^{(t+1)}_{l\to s,m_l}\cos\frac{2\pi m_l}{q}} = \sum_{\{m_{\partial l\setminus s}\}} \exp\left(h_s \left(\cos\frac{2\pi(m_s - \sum_{s'\in\partial l\setminus s} m_{s'})}{q} - 1 \right) + \sum_{s'\in\partial l\setminus s} \hat{\tilde{g}}^{(t)}_{s'\to l,m_{s'}} \right) \tag{4.29}$$

$$\hat{\tilde{g}}^{(t)}_{s\to l,m_l}\cos\frac{2\pi m_l}{q} = h_s\left(\cos\frac{2\pi m_l}{q} - 1\right) + \sum_{l'\in\partial s\setminus l}\tilde{g}^{(t)}_{l'\to s,m_l}\cos\frac{2\pi m_l}{q} \tag{4.30}$$

Expressions similar to Eqs. (4.5) and (4.11) for the partition function with belief propagation messages and dual belief propagation messages could be derived for the Potts model.

4.4 Conclusion

I have demonstrated a combination of duality and belief propagation to produce a new message passing procedure. This "dual belief propagation" algorithm has the same advantages, drawbacks, and restrictions as normal belief propagation, with the additional problem of signed messages if there is frustration in the original model. Applying it to models that do not have this problem gives approximations that are similar in accuracy to those of normal belief propagation. Potentially, the normal and dual methods could complement eachother to provide additional accuracy and insight into the physics of a network problem.

First, the sign problem must be rectified. Signed messages may not be a problem by itself, but the update equations rely upon normalized messages, and signed messages could cause divergences and strange behavior in the updates. It is possible introducing auxiliary messages representing the signs could serve as remedy. By itself, it would be stimulating to examine the structure of belief propagation as $h_s \to 0$. In this limit, the structure of the network changes, and presumably the messages at zero external field are related to the messages at nonzero field in some nontrivial way.

After this, there are many topics discussed for belief propagation that could be simply extended to apply to dual belief propagation. Of course, it is desirable to apply dual belief propagation to problems that belief propagation deals with, such as SAT and decoding. Considering the focus on the organization of the fixed point solution space for belief propagation [6], it would be enlightening to examine if the organization of the normal message space is reflected in the organization in the dual message space. Phase transitions must occur in the normal and dual models simultaneously, and belief propagation often exposes these transitions with a transition in fixed point message structure. These transitions and are likely also visible in the dual belief propagation fixed point analysis, giving additional insight into the phases and transition. What form does replica symmetry breaking take after duality?

Chapter 5

Two-component Coulomb glass in insulators with a local attraction

Disordered films with superconducting correlations host an amazing variety of interesting phenomena such as superconductor-insulator transitions tuned by disorder or an external magnetic field, with rather unusual transport properties [50, 51, 52, 53, 54, 55, 56, 57, 58, 59]. These phenomena led to a lot of interesting theoretical work [60, 61, 62, 63, 64, 65, 66, 67], which were to a great extent spurred by an experimental feature of many strongly disordered films with superconducting correlations - a giant magnetoresistance peak - for which a full theoretical understanding is still lacking. Nevertheless, the presence of this peak, along with Hall measurements, suggested that pairing survives well into the insulating regime, though it becomes localized, and this pairing is only gradually destroyed by the increasing magnetic field, as understood in Ref. [51]. Very recently, the presence of localized pairs has been verified by STM spectroscopy [58], showing the absence of coherence peaks in the tunneling density of states despite the presence of a superconducting gap – a fact predicted theoretically earlier [63]. The experimental evidence at hand support a distinct transition from a Bose insulating phase to a Fermi insulator and

clearly require a detailed study of the strongly insulating regime which incorporates survival of the localized pairs.

An important and well studied semiconductor material that follows the above methodology is InO_x. Despite its complex band structure and the uncertainties that come from it, the carriers in InO_x are widely believed to originate from oxygen vacancies, partially compensated by the triply-negatively-charged indium vacancies. The formation energy of oxygen vacancies with different charge was calculated recently in the ab initio study of Ref. [68], finding that a doubly-charged vacancy has low formation energy in a crystalline environment (in zero field). The next best state energetically is an empty site, while a single occupied site has large energy. This local 'pair' formation possibly underlies the superconductivity in this system, similarly to compounds like PbTe [69], where local negative U interactions could to lead to a non-standard phase, superconductivity of preformed hard core bosons [70].

Motivated by this evidence of local electron-electron attraction, in this chapter we propose a new two-component Coulomb glass model that combines strong disorder and long-range Coulomb repulsion with the additional possibility of local pockets of a short-range inter-electron attraction. This model hosts a variety of interesting phenomena, in particular a crucial modification of the Coulomb gap previously believed to be universal. Tuning the short-range interaction to be repulsive, we find non-monotonic humps in the density of states within the Coulomb gap. We further study variable-range hopping transport in such systems by extending the standard resistor network approach to include the motion of both single electrons and local pairs. In certain parameter regimes the competition between these two

types of carriers results in a distinct peak in resistance as a function of the local attraction strength, which can be tuned by a magnetic field.

Seeing the power of the duality used by Fisher and Lee to relate a two-dimensional disordered lattice of bosons to the flux-lattice of a fictitious superconductor [13], it would be beneficial to study the coexistence of single electrons and local bosonic pairs in this two-component Coulomb glass with use of a similar duality. Underlying this duality is the $x - y$ duality discussed in Chapter 2, coupled with the Villain approximation [71, 16]. A study of these disordered films with this kind of duality could both link the fermionic and bosonic variables in the same framework and, if applied to a more involved model that could study the superconducting side of the superconductor-insulator transition, could connect the insulating and superconducting phases.

We do not present such a duality in this thesis, but we will rigorously analyze the two-component Coulomb glass. Section 5.1 serves as a motivation to introduce the two-component Coulomb glass model to study disordered films, while Section 5.2 introduces and describes the components of the model. In Section 5.3, we describe the physics of the model while parameters are varied from the perspective of single particle density of states. In Sections 5.4 and 5.5, we introduce a generalization of Miller-Abrahams resistor networks to numerically measure the resistance of a sample of the film under the influence of single electrons and localized pairs. We summarize our findings in Section 5.6. [1]

[1] The content of this chapter is based on work done in collaboration with Anirban Gangopadhyay, Victor Galitski and Markus Müller [15].

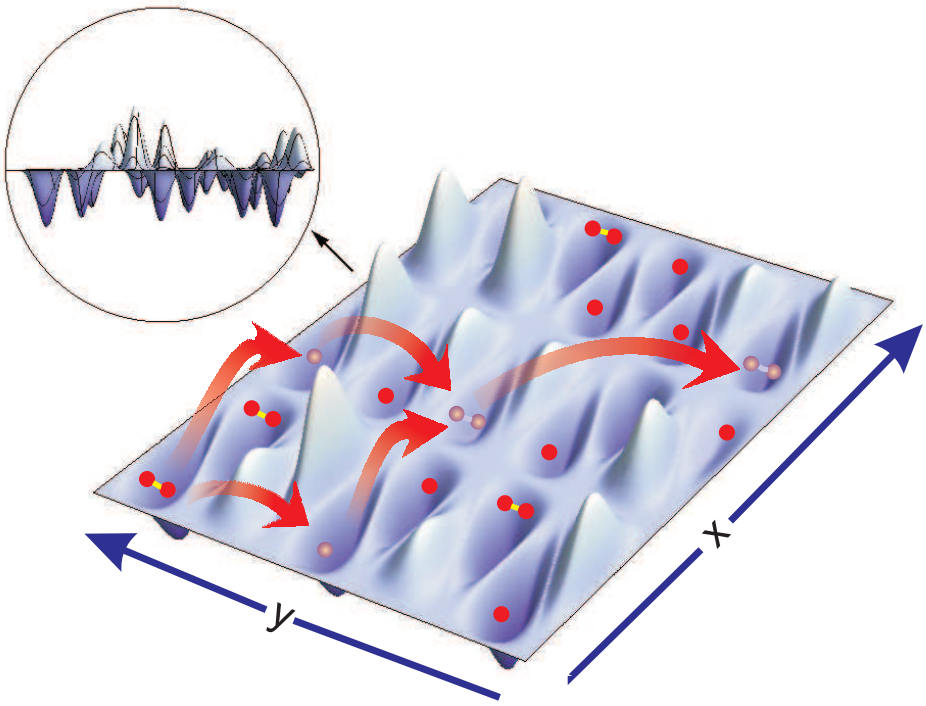

Figure 5.1: Illustration of the two-component model: The energy landscape is due to the combination of on-site disorder and Coulomb interactions. The arrows indicate typical hopping processes relevant for the complex low T transport in the two-component Coulomb glass.

5.1 Motivation for two-component Coulomb glass

A serious obstacle for the theory of such insulators is need to treat strong disorder and electron pairing effects on equal footing. In addition, it has been indicated in recent experiments that the often neglected long range unscreened Coulomb interactions play an important role in several materials. The change of resistance in strong disorder with respect to temperature [52], or on the large-field side of

the mangetoresistance peak in TiN [57] and InO_x [2], is often well described by the Efros-Shklovskii law. This suggests variable-range-hopping (VRH) in the presence of a Coulomb gap [72]. A negative-U Hubbard model approach can handle local pairing attractions and have been studied quite extensively previously, focusing on the superconductor-insulator transition [73, 74, 63], but these studies have mostly neglected long range Coulomb interactions. In contrast, we include Coulomb interactions and in fact put focus on insulating regimes where Coulomb interactions are crucial and compete in a non-trivial way with the local attraction. This study also has implications on Coulomb glasses in granular materials, where multiple occupation is allowed. These have recently been analyzed in related works [75, 76].

These experimental motivations lead to a lattice model that contains all of the various possible ingredients of the actual materials while remaining elementary enough to easily study: strong disorder, local attraction of electrons (favoring double occupancy), and long-range Coulomb interactions, in addition to quantum transport captured by nearest-neighbor hopping. The corresponding hamiltonian can be written in a general form, with the tunable parameters (disorder strength W and magnetic field $\mathbf{B}$) explicit.

$$\hat{H} = \sum_i \phi_i \hat{n}_i + \frac{1}{2}\sum_{i,j} \frac{e^2}{r_{ij}}(\hat{n}_i - \nu)(\hat{n}_j - \nu) + \sum_i \frac{1}{2} U_i(\mathbf{B})\hat{n}_i(\hat{n}_i - 1) \qquad (5.1)$$
$$+ \sum_{\langle ij \rangle} \left(t_{ij}^{(1)} e^{\theta_{ij}(\mathbf{B})} \hat{c}_i^\dagger \hat{c}_j + t_{ij}^{(2)} e^{2\theta_{ij}(\mathbf{B})} \hat{c}_i^\dagger \hat{c}_i^\dagger \hat{c}_j \hat{c}_j + \text{h.c.} \right)$$

In the above, $\phi_i = O(W)$ is the random on-site potential due to the disorder,

[2] B. Sacépé, private communication.

for example randomly positioned dopants. $\frac{e^2}{r_{ij}}$ is an unscreened Coulomb repulsion between localized carriers, and $U_i(\mathbf{B})$ is a pairing interaction renormalized by the Coulomb repulsion between charges on the same site (i.e. within one lattice spacing). We assume that this interaction is tunable somehow, particularly by the magnetic field. The last two terms represent quantum hopping of the single electrons and of the pairs of electrons. While pair hopping can be stated as a second order process of single electron hopping, the relationship between a single electron's and pair's tunneling amplitudes, $t_{ij}^{(1)}$ and $t_{ij}^{(2)}$, may not be simple in real materials. The details of the local electronic structure that is responsible for the negative U interaction may determine the behavior of the two amplitudes. So we allow the amplitudes to be independent. When we will discuss transport, the two hopping amplitudes are set as independent phenomenological parameters, and they determine two independent localization lengths for single electrons and localized pair excitations. The magnetic field enters the hopping strictly through the phase factors $\theta_{ij}(\mathbf{B})$.

Solving the full quantum hamiltonian would be extremely ambitious. We instead isolate individual aspects of this complex problem and deal with them in turn. We make two simplifying assumptions — first, we focus on the regime of these films where the electron pairs are indeed formed locally, but are far from condensating. Technically, we treat the hopping terms in the hamiltonian with the approximation $t_{ij}^{(1,2)} \ll \max(W, \frac{e^2}{a})$ (a being the lattice constant). This restricts us to a classical model where transport occurs with thermally-induced variable-range hopping. This is closely analogous to the standard analysis of doped semiconductors [77]. When discussing variable-range hopping, we take hopping terms into account using the

(average) localization lengths, ξ_1 and ξ_2, which are a result of the magnetic field dependent hoppings $t^{(1,2)}$. Unfortunately, this explicitly prevents capture of superconductivity within the model. Still, interesting physical phenomena observed in experiments, like the giant magnetoresistance peak, often occur rather deep into the insulating phase [59], where this strongly localized approach we use is meaningful.

Second, we assume that the magnetic field only tunes the local pairing interaction. Assuming a monotonic decrease of the pairing strength U with increasing magnetic field is perfectly reasonable. However, the magnetic field also effects the hopping (orbital effects), and this effect is not included in this thesis. We focus entirely on changing the pairing interaction while studying various physical observables, such as the density of states and longitudinal resistance. Collaborators A. Gangopadhyay, V. Galitski, and M. Müller study the magnetic field dependence introduced by the phases in the hopping terms in Eqn. 5.1 through explicit evaluation of the **B**-dependence of the localization lengths $\xi_{1,2}$ in Ref. [78]. In reality both effects are present simultaneously, and we find that they both contribute to a non-monotonic magnetoresistance.

5.2 Model

This two-component Coulomb glass model will be shown to feature a significantly richer variety of phenomena than the canonical Efros-Shklovskii model. We will now focus on it. The Efros Shklovskii model considers a lattice of sites, i, with random on-site energies, ϕ_i, for electrons populated to a filling factor, ν. Each site

i can host either $n_i = 0$ or $n_i = 1$ electrons. Those (classical) electrons interact through an unscreened Coulomb interaction e^2/r, and the disorder is distributed over a typical range W, e.g., uniformly in $\phi_i \in [-W, W]$. An important hallmark of this model is the soft Coulomb gap in the density of states (DOS), $\rho(E)$, close to the Fermi level. For many materials with compensated doping, including InO_x, the disorder is strong compared to the Coulomb interaction, i.e. $W \gg e^2/a$, where a is the typical distance between neighboring electrons. In that case, the Coulomb gap is theoretically predicted [79, 80] and empirically found [81] to be essentially universal at low energies: in two dimensions, $\rho(E)$ exhibits linear variation, $\rho(E) = \frac{\alpha}{e^4}|E|$. The co-efficient α is basically independent of the lattice type, the filling factor, and the details of the disorder potential [82, 83]. We calculate a value $\alpha \approx 0.35 \pm 0.01$ consistent with previous numerical studies [81, 84], but substantially smaller than Efros' simple analytical estimate $2/\pi$ [77].

This standard Coulomb gap has a direct implication for the resistance and appears in transport as a stretched exponential resistance of the form

$$R(T) \sim R_0 \exp\left(\frac{T_0}{T}\right)^{\frac{1}{2}}. \tag{5.2}$$

The co-efficient in the exponent,

$$T_0 = C\frac{e^2}{\sqrt{\alpha}\xi_1}, \tag{5.3}$$

involves one additional parameter: the average localization length for the electron wavefunctions, ξ_1, apart from a numerical constant whose value $4 \lesssim C \lesssim 5$ can be approximated from a percolation analysis of Miller-Abrahams random resistor networks [77] and from Monte Carlo simulations [85].

Now we develop upon the Efros-Shklovskii model by allowing local pairs, or double occupancy on the sites, (see Fig. 5.1 for an illustration). We use the classical hamiltonian

$$H = \sum_i \phi_i n_i + \frac{1}{2} \sum_{j \neq i} \frac{e^2}{r_{ij}} (n_i - \nu)(n_j - \nu) + \sum_i \frac{U_i}{2} n_i (n_i - 1), \tag{5.4}$$

where $n_i \in \{0, 1, 2\}$. The local pairing energies, U_i, for doubly occupied sites will be our control parameters representing the magnetic field that drives the crossover from the regime where electrons dominate ($U \gg 0$) to where pairs dominate ($U \ll 0$). For intermediate U, there are a mixture of gapless single electron and pair states. This range exhibits distinct features that can be realized in experiments. Note that our model (5.4) is also of interest for semiconductors in which doubly occupied sites (the upper Hubbard band) play a significant role [86, 87]. Many of the effects found here can be seen in modified form to granular systems as well. As mentioned above, the hopping will be reintroduced later as a perturbation to describe transport.

5.3 Single site density of states

5.3.1 Definitions

We start with the static properties of the two-component electron glass, in particular the single site density of states (DOS) within typical metastable states.

We start with a local energetically minimum state, a classical occupancy configuration which is energetically stable with respect to single electrons or pairs moving around, as well as with respect to the formation of local pairs by combining two

single electrons, or the reverse disintegration process. Let S_n be the set of sites with occupancy $n \in \{0, 1, 2\}$ in this local minimum configuration. We define the total energy to add (remove) a single electron on site i as E_i^{1+} (E_i^{1-}), and to add (remove) a pair excitation as E_i^{2+} (E_i^{2-}). We define the density of states for electron (pair) excitations, $\rho_{m=1(2)}$, as

$$\rho_m(E) = \frac{1}{N} \sum_{i \in \Sigma_m^+} \delta(E - E_i^{m+}) + \frac{1}{N} \sum_{i \in \Sigma_m^-} \delta(E - E_i^{m-}) \tag{5.5}$$

where N is the number of lattice sites, $\Sigma_1^+ = S_0 \cup S_1$, $\Sigma_1^- = S_1 \cup S_2$, $\Sigma_2^+ = S_0$, and $\Sigma_2^- = S_2$.

In the model without double occupancy ($U \to \infty$), imposing stability with respect to all possible single-electron moves,

$$E_i^{1+} - E_j^{1-} - e_{ij} \geq 0, \quad e_{ij} \equiv \frac{e^2}{r_{ij}}, \tag{5.6}$$

induces the Coulomb gap in the DOS. Multi-particle constraints beyond this impose weaker conditions. They do not significantly affect the DOS at low energy. In contrast, we will show that the presence of double occupancies results in an important additional constraints, which affects the Coulomb gap very significantly.

We will describe the evolution of the DOS as the attraction strength is tuned. Clearly, for strongly repulsive U, when all double occupancies are forbidden, the system exhibits no pairing and reduces to the standard Efros-Shklovskii model, for which the single particle DOS ρ_1 contains the canonical Coulomb gap with slope $\alpha \approx 0.35$.

5.3.2 Spatially Uniform Interaction - Anomalous Coulomb gap

The case of a uniform pair interaction, $U_i = U \; \forall i$, is described essentially analytically. Fig. 5.2 illustrates the evolution of the DOS's with local interaction strength U, which were obtained from numerical simulations described below. In the attractive case, $U < 0$, all electrons are paired in the local minimum we have acquired; sites are either empty or host two electrons. This is because any site with a single electron could lower the energy by adding to it another, distant electron. The pair DOS, $\rho_2(E)$, is linear at low energy, with the canonical slope $\frac{\alpha}{(2e)^4}$ corresponding to an Efros-Shklovskii model with fundamental charges $2e$. This can be seen from the pair stability constraint analogous to Eq. (5.6),

$$E_i^{2+} - E_j^{2-} - 4e_{ij} \geq 0. \tag{5.7}$$

This condition is strong enough to automatically ensure the stability with respect to single electron moves and pair formation/disintegration, and it is the dominant stability condition determining the low-energy pair-DOS and indeed the single-DOS as well. These assertions are easy to check, since single particle excitations are given by

$$E_i^{1\pm} = \frac{E_i^{2\pm} \pm |U|}{2}, \quad (U < 0), \tag{5.8}$$

on empty and occupied sites, respectively, gapped up to energies $E_g = \frac{|U|}{2}$. A single particle move does not only cost more than a pair moving in terms of onsite energy per particle, but also gives back less in terms of the Coulomb polaronic interaction term e_{ij}. Similarly, it can be seen that if (5.7) is satisfied, it is always unfavorable

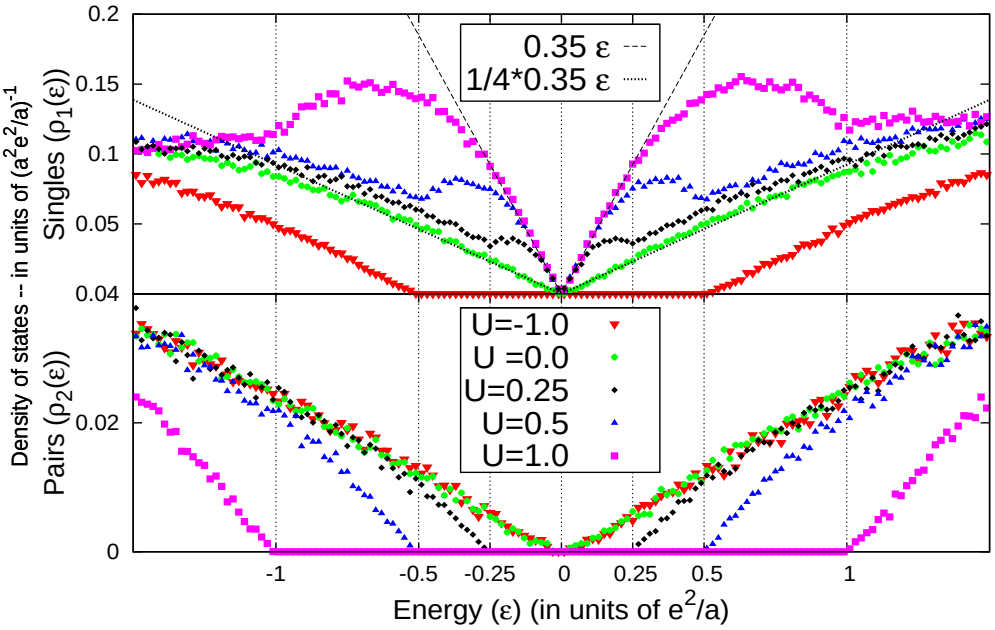

Figure 5.2: DOS for different uniform interaction U. $U = 0$ is a critical point at which both $\rho_{1,2}$ have a linear pseudogap. The slope of the single particle DOS ρ_1 is suppressed to $\alpha/4e^4$. For net repulsion, $U > 0$, ρ_1 has the canonical slope α/e^4 at lowest energy, followed by a hump at the scale $U/2$, crossing over to the critical slope, while pairs are gapped up to $E = U$. For $U < 0$, single electrons have a hard gap $|U|/2$, while pairs are pseudogapped with slope $\alpha/16e^4$. Note: For these plots, the chemical potential was explicitly zeroed when averaging the DOS over the various initial occupancy-distributions

for a to pair disintegrate into two electrons, partly because the attraction energy U is lost, and partly because the polaronic energy gained is not as much. The single particle excitation energy (5.8) implies that the single particle DOS is given by

$$\rho_1(E) = 2\rho_2(2E - \mathrm{sgn}(E)|U|), \quad (U < 0). \tag{5.9}$$

It follows that the single-DOS at energies just beyond the hard gap occur as $\rho_1(E) = \frac{\alpha}{4e^4}(|E| - E_g)$.

The point of no net interaction, $U = 0$, is a critical point, where both $\rho_1(E)$ and $\rho_2(E)$ have a soft gap. However, it is curious that the slope of ρ_1, universal in the Efros-Shklovskii model, is reduced by a factor of 4 from its otherwise universal value α/e^4, as if it were the Coulomb gap of a system with effective charge $e^* = \sqrt{2}$e. This geometric mean of 2e and 1e arises because the gap is determined completely from the pair constraints (5.7), but probed by 1e excitations. Indeed, for every pair of sites across which a pair can move, a single excitation constraint can be obtained by inserting (5.8) for $U = 0$ into (5.7).

$$E_i^{1+} - E_j^{1-} - 2e_{ij} \geq 0. \tag{5.10}$$

This is indeed more stringent than Eq. (5.6).

This is just slightly more complicated on the repulsive side, $U > 0$. Here, pairs are gapped up to energy $E_g = U$. This follows simply from the fact that on empty sites, one has

$$E^{2+} = 2E^{1+} + U, \quad U > 0, \tag{5.11}$$

with $E^{1+} > 0$, and an analogous single particle relation for doubly occupied sites. Indeed, to accomodate a pair, the potential well must be at least as deep as $-U$, and

this will ensure that the second electron is only loosely bound. The minimum energy required to remove the pair from such a well is U. Empty sites have a corresponding argument, and adding a pair costs at least the repulsion U of the second electron.

On the other hand, the single-DOS remains ungapped for repulsive U. At low energies ($|E| \ll \frac{U}{2}$), a single-electron Coulomb gap with the Efros-Shklovskii universal slope $\frac{\alpha}{e^4}$ emerges, as the vast majority of stability constraints involving sites at these energies are single-electron constraints. At larger energies, $|E| \gg U$, one can ignore the pairing energy U in the stability constraints, which reduce again to Eqs. (5.7,5.10) and thus lead to the slope reduced by geometric mean charges, $\frac{\alpha}{4e^4}$ (for E below the Coulomb gap, $E_{\rm Cb} \sim \frac{(e^2/a)^2}{W}$). This immediately leads to the prediction: in the repulsive case, there is a non-monotonicity in the single-DOS at intermediate energies, $U/2 \leq |E| \leq U$, $\rho_1(E)$, as confirmed by the numerical data in Fig. 5.2.

Let us now characterize the single-DOS for repulsive U in more detail. At small positive $E \ll U$, empty and singly occupied sites contribute equally to $\rho_1(E)$. Similarly, for negative energies it receives equal contribution from singly occupied and doubly occupied sites . If we denote these respective contributions as $\rho_1^{(0)}(E)$, $\rho_1^{(1)}(E)$ and $\rho_1^{(2)}(E)$ (superscripts denoting occupancies), we find that

$$\begin{aligned} \rho_1^{(1)}(E) &\approx \rho_1^{(0)}(E) \approx \frac{\alpha}{2}\frac{E}{e^4}, \quad 0 < E \ll U, \\ \rho_1^{(1)}(E) &\approx \rho_1^{(2)}(E) \approx \frac{\alpha}{2}\frac{|E|}{e^4}, \quad 0 < -E \ll U, \end{aligned} \tag{5.12}$$

i.e., the ground state occupation is nearly uncorrelated with the excitation energy.

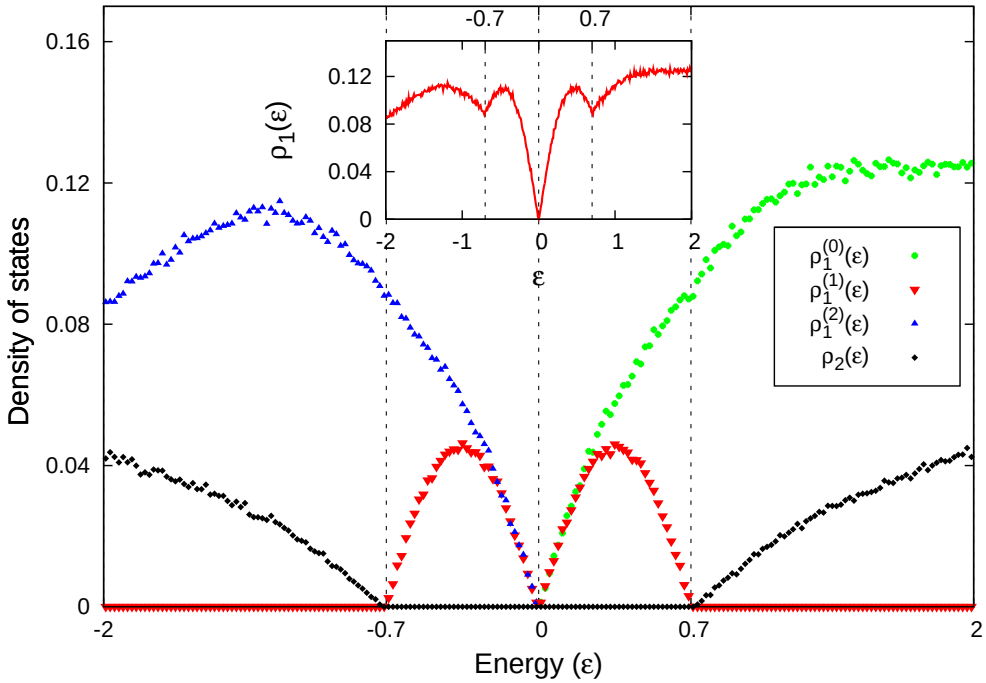

Figure 5.3: Breakup of the single-particle DOS $\rho_1(\epsilon) = \rho_1^{(0)}(\epsilon) + \rho_1^{(1)}(\epsilon) + \rho_1^{(2)}(\epsilon)$ for repulsive $U = 0.7$, split according to the site occupancies, as described by Eqns. 5.12, 5.13 and 5.14. Note that the contribution to $\rho_1(\epsilon)$ from singly-occupied sites ends at $\epsilon = \pm U$, exactly where the pair-DOS $\rho_2(\epsilon)$ begins. Inset: The dip in $\rho_1(\epsilon)$ corresponds to $\rho_1^{(1)}(\epsilon)$ going to zero at $\epsilon = \pm U$, as emphasized by the dashed lines.

From Eq. (5.11) it follows that the pair DOS $\rho_2(E)$ satisfies

$$\rho_2(E) = \frac{1}{2}\rho_1^{(0)}\left(\frac{E - \mathrm{sgn}(E)U}{2}\right). \tag{5.13}$$

Thus, from (5.12), beyond the pair gap it starts off as

$$\rho_2(E) \approx \frac{\alpha}{8}\frac{|E| - U}{e^4}, \tag{5.14}$$

as can be seen in Fig. 5.3

The contribution to $\rho_1(E)$ from singly occupied sites is restricted within the energy range $|E| \leq U$, since otherwise spontaneous particle rearrangements would

occur and the system would not be at a local minimum. Further, the single-DOS satisfies the simple relation

$$\rho_1^{(1)}(E) = \rho_1^{(1)}(E-U), \quad 0 < E < U, \tag{5.15}$$

which expresses a relationship between particle addition and removal for singly-occupied sites: $E^{1+} = E^{1-} + U > 0$. This implies in particular that

$$\rho_1^{(1)}(|E| \to U) \approx \frac{\alpha(U-|E|)}{2e^4}. \tag{5.16}$$

The DOS tends to zero at $E = \pm U$, and has a maximum around $E = \pm U/2$.

At the same time, the other contributions to the single-DOS, $\rho_1^{(0,2)}(E)$, do not exhibit any distinct features at energies of order U, except for a smooth roll-over from $\sqrt{2}e$ charge slope $\frac{\alpha}{2e^4}$ at $|E| \ll U$ (cf. Eq. (5.12)), to a slope that approaches $\frac{\alpha}{4e^4}$, the universal slope in the single component model, for $|E| > U$. As a result, the full single particle DOS, which is the sum of these two contributions, exhibits a local maximum around $E = \pm U/2$ and a local minimum around $E = \pm U$, essentially reflecting the properties of $\rho_1^{(1)}(E)$ imposed by the extra pair constraint (5.15). Similar physics was uncovered recently in granular systems [75], where the occupancy of sites is nearly unlimited. In this case Eq. (5.15) applies to the entire 1-particle DOS, and imposes mirrored Coulomb gaps.

Despite a lack of quantum fluctuations, the above evolution of the DOS has much in common with quantum critical phenomena [88], where U is a detuning parameter from criticality. Critical behavior is restored at larger energies $|E| \gg |U|$, with linear DOS-s and the anomalous slope of $\rho_1(E)$. The non-critical phase appears at low energies, where one type of carriers is gapped out, while the other type

exhibits a universal Coulomb gap. Also note that the features of the DOS and the underlying mechanisms found here are similar to those in a model of strongly and weakly interacting two-level systems that was recently proposed [89].

5.3.3 Numerical simulations

To further study the properties of this density of states DOS, and those in the case of random local interaction U_i, we performed numerical studies. To study metastable states, we generate a random disorder and start from a random configuration of occupancies, n_i ($\in \{0, 1, 2\}$), on a half-filled triangular lattice of size 200×200. A triangular lattice is used, with commensurate filling, so as not to introduce extra strain in the system in the limit of weak disorder (note that if the filling is not commensurate there is still some strain from the lattice). However, we focus on strong disorder where variable range hopping should dominate and the effect of the lattice type is expected to be small.

To be specific, we measure all distances in terms of lattice constant a, and in our finite-sized samples, the intersite distance r_{ij}, has been chosen as the minimum distance on a torus to reflect periodic boundary conditions. Energies are measured with reference to chemical potential μ, using the nearest neighbor Coulomb repulsion $\frac{e^2}{a}$ as units. The chemical potential is determined as the average of the smallest energies to add and remove an extra particle the metastable state. Pair energies are measured from the reference energy 2μ.

We choose the on-site disorder ϕ_i to be randomly distributed in the interval

$[-W, W]$. It is well-known that a disorder of order unity or more is required for the DOS to tend to an essentially universal Coulomb gap, $\rho_1(E) = \frac{\alpha|E|}{e^4}$ at low energies. In our model, since a site is allowed to have double occupancy, the strong disorder condition is met when W exceeds the typical nearest neighbor interaction of two doubly occupied sites. In order to find DOS features which approach a universal limit, we therefore chose to work with disorder $W = 4$. At substantially weaker disorder the low energy DOS's were indeed found to be non-universal.

Using a similar protocol to that described in [83], we move toward a local minumum of energy by allowing the re-distribution of occupancies through single particle moves, pair moves and pair dissociation/formation — the last within a restricted spatial range to ease computation — that lower the total energy of the system. For these purposes, it is important to recall that the appearance of the Coulomb gap in the single particle DOS is not dependent upon stability with respect to multi-particle moves, and is not very sensitive to them, since the single-particle moves impose the strongest stability constraints [83]. From similar reasoning, the universal features in the density for uniform U as considered above come from both single particle and pair stability constraints. It is reasonable to expect that the single particle and pair moves, and possibly the dissociation/formation, considered above impose the strongest stability conditions and determine the essential features of the single site DOS's $\rho_1(E)$ and $\rho_2(E)$.

Multiparticle processes beyond pair-related moves may relax the system into lower energy metastable states, but such states are expected to have similar single site density of states and transport properties. The single site DOS was obtained by

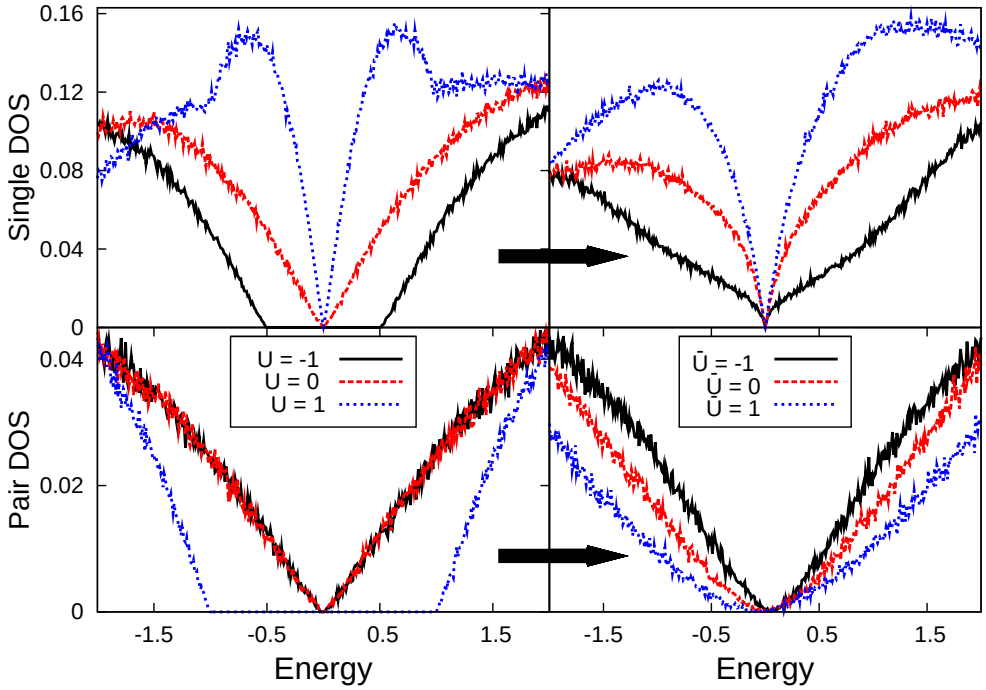

Figure 5.4: Comparison of the DOS's for constant (left) and random (right) U, with strong scatter $\Delta U = 2$ (units of e^2/a). The sharp gaps and humps are smoothed out by disorder, but the overall trend of increase/decrease of the low energy DOS remain intact.

calculating the histogram of the energies to add or subtract an electron or pair from each site, cf. Eq. (5.5). These DOS-s were averaged over many different disorder realizations, typically of the order of 100 for the 200×200 sized systems.

5.3.4 Spatially Disordered Interaction

It is more realistic that in a disordered system the pairing energies U_i are non-uniform rather than constant. Fig. 5.4 compares the DOS's for U_i constant vs. distributed randomly in energy range $\left[\overline{U} - \Delta U, \overline{U} + \Delta U\right]$. The sharp features of

Fig. 5.2 are smoothened. The gaps in $\rho_1(E)$ and $\rho_2(E)$ for repulsive and attractive U_i respectively, are smeared out in the random case. Low energy single electron states bleed into the gap in the single-DOS as soon as there are positive U_i. The density of such states grows as $\overline{U}$ increases, and eventually saturates to the standard linear pseudogap with slope $\frac{\alpha}{e^4}$. Closely analogous considerations apply to $\rho_2(E)$ with decreasing $\overline{U}$. The detailed behavior when $|U| \leq \Delta U$ is presumably non-universal.

The intriguing non-monotonic humps in ρ_1 discussed in Section 5.3.2 will persist only if ΔU is sufficiently smalller than $\overline{U}$. Crystalline samples are probably more amiable to finding them, where the local environment of different impurities are similar, giving rise to a narrow scatter ΔU.

5.4 Resistor network mapping

5.4.1 Choice of parameters

Considering insulators, a smaller density of states is usually reflected in an exponentially increased resistance. It is relevant to ask what happens in the "mixed regime" of our model, where both kinds of excitations show a soft gap. If charge transport was dominated by one of single or pair moves, we would expect an increasing resistance when approaching the mixed regime from the side of the dominant carrier, since the DOS said carrier type diminishes. But transport is more complicated in this two-component Coulomb glass than two separate transport channels. Electron and pair hops are not independent in the sample, and instead combine to form a network of interconnected pair and single electron moves, as illustrated in

Fig. 5.1. A transport path could jump between single electrons and pairs as carrier, and transport is a complex functional of the *combined* DOS. To study this insulating regime, we have generalized the construction of a network of Miller-Abrahams resistors [90] to include both of these pair and single particle processes in a single network of resistances. In order to elucidate the interplay between the two types of transport, we neglect spin blocking effects in the resistor network [86] (as may be justified in strong spin orbit coupled materials). The elementary hopping resistances were evaluated in a mean field fashion for selected metastable states [77, 91].

The problem of finding the conductivity of a sample, where transport takes place via variable range hopping of electrons from singly occupied sites to empty sites, can be mapped to an equivalent random resistor network problem (see Ref. [90] for a derivation). The sites in the hopping problem can be mapped to the vertices of this network, and the inter-site transition rates to the resistances linking these vertices. Finding the sample resistance simply reduces to calculating the effective resistance of the resistor network through a percolation approach [92] .

It is possible to formulate a similar resistor network mapping for variable range hopping on a lattice where double occupancy is also allowed. Early efforts in this direction were made by Kamimura et. al. [86]. There, the inter-site transition rate was taken as the sum of transition rates of the four types of possible single-particle processes characterized by the occupancies of the initial and final sites before the hop.

In our two-component model, we additionally introduce the pair-hopping channel between two sites alongside the single-particle transport channels considered in

Ref. [86]: in what follows, we refer to these single particle and pair hops as first order processes. However, with this introduction, two-particle processes involving more than two sites may also become significant as they can potentially provide lower-resistance alternatives to two-site pair hops between given initial and final occupancy configurations. For brevity, we refer to these two-particle processes involving more than two-sites as second order processes with the idea that two such processes in succession (a pair breaking followed by re-joining of the lone electrons) can constitute a first order process (a pair hop). The book-keeping of the different allowed hopping processes based on occupancies prescribed in Ref. [86] becomes increasingly more cumbersome, as one takes into account these second-order processes.

Hence we formulate a more general prescription for the resistor-network mapping which can be naturally extended to include higher-order processes. We first describe this reformulation for the case where transport takes place through first-order processes only. In this case, each vertex of the equivalent resistor network corresponds to a node (i, n_i) defined as a site i together with its occupancy $n_i \geq 1$. In a system with double occupancies allowed each site gives rise to two nodes, while upon eliminating double occupancies the only nodes are $(i, n_i = 1)$, which reduce to the standard Miller-Abrahams network. The various first order processes between the two given sites i and j correspond to resistances between different nodes of the network: see Table 5.1 for a full description of the four possibilites. In networks without double occupation, there is only one resistance, $R(i, 1; j, 1)$ associated to a given pair of sites.

In defining the resistances, it must be noted that, while the occupancies of the two sites under consideration assume all possible nonzero values, the occupancies of all other sites are frozen at a pseudoground state.

The expression for the transition rate given below in Eq. (5.17) for a single-particle hopping process of one type tallies exactly with that of Ref. [86]. Since it is expected that, between two given sites, one type of single-hopping process is energetically favoured, and thus has least resistance compared to the other three, the two resistor network mappings might be naively expected to produce identical results if only single-particle hops are considered. However, a conducting path through the sample constructed from these least resistance hops might involve an occupancy mismatch between resistors sharing the same site: the alternative paradigm for the resistor-network construction in our approach, which includes occupancy in the definition of vertices of the network, excludes this possiblity.

In the Section 5.4.3, we will include second-order processes by a simple extension of the definition of nodes from (i, n_i) to (i, n_i, j, n_j). Herein lies the utility of our approach : higher-order multi-particle processes can be easily incorporated simply by extending the dimensions of the node-network. The resistor network constructed from these generalized nodes contain the first-order processes as a subnetwork as will be described in detail later. This extension, however, makes calculation of the effective resistance very expensive. Therefore, in the Section 5.4.4, we describe possible approximations so that we can investigate the effect of pair breaking/formation (crucial two-particle moves in the regime where both single and pair DOS's are ungapped) despite staying within a simpler resistor-network with nodes of the form

Initial occupancies		Resistance $R(i, n_i; j, n_j)$
Site i	Site j	
1	0	$R(i, 1; j, 1)$
1	1	$R(i, 1; j, 2)$
2	0	$R(i, 2; j, 1)$
2	1	$R(i, 2; j, 2)$

Table 5.1: Description of possible single-particle hopping processes between two sites i and j through resistances constructed from nodes of the form (i, n_i) : it is important to note that n_i is the occupancy of site i *before* the hop, while n_j is the occupancy of site j *after* the hop

(i, n_i).

5.4.2 Resistor-network construction for first order processes

Let us now describe the resistor network mapping in detail. We start by reaching the pseudo-ground state, used earlier to extract the density of states.

In zero field, the time-averaged rate of transfer of electrons through single-hops from node (i, n_i) to node (j, n_j) (note: n_i is the occupancy of site i *before* the hop, while n_j is the occupancy of site j *after* the hop) is given by

$$\Gamma^0(i, n_i; j, n_j) = e^{-\frac{2r_{ij}}{\xi}} \frac{e^{-\beta E(n_i, n_j - 1)}}{Z} P(i, n_i; j, n_j) \tag{5.17}$$

Here $r_{ij} = |r_j - r_i|$ is the distance between the sites and ξ is the localization length of the electronic wavefunctions for single-hops and that of the pairs for pair

hops. $E(n_i, n_j - 1)$ is the total energy of the system when sites i and j have occupation numbers n_i and $n_j - 1$, respectively (for all sites other than i and j, we use the pseudo-ground state occupancies). Thus, the term $\mathrm{e}^{-\beta E}/Z$ acts as a Boltzmann probability for the initial occupancy-configuration of the two sites. $P(i, n_i; j, n_j)$ is the amplitude for phonon emission or absorption, as the electron system changes from a configuration with (i, n_i), $(j, n_j - 1)$ to one with $(i, n_i - 1)$, (j, n_j). Up to pre-exponential factors, which we approximate by a uniform value here (set to 1 by a choice of unit of time), $P(i, n_i; j, n_j)$ is given by $N(i, n_i, j, n_j) \equiv \frac{1}{e^{\beta|\Delta E(i,n_i;j,n_j)|}-1}$ (absorption, $\Delta E > 0$) and $(1 + N(i, n_i, j, n_j))$ for emission, $\Delta E < 0$. The two cases can be combined into a single expression for $P(i, n_i; j, n_j)$:

$$P(i, n_i; j, n_j) = \left| \frac{1}{e^{\beta \Delta E(i,n_i;j,n_j)} - 1} \right| \tag{5.18}$$

Finally, Z is the "two-site partition function" expressed as

$$Z = \sum_{n_i, n_j} e^{-\beta E(n_i, n_j)}$$

where the sum over n_i and n_j runs from 0 to 2.

A pair-hop from site i to site j corresponds to a resistor between nodes $(i, n_i = 2)$ and $(j, n_j = 2)$ defined by analogy to Eqn. 5.17 as

$$\Gamma^0(i, n_i; j, n_j) = e^{-\frac{2r_{ij}}{\xi}} \frac{e^{-\beta E(n_i, n_j - 2)}}{Z} P(i, n_i; j, n_j) \tag{5.19}$$

where the variables are the same as defined in Eqn. 5.17, except that $E(n_i, n_j - 2)$ is the total energy of the system when sites i and j have occupancies $n_i = 2$ and $n_j - 2 = 0$, respectively. The nodes connected by this pair hop, namely $(i, n_i = 2)$

and $(j, n_j = 2)$, are also connected by a single hop (the last hop described in Table 5.1) and we choose the smaller of the two resistance values as the effective resistance between the vertices.

Note the difference of Eqn. 5.17 from a similar expression for the time-averaged rate of transfer given in Ref. [77]. In the latter, for a single hop from i to j, the probability of occupancy $n_i = 1, n_j = 0$ is implemented through a product of the individual occupancy probabilities as $f_i(1 - f_j)$, where f_i is the Fermi distribution function, instead of the Boltzmann probability-term $\mathrm{e}^{-\beta E}/Z$ included here. In such an approach, detailed balance in the absence of an electric field can be obtained only by dropping the "polaron term" $\frac{e^2}{r_{ij}}$ (particle-hole interaction) when calculating $\Delta E(i, n_i; j, n_j)$. This simplification leads to a somewhat different value of the Efros-Shklovskii temperature T_0 (cf. Eqn. 5.23) as compared to the treatment used here [3]. However, the particle-hole interaction may play a more vital role in our model, where we include electron pairs: the increased importance arises from the fact that the polaron term is equal to $\frac{e^2}{r_{ij}}$ for a single hop but 4 times that for a pair hop and thus may be rather significant in the mixed regime favoring pair transport as a whole over pair breaking.

As a check of our prescription, it can be easily verified that upon barring the polaron term, in the limit of large U where pair formation/transport is hindered, the expression for the transition rate given in Eqn. 5.17 completely agrees with the resistor network construction used in Ref. [77] (see also Ref. [91]).

In the presence of a weak electric field, one can associate a resistor between

[3]Private communication with J. Bergli

the nodes (i, n_i) and (j, n_j) with resistance value given by

$$R(i, n_i; j, n_j) = R(j, n_j; i, n_i) = \frac{kT}{e^2\Gamma^0(i, n_i; j, n_j)}, \tag{5.20}$$

which is guaranteed to be nondirectional due to detailed balance.

5.4.3 Extension to include second order processes

The resistor network mapping stated above allows a generalization to include second order processes. Below, we shall describe an extension to include generic two-particle hops and then make approximations to a specific set of two particle hops: namely, ones that involve the formation or distintegration of a pair.

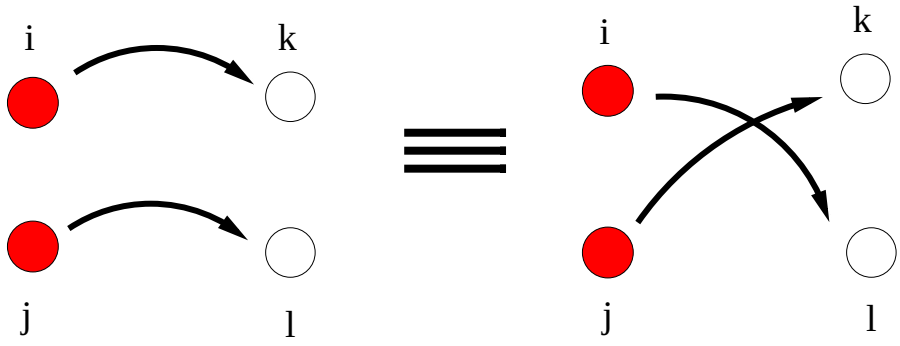

Figure 5.5: One activated resistance, $R(i, n_i, j, n_j; k, n_k, l, n_l)$, corresponds to the above two hops, as the particles are indistinguishable from each other

To describe a second-order process, we extend the notion of a node: now, each node is defined by a *pair* of sites with corresponding occupancies as (i, n_i, j, n_j), subject to the constraint $n_i + n_j \geq 2$. For a two-particle move from node (i, n_i, j, n_j) to (k, n_k, l, n_l) (note: n_i and n_j are the occupancies of sites i and j *before* the two-particle hop has taken place, while n_k and n_l are the occupancies of sites k and l

after), the time-averaged rate of transfer of electrons can be given by the equation

$$\Gamma^0(i,n_i,j,n_j;k,n_k,l,n_l) = e^{-\frac{2r_{ij;kl}}{\xi}} \frac{e^{-\beta E(n_i,n_j,n_k-1,n_l-1)}}{Z} P(i,n_i,j,n_j;k,n_k,l,n_l) \tag{5.21}$$

Here, $r_{ij;kl}$ is taken as $\min[r_{ik}+r_{jl}, r_{il}+r_{jk}]$, keeping in mind the indistinguishability of the hops $(i,n_i,j,n_j) \rightarrow (k,n_k,l,n_l)$ and $(i,n_i,j,n_j) \rightarrow (l,n_l,k,n_k)$ (see figure 5.5).

Although such a two-particle move potentially includes formation or breaking of pairs, we choose for ξ the localization length of the single-electron wave function in all cases except for pair moves, $(i,n_i,i,n_i) \rightarrow (k,n_k,k,n_k)$ (see Ref. [93] for a discussion of such issues in the context of the Hubbard model). Other than pair hops, the remaining first order processes are treated as a subnetwork of this network of generalized nodes in the following sense — a single particle-hop from site (i,n_i) to (j,n_j) involves a "trace" over resistances of the form $R(i,n_i,k,n_k;j,n_j,k,n_k)$ with the indices k and n_k running over all sites and occupancies respectively. In practice, in the percolation approach, this "trace" is performed by "activating" resistances $R(i,n_i,k,n_k;j,n_j,k,n_k)$ for all k and n_k simultaneously for the single-particle hop $(i,n_i) \rightarrow (j,n_j)$ since these resistances all have the same magnitude and are distinguished only by their location in node-space.

The probability for absorbing (or emitting) a phonon with the required energy is in analogy to Eq. (5.18):

$$P(i,n_i,j,n_j;k,n_k,l,n_l) = \left| \frac{1}{e^{\beta \Delta E(i,n_i,j,n_j;k,n_k,l,n_l)} - 1} \right| ;,$$

and the resistance associated with the link between the nodes is :

$$R(i, n_i, j, n_j; k, n_k, l, n_l) = \frac{kT}{e^2\Gamma^0(i, n_i, j, n_j; k, n_k, l, n_l)} \tag{5.22}$$

5.4.4 Simplified algorithms to include pair breaking/formation

The percolation algorithm with generalized nodes (i, n_i, j, n_j) takes into account all possible second order processes whose number grows roughly as the fourth power of the system size. This constrains the approach to small system sizes (upto a 50 X 50 lattice). Nevertheless, performing the full percolation analysis with the generalized node-network for such small sizes allows comparison with certain approximate networks that we describe below. Once the results are seen to agree well, these latter approximate networks can then be used on larger systems to calculate the effective network-resistance through the percolation approach [92].

The first simplification comes from the sparsity of the generalized resistor network in the sense that low-resistance second order processes (within percolation threshold) involve only small-range hops. As a result, for instance, for a 50 X 50 lattice, while calculating the resistances for second order processes, we can restrict the range of hopping to within 5 sites only.

Since, for reasons discussed above, we intend to focus on pair formation/disintegration only, another approximation is to retain only those resistances that correspond to this class of second-order processes. In the node-language, these resistances are of the form $R(i, n_i, j, n_j; k, n_k, k, n_k)$ (pair-formation) and $R(k, n_k, k, n_k; i, n_i, j, n_j)$ (pair-breaking). This imposes storage requirements which go as a third power of

the system size.

Moreover, with this simplification, we can revert to using nodes of the form (i, n_i) for first order processes. In effect, we are thus using a mixed definition of nodes, depending on the order of the process, which is unsuitable for a textbook percolation analysis. Let us therefore describe the percolation approach in slightly more technical detail. We activate the resistances in increasing order of magnitude, irrespective of the order of the process (and thus the definition of the nodes connected).

During the percolation analysis ("activating" resistances in ascending order of magnitude till a percolating cluster is obtained), we use the stored second-order resistances to look for an effective 'short cut' pair transport in the following way. While activating, say a resistance corresponding to pair breaking originating from site i, we check if the lone electrons from this pair breaking are connected through a path of already activated first order single-particle processes to another already activated second-order pair-formation at some other site j. If such a path exists, we refer to it as a 'short cut' pair-transport. In this case, we treat the nodes (i, n_i) and (j, n_j) as if connected through a pair-hop with the equivalent resistance equal to the last-resistance activated in this 'short-cut' path. The advantage of this approximate approach is that the percolation criterion is being applied effectively only to the simpler node-network with nodes of the form (i, n_i) and thus involves a drastic reduction in computation time.

With these two simplifications, it is feasible to study transport for system sizes as high as 200 X 200. This approximate algorithm has been used to check the

robustness of the results with inclusion of the higher-order processes. We find that as anticipated, the resistance for intermediate U (between single and pair-dominated regimes) does decrease due to the extra channels thus included. However, the results described below — notably the nonmonotonicity of the resistance— remain present.

5.5 Transport

An important point to note from the previous section is the exponential dependence of the effective resistances of the network on temperature and localization length. This restricts the accessed range of energies and typical hopping distances of the electrons participating in the network and using a percolation argument [92], one can determine the functional dependence of the resistance on temperature.

In presence of the Coulomb gap in the DOS and if one of single particles or pairs dominates the low temperature transport, the resistance is of Efros-Shklovskii type, cf. Eq. (5.2), with

$$R(T) \sim R_0 \exp\left(\frac{T_0}{T}\right)^{\frac{1}{2}}, \qquad T_0^{(i)} = C\frac{Q_i^2 e^2}{\sqrt{\alpha}\xi_i}, \tag{5.23}$$

where $Q_{i=1,2}$ is the charge of the carriers in units of e, ξ_i their average localization length and $C \approx 4-5$. These localization lengths may be evaluated by examining the elementary localized excitations above the ground state, with spatial dependence is governed by hopping in Eq. 5.1. The magnetic field affects the localization length through phases in the hopping terms and the resulting interference effects, as discussed e.g. in Ref. [94, 95]. For the purpose of analyzing the effect of changing U, however, we assume the localization lengths to be constant. This may accurately

describe an experimental situation in which the local pairing interaction U can be tuned (by chemical modifications or gating), without affecting localization lengths. In contrast, the case where U is tuned by a magnetic field will have the effects described below necessarily superposed over quantum interference effects, which sensitively affect $\xi_{1,2}$ and may dominate the effects which we address below.

As we shall explain, under certain circumstances we obtain a nonmonotonicity in resistance as a function of interaction U. This is most prominent when we establish the relationship $\frac{\xi_2}{\xi_1} = 4$, for which the Efros-Shklovskii temperatures $T_0^{(1,2)}$ are the same in both single-electron and pair-transport and resulting in a curious competition when U is tuned across zero. In reality, ratio ξ_2/ξ_1 varies greatly across phase diagrams of disordered films with superconducting correlations. The localization length of preformed pairs must diverge superconductor transition, while ξ_1 will remain non-critical [96, 97]. However, ξ_2 is expected to become shorter than ξ_1 far in the insulating phase, because of suppressed pair tunneling. A regime where $\xi_2 > \xi_1$ should therefore certainly exist, and below we consider the particular case $\xi_2/\xi_1 = 4$. We should keep in mind however that this large ratio presumably implies strong quantum fluctuations, due to hopping terms. With this caveat, our essentially classical description of the two-component Coulomb glass presented here should be taken as a phenomenological approach to examine Coulomb frustration effects in a system with variable range hopping transport and competing carriers.

5.5.1 Results

In the case of uniform U, there is no genuine mixed transport regime at low T, since one of the two carrier types is always gapped. Here we find the resistance, $R(U;T)$, to be a flat, i.e., constant function of U within error bars. Despite a suppression of $\rho_1(E)$ (see Fig. 5.2), the resistance does not increase significantly. As U is decreased, the pairs "fill in" the resulting gap left in the transport channels, keeping the resistance essentially constant.

However, the situation is more interesting with random U_i, where a genuine mixed two-component carrier regime exists. If pairs and single-electron excitations have strongly disparate localization lengths, $R(\overline{U})$ is still monotonic under tuning of $\overline{U}$, essentially reflecting the evolution of the DOS of the less localized carrier type. However, the two carriers do compete significantly in an intermediate regime. Indeed, we find an interesting non-monotonicity in the resistance; see the top panel in Fig. 5.6. At low temperatures we find a relatively significant peak in the resistance, centered around a small $\overline{U} < 0$. This feature is even slightly enhanced by increasing the randomness ΔU. Transport in the peak region is partially by pairs, which break up and propagate as single electrons, and then recombine again. The maximum of resistance occurs when roughly an equal number of single and pair hops form the critical links of the percolation network, see Fig. 5.6.

A plausible qualitative explanation for the numerically observed peak is the following: it is difficult to connect regions in which pair or single electron transport is favored, as opposed to the regimes $|\overline{U}| \gg 1$, where transport is dominated by

one type of carrier only. In other words, the mixed regime suffers from "contact resistances" between pair- and single-dominated parts of the resistor network in the following sense. A piece of transport path of pairs, must be connected to *two* good strands of single particle transport, and likewise a single particle transport path must find another one to continue as a pair path. Both links require some matching which tends to increase the overall resistance. This phenomenological explanation of our numerical observations bears some resemblance with the idea that superconducting islands may act as weak links in a single-electron-dominated transport regime on the insulating side of the SIT, as proposed in [74].

5.6 Conclusion

The two-component Coulomb glass model should be realized in disordered materials with strong tendency for local attractions (negative Hubbard U), though it also predicts interesting effects in cases where local pair interactions are repulsive, but moderate enough so that multiple occupancy is still possible. The occurrence of negative U interactions is likely to correspond with a bosonic superconductor-to-insulator transition upon further reduction of the disorder. In these types of samples, $\overline{U}$ may also be tuned by an external magnetic field, having a depairing effect on the electrons. However, since such a magnetic field also sensitively affects localization lengths, it is desirable to be able to use other means to influence the local interactions, as well (such as pressure, chemical doping etc), to separate the effects of pairing/depairing from localization length effects. If the disorder in the

local U's is large enough, we find a regime around $\overline{U} \approx 0$, where pairs and single electrons both contribute to transport, and a non-monotonic resistance as a function of U results, including an exponentially large peak as seen in experiment.

For the strongly localized, classical limit of the two-component Coulomb glasses we found several interesting phenomena exhibited in the low energy density of states. In particular, we find that a tendency for local attraction leads to a suppression of the density of states *beyond* the standard Efros-Shklovskii Coulomb gap. When local attraction and Coulomb repulsion on site balance out, $U = 0$, we find that the 2d Coulomb gap is reduced by a factor of 4 from its canonical value, as if a canonical Efros-Shklovskii model but with charges of $\sqrt{2}e$. More generally, if multiple charging of the same site (without additional local charging energy) with M charges were allowed, one would find a suppression by a factor of M^2.

For the case of moderately repulsive $U > 0$, if the randomness in interaction energy ΔU is small compared to the average interaction $\overline{U}$, our model shows non-monotonic humps in the single particle density of states. If pair transport is suppressed due to strong localization, this non-monotonicity in the repulsive case $U > 0$ should show up as a kink in the resistance around temperature $T_* \approx (U/2)^2/(Ce^2/\xi)$, where it crosses over from an Efros-Shklovskii law with a higher T_0 to a less steep resistance and half T_0 at lower T. The humps in ρ_1 should leave traces in AC measurements of the DOS [98], or more direct measurements such as photoemission or tunneling from a broad junction [99]. These DOS features may also be relevant for experiments of memory effects in deep insulators [100, 101], where pair-ready sites with repulsive interactions are known to be present [87].

A measurement of the pair density of states, particularly on the attractive side $U < 0$ could be made through measurement of tunnelling conductance from a (wide) superconducting probe, similar to experiments performed by Dynes et al [102].

In this work we have taken localization lengths to be independent of the tuning of the local interaction strength. However, if the interaction is tuned by magnetic field, a full description needs to take such quantum effects into account, changing the localization lengths as well. It has been argued [94, 95] that the magnetic dependence of localization lengths of pairs and electrons are opposite, which is likely important for a strong magnetoresistance peak. Here we show that the complex energetics and transport phenomena in the two-component Coulomb glass can enhance such a peak.

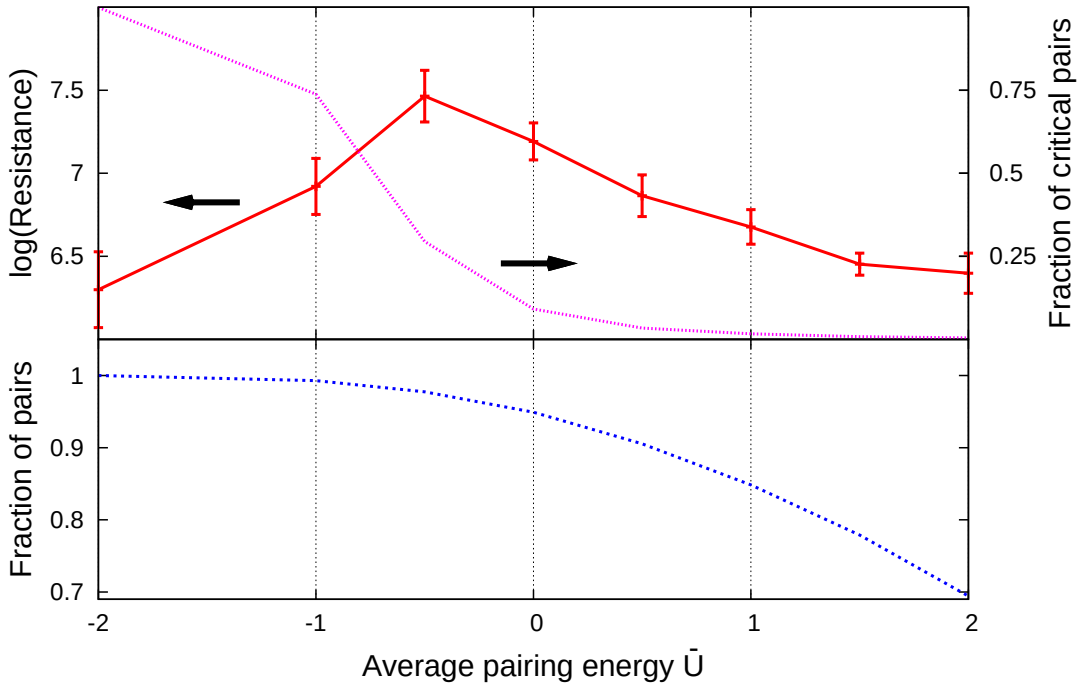

Figure 5.6: Top panel: Peak in the resistance upon tuning $\overline{U}$ at $T = 0.04$ vs. the average interaction $\overline{U}$ (with $\xi_2 = 4\xi_1$); and the fraction of pair hops in the percolating cluster (with resistances within 30% of the percolating resistance). The resistance peaks when roughly half of the critical resistors are pair and single moves, resp. Lower panel: fraction of paired electrons in a typical metastable state. This fraction smoothly decreases across the "mixed regime", since the bulk of such pairs is inactive in transport.

Chapter 6

Conclusions and future work

The common thread linking this thesis is that of classical duality on networks. The simple inclusion of an external field allows for the duality transformation to ignore a model's lattice, in contrast to the explicit lattice dependence without the field. This effect can be seen in any dimension and many kinds of interaction. I have explicitly demonstrated it for the Ising model with many-body interactions, the vector Potts model, and the $x - y$ model in Chapter 2.

This lattice independence allows for the possibility of using duality to study models that normally could not be considered. These include obvious models, such as the Ising model on a random network or with long range interaction, but there are also other very interesting models that have not been studied with duality before at all, such as the boolean satisfiability problem. Computer science problems like these can often be stated as some limit of a statistical mechanics problem, so a known duality for the statistical mechanics problem can lead directly to a duality for the associated computational problem. I showed this in Chapter 3 with the boolean satisfiability problem.

The ability to apply duality to networks also makes integration possible between duality and powerful network techniques. Among these, belief propagation pairs particularly well, resulting in another message passing algorithm, dual belief propagation. While obstacles hinder immediate progress, discussed in Chapter 4, dual belief propagation could prove a valuable tool.

In chapter 5, I presented the two-component Coulomb glass model and in depth density of states and transport analysis. Duality could give an important physical interpretation to this model. In particular, the magnetoresistance peak evident in this model can be explained as a result of the conduction being carried by electrons on one side of the peak and by localized pairs on the other, while in between the single and pair transport networks compete, raising the total resistance. Fisher and Lee showed a similar disordered model of bosons alone was equivalent to a model of fermions in a fictitious superconductor. It would be intriguing if a duality or self-duality could be uncovered for the two-component Coulomb glass or an appropriate generalization, perhaps easily reproducing the magnetoresistance peak. The model is an Ising-like model with a random external field and long range interactions, so duality would necessarily be of the sort discussed in this thesis. However, the difficulties inherent to this duality, in particular frustration in the original model leading to sign flips in the dual model, prevent application currently.

Lattice independent duality could assist the study of even the simple classical models normally considered by duality. There is much interest in e.g. the Ising and Potts models on random graphs and complex networks. In particular, studying the interaction between network topology and duality could lead to profitable research.

One obstacle that needs to be tackled is the sign flips in the dual model from frustration in the original. The relationship between frustration and sign flipping in the dual model is curious by itself, but it also hinders further progress in many areas since so many methods rely on the non-negativity of the Boltzmann terms. Perhaps a subtle redefinition of the parameters or the introduction of an auxiliary field could remove the obstacle, or perhaps current methods could be adapted to account for this possible negativity.

The crossover to zero external field should be studied carefully. The transitions from lattice independence for $h_s > 0$ to lattice dependent duality at $h_s = 0$ to lattice independence with sign flipping for $h_s < 0$ are intriguing. Letting $h_s \to 0_+$ for a single site will cause a deformation in the lattice, partially introducing the plaquette variables of normal duality. Studying this transformation, perhaps in relation to star-mesh transformations [103, 104], could shed light on the behavior of duality at and around zero external field, locally and globally.

Integrating this work with bond-algebraic duality [5] is a likely next step for studying quantum models. Simply looking at bond-algebraic duality for the quantum Ising model with a perpendicular magnetic field component as well as transverse, $\mathcal{H} = \sum_s (h_s^x \sigma_s^x + h_s^z \sigma_s^z) + \sum_l J_l \sigma_{l1} \sigma_{l2}$ would make for a sound beginning. One could examine the lattice dependence, while the limits $h_s^x \to 0$ and $h_s^z \to 0$ would provide sanity checks by sending the transformation to the quantum Ising chain bond-algebraic duality and the lattice independent Ising duality seen in this thesis, respectively.

Beyond this, it would be interesting to study the effects of quantum duality

through the Hubbard-Stratonovich (HS) and similar transformations in the path integral formalism. The HS transformation can already be seen as a duality, with the bosonic HS auxiliary field acting as dual variables. The original variable's interactions become mass-like terms for the HS fields, while the dynamics become dynamic interactions for the HS fields. To showcase, I will represent the normal Ising model in an external field using the path integral formalism and Grassmann fields and then apply the HS transformation and integrate out the original variables.

Let $c_s(\tau)$ be Grassmann fields at imaginary time τ for fermions representing the Ising spins: $\sigma_s \sim \frac{1}{2}(1 - \overline{c}_s c_s)$. Let $\psi_l(\tau) = \begin{pmatrix} c_{l1}(\tau) \\ \\ c_{l2}(\tau) \end{pmatrix}$ be a vector of the fields involved in the interaction at link l, and let $\overline{\psi}_l(\tau)$ be the Grassmann conjugate of this vector.

$$\mathcal{Z} = \left(\int \prod_s \mathcal{D}\overline{c}_s \, \mathcal{D}c_s \right) \exp\left\{ - \int \mathrm{d}\tau \left[\sum_s \overline{c}_s(\partial_\tau - \mu_s)c_s - \sum_l \tfrac{1}{2}\lambda_l \overline{\psi}_l \psi_l \overline{\psi}_l \psi_l \right] \right\} \quad (6.1)$$

Setting $\lambda_l = \frac{1}{4}J_l$ and $h_s = -\frac{1}{2}(h_s + \sum_{l \in \partial s} J_l)$ returns this to the Ising model considered in Chapter 2. The HS transformation consists of introducing a bosonic path integral that decouples interactions. I will decouple in the density channel here for simplicity. Given a 2-vector of fermions $\psi = \begin{pmatrix} \psi_1 \\ \\ \psi_2 \end{pmatrix}$ and its Grassmann conjugate $\overline{\psi}$, a fermion interaction may be decoupled with:

$$\int \frac{\mathrm{d}\overline{\Delta}\,\mathrm{d}\Delta}{2\pi\lambda/\,\mathrm{d}\tau} \exp\left\{ - \,\mathrm{d}\tau \left[\frac{\overline{\Delta}\Delta}{2\lambda} - \tfrac{1}{2}(\overline{\Delta} + \Delta)\overline{\psi}\psi \right] \right\} = \mathrm{e}^{\frac{1}{2}\,\mathrm{d}\tau\lambda\overline{\psi}\psi\overline{\psi}\psi} \quad (6.2)$$

This integral removes the quartic fermionic interaction and replaces with an interaction between each fermion and the bosonic field. I now use this to decouple the

link interactions in the model.

$$\begin{aligned}\mathcal{Z} &= \left(\int \prod_s \mathcal{D}\overline{c}_s\, \mathcal{D}c_s\right)\left(\int \prod_l \frac{\mathcal{D}\overline{\Delta}_l\, \mathcal{D}\Delta_l}{\det_\tau\left[2\pi\lambda_l/\,\mathrm{d}\tau\right]}\right) \\ &\quad \times \exp\left\{-\int \mathrm{d}\tau\left[\sum_s \overline{c}_s(\partial_\tau - \mu_s)c_s + \sum_l \frac{\overline{\Delta}_l\Delta_l}{2\lambda_l} - \sum_l \overline{\Delta_l\psi_l}\psi_l\right]\right\} \\ &= \left(\int \prod_l \frac{\mathcal{D}\overline{\Delta}_l\, \mathcal{D}\Delta_l}{\det_\tau\left[\pi\lambda_l/\,\mathrm{d}\tau\right]}\right)\left(\prod_s \det_\tau\left[\partial_\tau - \mu_s - \sum_{l\in\partial s}(\overline{\Delta_l} + \Delta_l)\right]\right)\mathrm{e}^{-\int \mathrm{d}\tau \sum_l \frac{\overline{\Delta}_l\Delta_l}{2\lambda_l}}\end{aligned} \tag{6.3}$$

As usual, the new dual variables sit on the links of the lattice with "an external field" related to the original interactions, while the new dual interactions take place across the original links with an interaction related to the original external fields. The difference in this case is that the dynamics of the dual fields are derived from the dynamics of the original variables and so the dual field dynamics and interactions remain together at the sites of the orignal lattice, complicating things. The difficulty may be mitigated in this case, but otherwise remains for models that do not reduce to a simple classical model. This is where mean field and other approximations come in.

While not a problem, it is strange that this path integral duality gives bosonic dual variables when classically the dual variables are Ising spins again, more naturally related to fermions. To address this, I would like to introduce a fermionic Hubbard-Stratonovich transformation; in this way the dual variables will be of fermionic nature just like the original. Unfortunately, the simplest fermionic Hubbard-Stratonovich transformation involves two auxiliary Grassmann fields, unlike the single auxiliary field in the bosonic HS transformation.

Let $\eta = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ be the 2-vector of auxiliary Grassmann fields, and introduce

free parameter g and 2×2 traceless hermittian matrix $\mathcal{U}$. It is my understanding that these parameters are analagous to the freedom to choose a channel in which to decouple in the bosonic HS transformation. I also introduce a traceless matrix $\hat{\mathcal{X}}$ that is perpendicular to $\mathcal{U}$ and has $\det \hat{\mathcal{X}} = -1$. The following integral can simultaneously decouple quartic interactions and fermion hopping across links.

$$\begin{aligned} \int \frac{\mathrm{d}\overline{d}_1 \, \mathrm{d}d_1 \, \mathrm{d}\overline{d}_2 \, \mathrm{d}d_2}{\det_2 [\, \mathrm{d}\tau(g+\mathcal{U})]} \exp\Big\{ - \mathrm{d}\tau \Big[& - \overline{\psi}(g+\mathcal{U})\eta - \overline{\eta}(g+\mathcal{U})\psi \\ & + \overline{\eta}(g+\mathcal{U})\eta - \tfrac{1}{2}\lambda\overline{\eta}\left(1+\frac{g}{\mathcal{U}}\right)\eta\overline{\psi}(1+\hat{\mathcal{X}})\psi \Big] \Big\} \\ = \mathrm{e}^{-\mathrm{d}\tau\left[-\overline{\psi}(g+\mathcal{U})\psi - \frac{1}{2}\mathrm{d}\tau\lambda\overline{\psi}\psi\overline{\psi}\psi\right]} & \end{aligned} \tag{6.4}$$

I will now apply this to a model with hopping ($\overline{\psi}_l \mathcal{V}_l \psi_l$ with matrix $\mathcal{V}_l = v_x\sigma_x + v_y\sigma_y$) and interaction ($\frac{1}{2}\lambda_l\overline{\psi}_l\psi_l\overline{\psi}_l\psi_l$) across links. I will set $\mathcal{U}_l = \mathcal{V}_l$ to remove the hopping term between the original variables, and I will set $\hat{\mathcal{X}} = \sigma_z$ so that the original variables may be locally integrated out.

$$\begin{aligned} \mathcal{Z} &= \left(\int \prod_s \mathcal{D}\overline{c}_s \, \mathcal{D}c_s\right) \exp\Big\{ - \int \mathrm{d}\tau \Big[\sum_s \overline{c}_s(\partial_\tau - \mu_s)c_s \\ &\qquad - \sum_l \overline{\psi}_l \mathcal{V}\psi_l - \sum_l \tfrac{1}{2}\lambda_l\overline{\psi}_l\psi_l\overline{\psi}_l\psi_l \Big] \Big\} \\ &= \left(\int \prod_s \mathcal{D}\overline{c}_s \, \mathcal{D}c_s\right) \left(\int \prod_l \frac{\mathcal{D}^2\overline{\eta}_l \, \mathcal{D}^2\eta_l}{\det_{2,\tau}[\, \mathrm{d}\tau(g_l + \mathcal{V}_l)]}\right) \\ &\quad \times \exp\Big\{ - \int \mathrm{d}\tau \Big[\sum_s \overline{c}_s(\partial_\tau - \mu_s + \sum_{l\in\partial s} g_l)c_s \\ &\qquad + \sum_l \left(\overline{\eta}_l(g_l+\mathcal{V}_l)\eta_l - \overline{\eta}_l(g_l+\mathcal{V}_l)\psi_l - \overline{\psi}_l(g_l+\mathcal{V}_l)\eta_l\right) \\ &\qquad - \sum_l \tfrac{1}{2}\lambda_l\overline{\eta}_l\left(1+\frac{g_l}{\mathcal{V}_l}\right)\eta_l\overline{\psi}_l(1+\hat{\mathcal{X}})\psi \Big] \Big\} \end{aligned} \tag{6.5}$$

At this point, there are no terms that involve c_s for different sites, so all c_s may be

locally integrated out, leaving a theory dependent only on the link variables η_l.

The obstacles inhibitting progress from this point are probably greater than for the bosonic HS transformation, as many standard approximations rely on the simple bosonic nature of the dual fields. However, in this case the link variables are fermionic, if doubly so, and this type of transformation appears to be novel and begs a more sophisticated and in depth study, not only because of its connection to classical dualities but also for its capacity to provide approximations for physical observables that are separate from the standard bosonic HS approximations. For example, a simple approximation to decouple the quartic or higher terms of η_l induced by integrating out the c_s is to set $\overline{\eta}_l(1+\frac{g_l}{\mathcal{V}_l})\eta_l \approx \frac{1}{\beta}\int \mathrm{d}\tau \overline{\eta}_l(1+\frac{g_l}{\mathcal{V}_l})\eta_l$. This mirrors the bosonic mean field approximation (in the density channel) $\Delta \approx \frac{1}{\beta}\int \mathrm{d}\tau \Delta$.

Beyond expanding upon the network duality results presented in this thesis, the SAT duality could also benefit from further research. Of course, this topic needs work simply to make it applicable. Solving the Diophantine equations should be simplified by taking into account the particular form of the dual SAT matrix. It would be extremely helpful to characterize each solution to the Diophantine equations by its corresponding term in the partition function summation and profit from being able to perform the summation without having to store the solutions and possibly from being able to deal with the SAT problem instead of #SAT.

Other computer science problems could benefit from network duality, including the subset sum and travelling salesman problems. With the Potts model's connection with signal reconstruction [105], there are numerous potential applications for Potts duality on a network and dual belief propagation for the Potts model.

Dual belief propagation needs work to function well, primarily solving the signed messages problem. After this, however, duality and belief propagation should cooperate well. Both methods are applicable to any number of models, and both work on the scale of the Boltzmann terms and partition function. Integrating the two could potentially produce twice as many message passing algorithms.

It would be prudent to examine the effects of replica symmetry breaking in dual belief propagation. Many models, including boolean satisfiability, do not give correct results for belief propagation when replica symmetric assumptions are made. It is possible that such assumptions may work for dual belief propagation when they do not for the normal algorithm, but it seems unlikely given that the assumptions are closely linked with the structure of the network, and duality in an external field does not change this structure. They are also linked with the range of correlations, and duality often changes this range, so it is still worth investigating.

There has been exciting recent work on the idea of belief propagation for quantum models as well, aptly named quantum belief propagation [106, 107, 108]. Be it bond-algebraic duality or the path integral duality discussed at length above, combining quantum duality with quantum belief propagation is an intriguing direction for research. In addition to examining traditional quantum statistical mechanics problems, one coule study the problems of quantum computational complexity, in such classes as BQP and QMA.

Finally, an in depth application of duality on the two-component Coulomb glass model from Chapter 5 is desirable. Of course, the frustration inevitably caused by the random magnetic field will cause sign flip terms in the dual model. More,

the long range hoppings of electrons and localized pairs is crucial to the resistivity behavior and the physics of the model. It is unknown how to directly include these hoppings in the desired duality without applying a quantum duality to the quantum generalization of the considered model. A first step may be to apply duality to the model without hopping and then see what occurs in the dual model after such a hopping.

It would also be beneficial to extend the two-component Coulomb glass model to be able to study the superconductor-insulator transition, as well as give more correct values to the height of the magnetoresistance peak. Work in this direction has been performed by my collaborators for the two-component Coulomb glass work, Anirban Gangopadhyay, Victor Galitski, and Markus Müller [78], by studying the change in localization length with respect to magnetic field for the hard-core bosons on the side of the peak governed by localized pairs of electrons.

Bibliography

[1] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. part i. *Phys. Rev.*, 60:252–262, Aug 1941.

[2] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, Feb 1944.

[3] Zohar Nussinov and Gerardo Ortiz. Bond algebras and exact solvability of hamiltonians: Spin $s = \frac{1}{2}$ multilayer systems. *Phys. Rev. B*, 79:214440, Jun 2009.

[4] E. Cobanera, G. Ortiz, and Z. Nussinov. Unified approach to quantum and classical dualities. *Phys. Rev. Lett.*, 104:020402, Jan 2010.

[5] Emilio Cobanera, Gerardo Ortiz, and Zohar Nussinov. The bond-algebraic approach to dualities. *Advances in Physics*, 60(5):679–798, 2011.

[6] Marc Mézard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press Inc., New York, NY, USA, 2009.

[7] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM.

[8] Leonid A Levin. Universal sequential search problems. *Problemy Peredachi Informatsii*, 9(3):115–116, 1973.

[9] Tommy R. Jensen and Bjarne Toft. *Introduction to Graph Coloring.* John Wiley & Sons, Inc., 1994.

[10] Gilbert Laporte. The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59(2):231 – 247, 1992.

[11] RicardV. Sol and Sergi Valverde. Information theory of complex networks: Onevolution and architectural constraints. In Eli Ben-Naim, Hans Frauenfelder, and Zoltan Toroczkai, editors, *Complex Networks*, volume 650 of *Lecture Notes in Physics*, pages 189–207. Springer Berlin Heidelberg, 2004.

[12] Melanie Mitchell. Complex systems: Network thinking. *Artificial Intelligence*, 170(18):1194 – 1212, 2006. Special Review Issue.

[13] Matthew P. A. Fisher and D. H. Lee. Correspondence between two-dimensional bosons and a bulk superconductor in a magnetic field. *Phys. Rev. B*, 39:2756–2759, Feb 1989.

[14] J. Mitchell, B. Hsu, and V. Galitski. Kramers-wannier duality of statistical mechanics applied to the boolean satisfiability problem of computer science. *ArXiv e-prints*, October 2013.

[15] Joe Mitchell, Anirban Gangopadhyay, Victor Galitski, and Markus Müller. Two-component coulomb glass in insulators with a local attraction. *Phys. Rev. B*, 85:195141, May 2012.

[16] Robert Savit. Duality in field theory and statistical systems. *Rev. Mod. Phys.*, 52:453–487, Apr 1980.

[17] John B. Kogut. An introduction to lattice gauge theory and spin systems. *Rev. Mod. Phys.*, 51:659–713, Oct 1979.

[18] R Savit. Vortices and the low-temperature structure of the x-y model. 17:1340–1350, 1978.

[19] Z. Nussinov, G. Ortiz, and M.-S. Vaezi. Partial solvability from Duality Transformations. *ArXiv e-prints*, November 2013.

[20] G. Ortiz, E. Cobanera, and Z. Nussinov. Dualities and the phase diagram of the p-clock model. *Nuclear Physics B*, 854(3):780 – 814, 2012.

[21] F Y Wu. Self-dual property of the potts model in one dimension. *arXiv : cond-mat/9805301v2*, 1998.

[22] L Turban. Phase transitions in a generalised two-dimensional ising gauge plus matter theory. 17:419–426, 1999.

[23] Z Glumac and K Uzelac. The partition function zeros in the one-dimensional q-state potts model. 27:7709–7717, 1994.

[24] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift fr Physik*, 31(1):253–258, 1925.

[25] Marco Astorino and Fabrizio Canfora. Duality and fisher zeros in the two-dimensional potts model on a square lattice. *Phys. Rev. E*, 81:051140, May 2010.

[26] J M Kosterlitz and D J Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181, 1973.

[27] Jorge V. José, Leo P. Kadanoff, Scott Kirkpatrick, and David R. Nelson. Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model. *Phys. Rev. B*, 16:1217–1241, Aug 1977.

[28] H. A. Fertig. Deconfinement in the two-dimensional XY model. *Phys. Rev. Lett.*, 89:035703, Jun 2002.

[29] Victor Matveev and Robert Shrock. On properties of the ising model for complex energy/temperature and magnetic field. *Journal of Physics A: Mathematical and Theoretical*, 41(13):135002, 2008.

[30] H.A. Fertig and Kingshuk Majumdar. Vortex deconfinement in the {XY} model with a magnetic field. *Annals of Physics*, 305(2):190 – 227, 2003.

[31] Dimitris Achlioptas. *Random Satisfiability*, chapter 8, pages 245–270. Volume 185 of Biere et al. [109], February 2009.

[32] Fabrizio Altarelli, Rémi Monasson, Guilhem Semerjian, and Francesco Zamponi. A review of the statistical mechanics approach to random optimization problems. *CoRR*, abs/0802.1829, 2008.

[33] Roberto Sebastiani and Armando Tacchella. *SAT Techniques for Modal and Description Logics*, chapter 25, pages 781–824. Volume 185 of Biere et al. [109], February 2009.

[34] Sanjeev Arora and Boaz Barak. *Complexity Theory: A Modern Approach*. Cambridge University Press, 2009.

[35] M. Mézard and G. Parisi. Mean-field theory of randomly frustrated systems with finite connectivity. *EPL (Europhysics Letters)*, 3(10):1067, 1987.

[36] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812, Jan 2002.

[37] R. Monasson and R. Zecchina. Statistical mechanics of the random K-satisfiability model. *Phys. Rev. E*, 56(2):1357–1370, 1997.

[38] Andrea Montanari, Federico Ricci-Tersenghi, and Guilhem Semerjian. Clusters of solutions and replica symmetry breaking in random k-satisfiability. *J. Stat. Mech.*, 2008:04004, Apr 2008.

[39] B. M. E. Moret. Planar nae3sat is in p. *SIGACT News*, 19(2):51–54, June 1988.

[40] Thom Mulders and Arne Storjohann. Diophantine linear system solving. In *Proceedings of the 1999 international symposium on Symbolic and algebraic computation*, ISSAC '99, pages 181–188, New York, NY, USA, 1999. ACM.

[41] Gerold Jäger and Clemens Wagner. Efficient parallelizations of hermite and smith normal form algorithms. *Parallel Comput.*, 35(6):345–357, June 2009.

[42] P. W. Kasteleyn. The statistics of dimers on a lattice : I. the number of dimer arrangements on a quadratic lattice. *Physica*, 27:1209–1225, December 1961.

[43] H. N. V. Temperley and Michael E. Fisher. Dimer problem in statistical mechanics-an exact result. *Philosophical Magazine*, 6(68):1061–1063, 1961.

[44] L. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing*, 31(4):1229–1254, 2002.

[45] L.G. Valiant. Holographic algorithms. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 306–315, 2004.

[46] Robert Shrock and Yan Xu. Exact results on potts model partition functions inageneralized external field and weighted-set graph colorings. *Journal of Statistical Physics*, 141(6):909–939, 2010.

[47] Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2(5), 2014.

[48] C. R Laumann, R Moessner, A Scardicchio, and S. L Sondhi. Phase transitions and random quantum satisfiability. *Quant. Inf. and Comp.*, 10(1):0001, Mar 2010.

[49] Benjamin Hsu, C. R Laumann, A. M. Lauchli, R Moessner, and S. L Sondhi. Approximating random quantum optimization problems. *Phys. Rev. A*, 87:062334, 2013.

[50] A. F. Hebard and M. A. Paalanen. Magnetic-field-tuned superconductor-insulator transition in two-dimensional films. *Phys. Rev. Lett.*, 65:927, 1990.

[51] M. A. Paalanen, A. F. Hebard, and R. R. Ruel. Low temperature insulating phase of uniformly disordered two-dimensional superconductors. *Phys. Rev. Lett.*, 69:1604, 1992.

[52] D. Shahar and Z. Ovadyahu. Superconductivity near the mobility edge. *Phys. Rev. B*, 46:10917, 1992.

[53] V. F. Gantmakher, M. V. Golubkov, V. T. Dolgopolov, G. E. Tsydynzhapov, and A. A. Shashkin. *JETP Lett.*, 68:337, 1998.

[54] G. Sambandhamurthy, L.W. Engel, A. Johansson, and D. Shahar. Superconductivity-related insulating behaviour. *Phys. Rev. Lett.*, 92:107005, 2004.

[55] G. Sambandhamurthy, L. W. Engel, A. Johansson, E. Peled, and D. Shahar. Experimental evidence for a collective insulating state in two-dimensional superconductors. *Phys. Rev. Lett.*, 94:017003, 2005.

[56] H. Q. Nguyen, S. M. Hollen, M. D. Stewart Jr., J. Shainline, Yin Aijun, J. M. Xu, and J. M. Valles Jr. Observation of giant positive magnetoresistance in a cooper-pair insulator. *Phys. Rev. Lett.*, 103:157001, 2009.

[57] T. I. Baturina, A. Yu. Mironov, V. M. Vinokur, M. R. Baklanov, and C. Strunk. Hyperactivated resistance in tin films on the insulating side of the disorder-driven superconductor-insulator transition. *Pis' ma v ZhETF*, 88(11):867–872, 2008.

[58] Benjamin Sacépé. Localization of preformed cooper pairs in disordered superconductors. *Nature Phys.*, 7:239–244, 2011.

[59] V. F. Gantmakher and V. T. Dolgopolov. Superconductor-insulator quantum phase transition. *Physics-Uspekhi*, 53(1):1–49, 2010.

[60] M.P.A. Fisher. *Phys. Rev. Lett.*, 65:923, 1990.

[61] A.M. Finkelstein. *JETP Lett.*, 45:46, 1987.

[62] M. V. Feigel'man, L. B. Ioffe, V. E. Kravtsov, and E. Cuevas. Fractal superconductivity near localization threshold. *Annals of Physics*, 325(7):1390–1478, 2010.

[63] A Ghoshal, M. Randeria, and N. Trivedi. *Phys. Rev. B*, 65:014501, 2001.

[64] V. M. Galitski and A. I. Larkin. *Phys. Rev. B*, 63:174506, 2001.

[65] Y. Dubi, Y. Meir, and Y. Avishai. *Nature*, 449:876, 2007.

[66] M. Müller and B. I. Shklovskii. *Phys. Rev. B*, 79:134504, 2009.

[67] V. L. Pokrovsky, G. M. Falco, and T. Nattermann. Phase diagram of electron systems near the superconductor-insulator transition. *Phys. Rev. Lett.*, 105:267001, 2010.

[68] P. Reunchan, X. Zhou, S. Limpijumnong, A. Janotti, and C. G. Van de Walle. Vacancy defects in indium oxide: An ab-initio study. *Current Applied Physics*, 11:296–300, 2011.

[69] P. J. Baker, R. J. Ormeno, C. E. Gough, Y. Matsushita, and I. R. Fisher. Microwave surface impedance measurements of $tl_xpb_{1-x}te$: A proposed negative-u induced superconductor. *Phys. Rev. B*, 81:064506, 2010.

[70] J. Schmalian and M. Dzero. *Phys. Rev. Lett.*, 94:157003, 2005.

[71] Villain, J. Theory of one- and two-dimensional magnets with an easy magnetization plane. ii. the planar, classical, two-dimensional magnet. *J. Phys. France*, 36(6):581–590, 1975.

[72] A. L. Efros and B. I. Shklovskii. Coulomb gap and low temperature conductivity of disordered systems. *J. Phys. C*, 8:49–51, 1975.

[73] M. V. Feigel'man, L. B. Ioffe, and M. Mézard. Superconductor-insulator transition and energy localization. *Phys. Rev. B*, 82:184534, 2010.

[74] Y. Dubi, Y. Meir, and Y. Avishai. Theory of the magnetoresistance of disordered superconducting films. *Phys. Rev. B*, 73:054509, 2006.

[75] Tianran Chen, Brian Skinner, and B. I. Shklovskii. Coulomb gap triptych in a periodic array of metal nanocrystals. *Phys. Rev. Lett.*, 109:126805, Sep 2012.

[76] Brian Skinner, Tianran Chen, and B. I. Shklovskii. Theory of hopping conduction in arrays of doped semiconductor nanocrystals. *Phys. Rev. B*, 85:205316, May 2012.

[77] A. L. Efros and B. I. Shklovskii. *Electronic properties of doped semiconductors.* Springer Berlin, 1984.

[78] Anirban Gangopadhyay, Victor Galitski, and Markus Müller. Magnetoresistance of an anderson insulator of bosons. *Phys. Rev. Lett.*, 111:026801, Jul 2013.

[79] F .G. Pikus and A. L. Efros. Critical behaviour of density of states in disordered systems with localized electrons. *Phys. Rev. Lett.*, 73:3014, 1994.

[80] M. Müller and S. Pankov. Mean field theory for the three-dimensional coulomb glass. *Phys. Rev. B*, 75:144201, 2007.

[81] A. Möbius, P Karmann, and M Schreiber. Coulomb gap revisited - a renormalization approach. *J. Phys.: Conf. Ser.*, 150:022057, 2009.

[82] E. I. Levin, V. L. Nguen, B. I. Shklovskii, and A. L. Efros. *Sov. Phys. JETP*, 65:842, 1987.

[83] S. D. Baranovskii, A..L. Efros, B. L. Gelmont, and B. I. Shklovskii. Coulomb gap in disordered systems : computer simulation. *J. Phys. C*, 12:1023–1034, 1979.

[84] Martin Goethe and Matteo Palassini. Phase diagram, correlation gap, and critical properties of the coulomb glass. *Phys. Rev. Lett.*, 103:045702, Jul 2009.

[85] E. Bardalen, J. Bergli, and Y.M. Galperin. coulomb glasses: a comparison between mean-field and monte carlo results. *Phys. Rev. B*, 85:155206, 2012.

[86] H. Kamimura and A. Kurobe. Theoretical model on the interplay of disorder and electron correlations. *J. Phys. Soc. Japan*, 51:1904, 1982.

[87] A. Vaknin, A. Frydman, Z. Ovadyahu, and M. Pollak. *Phys. Rev. B*, 54:13604, 1996.

[88] Subir Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 2000.

[89] M. Schechter and P. C. E. Stamp. Inversion symmetric two-level systems and the low-temperature universality in disordered solids. *Phys. Rev. B*, 88:174202, Nov 2013.

[90] Allen Miller and Elihu Abrahams. Impurity conduction at low concentrations. *Phys. Rev.*, 103:745, 1960.

[91] A. Amir, Y. Oreg, and Y. Imry. Variable range hopping in the coulomb glass. *Phys. Rev. B*, 80:245214, 2009.

[92] V. Ambegaokar, B. Halperin, and J. Langer. *Phys. Rev. B*, 4:2612, 1971.

[93] R. N. Bhatt and Erik Nielsen. Search for ferromagnetism in doped semiconductors in the absence of transition metal ions. *Phys. Rev. B*, 82:195117, 2010.

[94] H.L. Zhao, B.Z. Spivak, M. Gelfand, and S. Feng. *Phys. Rev. B*, 44:10760, 1991.

[95] Markus Mller. Magnetoresistance and localization in bosonic insulators. *EPL (Europhysics Letters)*, 102(6):67008, 2013.

[96] P. W. Anderson. *J.Phys. Chem. Solids*, 11:26, 1959.

[97] Michael Ma and Patrick A. Lee. Localized superconductors. *Phys. Rev. B*, 32:5658, 1985.

[98] N. P. Armitage, E. Helgren, and G. Gruner. *Phys. Rev. B.*, 69:014201, 2004.

[99] J. G. Massey and M. Lee. *Phys. Rev. Lett.*, 75:4266, 1995.

[100] A. Vaknin, Z. Ovadyahu, and M. Pollak. *Phys. Rev. B*, 65:134208, 2002.

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

[102] O. Naaman, W. Tezier, and R. C. Dynes. *Phys. Rev. Lett.*, 87:097004, 2001.

[103] M. van Lier and R. Otten. Planarization by transformation. *Circuit Theory, IEEE Transactions on*, 20(2):169–171, Mar 1973.

[104] E.B. Curtis, D. Ingerman, and J.A. Morrow. Circular planar graphs and resistor networks. *Linear Algebra and its Applications*, 283(13):115 – 150, 1998.

[105] F Friedrich, A Kempe, V Liebscher, and G Winkler. Complexity penalized m-estimation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.

[106] M. B. Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. *Phys. Rev. B*, 76:201102, Nov 2007.

[107] Florent Krzakala, Alberto Rosso, Guilhem Semerjian, and Francesco Zamponi. Path-integral representation for quantum spin models: Application to the quantum cavity method and monte carlo simulations. *Phys. Rev. B*, 78:134428, Oct 2008.

[108] C. Laumann, A. Scardicchio, and S. L. Sondhi. Cavity method for quantum spin glasses on the bethe lattice. *Phys. Rev. B*, 78:134424, Oct 2008.

[109] Armin Biere, Marijn J. H. Heule, Hans van Maaren, and Toby Walsh, editors. *Handbook of Satisfiability*, volume 185 of *Frontiers in Arificial Intelligence and Applications*. IOS Press, February 2009.