Dynamics for the Random Cluster Model

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Christopher Kacwin

aus

Hamburg

Bonn 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Michael Griebel

2. Gutachter: Prof. Dr. Marc Alexander Schweitzer

Tag der Promotion: 03. August 2021

Erscheinungsjahr: 2021

Zusammenfassung

Das „random cluster“-Modell ist ein zentrales statistisches Modell mit Anwendungen in der Physik, der Wahrscheinlichkeitstheorie und der Graphentheorie. Für einen Graphen $G = (V, E)$ betrachtet man eine Wahrscheinlichkeitsverteilung auf Teilkantenmengen $A \subseteq E$ gegeben durch

$$\mu(A) = Z^{-1} p^{|A|} (1-p)^{|E \setminus A|} q^{c(A)}$$

mit Parametern $p \in (0,1)$ und $q \in (0,\infty)$. Hierbei ist $c(A) \in \mathbb{N}$ die Anzahl der Zusammenhangskomponenten des Teilgraphen (V, A) (die sogenannten „cluster“), und Z ist eine Normalisierungskonstante gegeben durch

$$Z = \sum_{A \subseteq E} p^{|A|} (1-p)^{|E \setminus A|} q^{c(A)} .$$

Bei diesem Modell ist der Phasenübergang für bestimmte Graphen besonders interessant: Bei kleinem p befindet sich das Modell in der subkritischen Phase, charakterisiert durch eine hohe Zahl von Zusammenhangskomponenten mit geringer Ausdehnung. Bei großem p befindet sich das Modell in der superkritischen Phase, hier ist die Existenz einer globalen, den ganzen Graphen durchdringenden Komponente bezeichnend. Die Phasen sind getrennt durch einen kritischen Punkt $p_c = p_c(q, G)$, und die Eigenschaften des Modells an diesem kritischen Punkt sind Gegenstand intensiver Forschung.

Ein effektives Instrument zur Untersuchung solcher Modelle sind Markov-Ketten, die gegen die gewünschte Wahrscheinlichkeitsverteilung konvergieren. Diese werden unter anderem benötigt, um praktische Simulationen des Modells durchzuführen, da es normalerweise nicht möglich ist, Zustände gemäß der gewünschten Verteilung direkt zu erzeugen. In dieser Arbeit befassen wir uns mit verschiedenen Markov-Ketten für das „random cluster“-Modell mit besonderem Augenmerk auf der Konvergenzgeschwindigkeit, in Abhängigkeit von der vorherrschenden Phase des Modells.

Hierfür werden zunächst grundlegende Techniken im Zusammenhang mit Markov-Ketten zusammengetragen. Wir leiten Kenngrößen für die Konvergenzgeschwindigkeit von Markov-Ketten und Varianzabschätzungen für Markov-Chain-Monte-Carlo-Verfahren her, und bringen diese in Zusammenhang mit der Kopplung von Markov-Ketten.

Danach folgt eine umfassende Einführung in das „random cluster“-Modell und verwandte klassische Modelle. Wir betrachten Monotonie-Eigenschaften des Modells auf endlichen und unendlichen Graphen. Für den $\mathbb{Z}^2$-Gittergraphen tragen wir die wichtigsten Resultate über den Phasenübergang des Modells zusammen. Hierbei sind die Konsequenzen für rechteckige Teilgraphen von $\mathbb{Z}^2$ für das weiterführende Vorgehen von besonderer Bedeutung.

Nach einer Erörterung der für uns interessanten Markov-Ketten betrachten wir Beweistechniken für monotone und lokale Markov-Ketten. Diese sind vor allem in der subkritischen Phase anwendbar und führen zu optimalen Konvergenz-Resultaten. Zudem leiten wir eine Technik her, die es mit Hilfe von „speed of disagreement percolation“-Resultaten erlaubt, von beliebig polynomieller auf optimale Konvergenzgeschwindigkeit zu schließen. Die „speed of disagreement percolation“-Resultate sind kombinatorischer Natur und deren Herleitung hängt oft nicht mit der Phase des Modells zusammen. In der superkritischen Phase sind lokale Markov-Ketten oft ineffektiv. Wir zeigen, dass man die Techniken und Konvergenzresultate auf die superkritische Phase verallgemeinern kann, wenn man sich darauf beschränkt, die Markov-Kette in der Nähe der Zielverteilung starten zu lassen.

Im nächsten Kapitel führen wir das „down-up“-Modell ein. Dieses neue Modell ist eng mit dem „random cluster“-Modell verwandt und kann als eine nichtlokale Verallgemeinerung des Ising-Modells angesehen werden. Wir zeigen, dass dieses Modell eine starke monotone Struktur hat und nutzen die Nähe zum „random cluster“-Modell um Phaseneigenschaften herzuleiten. Zudem führen wir monotone Markov-Ketten für dieses Modell ein und zeigen exemplarisch für die „monotone down-up dynamics“-Markov-Kette, dass sich die Techniken aus dem vorherigen Kapitel hierauf übertragen lassen.

Zum Schluss betrachten wir die Swendsen-Wang-Markov-Kette für das „random cluster“-Modell. Diese ist weder lokal noch monoton, deshalb lassen sich die Techniken aus den vorherigen Kapiteln nicht übertragen. Bezüglich der subkritischen Phase sind in den letzten Jahren dennoch Fortschritte erzielt worden. Wir tragen Konvergenzresultate für diese Phase zusammen und richten danach unser Augenmerk auf die noch offene Frage nach der Konvergenzgeschwindigkeit in der superkritischen Phase. Hierzu führen wir eine zum Teil neue Kopplung der Markov-Kette ein und zeigen mit numerischen Simulationen, dass diese optimale Konvergenzgeschwindigkeit nahelegt. Zudem belegen wir numerisch, dass die Kopplung eine Lokalitätseigenschaft besitzt. Dazu leiten wir ein „speed of disagreement percolation“-Resultat her, diesmal allerdings in Abhängigkeit von der superkritischen Phase.

Contents

1 Introduction **1**

2 Markov Chain Monte-Carlo Methods **7**
2.1 Monte-Carlo integration 7
2.2 Perturbed Monte-Carlo integration 8
2.3 Markov chain Monte-Carlo integration 11
2.4 Mixing time 16
2.5 Markov chain couplings 19
2.6 Grand couplings 21
2.7 Path coupling 23

3 The Fortuin-Kasteleyn Random Cluster Model **27**
3.1 Monotonicity of the random cluster model 28
3.2 The q-state Potts model 31
3.3 Monotonicity of the Ising model 34
3.4 Infinite volume measures 36
3.5 The critical point 40

4 Markov chains for the Fortuin-Kasteleyn Random Cluster Model **43**
4.1 Glauber dynamics 43
4.2 Block dynamics 44
4.3 Swendsen-Wang dynamics 45
4.4 Monotone Swendsen-Wang dynamics 45
4.5 Heatbath dynamics 48
4.6 Spin block dynamics 49

5 State of the Art **51**
5.1 Spin block dynamics 58
5.2 The Ising model case 60
5.3 The Fortuin-Kasteleyn random cluster setting 64
5.4 Speed of disagreement percolation 72
5.5 The supercritical phase 80

6 The Down-Up Model **91**
6.1 Monotonicity with respect to boundary conditions 92
6.2 Monotonicity with respect to increasing events 100
6.3 Phase properties in the $\mathbb{Z}^2$-planar case 101

6.3.1 Monotone down-up dynamics in the subcritical phase 103
6.3.2 Monotone down-up dynamics in the supercritical phase . . . 106

7 Couplings for the Swendsen-Wang Dynamics 109
7.1 Grand coupling 109
7.2 Spectral gap comparison results 110
7.3 Spectral gap results in the subcritical phase 111
7.4 Recent mixing time results 112
7.5 Experimental results in the supercritical phase 113
7.5.1 Coupling time measurements 113
7.5.2 Limit distribution and locality of the coupling 119
7.6 The H^*-coupling in the supercritical phase 125

8 Conclusion 127
8.1 Summary 127
8.2 Outlook 128

Bibliography 131

1 Introduction

Statistical mechanics is a field of research which has gained increasing importance over the course of the last 150 years. Originally introduced by James Clerk Maxwell [Max60] and Ludwig Boltzmann [Bol96] to explain kinetic properties of gas, the concepts were generalized by J. Williard Gibbs [Gib02] for general mechanical systems. Today, the application fields include high energy physics, weather forecasting, bioinformatics and game theory. It is effective in dealing with models that have a very large amount of microscopic agents from which macroscopic features emerge, e.g. gas particles. To track the motion of such particles in a deterministic way quickly becomes unfeasible for increasing system sizes, even with current technology. Instead, the state of a given system is postulated to be a random variable X, with an associated probability measure

$$\mu(X) = \frac{1}{Z} \exp(-\beta S(X))$$

which is called Boltzmann distribution. The action term $S(X)$ describes the energy of a given state, and the parameter β acts as an inverse temperature, i.e. at high β (low temperature) the variation of the model is reduced. The normalization constant

$$Z = \sum_X \exp(-\beta S(X))$$

is called partition function. This model is motivated by the simple assumptions that the probability of a given state entirely depends on its energy, and that the probability ratio of two states

$$\frac{\mu(X)}{\mu(Y)} = \exp(-\beta(S(X) - S(Y)))$$

only depends on the difference of their energies. States with high energy terms are less probable, the system tries to minimize its energy in a probabilistic manner. This is in accordance to the "least action principle" in physics which is often used to derive the deterministic equations for an evolving system.
Even though the model is probabilistic in nature, it is still possible to make predictive statements using analytic tools and numerical simulations. The main interests typically can be formulated as questions like "What is the probability of event $\mathcal{A}$?"

or "What is the expected value of quantity f?", which leads to the calculation

$$\mathbb{E}[f] = \sum_X f(X)\mu(X)\,,$$

with f being called observable. On one hand, this sum has a lot in common with the partition function, and often it is possible to gain insights into the model if the partition function is well-understood analytically. On the other hand, numerical approximations using Monte-Carlo integration methods are possible under certain circumstances.

From the Ising model...

In 1924 the mathematician and physicist Ernst Ising attained his Ph.D. solving the what is now called 1D-Ising model [Isi25], laying the ground work for one of the most famous models of statistical mechanics. In an attempt to understand ferromagnetic interaction of particles, he assigned to each integer $k \in \{1, \dots, N\}$ a spin variable $x(k) \in \{-1, +1\}$ together with an action term

$$S(x) = -\sum_{k=1}^{N-1} J_k x(k)x(k+1)$$

that favors alignment of spins. The positive weights J_k represent the bond strength of the ferromagnetic interaction between the k-th and $(k+1)$-th particle. Ising showed that the resulting long-distance interaction of particles decays exponentially fast with increasing distance, for all inverse temperatures $\beta > 0$ and increasing system size N. The model therefore has only one phase, i.e. a definite characteristic behavior. From these considerations, Ising incorrectly concluded that the Ising model must have this property for all dimensions.
In 1936, Rudolf Peierls proved in [Pei36] that the 2-dimensional Ising model undergoes a phase transition between an ordered and a disordered phase. For small β, the long range interaction between spins decays exponentially fast with increasing distance, which results in a disordered state resembling white noise. For large β, long range interactions do not decay to zero, resulting in an ordered state where most spins point in the same direction. These phases are strictly separated by the critical inverse temperature β_c. The striking feature of the model is given by the fact that it accurately predicts the behavior of ferromagnetic materials under varying temperature. When heated to high temperatures, magnets lose their macroscopic magnetic properties due to the dominating kinetic energy of the particles. When cooled down, at a certain critical point the particles realign into a specific direction, a phenomenon called spontaneous magnetization.
In 1944, Lars Onsager [Ons44] solved the 2-dimensional Ising model on a discrete torus grid without external field, giving formulas for correlation functions and the free energy. In 1949, he announced the formula for the spontaneous magnetization without proof, which was delivered by Chen-Ning Yang [Yan52] and Gábor Szegő [Sze52] two years later.

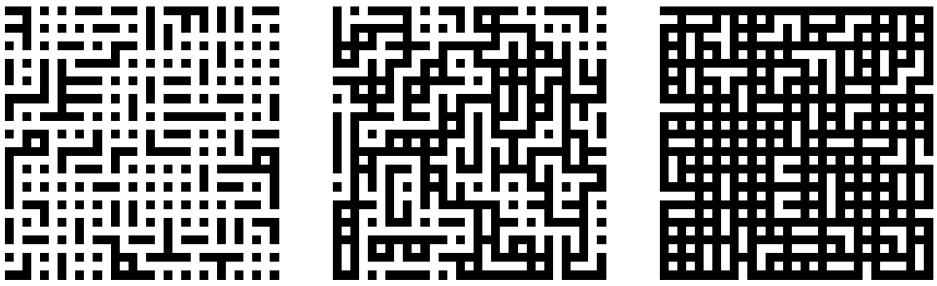

Figure 1.1: Random cluster model simulations on a 16×16 grid with $q = 2$ and $p = p_c - 0.2$ (left), $p = p_c$ (middle) and $p = p_c + 0.35$ (right). The cluster structure (black) shows the typical behavior of the model for the different phases. In the subcritical phase clusters are small, while in the supercritical phase there is a dominating global cluster.

...to the random cluster model

In 1969, Kees Fortuin and Piet Kasteleyn [FK72] introduced the random cluster model to study combinatorial structures and electrical networks. This model can be thought of as a unification of the percolation model, Ising model and Potts model. Let $G = (V, E)$ be an undirected, finite and connected graph. The set of possible states $\mathbb{X} = \{A \subseteq E\}$ is given as the set of all subsets of E. For such a state A, we say that an edge $e \in E$ is *open* if $e \in A$, otherwise it is *closed*. Moreover, let $c(A)$ be the number of connected components of A, also called (open) *clusters*. The random cluster measure on $\mathbb{X}$ with edge parameter $p \in [0, 1]$ and cluster weight $q \in (0, \infty)$ is then given by

$$\mu(A) = \mu_{G,p,q}(A) = Z_{G,p,q}^{-1} \, p^{|A|} \, (1-p)^{|E \setminus A|} \, q^{c(A)} \, ,$$

where the partition function $Z = Z_{G,p,q}$ is chosen such that μ is a probability measure. It is easy to see that for $q = 1$, this model reduces to edge percolation on G and one has $Z_{G,p,1} = 1$. For $q \in \{2, 3, \ldots\}$, this model is closely related to the q-state Potts model, with the special case of $q = 2$ being the Ising model. Due to the explicit nature of Onsager's proofs for the Ising model, many questions pertaining the random cluster model could not be answered in a satisfying way until recently. One of the main problems was the rigorous definition of the occurring phase transition of the 2-dimensional random cluster model, and the particular behavior of the model at its critical point, see Figure 1.1.
This field of research retained its popularity into the present day, as can be seen from the sheer number of publications on the topic each year. The breakthrough came in 2010 with the work of Vincent Beffara and Hugo Duminil Copin [BDC12] showing that the model has a sharp phase transition at the critical point, and that the subcritical phase has the 'exponential decay of correlations' property. In the same year, Stanislaw Smirnov was awarded the Fields medal "for the proof of conformal invariance of percolation and the planar Ising model in statistical physics", establishing a beautiful link between critical statistical mechanics and conformal invariance, a property that is conjectured to hold for a variety of statistical model.

Markov chain Monte-Carlo integration

Often, it is not possible to solve statistical models analytically. This stems from the fact that for a model with microscopic entities, the system size has to be very large to observe emerging macroscopic properties, vastly increasing the computational complexity. For instance, the 16x16 grid graph from Figure 1.1 has 480 edges, which is a very small number of variables for typical applications. This however already results in a state space $\mathbb{X}$ with 2^{480} elements, relinquishing all hope to make accurate computations. Moreover, simulating the distribution directly fails for the same reasons.
The state of the art method to approximate observables in this case is given by the Markov chain Monte-Carlo method. Introducing a Markov chain $(A^t)_{t\in\mathbb{N}}$ which has μ as stationary distribution is often possible, and this allows us to generate samples from μ in a dynamic way. The main questions in this regard are:

- How fast does $(A^t)_{t\geq 0}$ approach to equilibrium?
- How big is the correlation between subsequent samples A^t and A^{t+k} for $k \in \mathbb{N}$?

The answers to these questions can be given in terms of the *mixing time* and *spectral gap* of the Markov chain. If we say that for the initial state A^0 the random variable A^t has distribution μ^t, the mixing time is defined as

$$\tau(\epsilon) = \min\left\{t\colon \sup_{A^0} \|\mu - \mu^t\|_{TV} \leq \epsilon\right\},$$

i.e. it is given by the minimal time t at which the distribution of A^t is close to equilibrium regardless of initial value, measured in total variation distance. The spectral gap of a Markov chain is given by

$$(1 - |\lambda_2|)^{-1},$$

where λ_2 is the Eigenvalue of the Markov chain that has the biggest absolute value aside from 1. It can be shown that variance estimates concerning the Markov chain Monte-Carlo integration method are directly affected by the spectral gap.
For statistical models, it is important to characterize the dependence of these quantities on the system size. If we consider the random cluster model for instance, the mixing time of a Markov chain might grow exponentially fast with respect to the number of edges of the graph G. The task of generating samples close to equilibrium will be unfeasible for already moderate system sizes in this case. The study of Markov chains as well as corresponding mixing times and spectral gaps is therefore vital to the field of statistical mechanics.
A famous example has been introduced by Robert Swendsen and Jian-Sheng Wang [SW87] in 1987. The now called Swendsen-Wang algorithm exploits the connection between the random cluster model and the corresponding Potts model, creating a Markov chain that switches between both representations in a computationally convenient way. In contrast to local Markov chains that attempt to evolve variables one at a time, the Swendsen-Wang Markov chain changes the given state in a global way, which seems to be one reason for its good scaling properties with respect to

the system size. The algorithm has been studied extensively in the last few decades, because it is a major hope to many other fields dealing with statistical mechanics that up to this day lack a convenient way to compute large-scale simulations.
The main effort of this thesis is to understand Markov chains in the random cluster model setting, particularly in the 2-dimensional case. The publication [BDC12] gave the ground work for many results to follow during the last 10 years, mainly concerning the subcritical phase of the model. For the Swendsen-Wang algorithm, many questions still remain unanswered, and this thesis will contribute to its investigation both in theory and with numerical experiments.

Outline of this thesis

The remainder of this thesis is organized as follows:

- In the second chapter, we derive basic tools needed to understand Markov chains and Markov chain Monte-Carlo methods. Key characteristics like the spectral gap and mixing time are introduced, and put into context with various coupling techniques such as the 'coupling from the past' method and the path coupling theorem.

- Chapter 3 gives a thorough introduction to the random cluster model. We deduce monotonicity properties of the model on finite and infinite graphs, using Glauber dynamics arguments. Recent results concerning the phase transition of the model in 2 dimensions are stated. We elaborate on the connection to percolation, the Ising model and the Potts model.

- In Chapter 4, we summarize the Markov chains that are interesting in the context of the random cluster model. These include Glauber dynamics, heatbath dynamics, alternate scan dynamics and the Swendsen-Wang dynamics as well as its monotone version.

- Chapter 5 reviews recent results and proof techniques for a variety of the just mentioned Markov chains, mainly using the monotonicity property of the model together with exponential decay of correlations in the subcritical phase. We introduce 'speed of disagreement percolation' estimates for coupled Markov chains and show that a bootstrap argument can often give optimal mixing time results for Markov chains with polynomial mixing time. Moreover, some of the techniques can be used in the supercritical phase as well, leading to new results for monotone dynamics in this case.

- As an application, we introduce a new model in Chapter 6 and apply the methods from Chapter 5 to derive model properties such as local SSM and a local to global coupling result. This down-up model is closely related to the random cluster model and therefore has potential to deliver new insights. We show that the model has a rich partial order structure, and investigate monotone Markov chains for this model, including a generalization of the monotone Swendsen-Wang dynamics. For this Markov chain we expemplary show mixing time results in the subcritical and supercritical phase.

- The original Swendsen-Wang dynamics Markov chain does not exhibit the monotonicity property, therefore we review it separately in Chapter 7. We state recent results concerning the mixing time of this Markov chain, as well as coupling results obtained from a grand coupling argument. In the supercritical phase, the theoretical results are not optimal, and we investigate certain couplings that are promising in this regard. We show a typical 'speed of disagreement percolation' bound for a Swendsen-Wang dynamics coupling in the supercritical phase. Moreover, we provide extensive numerical evidence that the discussed couplings perform exceedingly well, in accordance to the conjecture about the mixing time of the Swendsen-Wang dynamics in the supercritical phase.

- We conclude the thesis with a summary and an outlook regarding possible extensions of this work in Chapter 8.

2 Markov Chain Monte-Carlo Methods

In this chapter we will introduce the Markov chain Monte-Carlo method and related notions, see for instance [LP17]. Throughout the chapter, we will consider a finite set $\mathbb{X}$ called *state space*, together with a probability mass function $\mu\colon \mathbb{X} \to (0,1]$ such that

$$\sum_{X\in\mathbb{X}} \mu(X) = 1\,.$$

We call a function $f\colon \mathbb{X} \to \mathbb{R}$ an *observable* and our main goal is to compute the expected value of an observable:

$$\mathbb{E}[f] = \sum_{X\in\mathbb{X}} f(X)\mu(X)\,.$$

We will focus on probabilistic methods to obtain an approximation of this quantity.

2.1 Monte-Carlo integration

The easiest method one can think of would be the standard Monte-Carlo estimator

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i)\,,$$

where the X_i's are independent samples drawn according to μ. It is easy to see that this estimator is unbiased, i.e. the expected value is exactly $\mathbb{E}[f]$. The quality of this estimator can be measured by its variance

$$\mathbb{E}\left[\left(\mathbb{E}[f] - \frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2\right]\,.$$

Normally one lacks the possibility to simulate μ directly like this, but for the sake of completeness we include

Theorem 2.1 (Monte-Carlo integration)**.** *For independent and identically distributed random variables $X_i \sim \mu$, $i = 1, \dots, N$ it holds*

$$\mathbb{E}\left[\left(\mathbb{E}[f] - \frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2\right] = \frac{1}{N}\mathrm{Var}[f]\,,$$

with $\mathrm{Var}[f] = \mathbb{E}\left[(\mathbb{E}[f] - f)^2\right]$.

Proof. We write $\mathbb{E}[f] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[f]$ and rearrange:

$$\mathbb{E}\left[\left(\mathbb{E}[f] - \frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2\right] = \frac{1}{N^2}\,\mathbb{E}\left[\left(\sum_{i=1}^{N}(\mathbb{E}[f] - f(X_i))\right)^2\right].$$

This expression is the variance applied to the sum of independent identically distributed random variables with mean zero, therefore we get

$$\mathbb{E}\left[\left(\mathbb{E}[f] - \frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2\right] = \frac{1}{N}\mathbb{E}\left[(\mathbb{E}[f] - f)^2\right] = \frac{1}{N}\mathrm{Var}[f]\,.$$

□

We see that the variance of the Monte-Carlo estimator does neither depend on smoothness properties of f, nor on the dimensionality of the underlying space $\mathbb{X}$. This is especially useful for high-dimensional integration problems, as is often the case in statistical physics. The rate at which the variance decreases is rather slow, however in most cases one cannot hope to perform better.

2.2 Perturbed Monte-Carlo integration

We turn to the case where it is not possible to simulate μ in practice. Let us assume that we can draw samples from an approximation μ_ϵ of μ. How does the corresponding estimator behave? Let us first take a look at the expectation. For $X \sim \mu$ and $Y \sim \mu_\epsilon$, one has

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \mathbb{E}\left[|f(X) - f(Y)|\right] \leq \mathbb{P}[X \neq Y] \sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|\,.$$

Because we have defined X and Y only up to their respective distribution, the quantity

$$\mathbb{P}[X \neq Y]$$

does not make any sense right now. If we however consider (X, Y) as a random variable with joint distribution ν such that $X \sim \mu$ and $Y \sim \mu_\epsilon$, we can compute

$$\mathbb{P}[X \neq Y] = \sum_{X,Y\in\mathbb{X}} \nu(X, Y)\mathbb{1}[X \neq Y]\,.$$

Such a pair (X, Y) is called *coupling* of X and Y. Our above calculation holds for all possible couplings of X and Y, so it is natural to ask which coupling minimizes $\mathbb{P}[X \neq Y]$. First, it is easy to see that for any event $\mathcal{A} \subseteq \mathbb{X}$, one has

$$\mathbb{P}[X \neq Y] \geq \mathbb{P}[X \in \mathcal{A}, Y \notin \mathcal{A}] \geq \mathbb{P}[X \in \mathcal{A}] - \mathbb{P}[Y \in \mathcal{A}] = \mu[\mathcal{A}] - \mu_\epsilon[\mathcal{A}]\,,$$

which gives

$$\mathbb{P}[X \neq Y] \geq \max_{\mathcal{A} \subseteq \mathbb{X}} (\mu[\mathcal{A}] - \mu_\epsilon[\mathcal{A}]) = \|\mu - \mu_\epsilon\|_{TV}\,.$$

So we have a lower bound which is exactly the total variation distance between μ and μ_ϵ. But can we realize this lower bound with a coupling? The answer is yes, and it is constructive. Let $\mathcal{A}$ be the set of states such that

$$Z \in \mathcal{A} \Leftrightarrow \mu(Z) \leq \mu_\epsilon(Z)\,.$$

Additionally define a probability measure ν_+ on $\mathcal{A}$ which is proportional to $\mu_\epsilon - \mu$ on $\mathcal{A}$, and a probability measure ν_- on $\mathbb{X} \setminus \mathcal{A}$ proportional to $\mu - \mu_\epsilon$. Then we can define a coupling procedure as follows:

- Pick a random number $p \in [0, 1]$, uniformly distributed.
- If $p \leq \mu[\mathcal{A}]$, sample X from $\mu(\cdot \mid X \in \mathcal{A})$ and set $Y = X$.
- If $\mu[\mathcal{A}] < p < \mu_\epsilon[\mathcal{A}]$, sample X from ν_- and Y from ν_+.
- If $\mu_\epsilon[\mathcal{A}] \leq p$, sample Y from $\mu_\epsilon(\cdot \mid Y \notin \mathcal{A})$ and set $X = Y$.

It is easy to see that for this coupling, one has

$$\mathbb{P}[X \neq Y] = \mu_\epsilon[\mathcal{A}] - \mu[\mathcal{A}] \leq \|\mu - \mu_\epsilon\|_{TV}\,.$$

The distribution of X is given as follows: for $Z \in \mathcal{A}$, one has

$$\mathbb{P}[X = Z] = \mu[\mathcal{A}]\mu(Z \mid Z \in \mathcal{A}) = \mu(Z)\,.$$

For $Z \notin \mathcal{A}$, one has

$$\begin{aligned}
\mathbb{P}[X = Z] &= (\mu_\epsilon[\mathcal{A}] - \mu[\mathcal{A}])\nu_-(Z) + (1 - \mu_\epsilon[\mathcal{A}])\mu_\epsilon(Z \mid Z \notin \mathcal{A}) \\
&= (\mu_\epsilon[\mathcal{A}] - \mu[\mathcal{A}])\frac{\mu(Z) - \mu_\epsilon(Z)}{\mu[\mathbb{X} \setminus \mathcal{A}] - \mu_\epsilon[\mathbb{X} \setminus \mathcal{A}]} + \mu_\epsilon(Z) \\
&= \mu(Z) - \mu_\epsilon(Z) + \mu_\epsilon(Z) \\
&= \mu(Z)\,.
\end{aligned}$$

So we have $X \sim \mu$, and a similar computation gives $Y \sim \mu_\epsilon$. We have shown that

$$\max_{\mathcal{A} \subseteq \mathbb{X}} (\mu[\mathcal{A}] - \mu_\epsilon[\mathcal{A}]) = \min_{(X,Y)} \mathbb{P}[X \neq Y \mid X \sim \mu, Y \sim \mu_\epsilon]\,,$$

where the minimum on the right hand side is over all couplings of X and Y. Using these notions, we can prove a variance estimate for the estimator which uses the perturbed measure μ_ϵ.

Theorem 2.2 (Perturbed Monte-Carlo integration)**.** *Let μ, μ_ϵ be probability measures on $\mathbb{X}$. For $i = 1, \dots, N$, let Y_i be i.i.d. random variables with law μ_ϵ, and let X be a random variable with law μ. Then one has*

$$\begin{aligned} & \sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \tfrac{1}{N}\sum_{i=1}^N f(Y_i)\right)^2\right]} \\ \leq \; & \tfrac{1}{\sqrt{N}}\sqrt{\mathrm{Var}[f]} + \left(\sqrt{\tfrac{\|\mu-\mu_\epsilon\|_{TV}}{N}} + \|\mu - \mu_\epsilon\|_{TV}\right) \sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)| \,. \end{aligned}$$

Proof. For $i = 1, \dots, N$ let X_i be random variables with law μ and couple (X_i, Y_i) such that $\mathbb{P}[X_i \neq Y_i] = \|\mu - \mu_\epsilon\|_{TV}$. By considering the unbiased estimator $\frac{1}{N}\sum f(X_i)$, we can estimate

$$\begin{aligned} & \sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \tfrac{1}{N}\sum_{i=1}^N f(Y_i)\right)^2\right]} \\ = \; & \sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \tfrac{1}{N}\sum_{i=1}^N f(X_i) + \tfrac{1}{N}\sum_{i=1}^N f(X_i) - \tfrac{1}{N}\sum_{i=1}^N f(Y_i)\right)^2\right]} \\ \leq \; & \tfrac{1}{\sqrt{N}}\sqrt{\mathrm{Var}[f]} + \sqrt{\mathbb{E}\left[\left(\tfrac{1}{N}\sum_{i=1}^N f(X_i) - \tfrac{1}{N}\sum_{i=1}^N f(Y_i)\right)^2\right]} \\ = \; & \tfrac{1}{\sqrt{N}}\sqrt{\mathrm{Var}[f]} + \tfrac{1}{N}\sqrt{\mathbb{E}\left[\left(\sum_{i=1}^N [f(X_i) - f(Y_i)]\right)^2\right]} \,. \end{aligned}$$

Here we used the triangle inequality together with Theorem 2.1. In the second summand, the terms $f(X_i) - f(Y_i)$ are random variables which are independent, and each of them has probability $p = \|\mu - \mu_\epsilon\|_{TV}$ to be non-zero. Therefore, we can treat this term similar to a binomial distribution, where each non-zero contribution can be estimated by $\sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|$. We get

$$\begin{aligned} & \mathbb{E}\left[\left(\sum_{i=1}^N [f(X_i) - f(Y_i)]\right)^2\right] \\ \leq \; & \sum_{i=0}^N \binom{N}{i} p^i (1-p)^{N-i} \left(i \cdot \sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|\right)^2 \\ = \; & \left(\sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|\right)^2 \sum_{i=0}^N \binom{N}{i} p^i (1-p)^{N-i} i^2 \\ = \; & \left(\sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|\right)^2 (Np + N(N-1)p^2) \,. \end{aligned}$$

Combining this with the upper calculation gives

$$\begin{aligned} & \sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \tfrac{1}{N}\sum_{i=1}^{N} f(Y_i)\right)^2\right]} \\ \leq \quad & \sqrt{\tfrac{\mathrm{Var}[f]}{N}} + \left(\sqrt{\tfrac{\|\mu-\mu_\epsilon\|_{TV}}{N}} + \|\mu-\mu_\epsilon\|_{TV}\right) \sup_{W,Z\in\mathbb{X}} |f(W)-f(Z)| \,. \end{aligned}$$

□

We see that this estimate reduces to the estimate from Theorem 2.1 in the case $\mu = \mu_\epsilon$. Moreover, for $N \to \infty$ we get the remainder

$$\|\mu - \mu_\epsilon\|_{TV} \sup_{W,Z\in\mathbb{X}} |f(W) - f(Z)|$$

which is a reasonable bound in this situation.

2.3 Markov chain Monte-Carlo integration

We turn our attention to estimators which are based on Markov chains. Let

$$P\colon \mathbb{X}\times\mathbb{X} \to [0,1]$$

be a probability matrix. We can define a discrete-time Markov chain with initial value $X^0 \in \mathbb{X}$ via

$$X^t \sim P(X^{t-1}, \cdot) \text{ for } t \in \mathbb{N} \,.$$

We say that the Markov chain $(X^t)_{t\geq 0}$ (or P) is reversible with respect to μ if P satisfies

$$\mu(X)P(X,Y) = \mu(Y)P(Y,X)$$

for all $X, Y \in \mathbb{X}$. It is easy to see that $\mu P = \mu$ in this case (if we interpret μ as a row vector) and that P^t is also reversible with respect to μ for any $t \in \mathbb{N}$. Moreover, reversible Markov chains are easy to analyze because the corresponding linear operator

$$f \mapsto Pf = \sum_{X\in\mathbb{X}} P(\cdot, X) f(X)$$

is self-adjoint with respect to μ:

$$\begin{aligned} (Pf, g)_\mu &= \sum_{X\in\mathbb{X}} Pf(X)g(X)\mu(X) = \sum_{X\in\mathbb{X}}\sum_{Y\in\mathbb{X}} P(X,Y)f(Y)\, g(X)\mu(X) \\ &= \sum_{X\in\mathbb{X}}\sum_{Y\in\mathbb{X}} P(Y,X)f(Y)\, g(X)\mu(Y) = \sum_{Y\in\mathbb{X}} f(Y)Pg(Y)\mu(Y) = (f, Pg)_\mu \,. \end{aligned}$$

This means that we can find an orthogonal basis $v_1, v_2, \ldots, v_k$ of $\{f\colon \mathbb{X}\to\mathbb{R}\}$ such that

$$(P^t v_i, v_j)_\mu = \begin{cases} \lambda_i^t (v_i, v_i)_\mu & i = j \\ 0 & i \neq j \end{cases}$$

with real Eigenvalues λ_i for $i = 1, \dots, k$. Without loss of generality we can choose v_1 defined via $v_1(X) = 1$ for all $X \in \mathbb{X}$, with $\lambda_1 = 1$ as is common for stochastic matrices. The Eigenvalues λ_i of P moreover satisfy $|\lambda_i| \leq 1$ due to the Gerschgorin circle theorem. We assume without loss of generality that $|\lambda_i| \geq |\lambda_j|$ for $i \geq j$. All of this will play a crucial role for the next estimator.

Theorem 2.3 (Idealized Markov chain Monte-Carlo integration)**.** *Let $X, X^0 \sim \mu$ and $X^t \sim P(X^{t-1}, \cdot)$ for $t \in \mathbb{N}$, where P is a probability matrix on $\mathbb{X}$ that is reversible with respect to μ. Assume that $|\lambda_2| < 1$. Then we have*

$$\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \frac{1}{N}\sum_{t=0}^{N-1} f(X^t)\right)^2\right] \leq \frac{1}{N}\mathrm{Var}[f]\left(1 + \frac{2}{1-|\lambda_2|}\right).$$

Proof. As in Theorem 2.1 we expand $\mathbb{E}[f(X)] = \frac{1}{N}\sum_{t=0}^{N-1}\mathbb{E}[f(X)]$ and get

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \frac{1}{N}\sum_{t=0}^{N-1} f(X^t)\right)^2\right] \\
= \;& \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{t=0}^{N-1}[\mathbb{E}[f(X)] - f(X^t)]\right)^2\right] \\
= \;& \frac{1}{N^2}\sum_{s,t=0}^{N-1}\mathbb{E}\left[(\mathbb{E}[f(X)] - f(X^s))(\mathbb{E}[f(X)] - f(X^t))\right] \\
= \;& \frac{1}{N}\mathrm{Var}[f] + \frac{2}{N^2}\sum_{s=0}^{N-1}\sum_{t=s+1}^{N-1}\mathbb{E}\left[(\mathbb{E}[f(X)] - f(X^s))(\mathbb{E}[f(X)] - f(X^t))\right].
\end{aligned}$$

Here, X^s and X^t are not independent, we have that $X^s \sim \mu P^s = \mu$ and $X^t \sim P^{t-s}(X^s, \cdot)$. This means that the dependence between X^s and X^t is the same as the dependence between X^0 and X^{t-s}. Therefore, we can write

$$\begin{aligned}
&\frac{2}{N^2}\sum_{s=0}^{N-1}\sum_{t=s+1}^{N-1}\mathbb{E}\left[(\mathbb{E}[f(X)] - f(X^s))(\mathbb{E}[f(X)] - f(X^t))\right] \\
= \;& \frac{2}{N^2}\sum_{s=0}^{N-1}\sum_{t=s+1}^{N-1}\mathbb{E}\left[(\mathbb{E}[f(X)] - f(X^0))(\mathbb{E}[f(X)] - f(X^{t-s}))\right] \\
= \;& \frac{2}{N^2}\sum_{t=1}^{N-1}(N-t)\mathbb{E}\left[(\mathbb{E}[f(X)] - f(X^0))(\mathbb{E}[f(X)] - f(X^t))\right].
\end{aligned}$$

Fix $t \in \mathbb{N}$, write $\mathbb{E}[f(X)] = M$ and consider the covariance term

$$\mathbb{E}\left[(M - f(X^0))(M - f(X^t))\right].$$

We know that $X^t \sim P^t(X^0, \cdot)$, and therefore this expression becomes

$$\begin{aligned} & \sum_{X,Y \in \mathbb{X}} \mathbb{P}[X^0 = X \text{ and } X^t = Y](M - f(X))(M - f(Y)) \\ = & \sum_{X,Y \in \mathbb{X}} \mu(X) P^t(X,Y)(M - f(X))(M - f(Y)) \\ = & (P^t(M-f)(\cdot), (M-f)(\cdot))_\mu \,. \end{aligned}$$

We can decompose the function $(M - f)(\cdot)$ with the Eigenbasis $v_1, \ldots, v_k$

$$(M-f)(\cdot) = \sum_{i=1}^{k} \frac{((M-f)(\cdot), v_i)_\mu}{(v_i, v_i)_\mu} v_i$$

and see that the first term of this sum is zero:

$$((M-f)(\cdot), v_1)_\mu = \sum_{X \in \mathbb{X}} (M - f(X))\mu(X) = M - M = 0\,.$$

Using the orthogonality of the Eigenvectors gives

$$\begin{aligned} (P^t(M-f)(\cdot), (M-f)(\cdot))_\mu &= \sum_{i=2}^{k} \lambda_i^t \frac{((M-f)(\cdot), v_i)_\mu^2}{(v_i, v_i)_\mu} \\ &\leq |\lambda_2|^t \sum_{i=2}^{k} \frac{((M-f)(\cdot), v_i)_\mu^2}{(v_i, v_i)_\mu} \\ &= |\lambda_2|^t ((M-f)(\cdot), (M-f)(\cdot))_\mu \\ &= |\lambda_2|^t \mathrm{Var}[f]\,. \end{aligned}$$

Plugging in this result, we get

$$\begin{aligned} \mathbb{E}\left[\left(\mathbb{E}[f(X)] - \frac{1}{N}\sum_{t=0}^{N-1} f(X^t)\right)^2\right] &\leq \frac{1}{N}\mathrm{Var}[f] + \frac{2}{N^2}\mathrm{Var}[f] \sum_{t=1}^{N-1} (N-t)|\lambda_2|^t \\ &= \frac{1}{N}\mathrm{Var}[f]\left(1 + 2\sum_{t=1}^{N-1}\left(1 - \frac{t}{N}\right)|\lambda_2|^t\right) \\ &\leq \frac{1}{N}\mathrm{Var}[f]\left(1 + 2\sum_{t=1}^{N-1}|\lambda_2|^t\right) \\ &\leq \frac{1}{N}\mathrm{Var}[f]\left(1 + \frac{2}{1-|\lambda_2|}\right)\,. \end{aligned}$$

□

The quantity $(1 - |\lambda_2|)^{-1}$ is called *spectral gap* of P. Note that in the above calculation, the case $|\lambda_2| = 1$ is excluded. The variance of this estimator might not decrease at all in this case (take for instance P as the identity matrix).
For a fixed $s \in \mathbb{N}$, define

$$d(P^s, \mu) = \sup_{X \in \mathbb{X}} \|P^s(X, \cdot) - \mu\|_{TV}$$

as the worst case total variation after s transition steps with respect to the initial value X. This quantity is decreasing in s provided that $|\lambda_2| < 1$. To see this, consider a probability measure ν on $\mathbb{X}$ and the probability measure

$$\nu P(X) = \sum_{Y \in \mathbb{X}} \nu(Y) P(X, Y) \,.$$

We know that $\mu P = \mu$ due to the reversibility property, therefore μ is the left Eigenvector of P with Eigenvalue $\lambda_1 = 1$ (unique up to scaling). The map $\nu \mapsto \nu P$ is therefore a contraction on the set of probability measures with unique fixpoint. So far, Theorems 2.1, 2.2 and 2.3 treat situations which are idealized in a certain way. The next theorem addresses the commonly used estimator.

Theorem 2.4 (Markov chain Monte-Carlo integration)**.** *Let $X \sim \mu$, $X^0 \in X$ and $X^t \sim P(X^{t-1}, \cdot)$ for $t \in \mathbb{N}$, where P is a probability matrix on $\mathbb{X}$ that is reversible with respect to μ. Assume that $|\lambda_2| < 1$, and let $N = ns$ with $n, s \in \mathbb{N}$ such that $d(P^s, \mu) \leq \frac{1}{6}$. Then we have*

$$\begin{aligned} & \sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \frac{1}{N}\sum_{t=s}^{N+s-1} f(X^t)\right)^2\right]} \\ \leq \quad & \sqrt{\frac{1}{N}\mathrm{Var}[f]\left(1 + \frac{2}{1-|\lambda_2|}\right)} + \frac{\sqrt{2d(P^s,\mu)}}{n} \sup_{W,Z \in \mathbb{X}} |f(W) - f(Z)| \,. \end{aligned}$$

Proof. For $k \in \{s, 2s, 3s, \ldots, ns\}$, let Y^k be a random variable coupled to X^k in the following way: For $k = s$, let Y^s have law μ and couple (X^s, Y^s) such that

$$\mathbb{P}[X^s \neq Y^s] \leq d(P^s, \mu) \,.$$

For $k > s$, consider two cases. If $X^{k-s} = Y^{k-s}$, simply couple $X^k = Y^k$. If not, assign to Y^k the law $P(Y^{k-s}, \cdot)$ and couple (X^k, Y^k) such that

$$\mathbb{P}[X^k \neq Y^k] \leq d(P^s, \mu) \,.$$

We proceed to fill in the gaps for a given $k \in \{s, 2s, 3s, \ldots, ns\}$. Define random variables $Y^{k+1}, \ldots, Y^{k+s-1}$ which are coupled to $X^{k+1}, \ldots, X^{k+s-1}$ as follows: If $X^k = Y^k$, set $Y^{k+t} = X^{k+t}$ for $t \in \{1, \ldots, s-1\}$. If not, we consider $(Y^{k+1}, \ldots, Y^{k+s-1})$ as a possible outcome of the Markov chain with transition matrix P conditioned on the fact that it starts in Y^k and ends in Y^{k+s}. The set of random variables $(Y^t)_{t \geq s}$ forms a Markov chain with transition matrix P and initial law $Y^s \sim \mu$. Therefore we can use the triangle inequality together with Theorem

2.3:

$$\begin{aligned}
&\sqrt{\mathbb{E}\left[\left(\mathbb{E}[f(X)] - \frac{1}{N}\sum_{t=s}^{N+s-1} f(X^t)\right)^2\right]} \\
\leq\ &\sqrt{\frac{1}{N}\mathrm{Var}[f]\left(1+\frac{2}{1-|\lambda_2|}\right)} + \sqrt{\mathbb{E}\left[\left(\frac{1}{N}\sum_{t=s}^{N+s-1}(f(X^t)-f(Y^t))\right)^2\right]}.
\end{aligned}$$

The chain $(Y^t)_{t\geq s}$ moreover satisfies

$$X^k = Y^k \Rightarrow X^{k+t} = Y^{k+t} \text{ for } t \geq 0\,,$$

as well as

$$\mathbb{P}[X^k \neq Y^k \mid X^{k-s} \neq Y^{k-s}] \leq d(P^s, \mu)$$

for $k \in \{2s, 3s, \ldots, ns\}$. Therefore, we have that

$$\mathbb{P}[X^k \neq Y^k] \leq d(P^s,\mu)^k \text{ for } k \in \{s, 2s, 3s, \ldots, ns\}\,.$$

Now we can estimate

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{N}\sum_{t=s}^{N+s-1}(f(X^t)-f(Y^t))\right)^2\right] \\
=\ &\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\sum_{t=0}^{s-1}(f(X^{is+t})-f(Y^{is+t}))\right)^2\right] \\
\leq\ &\frac{1}{N^2}\sum_{i=1}^{n}\mathbb{P}[X^{(n+1-i)s} \neq Y^{(n+1-i)s}]((n+1-i)s)^2 \sup_{W,Z\in\mathbb{X}}|f(W)-f(Z)|^2 \\
\leq\ &\frac{1}{N^2}\sum_{i=1}^{n} d(P^s,\mu)^i (is)^2 \sup_{W,Z\in\mathbb{X}}|f(W)-f(Z)|^2 \\
=\ &\frac{s^2 \sup_{W,Z\in\mathbb{X}}|f(W)-f(Z)|^2}{N^2}\sum_{i=1}^{n} d(P^s,\mu)^i i^2 \\
\leq\ &\frac{\sup_{W,Z\in\mathbb{X}}|f(W)-f(Z)|^2}{n^2} 2d(P^s,\mu)
\end{aligned}$$

for $d(P^s, \mu)$ small enough. Combining this with the above estimate we get the result. □

We see that the initial bias of the estimator depends on $d(P^s,\mu)^{1/2}$ and decays with $\frac{1}{n}$. To fully understand this bound, we still need to clarify how $d(P^s,\mu)$ decays with respect to s.

2.4 Mixing time

To see this, we first show an alternative definition of the total variation. For two probability measures ν, μ on $\mathbb{X}$, it is easy to see that the event $\mathcal{A} = \{X \colon \nu(X) > \mu(X)\}$ maximizes the expression

$$\nu[\mathcal{A}] - \mu[\mathcal{A}]\,,$$

and this can also be written as

$$\nu[\mathcal{A}] - \mu[\mathcal{A}] = \sum_{X \in \mathcal{A}} (\nu(X) - \mu(X)) = \sum_{X \in \mathcal{A}} |\nu(X) - \mu(X)|\,.$$

The dual event $\mathcal{B} = \{X \colon \nu(X) \leq \mu(X)\}$ then satisfies

$$\nu[\mathcal{B}] - \mu[\mathcal{B}] = -\sum_{X \in \mathcal{B}} |\nu(X) - \mu(X)| = -\sum_{X \in \mathcal{A}} |\nu(X) - \mu(X)|$$

due to the fact that $\nu[\mathcal{A}] + \nu[\mathcal{B}] - \mu[\mathcal{A}] - \mu[\mathcal{B}] = 0$. This implies that

$$\nu[\mathcal{A}] - \mu[\mathcal{A}] = \frac{1}{2} \sum_{X \in \mathbb{X}} |\nu(X) - \mu(X)| = \frac{1}{2} \|\nu - \mu\|_1$$

and therefore

$$\|\nu - \mu\|_{TV} = \frac{1}{2} \|\nu - \mu\|_1\,.$$

For $\nu = \delta_Z$ we can calculate

$$\begin{aligned}
\|P^s(Z,\cdot) - \mu\|_1 &= \|\nu P^s - \mu\|_1 = \|(\nu - \mu)P^s\|_1 \\
&= \sum_{X \in \mathbb{X}} \left| \sum_{Y \in \mathbb{X}} (\nu(Y) - \mu(Y)) P^s(Y,X) \right| \\
&= \sum_{X \in \mathbb{X}} \left| \sum_{Y \in \mathbb{X}} \left(\frac{\nu(Y)}{\mu(Y)} - 1 \right) \mu(Y) P^s(Y,X) \right| \\
&= \sum_{X \in \mathbb{X}} \left| \sum_{Y \in \mathbb{X}} \left(\frac{\nu(Y)}{\mu(Y)} - 1 \right) \mu(X) P^s(X,Y) \right| \\
&= \sum_{X \in \mathbb{X}} \mu(X) \left| \sum_{Y \in \mathbb{X}} \left(\frac{\nu(Y)}{\mu(Y)} - 1 \right) P^s(X,Y) \right| \\
&= \sum_{X \in \mathbb{X}} \mu(X) \left| \sum_{Y \in \mathbb{X}} \frac{\nu(Y)}{\mu(Y)} P^s(X,Y) - 1 \right| \\
&= \sum_{X \in \mathbb{X}} \mu(X) \left| \frac{P^s(X,Z)}{\mu(Z)} - 1 \right|,
\end{aligned}$$

leaving us with the identity

$$d(P^s, \mu) = \frac{1}{2} \sup_{Z \in \mathbb{X}} \left\{ \sum_{X \in \mathbb{X}} \mu(X) \left| \frac{P^s(X,Z)}{\mu(Z)} - 1 \right| \right\}.$$

Consider the Eigenvalue decomposition of the indicator function

$$\delta_Z = \sum_{i=1}^{k} \frac{(\delta_Z, v_i)_\mu}{(v_i, v_i)_\mu} v_i = \sum_{i=1}^{k} \frac{v_i(Z)\mu(Z)}{(v_i, v_i)_\mu} v_i \,.$$

Using this representation, we can further calculate

$$\begin{aligned} \left| \frac{P^s(X,Z)}{\mu(Z)} - 1 \right| &= \left| \frac{[P^s \delta_Z](X)}{\mu(Z)} - 1 \right| \\ &= \left| \frac{\sum_{i=1}^{k} \frac{v_i(Z)\mu(Z)}{(v_i,v_i)_\mu} [P^s v_i](X)}{\mu(Z)} - 1 \right| \\ &= \left| \sum_{i=1}^{k} \frac{v_i(Z)}{(v_i, v_i)_\mu} \lambda_i^s v_i(X) - 1 \right| \end{aligned}$$

and using the fact that the first Eigenvector is a constant vector, we get

$$\left| \frac{P^s(X,Z)}{\mu(Z)} - 1 \right| = \left| \sum_{i=2}^{k} \frac{v_i(X)v_i(Z)}{(v_i, v_i)_\mu} \lambda_i^s \right| .$$

From this, it is easy to see that

$$d(P^s, \mu) \leq C|\lambda_2|^s$$

for some constant C depending on μ and $\mathbb{X}$, which means that we have exponential convergence of $d(P^s, \mu) \to 0$ in the asymptotical sense. Precisely, if we assume that the Eigenfunctions are normalized, we get

$$\begin{aligned} & \left| \sum_{i=2}^{k} v_i(X) v_i(Z) \lambda_i^s \right| \\ \leq \quad & |\lambda_2|^s \sum_{i=2}^{k} |v_i(X)||v_i(Z)| \\ \leq \quad & |\lambda_2|^s \left(\sum_{i=2}^{k} v_i(X)^2 \right)^{1/2} \left(\sum_{i=2}^{k} v_i(Z)^2 \right)^{1/2} . \end{aligned}$$

Moreover, we can consider

$$\frac{1}{\mu(X)} = \frac{(\delta_X, \delta_X)_\mu}{\mu(X)^2} = \frac{\left(\sum_{i=1}^{k} \mu(X) v_i(X) v_i, \sum_{i=1}^{k} \mu(X) v_i(X) v_i \right)_\mu}{\mu(X)^2} = \sum_{i=1}^{k} v_i(X)^2$$

and therefore we finally arrive at

$$d(P^s, \mu) \leq \frac{1}{2} |\lambda_2|^s \sup_{Z \in \mathbb{X}} \sum_{X \in \mathbb{X}} \frac{\mu(X)}{\mu(X)^{1/2} \mu(Z)^{1/2}} \leq \frac{1}{2} |\lambda_2|^s \sup_{Z \in \mathbb{X}} \left\{ \mu(Z)^{-1} \right\} .$$

Often, this bound is not helpful because the constant is really large, but it is easy to see that $d(P^s, \mu) \leq 1$ for all $s \in \mathbb{N}$. Instead, we define the *mixing time*

$$\tau_{mix}(P, \epsilon) = \min \left\{ t \colon d(P^t, \mu) \leq \epsilon \right\}$$

and we will try to find good bounds for $\tau_{mix}(P, \epsilon)$. The estimate we have shown just before implies

Lemma 2.5. *Let P be a transition matrix that is reversible with respect to μ and satisfies $|\lambda_2| < 1$. Let*

$$\frac{1}{\mu_*} = \sup_{Z \in \mathbb{X}} \left\{ \mu(Z)^{-1} \right\} .$$

Then

$$\tau_{mix}(P, \epsilon) \leq \frac{|\log(2\epsilon\mu_*)|}{1 - |\lambda_2|} .$$

Proof. Consider

$$\frac{1}{2\mu_*} |\lambda_2|^s \leq \epsilon ,$$

which can be restated as

$$s \leq \frac{\log(2\epsilon\mu_*)}{\log(|\lambda_2|)} \leq \frac{|\log(2\epsilon\mu_*)|}{1 - |\lambda_2|} .$$

□

On the other hand, we can make a similar statement in the other direction.

Lemma 2.6. *Let P be a transition matrix that is reversible with respect to μ and satisfies $\frac{1}{2} \leq |\lambda_2| < 1$. Then*

$$\frac{1}{1 - |\lambda_2|} \leq \frac{2\tau_{mix}(P, \epsilon)}{|\log(2\epsilon)|} .$$

Proof. Consider $v_2 \colon \mathbb{X} \to \mathbb{R}$ to be the Eigenfunction of P with

$$Pv_2 = \lambda_2 v_2 .$$

Then we have $(v_1, v_2)_\mu = 0$ and

$$\begin{aligned} |\lambda_2^s v_2(Z)| &= |[P^s v_2](Z)| = \left| \sum_{X \in \mathbb{X}} P^s(Z, X) v_2(X) - \mu(X) v_2(X) \right| \\ &\leq \|P^s(Z, \cdot) - \mu\|_1 \|v_2\|_\infty \\ &\leq 2d(P^s, \mu) \|v_2\|_\infty . \end{aligned}$$

If we choose Z such that $|v_2(Z)| = \|v_2\|_\infty$ we get

$$|\lambda_2|^s \leq 2d(P^s, \mu)$$

which implies for the mixing time

$$\tau_{mix}(P, \epsilon) \log(|\lambda_2|) \leq \log(2\epsilon) .$$

Rearranging gives

$$\frac{\tau_{mix}(P,\epsilon)}{|\log(2\epsilon)|} \geq \frac{1}{|\log(|\lambda_2|)|} \geq \frac{1}{2(1-|\lambda_2|)}\,.$$

□

Note that there is no constant involved in this estimate, making it very useful. All bounds for the mixing time can be directly transferred to bounds for the spectral gap in this way.

2.5 Markov chain couplings

Often, the mixing time is not directly accessible, however one can bound it if there exists an amenable coupling for P. We call a transition matrix Q which acts on pairs of states (X,Y) a coupling for P if

$$Q((X,Y),(\cdot,Z)) = P(X,\cdot) \text{ for all } Y,Z \in \mathbb{X}\,,$$
$$Q((X,Y),(W,\cdot)) = P(Y,\cdot) \text{ for all } X,W \in \mathbb{X}\,.$$

For a Markov chain $(X^t,Y^t)_{t\geq 0}$ with initial state (X^0,Y^0) that satisfies

$$(X^{t+1},Y^{t+1}) \sim Q((X^t,Y^t),\cdot)$$

for $t \geq 0$, one then has that $(X^t)_{t\geq 0}$ and $(Y^t)_{t\geq 0}$ are Markov chains with transition matrix P. It is clear that this procedure defines a coupling for the probability measures $\delta_{X^0}P^t$ and $\delta_{Y^0}P^t$, which leads to

$$\mathbb{P}[X^t \neq Y^t] \geq \|\delta_{X^0}P^t - \delta_{Y^0}P^t\|_{TV}\,.$$

Moreover, one has

$$\begin{aligned}\|\delta_{X^0}P^t - \mu\|_{TV} &= \left\|\sum_{Y\in\mathbb{X}} \mu(Y)(\delta_{X_0}-\delta_Y)P^t\right\|_{TV} \\ &\leq \sum_{Y\in\mathbb{X}} \mu(Y)\|\delta_{X^0}P^t - \delta_Y P^t\|_{TV} \\ &\leq \sup_{Y^0\in\mathbb{X}} \|\delta_{X^0}P^t - \delta_{Y^0}P^t\|_{TV}\end{aligned}$$

which implies

$$d(P^t,\mu) \leq \sup_{X^0,Y^0\in\mathbb{X}} \mathbb{P}[X^t \neq Y^t]\,.$$

We can define the worst-case *coupling time* for Q as

$$\tau_{coup}(Q,\epsilon) = \inf\left\{t\colon \mathbb{P}[X^t \neq Y^t] \leq \epsilon \text{ for all } X^0,Y^0 \in \mathbb{X}\right\}\,,$$

which might be infinite, we however have the bound

$$\tau_{mix}(P,\epsilon) \leq \tau_{coup}(Q,\epsilon)$$

for all couplings Q of P. This means that to obtain a mixing time bound for P, it suffices to find a coupling time bound for some coupling Q of P which might be easier for certain Markov chains. Using the coupling idea, we can prove the following scaling bound for the mixing time.

Lemma 2.7. *Let $\tau = \tau_{mix} = \tau_{mix}(P, \epsilon)$. Then it holds $n\tau_{mix} \leq \tau_{mix}(P, \epsilon^n)$.*

Proof. Let X^0 be an arbitrary initial state, and let $Y^\tau \sim \mu$. We know that for any X^0 one has

$$\|\mu - P^\tau(X^0, \cdot)\|_{TV} \leq \epsilon\,,$$

so we can find a coupling (X^τ, Y^τ) with

$$\mathbb{P}[X^\tau \neq Y^\tau] \leq \epsilon\,.$$

Now, we can define couplings doing the following for $n \in \mathbb{N}$: If $X^{n\tau} = Y^{n\tau}$, apply P^τ to both such that $X^{(n+1)\tau} = Y^{(n+1)\tau}$. If not, draw $Y^{(n+1)\tau}$ according to μ and couple it to $X^{(n+1)\tau}$ like in the initial step (applying P^τ to $X^{n\tau}$). This coupling satisfies

$$\mathbb{P}[X^{n\tau} \neq Y^{n\tau}] \leq \epsilon^n\,,$$

and the existence of this coupling between μ and $P^{n\tau}(X^0, \cdot)$ gives

$$\|\mu - P^{n\tau}(X^0, \cdot)\|_{TV} \leq \epsilon^n$$

for all X^0. □

Couplings can also be used to define unbiased estimators. For this, we restrict ourselves to couplings Q of P which do not diverge after meeting once, i.e.

$$(Y, Z) \sim Q((X, X), \cdot) \Rightarrow Y = Z\,.$$

Any coupling Q can be modified to satisfy this property by considering

$$Q^*((X, Y), (W, Z)) = \begin{cases} Q((X, Y), (W, Z)) & X \neq Y \\ \mathbb{1}[W = Z]P(X, W) & X = Y \end{cases}.$$

This way, coupled Markov chains $(X^t, Y^t)_{t\geq 0}$ satisfy

$$X^s = Y^s \Rightarrow X^t = Y^t \text{ for all } t \geq s\,,$$

and $\mathbb{P}[X^t \neq Y^t]$ is monotonically decreasing in t. The following idea is due to [JOA19]. Let μ' be a law on $\mathbb{X}$ which is easy to simulate, and let $X^0 \sim \mu'$ and $Y^{-1} \sim \mu'$, as well as $Y^0 \sim P(Y^{-1}, \cdot)$. From here, we evolve the Markov chain $(X^t, Y^t)_{t\geq 0}$ according to Q until $X^t = Y^t$ happens. If $T = \tau_{coup}(Q, \epsilon)$ is finite, one has that $\mathbb{P}[X^{nT} \neq Y^{nT}] \leq \epsilon^n$, which means that the meeting time

$$t^* = \inf\{t > 0 \colon X^t = Y^t\}$$

is finite with probability 1. The estimator

$$f(Y^{-1}) + \sum_{i=0}^{t^*-1} [f(Y^t) - f(X^t)]$$

is unbiased under certain regularity conditions. For $Y \sim \mu$, consider the heuristic calculation

$$\begin{aligned}
\mathbb{E}[f(Y)] &= \lim_{t\to\infty} \mathbb{E}[f(Y^t)] \\
&= \mathbb{E}[f(Y^{-1})] + \sum_{t=0}^{\infty} \big(\mathbb{E}[f(Y^t)] - \mathbb{E}[f(Y^{t-1})]\big) \\
&= \mathbb{E}[f(Y^{-1})] + \sum_{t=0}^{\infty} \big(\mathbb{E}[f(Y^t)] - \mathbb{E}[f(X^t)]\big) \\
&=^{(*)} \mathbb{E}\left[f(Y^{-1}) + \sum_{t=0}^{\infty}(f(Y^t) - f(X^t))\right] \\
&= \mathbb{E}\left[f(Y^{-1}) + \sum_{t=0}^{t^*-1}(f(Y^t) - f(X^t))\right].
\end{aligned}$$

The marked equation, where an infinite sum and a limit process are interchanged, has to be justified. Sufficient conditions and variants of the method can be found in [JOA19].

2.6 Grand couplings

We might even go one step further and consider couplings that attempt to join Markov chains that start in all states $X \in \mathbb{X}$ simultaneously. To this end, we introduce $G = \{g\colon \mathbb{X} \to \mathbb{X}\}$ the set of rules (which is finite in our case) and a discrete probability measure ρ on G such that

$$g \sim \rho \Rightarrow g(X) \sim P(X, \cdot)\,.$$

This means that sampling from $P(X, \cdot)$ is the same as sampling a rule g with law ρ and taking $g(X)$ as the sample. Using this formalism, we can define a coupling of P for an arbitrary number n of initial states $(X_1^0, \ldots, X_n^0)$ by

$$\forall t > 0\colon g_t \sim \rho, \quad (X_1^t, \ldots, X_n^t) = (g_t(X_1^{t-1}), \ldots, g_t(X_n^{t-1}))\,.$$

For this so-called grand coupling of P, one can again ask how fast it couples. Note that for a subset $\mathcal{A} \subseteq \mathbb{X}$, the set $g(\mathcal{A}) = \{g(X)\colon X \in \mathcal{A}\}$ satisfies

$$|\mathcal{A}| \geq |g(\mathcal{A})| \text{ for all } g \in G\,.$$

It therefore makes sense to define

$$\tau_{coup}(\rho, \epsilon) = \inf\{t\colon \mathbb{P}[|g_t \circ \ldots \circ g_1(\mathbb{X})| \neq 1] \leq \epsilon\}$$

and due to a similar argumentation as above one has

$$\tau_{mix}(P, \epsilon) \leq \tau_{coup}(\rho, \epsilon)$$

for grand couplings ρ of P.
We will now introduce a method called *coupling from the past* (CFTP) due to Propp and Wilson [PW96]. For $j \in \mathbb{Z}$, let $g_j \sim \rho$ and define

$$G_t = g_0 \circ g_{-1} \circ \ldots \circ g_{-t+1} \text{ for } t > 0\,.$$

Then $G_t(\mathbb{X})$ represents all possible outcomes of the Markov chain at time 0, started with arbitrary initial condition at time $-t$. This set is monotonically decreasing with t, i.e. $G_t(\mathbb{X}) \subseteq G_s(\mathbb{X})$ for $s \leq t$. If we assume that $T = \tau_{coup}(\rho, \epsilon)$ is finite, it is easy to see that

$$\mathbb{P}[|G_{nT}(\mathbb{X})| \neq 1] \leq \epsilon^n$$

and especially one has that

$$G_\infty(\mathbb{X}) = \{X \in \mathbb{X} \colon X \in G_t(\mathbb{X}) \text{ for all } t > 0\}$$

is a single element with probability 1, and we therefore treat it as a random variable taking values in $\mathbb{X}$.

Lemma 2.8. *Let ρ be a grand coupling of P such that $T = \tau_{coup}(\rho, \epsilon)$ is finite. Then the set $G_\infty(\mathbb{X})$ contains exactly one element with probability 1, and this element is distributed according to μ.*

Proof. Let X have law μ. Then for all $t > 0$, one has that $G_t(X)$ has law μ and is included in $G_t(\mathbb{X})$. Due to the fact that $\mathbb{P}[|G_{nT}(\mathbb{X})| \neq 1] \leq \epsilon^n$, we have that

$$\mathbb{P}[G_{nT}(X) \neq G_\infty(\mathbb{X})] \leq \epsilon^n\,,$$

and therefore

$$t^* = \inf\{t > 0 \colon |G_t(\mathbb{X})| = 1\}$$

is finite with probability 1. So it follows

$$G_{t^*}(X) = G_\infty(\mathbb{X})\,,$$

which implies that $G_\infty(\mathbb{X})$ is distributed according to μ. □

Using the *coupling from the past* procedure, we can simulate the random variable $G_\infty(\mathbb{X})$ which has law μ:

1. Set $t \leftarrow 1$.

2. Sample g_{-t+1} from ρ.

3. Compute the set $G_t(\mathbb{X})$.
   If it is a single element, it equals $G_\infty(\mathbb{X})$.
   If not, set $t \leftarrow t + 1$ and go to Step 2.

It is important to note that the obtained sample only has law μ if the above algorithm has no stopping criterion other than finishing at Step 3. If $T = \tau_{coup}(\rho, \epsilon)$ is finite, this algorithm will terminate with probability 1, and the law of the termination time t^* can be estimated using

$$\mathbb{P}[|G_{nT}(\mathbb{X})| \neq 1] \leq \epsilon^n \,.$$

Provided that the computation of the sets $G_t(\mathbb{X})$ is feasible, this procedure allows us to implement the estimator from Theorem 2.3, with an initial simulation of $X^0 = G_\infty(\mathbb{X})$.

Theorem 2.9 (CFTP + MCMC)**.** *Let P be a transition matrix that is reversible with respect to μ and has $|\lambda_2| < 1$. Furthermore, let ρ be a grand coupling of P with finite coupling time $\tau_{coup}(\rho, \epsilon)$. Then the estimator*

$$\frac{1}{N} \sum_{t=0}^{N-1} f(X^t)$$

is unbiased, where X^0 is obtained from the CFTP procedure and X^t is simulated according to the law $P(X^{t-1}, \cdot)$. Its variance is the same as in Theorem 2.3.

The 'cost' t^* to sample from μ directly which is roughly $\tau_{coup}(\rho, \epsilon)$ can be compared to $\tau_{mix}(P, \epsilon)$ which essentially dictates the decay of the bias in Theorem 2.4.

2.7 Path coupling

So far, we have seen results for sampling accuracy, based on mixing and coupling times. One technique to obtain results on the coupling time itself is called *path coupling* which was introduced in [BD97]. The idea is to view a coupling for a Markov chain as a contraction on the configuration space, according to some carefully chosen metric. A fixpoint argument will then imply that applying the coupling repeatedly will result in a finite coupling time. We will introduce this method for couplings Q of P, it however also applies to grand couplings.
Consider $\mathbb{X}$ to be the vertex set of a finite graph, and consider some set of unoriented edges $\mathbb{Y} \subseteq \{\{X, X'\} \colon X, X' \in \mathbb{X}\}$ such that the graph $(\mathbb{X}, \mathbb{Y})$ is connected. We call a finite sequence $(X_1, \ldots, X_n)$ a path in $\mathbb{X}$ if $\{X_i, X_{i+1}\} \in \mathbb{Y}$ for $i = 1, \ldots n-1$ (with this definition every element X defines a path with length $n = 1$). Let $d \colon \mathbb{Y} \to \mathbb{R}_+$ be a pre-metric, i.e. a map such that every edge is a shortest path:

$$d(\{X, Y\}) = \inf \left\{ \sum_{i=1}^{n-1} d(\{X_i, X_{i+1}\}) \,\middle|\, (X = X_1, \ldots, X_n = Y) \text{ is a path} \right\}$$

for all $\{X, Y\} \in \mathbb{Y}$. Such a pre-metric can be extended to a metric on $\mathbb{X}$ via

$$d(X, Y) = \inf \left\{ \sum_{i=1}^{n-1} d(\{X_i, X_{i+1}\}) \,\middle|\, (X = X_1, \ldots, X_n = Y) \text{ is a path} \right\}$$

that satisfies the usual properties

$$\begin{aligned} d(X,X) &= 0 \text{ for all } X \in \mathbb{X}\,, \\ d(X,Y) &= d(Y,X) \text{ for all } X,Y \in \mathbb{X}\,, \\ d(X,Z) &\leq d(X,Y) + d(Y,Z) \text{ for all } X,Y,Z \in \mathbb{X}\,, \end{aligned}$$

making $(\mathbb{X}, d)$ a metric space. The next theorem shows that a coupling Q of P which is contractive for the edges in $\mathbb{Y}$ is contractive for the whole set $\mathbb{X}$.

Theorem 2.10 (Path coupling, [BD97])**.** *Let $(\mathbb{X}, \mathbb{Y})$ be an unoriented, connected graph and let $d\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ be a metric which is induced by a pre-metric on $\mathbb{Y}$. Let Q be a coupling for P that is defined on adjacent states $\{X,Y\} \in \mathbb{Y}$, and denote with (X',Y') the random variables distributed according to $Q((X,Y),\cdot)$. Assume that $0 \leq \alpha < 1$ exists such that for all $\{X,Y\} \in \mathbb{Y}$ one has*

$$\mathbb{E}[d(X',Y') \mid X,Y] \leq \alpha\, d(X,Y)\,.$$

Then Q extends to a coupling for arbitrary $X,Y \in \mathbb{X}$ satisfying the above inequality.

Proof. Let $X,Y \in \mathbb{X}$, and consider a shortest path $(Z_1,\ldots,Z_n)$ with $X = Z_1$ and $Z = X_n$. We can use the coupling Q on pairs (Z_i, Z_{i+1}) to obtain the sequence $Z'_1,\ldots,Z'_n$ with $Z'_1 = X'$ and $Z'_n = Y'$, hereby extending the coupling to arbitrary states X,Y. Then it holds

$$\begin{aligned} & \mathbb{E}\left[d(X',Y') \mid X,Y\right] \\ = \;& \mathbb{E}\left[\inf\left\{\sum_{i=1}^{m-1} d(W_i,W_{i+1}) \mid (W_1,\ldots,W_m) \text{ is a path from } X' \text{ to } Y'\right\} \mid X,Y\right] \\ \leq \;& \mathbb{E}\left[\sum_{i=1}^{n-1} d(Z'_i,Z'_{i+1}) \mid X,Y\right] \\ = \;& \sum_{i=1}^{n-1} \mathbb{E}[d(Z'_i,Z'_{i+1}) \mid Z_i,Z_{i+1}] \\ \leq \;& \alpha \sum_{i=1}^{n-1} d(Z_i,Z_{i+1}) \\ = \;& \alpha\, d(X,Y)\,. \end{aligned}$$

□

For all cases that we will consider, the metric $d(X,Y)$ takes values in $\mathbb{N}_0$.

Corollary 2.11. *Let the prerequisites of Theorem 2.10 hold, and let $d(X,Y)$ be a metric that takes values in $\mathbb{N}_0$. Let $D \in \mathbb{N}$ be the maximal distance between two*

elements in $\mathbb{X}$*. Then for any* $X^0, Y^0 \in \mathbb{X}$*, the Markov chains* $(X^t)_{t\geq 0}$ *and* $(Y^t)_{t\geq 0}$ *induced by the coupling* Q *satisfy*

$$\mathbb{P}[X^t \neq Y^t] \leq \epsilon \text{ if } t \geq \frac{\log D |\log \epsilon|}{|\log \alpha|}\,.$$

In other words, one has

$$\tau_{coup}(Q, \epsilon) \leq \frac{\log D |\log \epsilon|}{|\log \alpha|}\,.$$

Proof. From Theorem 2.10 it follows immediately that

$$\mathbb{E}[d(X^t, Y^t)] \leq D\alpha^t\,.$$

Due to the fact that $d(\cdot,\cdot)$ only takes discrete values, we can choose t such that $D\alpha^t \leq \epsilon$ and it follows that

$$\mathbb{P}[d(X^t, Y^t) \neq 0] \leq \epsilon\,,$$

but this is the same as $\mathbb{P}[X^t \neq Y^t] \leq \epsilon$. Rearranging the condition for t gives the result. □

We see that from the contraction property, a coupling time bound immediately follows. The path coupling theorem is important because in many cases it is easier to prove the contraction property for adjacent elements in the graph than for arbitrary elements.

3 The Fortuin-Kasteleyn Random Cluster Model

The random cluster model introduced by Fortuin and Kasteleyn [FK72] has been of central interest in statistical mechanics for the last few decades. The outline given here loosely follows [DC17]. See also [Gri02] for an exhaustive treatment.

Let $G = (V, E)$ be an undirected, finite and connected graph. We define the configuration space $\mathbb{X} = \{A \subseteq E\}$ as the set of all subsets of E. For an edge configuration A, we say that an edge $e \in E$ is *open* if $e \in A$, otherwise it is *closed*, see Figure 3.1. Moreover, let $c(A)$ be the number of connected components of A, also called (open) *clusters*. The random cluster measure on $\mathbb{X}$ with edge parameter $p \in (0, 1)$ and cluster weight $q \in (0, \infty)$ is then given by

$$\mu(A) = \mu_{G,p,q}(A) = Z_{G,p,q}^{-1} \, p^{|A|} \, (1-p)^{|E \setminus A|} \, q^{c(A)} \,,$$

where the partition sum $Z = Z_{G,p,q}$ is chosen such that μ is a probability measure. This model has a few special cases:

- For $q = 1$ the model is just Bernoulli percolation and one has $Z_{G,p,1} = 1$.
- For integers $q \geq 2$, the model has a strong relation to the q-state Potts model.
- Another reformulation for integer $q \geq 2$ is given by a closed loop model.
- For $p \to 0$ and $q/p \to 0$, the model converges to the uniform measure on all spanning trees of G.

During the course of this thesis, we will focus our attention on the integer $q \geq 2$ case, with G tending to $\mathbb{Z}^2$ (which we will have to define rigorously). However, many important properties of the model can be stated for $q \in [1, \infty)$ on arbitrary finite graphs.

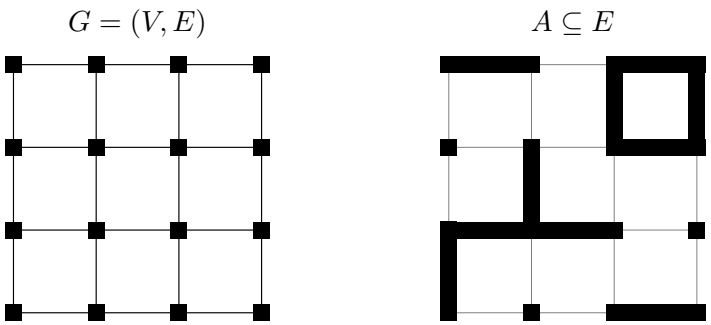

Figure 3.1: A 4×4 grid graph (left) and a corresponding edge configuration $A \subseteq E$ (right), drawn with bold edges and vertices to emphasize the cluster structure of A. In this case, $c(A) = 7$.

3.1 Monotonicity of the random cluster model

Let us start with the observation that the configuration space $\mathbb{X}$ exhibits a natural partial order given by inclusion:

$$A \leq B \Leftrightarrow A \subseteq B\,.$$

We call a subset $\mathcal{A} \subseteq \mathbb{X}$ an *increasing event* if for an element $A \in \mathcal{A}$ all supersets of A are also in $\mathcal{A}$. For $q \geq 1$, it is easy to see that the random cluster measure is monotone in p.

Lemma 3.1 (Monotonicity in p)**.** *Let $G = (V, E)$ be an undirected, finite and connected graph. Let $q \geq 1$ and $0 < p \leq p' < 1$. Then the measures $\mu = \mu_{G,p,q}$ and $\mu' = \mu_{G,p',q}$ satisfy*

$$\mu[\mathcal{A}] \leq \mu'[\mathcal{A}]$$

for all increasing events $\mathcal{A} \subseteq \mathbb{X}$.

Moreover, increasing events have a certain positivity property.

Lemma 3.2 (Fortuin-Kasteleyn-Ginibre inequality)**.** *Let $G = (V, E)$ be an undirected, finite and connected graph. Let $q \geq 1$ and $0 < p < 1$, and $\mu = \mu_{G,p,q}$ the random cluster measure. Then for two increasing events $\mathcal{A}$ and $\mathcal{B}$ one has*

$$\mu[\mathcal{A} \cap \mathcal{B}] \geq \mu[\mathcal{A}]\mu[\mathcal{B}]\,.$$

Note that the inequality reverses its sign if one of the events is decreasing, i.e. its complement is increasing. Last but not least, the random cluster measure is monotone with respect to boundary conditions.

Lemma 3.3 (Monotonicity with respect to boundary conditions)**.** *Let $G = (V, E)$ be an undirected, finite and connected graph, and $F \subseteq E$ a subset of edges. Let $C \subseteq F$ be a boundary condition for the measure*

$$\mu^{F,C}(A) = \mu_{G,p,q}(A \mid A \cap F = C)\,.$$

Then for any increasing event $\mathcal{A}$ and $C \subseteq C' \subseteq F$ one has

$$\mu^{F,C}[\mathcal{A}] \leq \mu^{F,C'}[\mathcal{A}]\,.$$

The statements are all about *stochastic domination*. For two probability distributions μ and μ', we write $\mu \leq \mu'$ if

$$\mu[\mathcal{A}] \leq \mu'[\mathcal{A}]$$

for all increasing events $\mathcal{A} \subseteq \mathbb{X}$. This relation defines a partial order on the set of probability distributions on $\mathbb{X}$.
To prove these three lemmata, we can use Markov chain techniques. One of the simplest Markov chains for the random cluster model are the so-called *Glauber dynamics*. In the literature, it is also sometimes called *heathbath process*, however we wish to reserve this term for the spin system setting. Given a state A as starting point, one Glauber dynamics step to obtain B can be described as follows:

1. Choose an edge $e \in E$ uniformly at random.
2. Choose $B = A \setminus \{e\}$ or $B = A \cup \{e\}$, according to the measure

$$\mu_{G,p,q}(B | B \setminus \{e\} = A \setminus \{e\})\,.$$

We will check the reversibility of this Markov chain with respect to $\mu_{G,p,q}$ in Chapter 4. The transition probability at Step 2 depends on the cluster structures of $A \setminus \{e\}$ and $A \cup \{e\}$. We call $e \in E$ *pivotal* to A if $c(A \setminus \{e\}) > c(A \cup \{e\})$.

Proof of Lemma 3.1. Let $X^0 = \emptyset$ and $Y^0 = E$ be the starting points of the Markov chains $(X^t)_{t\geq 0}$ and $(Y^t)_{t\geq 0}$, given by Glauber dynamics with parameter p and p'. Furthermore, we couple the chains in the following way:

- At time t, the edge $e \in E$ chosen in Step 1 is the same for both dynamics.
- The choice $X^{t+1} = X^t \setminus \{e\}$ or $X^{t+1} = X^t \cup \{e\}$ is determined by a parameter $r \in [0, 1]$ drawn uniformly at random. If

$$r < \mu_{G,p,q}[X^{t+1} = X^t \setminus \{e\} | (X^t \setminus \{e\}) \subseteq X^{t+1}]\,,$$

  we set $X^{t+1} = X^t \setminus \{e\}$, otherwise $X^{t+1} = X^t \cup \{e\}$.
- Y^{t+1} is determined analogously with respect to the measure $\mu_{G,p',q}$ and the *same* parameter r.

Using this coupling, one readily verifies that $X^t \subseteq Y^t$ for all $t \geq 0$: Let us assume that $X^{t-1} \subseteq Y^{t-1}$. Then, if e is pivotal to Y^{t-1}, it is also pivotal to X^{t-1}. So if an edge e pivotal to Y^{t-1} is chosen in Step 1 of the Glauber dynamics, we can calculate

$$\mathbb{P}[e \notin Y^t \mid Y^{t-1}] = \frac{(1-p')q}{p' + (1-p')q} \leq \frac{(1-p)q}{p + (1-p)q} = \mathbb{P}[e \notin X^t \mid X^{t-1}]\,.$$

This implies $X^t \subseteq Y^t$, because the converse can only happen if $e \notin Y^t$, but this implies $r < \mathbb{P}[e \notin Y^t \mid Y^{t-1}] \leq \mathbb{P}[e \notin X^t \mid X^{t-1}]$ and therefore $e \notin X^t$. If the edge e is not pivotal to Y^{t-1}, then

$$\mathbb{P}[e \notin Y^t \mid Y^{t-1}] = 1 - p' \leq 1 - p \leq \mathbb{P}[e \notin X^t \mid X^{t-1}]$$

where the last inequality is true whether e is pivotal to X^{t-1} or not. So again, we get $X^t \subseteq Y^t$.
Now, let $\mathcal{A}$ be an increasing event. Then we have $\mathbb{P}[X^t \in \mathcal{A}] \longrightarrow \mu_{G,p,q}[\mathcal{A}]$ and $\mathbb{P}[Y^t \in \mathcal{A}] \longrightarrow \mu_{G,p',q}[\mathcal{A}]$ because $(X^t)_{t\geq 0}$ and $(Y^t)_{t\geq 0}$ tend to their respective stationary distributions. Moreover, we know that $\mathbb{P}[X^t \in \mathcal{A}] \leq \mathbb{P}[Y^t \in \mathcal{A}]$ because $X^t \subseteq Y^t$ for all $t \geq 0$. Therefore, in the limit we get $\mu_{G,p,q}[\mathcal{A}] \leq \mu_{G,p',q}[\mathcal{A}]$. □

We can prove the other two lemmata similarly.

Proof of Lemma 3.2. We reformulate the statement as

$$\mu[\mathcal{B}] \leq \mu[\mathcal{B} \mid \mathcal{A}]$$

and construct a coupling using Glauber dynamics. Let $X^0 = Y^0 \in \mathcal{A}$ and define $(X^t)_{t\geq 0}$ to be the usual Glauber dynamics with respect to μ. The chain $(Y^t)_{t\geq 0}$ shall be given by Glauber dynamics with respect to $\mu(\cdot \mid \mathcal{A})$, i.e. at Step 2 of the usual dynamics the chain cannot leave the increasing set $\mathcal{A}$. The set $\mathcal{A}$ is 'simply connected' by Glauber dynamics steps, as all states can be reached from the maximal element $E \in \mathcal{A}$. We can couple $(X^t)_{t\geq 0}$ and $(Y^t)_{t\geq 0}$ in the same way as we did in the proof of Lemma 3.1. Using this coupling, we again see that $X^t \subseteq Y^t$ for all $t \geq 0$, because $(Y^t)_{t\geq 0}$ behaves the same as $(X^t)_{t\geq 0}$, only with an additional constraint to not remove an edge sometimes. We have that $\mathbb{P}[X^t \in \mathcal{B}] \longrightarrow \mu[\mathcal{B}]$ and $\mathbb{P}[Y^t \in \mathcal{B}] \longrightarrow \mu[\mathcal{B} \mid \mathcal{A}]$, as well as $\mathbb{P}[X^t \in \mathcal{B}] \leq \mathbb{P}[Y^t \in \mathcal{B}]$ for all $t \geq 0$. So in the limit, we obtain $\mu[\mathcal{B}] \leq \mu[\mathcal{B} \mid \mathcal{A}]$. □

Proof of Lemma 3.3. Let $X^0 = C$ and $Y^0 = C'$ and define $(X^t)_{t\geq 0}$ and $(Y^t)_{t\geq 0}$ via coupled Glauber dynamics with respect to the measures $\mu^{F,C}$ and $\mu^{F,C'}$. This boils down to the variant where only edges $e \in E\backslash F$ are chosen at Step 1. Then as above it can easily be verified that $X^t \subseteq Y^t$ for all $t \geq 0$. We have that $\mathbb{P}[X^t \in \mathcal{A}] \longrightarrow \mu^{F,C}[\mathcal{A}]$ and $\mathbb{P}[Y^t \in \mathcal{A}] \longrightarrow \mu^{F,C'}[\mathcal{A}]$, as well as $\mathbb{P}[X^t \in \mathcal{A}] \leq \mathbb{P}[Y^t \in \mathcal{A}]$ for all $t \geq 0$. So in the limit, we obtain $\mu^{F,C}[\mathcal{A}] \leq \mu^{F,C'}[\mathcal{A}]$. □

Lemma 3.1, 3.2 and 3.3 all use the (grand) monotone coupling for Glauber dynamics to obtain their respective statements in the limit $t \to \infty$. Using the coupling from the past (CFTP) procedure (see Chapter 2), we can derive even stronger results.

Theorem 3.4 (Monotone coupling realization)**.** *Let $p \leq p'$. Let $F \subseteq E$ be a boundary set and $C \subseteq D \subseteq F$ two boundary conditions. Let $\mathcal{A}$ be an increasing event such that $\{B\colon B \cap F = D, B \in \mathcal{A}\}$ is nonempty. Let $\mu = \mu^{F,C}_{G,p,q}$ and $\mu' = \mu^{F,D}_{G,p',q}(\cdot \mid \mathcal{A})$. Then we have*

$$\mu \leq \mu' .$$

Moreover, there exists a coupling for random variables (A, B) such that

$$A \sim \mu , B \sim \mu' , \mathbb{P}[A \leq B] = 1 .$$

Proof. Let

$$\mathbb{Y} = \{X \in \mathbb{X}\colon X \cap F = C\},\ \mathbb{Y}' = \{X \in \mathbb{X}\colon X \cap F = D , X \in \mathcal{A}\}$$

be the sets we define our Glauber dynamics on. We set up a grand coupling step for the Glauber dynamics in the usual way, i.e. the edge e chosen in Step 1 (uniformly from the set $E \setminus F$) is the same for all possible states the Markov chain is in. The result of the choice in Step 2 depends on a random variable r which is uniformly distributed on $[0, 1]$, also independent of the state of the chain. This procedure defines a map $(e, r) \to g$, where $g\colon \mathbb{Y} \to \mathbb{Y}$ is the rule described above for parameter p and C as boundary condition. Given parameter p', D as boundary condition, and the additional constraint that the Markov chain cannot leave $\mathcal{A}$, we call the obtained rule $h\colon \mathbb{Y}' \to \mathbb{Y}'$. From the previous lemmas we know that for $A \leq B$, $A \in \mathbb{Y}$ and $B \in \mathbb{Y}'$ one has $g(A) \leq h(B)$. Let us consider the CFTP procedure for both processes, where we couple g_{-t} and h_{-t} as described. Let $G^\infty(\mathbb{Y})$ and $H^\infty(\mathbb{Y}')$ be the corresponding results obtained from the CFTP procedures. The pair $(G^\infty(\mathbb{Y}), H^\infty(\mathbb{Y}'))$ is the coupling we are looking for: $G^\infty(\mathbb{Y})$ is distributed according to μ and $H^\infty(\mathbb{Y}')$ is distributed according to μ'. It remains to show that the coupling for the rules g_{-t} and h_{-t} implies $G^\infty(\mathbb{Y}) \leq H^\infty(\mathbb{Y}')$.
To see this, consider the set

$$G^t(\mathbb{Y}') = g_0 \circ \dots g_{-t+1}(\mathbb{Y}) .$$

We know that $G^\infty(\mathbb{Y}) \in G^t(\mathbb{Y})$ for any t, but from the definition it also follows that $G^t(C)$ is the lower bound for $G^t(\mathbb{Y})$. In the same way, we have $H^t(D \cup (E \setminus F)) \geq Z$ for all $Z \in H^t(\mathbb{Y}')$. Moreover, due to the coupling for the rules g and h we have that $G^t(C) \leq H^t(D \cup (E \setminus F))$. For $t \to \infty$, we know that $G^t(C) \to G^\infty(\mathbb{Y})$ and $H^t(D \cup E \setminus F) \to G^\infty(\mathbb{Y}')$. So in the limit, one has $G^\infty(\mathbb{Y}) \leq H^\infty(\mathbb{Y}')$. The existence of this coupling immediately implies $\mu \leq \mu'$. □

It is worth noting that this procedure to obtain the coupling is constructive.

3.2 The q-state Potts model

As we have seen, Glauber dynamics are a powerful tool to analyze the random cluster model. The next Markov chain will make the connection to the q-state Potts model

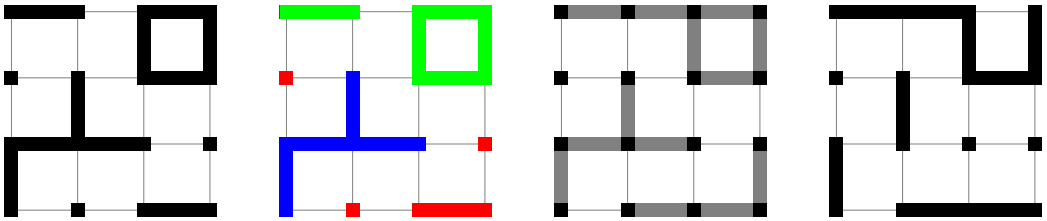

Figure 3.2: A visualization of the Swendsen-Wang Markov chain step. For an initial edge configuration (far left), assign to all clusters a spin (left, here visualized with $q = 3$ colors). All edges with aligning spin at the endpoints (right) are eligible for the resulting edge configuration (far right).

more evident. From now on, we consider $q \in \mathbb{N}$. The following procedure is called Edwards-Sokal coupling or Swendsen-Wang algorithm [SW87] (see Figure 3.2):

1. For a given state A, assign to each open cluster C of A a *spin* $s \in \{1, \dots, q\}$, uniformly at random.
2. Assign to each vertex $v \in V$ the spin of the open cluster C it belongs to. This map $\sigma \colon V \to \{1, \dots, q\}$ we call *spin configuration*.
3. For a spin configuration s, obtain the new state B via the rule

$$\mathbb{P}[e \in B] = \begin{cases} p & \text{both endpoints of } e \text{ have the same spin in } \sigma, \\ 0 & \text{else,} \end{cases}$$

considered independently for each $e \in E$.

We will now show the following:

Lemma 3.5. *Let $G = (V, E)$ be an undirected, finite and connected graph, and let $q \in \mathbb{N}$ and $0 < p < 1$. The Edwards-Sokal coupling is ergodic and reversible with respect to the random cluster measure $\mu = \mu_{G,p,q}$.*

Proof. Let $A, B \subseteq E$ be two edge configurations, and denote with $\mathbb{P}[A \to B]$ the probability to go from A to B via an Edwars-Sokal coupling step. The chain is ergodic because $\mathbb{P}[A \to B]$ is nonzero; if at Step 1 all clusters get the same color, and at Step 3 exactly the edges in B are chosen, we obtain any B from any A.
For reversibility we need to show that

$$\mu(A)\mathbb{P}[A \to B] = \mu(B)\mathbb{P}[B \to A]\,.$$

We calculate

$$\mathbb{P}[A \to B] = \sum_{\sigma} \mathbb{P}[A \to \sigma]\mathbb{P}[\sigma \to B]$$

and consider these probabilities separately. Let $A(\sigma) \subseteq E$ be the set of edges in E whose endpoints have the same spin. Then we have

$$\mathbb{P}[A \to \sigma] = \mathbb{1}[A \subseteq A(\sigma)]q^{-c(A)}\,.$$

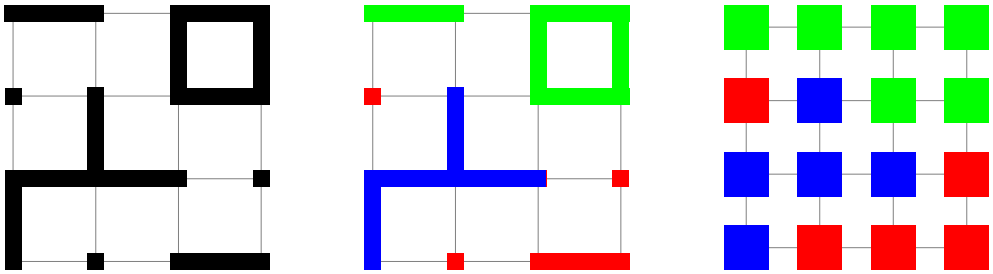

Figure 3.3: Obtaining a spin configuration ($q = 3$) from an edge configuration via steps 1 and 2 of the Edwards-Sokal coupling.

On the other hand,

$$\mathbb{P}[\sigma \to B] = \mathbb{1}[B \subseteq A(\sigma)]p^{|B|}(1-p)^{|A(\sigma)\setminus B|}\,,$$

so overall we get

$$\begin{aligned}
&\mu(A)\mathbb{P}[A \to B] \\
=&\,\mu(A)\sum_{\sigma}\mathbb{1}[A\cap B \subseteq A(\sigma)]p^{|B|}(1-p)^{|A(\sigma)\setminus B|}q^{-c(A)} \\
=&\,Z^{-1}\sum_{\sigma}(1-p)^{-|E\setminus A(\sigma)|}\mathbb{1}[A\cap B \subseteq A(\sigma)]p^{|B|}(1-p)^{|E\setminus B|}p^{|A|}(1-p)^{|E\setminus A|}\,.
\end{aligned}$$

This term is symmetric in A and B, so we get reversibility. □

This Markov chain is popular due to being easy to implement on one hand, as well as being a global algorithm on the other hand. It is believed (and this belief is heavily backed up by experiments) that it converges faster to the stationary distribution than local algorithms in many cases. However, it is also interesting in the theoretical context.

Lemma 3.6. *Let $A \sim \mu$. Then the measure $\nu(\sigma)$ which is obtained by applying Step 1 and 2 of the Edwards-Sokal coupling to A is the q-state Potts model distribution on spin configurations (see Figure 3.3)*

$$\nu(\sigma) = Z^{-1}(1-p)^{|E\setminus A(\sigma)|}\,.$$

Likewise, applying Step 3 to $\sigma \sim \nu$ results in an edge configuration $B \sim \mu$.

Note that we employ a different notation of the Potts model in contrast to the usual literature. It is more convenient for us to stay closer to the random cluster model.

Proof. We compute

$$\begin{aligned}\nu(\sigma) &= \sum_A \mu(A)\mathbb{P}[A \to \sigma] \\ &= \sum_A Z^{-1} p^{|A|}(1-p)^{|E\setminus A|} q^{c(A)} \mathbb{1}[A \subseteq A(\sigma)] q^{-c(A)} \\ &= Z^{-1} \sum_A p^{|A|}(1-p)^{|E\setminus A|} \mathbb{1}[A \subseteq A(\sigma)] \\ &= Z^{-1}(1-p)^{|E\setminus A(\sigma)|}\,.\end{aligned}$$

Applying Step 3 to $\sigma \sim \nu$ yields the original measure μ:

$$\begin{aligned}&\sum_\sigma \nu(\sigma)\mathbb{P}[\sigma \to B] \\ =&\sum_\sigma Z^{-1}(1-p)^{|E\setminus A(\sigma)|}\mathbb{1}[B \subseteq A(\sigma)]p^{|A|}(1-p)^{|A(\sigma)\setminus B|} \\ =&Z^{-1}\sum_\sigma \mathbb{1}[B \subseteq A(\sigma)]p^{|B|}(1-p)^{|E\setminus B|} \\ =&Z^{-1}p^{|B|}(1-p)^{|E\setminus B|}q^{c(B)} = \mu(B)\end{aligned}$$

□

The Edwards-Sokal coupling alternates between the random cluster model and the q-state Potts model. This can also be interpreted as a special type of Gibbs sampler.

Corollary 3.7 (Joint model)**.** *Consider the Joint model measure*

$$\rho(A,\sigma) = Z^{-1}p^{|A|}(1-p)^{|E\setminus A|}\mathbb{1}[A \subseteq A(\sigma)]\,.$$

Then one has

a)
$$\sum_\sigma \rho(A,\sigma) = \mu(A)\,.$$

b)
$$\sum_A \rho(A,\sigma) = \nu(\sigma)\,.$$

c) The Edwards-Sokal coupling applied to A generates $\sigma \sim \rho(\sigma \mid A)$ and then $B \sim \rho(B \mid \sigma)$.

3.3 Monotonicity of the Ising model

For $q = 2$, the q-state Potts model is also called Ising model, named after Ernst Ising [Isi25]. In this case, the model exhibits another monotonicity structure possible due

to its binary nature. We define a partial order on spin configurations $\sigma, \tau \colon V \to \{1,2\}$ via

$$\sigma \leq \tau \Leftrightarrow \forall v \in V \colon \sigma(v) \leq \tau(v)\,.$$

Many of the properties that follow are similar to those of the random cluster model, however this partial order is fundamentally different due to the spin symmetry of the Ising model (in the absence of boundary conditions). We will use another type of Glauber dynamics to prove the fundamental properties. Consider the procedure

1. For a given spin configuration σ, choose a vertex $u \in V$ uniformly at random.
2. Draw $r \in [0,1]$ from the uniform distribution and obtain the spin configuration σ' given by

$$\sigma'(v) = \begin{cases} \sigma(v) & v \neq u \\ 1 & v = u,\ r < \nu[\sigma'(u) = 1 \mid \forall v \neq u \colon \sigma'(v) = \sigma(v)] \\ 2 & \text{else}\,. \end{cases}$$

This so-called *heatbath dynamics* is clearly reversible. It is already formulated to obtain a grand coupling for the Ising model case, namely by choosing $u \in V$ in Step 1 and $r \in [0,1]$ in Step 2 independently from σ. It has the additional property of being completely local, because varying a spin configuration at spin u only has an influence on edges $e \in A(\sigma)$ if one of the endpoints of e is u. We proceed to prove an analogue to Theorem 3.4.

Theorem 3.8 (Monotone coupling for the Ising model)**.** *Let $G = (V,E)$ be a finite graph, $q = 2$ and $p \in (0,1)$. Let $W \subseteq V$ be a boundary set and $\psi, \chi \colon W \to \{1,2\}$ be boundary conditions with $\psi \leq \chi$. Let $\mathcal{A}$ be an increasing event from the set of spin configurations $\Omega = \{\sigma \colon V \to \{1,2\}\}$ with $\{\sigma \mid \forall w \in W \colon \sigma(w) = \chi(w),\ \sigma \in \mathcal{A}\}$ being nonempty. Consider the conditional Ising model distributions*

$$\nu = \nu(\sigma \mid \forall w \in W \colon \sigma(w) = \psi(w))$$

and

$$\nu' = \nu(\sigma \mid \forall w \in W \colon \sigma(w) = \chi(w),\ \sigma \in \mathcal{A})\,.$$

Then we have

$$\nu \leq \nu'\,,$$

and there exists a coupling (σ, τ) with

$$\sigma \sim \nu,\ \tau \sim \nu',\ \mathbb{P}[\sigma \leq \tau] = 1\,.$$

Note that we omit the monotonicity in p, which is not true in this case.

Proof. Let us assume first that the second statement is true. For an increasing event $\mathcal{B} \subseteq \Omega$, we have that

$$\nu[\mathcal{B}] = \mathbb{P}[\sigma \in \mathcal{B}] \leq \mathbb{P}[\tau \in \mathcal{B}] \leq \nu'[\mathcal{B}]\,,$$

showing that $\nu \leq \nu'$. To show the existence of the coupling (σ, τ), we proceed as in Lemma 3.2 and 3.3 to show that the grand coupling is monotone, and then use a CFTP approach to obtain the result. Let

$$\mathbb{Y} = \{\sigma \in \Omega \mid \forall w \in W \colon \sigma(w) = \psi(w)\}$$

and

$$\mathbb{Y}' = \{\tau \in \Omega \mid \forall w \in W \colon \tau(w) = \chi(w), \tau \in \mathcal{A}\}$$

be the domains of ν and ν', respectively. Consider the coupled heatbath process for $\sigma \in \mathbb{Y}$, $\tau \in \mathbb{Y}'$ and $\sigma \leq \tau$ with the additional constraint that at Step 2, the Markov chain does nothing if the proposed change would result in a configuration $\sigma' \notin \mathbb{Y}$ or $\tau' \notin \mathbb{Y}'$. If $u \in V$ is also in W, we have $\sigma' = \sigma$ and $\tau' = \tau$, and therefore $\sigma' \leq \tau'$. Let $u \in V \setminus W$ and $U \subseteq V$ be the set of vertices that share an edge with u. Then $\mathbb{P}[\sigma'(u) = 2]$ increases with increasing

$$|\{v \in U \colon \sigma(v) = 2\}|\,.$$

Because of $\sigma \leq \tau$ we get

$$|\{v \in U \colon \sigma(v) = 2\}| \leq |\{v \in U \colon \tau(v) = 2\}|$$

and the fact that $\mathbb{1}[\tau' \in \mathcal{A}]$ is an increasing function in $\tau'(u)$ overall implies

$$\mathbb{P}[\sigma'(u) = 2] \leq \mathbb{P}[\tau'(u) = 2]\,.$$

Therefore, at Step 2 of the coupled heatbath dynamics, the case $\sigma'(u) = 2$, $\tau'(u) = 1$ cannot happen and we get

$$\sigma \leq \tau \Rightarrow \sigma' \leq \tau'\,.$$

The rest of the proof is completely analogue to Theorem 3.4. The desired coupling is obtained by coupled CFTP procedures. □

3.4 Infinite volume measures

We will now turn towards one of the main features of the model, the phase transition. For this, we will restrict ourselves to the $\mathbb{Z}^2$ grid graph and subgraphs of it. Because $\mathbb{Z}^2$ is an infinite graph, we first need to rigorously introduce the random cluster measure on this graph. Consider $\mathbb{Z}^2$ as a graph embedded in $\mathbb{R}^2$, and let E_n be the set of edges where both endpoints $a, b \in \mathbb{R}^2$ satisfy $\|a\|_\infty, \|b\|_\infty \leq n$, with V_n being the corresponding vertex set. The subgraph induced by E_n is a bounded, rectangular grid graph with side-length $2n$ (Figure 3.4).

Theorem 3.9 (random cluster model on $\mathbb{Z}^2$)**.** *Let p, q be fixed and $\mathcal{A}_n$ an event that only depends on edges in E_n. Then*

$$\mu^0[\mathcal{A}_n] = \mu^0_{\mathbb{Z}^2,p,q}[\mathcal{A}_n] \coloneqq \lim_{m\to\infty} \mu_{E_m,p,q}[\mathcal{A}_n]$$

exists and defines a measure on such events. This measure extends to infinite events as well.

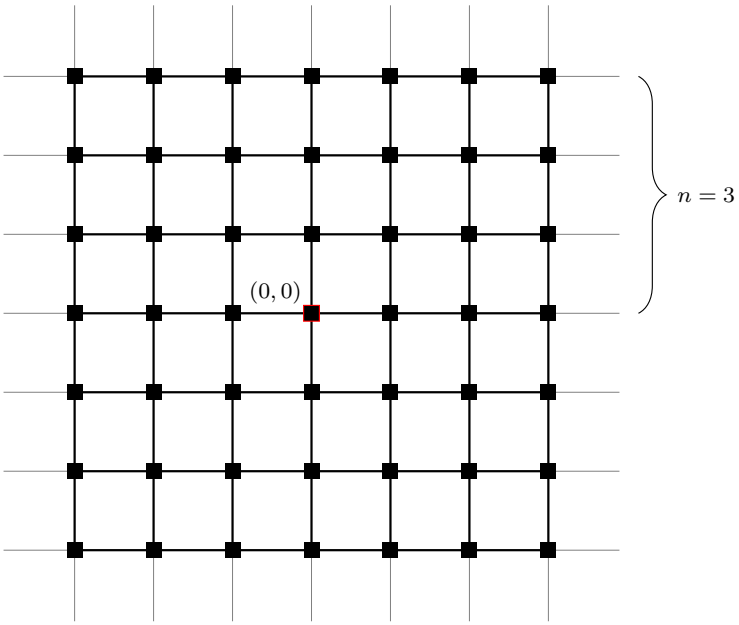

Figure 3.4: The graph E_n ($n = 3$) embedded in $\mathbb{Z}^2$.

The measure μ^0 is called infinite-volume random cluster measure with free boundary conditions. All edges in $\mathbb{Z}^2 \setminus E_n$ can be thought of as closed edges. The measure μ^1, which is obtained if the edges in $\mathbb{Z}^2 \setminus E_n$ are always open, is called infinite-volume random cluster measure with wired boundary conditions. Although it is not possible to define boundary conditions on $\mathbb{Z}^2$, there is a-priori no reason for these two measures to be the same, and we will see cases where they in fact differ.

Proof. For now, consider an event $\mathcal{A}_n$ which only depends on edges in E_n and is increasing. For $m \geq n$, let $\mu_m = \mu_{E_m,p,q}$. Note that the event $\mathcal{A}_n$ is defined for all models μ_m. Let furthermore $F_m = E_{m+1} \setminus E_m$ and $\mathcal{B}_m(B) = \{C \subseteq E_{m+1} \colon C \cap F_m = B\}$ for $B \subseteq F_m$. We have

$$\begin{aligned}
\mu_{m+1}[\mathcal{A}_n] &= \sum_{B \subseteq E_{m+1} \setminus E_m} \sum_{A \subseteq E_m} \mathbb{1}[A \cup B \in \mathcal{A}_n] \mu_{m+1}(A \cup B) \\
&= \sum_{B \subseteq E_{m+1} \setminus E_m} \sum_{A \subseteq E_m} \mathbb{1}[A \in \mathcal{A}_n] \mu_{m+1}[\mathcal{B}_m(B)] \mu_{m+1}(A \cup B \mid \mathcal{B}_m(B)) \\
&= \sum_{B \subseteq E_{m+1} \setminus E_m} \mu_{m+1}[\mathcal{B}_m(B)] \sum_{A \subseteq E_m} \mathbb{1}[A \in \mathcal{A}_n] \mu_m^{F_m,B}(A) \\
&\geq \sum_{B \subseteq E_{m+1} \setminus E_m} \mu_{m+1}[\mathcal{B}_m(B)] \sum_{A \subseteq E_m} \mathbb{1}[A \in \mathcal{A}_n] \mu_m^{F_m,\emptyset}(A) \\
&= \sum_{B \subseteq E_{m+1} \setminus E_m} \mu_{m+1}[\mathcal{B}_m(B)] \mu_m[\mathcal{A}_n] \\
&= \mu_m[\mathcal{A}_n] \,,
\end{aligned}$$

where we used the monotonicity of the model in boundary conditions for increasing events. This implies that the sequence $\{\mu_m[\mathcal{A}_n]\}$ converges, so $\mu^0[\mathcal{A}_n]$ is well-defined

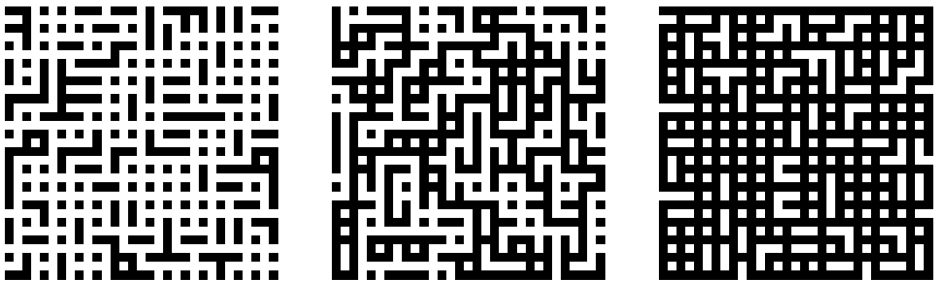

Figure 3.5: Random cluster model simulations on a 16×16 grid with $q = 2$ and $p = p_c - 0.2$ (left), $p = p_c$ (middle) and $p = p_c + 0.35$ (right). The cluster structure (black) shows the typical behavior of the model for the different phases.

for increasing events. Next, we show that all finite events $\mathcal{A}_n$ can be constructed from increasing ones. For any state $A \in E_n$, we know that $\{B\colon A \cap E_n \subseteq B \cap E_n\}$ and $\{B\colon A \cap E_n \subseteq B \cap E_n\} \setminus \{B\colon A \cap E_n = B \cap E_n\}$ are increasing sets. Therefore, we can define μ^0 for the event $\{B\colon A \cap E_n = B \cap E_n\}$ via

$$\mu^0[\{B\colon A \cap E_n \subseteq B \cap E_n\}] - \mu^0[\{B\colon A \cap E_n \subseteq B \cap E_n\} \setminus \{B\colon A \cap E_n = B \cap E_n\}]\,.$$

Any event that depends only on edges in E_n can be written as the disjoint union of such events, so we can extend μ^0 naturally to arbitrary events depending only on edges in E_n.
It remains to show that the measure extends to infinite events. Let X be the power set of E, where E is the set of edges in $\mathbb{Z}^2$, i.e. X is the set of all possible configurations on the infinite lattice. We define $\mu^0[X] = 1$ and $\mu^0[\emptyset] = 0$. Let $\mathcal{A} \subseteq X$ be a finite event. Then the event $X \setminus A$ is also a finite event. Moreover, the union of two finite events is finite. Therefore, the set consisting of finite events, the empty set and X form an algebra. Our considerations above have shown that μ^0 is a pre-measure on this algebra, and it is of course σ-finite. Therefore, by the Hahn-Kolmogorov-theorem, it uniquely extends to a measure on the σ-Algebra generated by finite events, and this measure is also a probability measure. □

The model has three distinct phases: the subcritical phase, the critical phase and the supercritical phase, see Figure 3.5. These can easily be observed while simulating the model at different parameter sets, and the nature of the phase transition has been conjectured for a very long time now. We will begin with the observation that the model is self-dual, which allows us to make a guess for the location of the critical point.
Let us consider the model μ_m on the graph E_m. This graph is planar, hence there exists a unique dual graph in the plane, call it D_m. For a configuration $A \subseteq E_m$, we can define a dual configuration $B \subseteq D_m$ as follows: An edge $e \in A$ is open if and only if its dual edge $d(e) \in B$ is not open in the dual model (see Figure 3.6). Given that A is distributed according to μ_m, what is the distribution of B? To see this, we consider the subgraph induced by $A = \emptyset$. This graph has $|V_m|$ vertices, zero edges $|A| = 0$, one face $f(A) = 1$, and $c(A) = |V_m|$ open clusters. So we get $|V_m| - c(A) - |A| + f(A) = -1$ for $A = \emptyset$. This identity actually holds for all

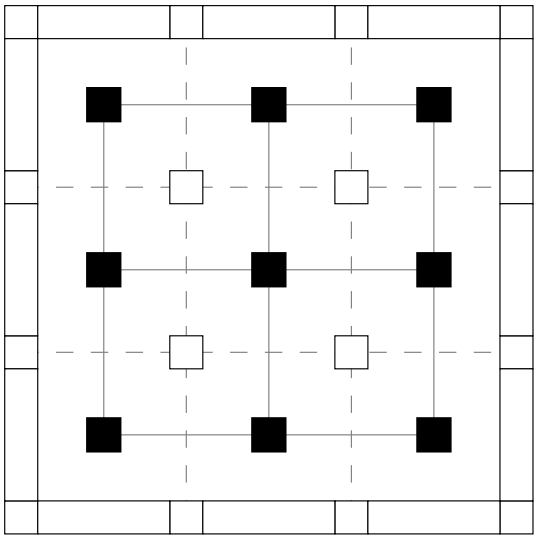

Figure 3.6: The dual model of the random cluster measure on E_n (black) is given on a rectangular graph with wired boundary conditions (white). Primal faces correspond to dual vertices, and the outer face is represented by the wired component of the dual graph.

$A \subseteq E_m$, which can be shown by induction. For an arbitrary A, let $e \notin A$. If e is not pivotal to A, then both endpoints of e already belong to the same cluster, so we have $c(A) = c(A \cup \{e\})$. This means that the face containing e is split into two faces such that $f(A) + 1 = f(A \cup \{e\})$. Thus, we have that $|A|$ increases by one and $f(A)$ increases by one, sustaining the identity. If e is pivotal to A, we get that $c(A) = c(A \cup \{e\}) + 1$. Moreover it is easy to see that the face containing e is not divided into two faces, such that the identity is also preserved. We have shown

Lemma 3.10. *For an edge configuration A of a finite, planar graph $G = (V, E)$, one has*

$$|V| - |A| + 1 = c(A) - f(A)\,,$$

where $c(A)$ is the number of open clusters and $f(A)$ is the number of faces of A.

It is easy to see that the faces of A can be identified with the open clusters of B such that $f(A) = c(B)$, see Figure 3.6. Thus, the expression $c(A) - c(B) + |D_m \setminus B|$ is constant and we can compute

$$\begin{aligned}
\mu_m(A) &\propto p^{|A|}(1-p)^{|E_m \setminus A|} q^{c(A)} \\
&\propto p^{|D_m \setminus B|}(1-p)^{|B|} q^{c(B)-|D_m \setminus B|} \\
&= (p/q)^{|D_m \setminus B|}\,(1-p)^{|B|} q^{c(B)} \\
&= (p/q + 1 - p)^{|D_m|} \left(\frac{(1-p)}{p/q+1-p}\right)^{|B|} \left(\frac{p}{p/q+1-p}\right)^{|D_m \setminus B|} q^{c(B)} \\
&\propto p_*^{|B|}(1-p_*)^{|D_m \setminus B|} q^{c(B)}
\end{aligned}$$

with $p_* = \frac{(1-p)}{p/q+1-p}$. We conclude that the distribution of B is given by the random cluster model on the dual graph, with same q and dual parameter p_*. How does the dual graph of E_m look like? We get a rectangular graph, where the outermost vertices are additionally connected to an outer vertex (representing the outer face of the primal graph). Considering the random cluster model on this graph, it is equivalent to the model on a graph where the outer edges are fixed to be always open; the outer open component acts as the outer vertex in this case (see Figure 3.6). From all these considerations, one can infer

Corollary 3.11 (Dual measure)**.** *Let $G = (V, E)$ be a finite, planar graph, $q \in \mathbb{N}$ and $p \in (0, 1)$. Furthermore, let G_* be the dual graph. An edge configuration $A \subseteq E$ is distributed according to $\mu = \mu_{G,p,q}$ if and only if the dual configuration B is distributed according to $\mu_* = \mu_{G_*,p_*,q}$ where $p_* = \frac{(1-p)}{p/q+1-p}$.*
For the $\mathbb{Z}^2$-graph, it follows that the measure $\mu^0_{\mathbb{Z}^2,p,q}$ has the dual $\mu^1_{\mathbb{Z}^2,p_*,q}$.

3.5 The critical point

One can ask what happens if we consider self-dual graphs, in the finite case or for $\mathbb{Z}^2$. In this case, the model should truly be self-dual if $p = p_*$. Moreover, one can state the following heuristic argument concerning the phase transition: If the model only has one critical point, it should be the self-dual point; otherwise the dual point to a critical point should also be critical in some sense. This was observed first by Kramers and Wannier [KW41], which stated an equivalent relation for p_* in the Ising model setting. Let us compute the self-dual point p_{sd}:

$$\begin{aligned} & p = p_* \\ \Leftrightarrow & p = \frac{(1-p)}{p/q+1-p} \\ \Leftrightarrow & p = \frac{\sqrt{q}}{\sqrt{q}+1}\,. \end{aligned}$$

For a long time, it was conjectured that the phase transition takes place at the self-dual point. What does this actually mean? One can observe that the model exhibits vastly different behavior dependent on which phase it is in, see Figure 3.5. We will now state the main results for the different phases of the random cluster model (without proof). In the following, let '$v \leftrightarrow w$' denote the event that v and w belong to the same open cluster in A. We start with the subcritical phase.

Theorem 3.12 (Beffara, Duminil Copin, [BDC12])**.** *Let $q \geq 1$ and $p < p_c = p_{sd}$. Then the measures $\mu^0_{\mathbb{Z}^2,p,q}$ and $\mu^1_{\mathbb{Z}^2,p,q}$ are equal and one has*

$$\mu_{\mathbb{Z}^2,p,q}[v \leftrightarrow w] \leq C(p,q)\exp(-c(p,q)\|v-w\|_2)$$

for some constants $c(p,q), C(p,q) > 0$ and the euclidean norm $\|\cdot\|_2$.

Already prior to this, it has been shown by Alexander that exponential decay of correlations in the infinite setting implies the same for finite sublattices.

Theorem 3.13 (Alexander, [Ale04]). *Let $q \geq 1$ and $p < p_c = p_{sd}$. For $n \in \mathbb{N}$, consider the random cluster model μ_n on E_n with any boundary condition. Then*

$$\mu_n \left[v \leftrightarrow w \text{ via an open path in } E_n \right] \leq C(p,q) \exp(-c(p,q) \|v - w\|_2) \,,$$

where the constants do not depend on the boundary condition or n.

Theorem 3.13 is interesting for us because we will deal with finite sublattices of $\mathbb{Z}^2$ a lot. It is important to note that the result also holds for more general shapes such as rectangular subgraphs of $\mathbb{Z}^2$.
In the supercritical phase, Theorem 3.12 holds for the dual model and has the following consequence, see for instance [Gri02].

Theorem 3.14 (Unique infinite volume cluster). *Let $q \geq 1$ and $p > p_c = p_{sd}$. Then the measures $\mu^0_{\mathbb{Z}^2,p,q}$ and $\mu^1_{\mathbb{Z}^2,p,q}$ are equal and one has*

$$\mu_{\mathbb{Z}^2,p,q} \left[\exists! \text{ unbounded open cluster } \right] = 1 \,,$$

as well as

$$\mu_{\mathbb{Z}^2,p,q} \left[v \leftrightarrow w \right] \geq B(p,q) > 0$$

for all $v, w \in \mathbb{Z}^2$.

The recent results for the critical case show that the random cluster model has a *continuous* phase transition for $q \in [1,4]$, where the infinite volume measure does not depend on the boundary conditions. It is *discontinuous* for $q > 4$, where care has to be taken when constructing the infinite volume measure from finite ones with specific boundary conditions.

Theorem 3.15 (Duminil Copin, Sidoravicius, Tassion [DCST16]). *Let $1 \leq q \leq 4$ and $p = p_c = p_{sd}$. Then the measures $\mu^0_{\mathbb{Z}^2,p,q}$ and $\mu^1_{\mathbb{Z}^2,p,q}$ are equal and one has*

$$\mu_{\mathbb{Z}^2,p,q} \left[\exists \text{ unbounded open cluster } \right] = 0 \,.$$

The correlation function

$$\mu_{\mathbb{Z}^2,p,q} \left[v \leftrightarrow w \right]$$

decays subexponentially with increasing euclidean distance $\|v - w\|_2$.

Theorem 3.16 (Duminil Copin, Gagnebin, Harel, Manolescu, Tassion [DCGH$^+$16])**.** *Let $q > 4$ and $p = p_c = p_{sd}$. Then the measures $\mu^0_{\mathbb{Z}^2,p,q}$ and $\mu^1_{\mathbb{Z}^2,p,q}$ are **not** equal. For free boundary conditions, one has*

$$\mu^0_{\mathbb{Z}^2,p,q}[v \leftrightarrow w] \leq C(p,q)\exp(-c(p,q)\|v-w\|_2)$$

for some constants $c(p,q), C(p,q) > 0$ and the euclidean norm $\|\cdot\|_2$. For wired boundary conditions, one has

$$\mu^1_{\mathbb{Z}^2,p,q}[\exists!\ \textit{unbounded open cluster}\ ] = 1\,,$$

as well as

$$\mu^1_{\mathbb{Z}^2,p,q}[v \leftrightarrow w] \geq B(p,q) > 0$$

for all $v, w \in \mathbb{Z}^2$. Other infinite volume measures can be constructed using different boundary conditions.

This has the following consequences for the q-state Potts model: In the critical case with $q \in \{1,2,3,4\}$, there is a unique infinite-volume measure ν, there is almost surely no unbounded cluster with neighboring vertices that all have the same spin, and $\nu[\sigma(v) = \sigma(w)]$ tends to $\frac{1}{q}$ subexponentially fast with increasing distance. For $q > 4$, the critical Potts model has many infinite-volume measures with behavior depending on the underlying boundary conditions.

4 Markov chains for the Fortuin-Kasteleyn Random Cluster Model

Throughout this chapter we will introduce Markov chains for the Fortuin-Kasteleyn random cluster model and the related q-state Potts model. Of special interest to us will be their scaling behavior for increasing graph sizes.

4.1 Glauber dynamics

Let $G = (V, E)$ be an undirected, finite and connected graph. We consider the random cluster probability measure $\mu = \mu_{G,p,q}$ with parameters $p \in (0, 1)$ and $q \in [1, \infty)$, defined on subsets $A \subseteq E$ given by

$$\mu(A) = \frac{1}{Z}\, p^{|A|}\, (1-p)^{|E \setminus A|}\, q^{c(A)}$$

with partition function Z. We have already seen the Glauber dynamics Markov chain (also called Sweeny's algorithm [Swe83]), which can be described via the procedure:

1. Choose an edge $e \in E$ uniformly at random.
2. Choose $A^{t+1} = A^t \setminus \{e\}$ or $A^{t+1} = A^t \cup \{e\}$, according to the measure

$$\mu\left(\cdot \mid A \setminus \{e\} = A^t \setminus \{e\}\right) .$$

This Markov chain is reversible: consider two states A, B which differ only in one edge $e \in E$ such that the probability to go from A to B and vice versa is not zero. We define $C = A \cap B$, and see that the probability to go from A to B is given by

$$|E|^{-1} \cdot \mu\left[D = B \mid D \setminus \{e\} = C\right] .$$

Moreover, we can write

$$\mu(A) = \mu[\{D \colon D \setminus \{e\} = C\}]\mu[D = A \mid D \setminus \{e\} = C]$$

such that the probability to sample A from μ and then go to B via a Glauber dynamics step is given by

$$|E|^{-1} \cdot \mu\left[D = B \mid D \setminus \{e\} = C\right] \mu[D = A \mid D \setminus \{e\} = C] \mu[\{D \colon D \setminus \{e\} = C\}] \,.$$

This expression is symmetric in A and B, which is the definition of reversibility.
We have seen in Lemma 3.1 that this Markov chain naturally preserves the given partial order $A \leq B \Leftrightarrow A \subseteq B$, meaning that there exists a (grand) coupling such that Markov chains with $A^s \leq B^s$ satisfy $A^t \leq B^t$ for all $t \geq s$. This even generalizes to comparable boundary conditions: If we consider $F \subseteq E$ to be fixed (i.e. at Step 1 of the dynamics, we choose edges from $E \setminus F$), we still get that $A^t \leq B^t$ for all $t \geq s$ provided that $A^s \leq B^s$ (see Lemma 3.3).
Let us consider the realizability of the Glauber dynamics Markov chain. At each step, one has to decide whether an edge $e \in E$ is pivotal or not. This can lead to the situation that the algorithmic cost of a Gauber dynamics step is not $\mathcal{O}(1)$, but rather depending on the underlying graph $G = (V, E)$. In [EmcW13], many algorithmic approaches to realize the Glauber dynamics have been analyzed and compared, coming to the conclusion that for subgraphs of $\mathbb{Z}^2$, the complexity can be controlled. In the subcritical and supercritical phase, the cluster structure of the underlying states usually allows fast computations in this regard.

4.2 Block dynamics

We can use a similar argumentation for the more general so-called *block dynamics*. Consider a probability distribution ω on the subsets of E, and define the procedure

1. Choose a subset $T \subseteq E$ according to the law ω.
2. Choose A^{t+1} according to the measure

$$\mu\left(A \mid A \setminus T = A^t \setminus T\right) \,,$$

in other words resample A^t on T according to the conditional measure with fixed values on $E \setminus T$.

This Markov chain, call it P, is reversible: For two states A, B the probability to go from A to B is given by

$$\mathbb{P}[A \to B] = \sum_{T \subseteq E} \omega(T) \mu\left(B \mid B \setminus T = A \setminus T\right) \,.$$

For a given T, if $\mu\left(B \mid B \setminus T = A \setminus T\right) \neq 0$ we can define $C = C(T) = A \setminus T = B \setminus T$. Therefore, we get

$$\begin{aligned} & \omega(T) \mu\left(B \mid B \setminus T = A \setminus T\right) \mu(A) \\ = \; & \omega(T) \mu\left(B \mid B \setminus T = C\right) \mu(A) \\ = \; & \omega(T) \mu\left(B \mid B \setminus T = C\right) \mu\left(A \mid A \setminus T = C\right) \mu[\{D \colon D \setminus T = C\}] \,, \end{aligned}$$

and it follows that the term $\mu(A)\mathbb{P}[A \to B]$ is symmetric in A and B, proving reversibility.

Moreover, there exists a monotone grand coupling for this Markov chain. The subset T chosen in Step 1 can be chosen independently of the state the Markov chain is in, and for two states $A \leq B$, the resampling in Step 2 can be done monotonically according to Theorem 3.4.
Of course, the block dynamics will behave very different for specific choices of ω. For instance, if ω chooses the whole set E with probability 1, it reduces to perfect sampling. Using Glauber dynamics with the CFTP procedure, it is possible to generate a sample from μ, however the algorithmic cost will depend on the coupling time for the Glauber dynamics.

4.3 Swendsen-Wang dynamics

The next Markov chain we will consider is the Swendsen-Wang dynamics [SW87]. We have already seen it in Lemma 3.5 and Corollary 3.7, so let us reintroduce it quickly. We restrict ourselves to integer values of q and define the procedure

1. For a given state A, assign to each open cluster C of A a *spin* $s \in \{1, \dots, q\}$, uniformly at random.

2. Assign to each vertex $v \in V$ the spin of the open cluster C it belongs to. This map $\sigma \colon V \to \{1, \dots, q\}$ we call *spin configuration*.

3. For a spin configuration σ, obtain the new state B via the rule

$$\mathbb{P}[e \in B] = \begin{cases} p & \text{both endpoints of } e \text{ have the same spin in } \sigma, \\ 0 & \text{else,} \end{cases}$$

   considered independently for each $e \in E$.

From Lemma 3.5 we know that this Markov chain is reversible and ergodic, and Lemma 3.6 tells us that spin configurations σ obtained from $A \sim \mu$ are distributed according to the q-state Potts measure

$$\nu(\sigma) = Z^{-1}(1-p)^{|E \setminus A(\sigma)|}.$$

Similarly,we can start with a spin configuration and apply steps 3, 1 and 2 to obtain a new spin configuration. Both Markov chains are easily realizable, the algorithmically difficult step is to find the cluster configuration for a given edge configuration, but this is possible in $\mathcal{O}(|E|)$ cost using standard graph search algorithms. For a global algorithm that potentially changes the whole system, this is acceptable and allows for an easy comparison to local algorithms.

4.4 Monotone Swendsen-Wang dynamics

In the Ising model case with $q = 2$ there exists a modification of this procedure that is monotone, introduced in [BCV18]. The partial order on spin configurations given by

$$\sigma \leq \tau \Leftrightarrow \sigma(v) \leq \tau(v) \text{ for all } v \in V$$

makes the configurations $\sigma \equiv 1$ and $\tau \equiv 2$ extremal and we have a similar situation as in the random cluster model case, with a unique minimal and maximal element being present. The following modification has been called *monotone Swendsen-Wang dynamics*:

1. For a spin configuration σ, obtain the edge configuration B via the rule

$$\mathbb{P}[e \in B \mid \sigma] = \begin{cases} p & \text{both endpoints of } e \text{ have the same spin in } \sigma\,, \\ 0 & \text{else,} \end{cases}$$

   considered independently for each $e \in E$.

2. Assign to each vertex $v \in V$ a new spin $\theta(v)$, chosen uniformly between 1 and 2.

3. Consider each open cluster C of B. If for all $v \in C$ the spins $\theta(v)$ align, set $\sigma'(v) = \theta(v)$ for all $v \in C$. Otherwise set $\sigma'(v) = \sigma(v)$ for $v \in C$. Then σ' is the new spin configuration.

This Markov chain is similar to the usual Swendsen-Wang dynamics, but clusters have to agree at each vertex on the spin they choose, otherwise there is no change. Intuition tells us that this Markov Chain should be slower than Swendsen-Wang dynamics, as clusters will have a decreasing chance of changing their spin with increasing size.
For reversibility, we need to compute

$$\mathbb{P}[\sigma \to \sigma'] = \sum_B \mathbb{P}[\sigma \to (\sigma, B)]\mathbb{P}[(\sigma, B) \to \sigma']\,.$$

The first probability is given by Step 1 of the dynamics and coincides with the usual Swendsen-Wang dynamics, namely

$$\mathbb{P}[\sigma \to (\sigma, B)] = \mathbb{1}[B \subseteq A(\sigma)]p^{|B|}(1-p)^{|A(\sigma)\setminus B|}\,.$$

The second probability depends on σ, B and σ'. Let $D = D(\sigma, \sigma')$ be the set of vertices with $\sigma(v) \neq \sigma'(v)$. Then we have

$$\mathbb{P}[(\sigma, B) \to \sigma'] = \mathbb{1}[B \subseteq A(\sigma')] \prod_C \left(\mathbb{1}[C \subseteq D]2^{-|C|} + \mathbb{1}[C \not\subseteq D]\left(1 - 2^{-|C|}\right)\right)\,,$$

with the product iterating over all open clusters of B. The first summand of that product describes the event that change takes place, i.e. all vertices have to agree. In the second summand, we need the event that at least one of the vertices of C chooses the old color, such that change is not taking place. The product is already symmetric in σ and σ', so we denote it by $\Pi(B, \sigma, \sigma')$ and conclude

$$\nu(\sigma)\mathbb{P}[\sigma \to \sigma'] = Z^{-1} \sum_B \mathbb{1}[B \subseteq A(\sigma)]p^{|B|}(1-p)^{E\setminus B|}\mathbb{1}[B \subseteq A(\sigma')]\Pi(B, \sigma, \sigma')\,.$$

This term is symmetric in σ and σ', proving reversibility.
Let us check the monotonicity of this procedure. The grand coupling for this Markov

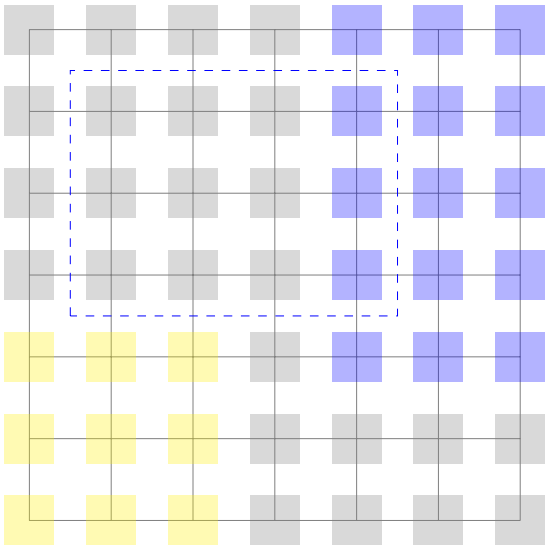

Figure 4.1: The situation described for the monotone Swendsen-Wang algorithm. The yellow spins correspond to $v \in V$ with $\sigma(v) = \tau(v) = 1$, the blue spins correspond to $v \in V$ with $\sigma(v) = \tau(v) = 2$ and the gray spins correspond to $v \in V$ with $\sigma(v) = 1$ and $\tau(v) = 2$ such that $\sigma \leq \tau$. The highlighted region corresponds to a $B(\tau)$-cluster that will change to spin 1. All $B(\sigma)$-clusters that intersect the highlighted region will either stay at spin 1 if intersecting the gray region, or change to spin 1 if contained in the blue region.

chain is straightforward: Choose the same random numbers to decide if an edge is in A at Step 1, and choose the same proposal spin configuration $\theta(v)$ in Step 2. For spin configurations $\sigma \leq \tau$, it is clear that $\sigma' \leq \tau'$ is satisfied if at Step 1 the same edge configuration is generated. We therefore have to check how the corresponding edge configurations $B(\sigma)$ and $B(\tau)$ relate to each other. Consider an open cluster C in $B(\tau)$ which has $\tau(v) = 2$ for all $v \in C$. Assume further that for all $v \in C$, we have $\theta(v) = 1$ such that $\tau'(v) = 1$, because otherwise $\sigma'(v) \leq \tau'(v)$ for all $v \in C$ immediately follows. We have to show that in this situation, one has $\sigma'(v) = 1$ for all $v \in C$. Due to θ agreeing on C on the spin $s = 1$, for all sites $v \in C$ one has that $\sigma(v) = 1$ implies $\sigma'(v) = 1$ because the corresponding $B(\sigma)$-cluster of v will not change spin. For a site $v \in C$ with $\sigma(v) = 2$, the corresponding $B(\sigma)$-cluster is contained in C: every boundary edge either comes from Step 1, sharing it with C, or it comes from a spin difference, i.e. a bordering vertex with spin 1. But $\sigma \leq \tau$ then implies that the $B(\sigma)$-cluster of v is contained in C. Therefore, all spins on this cluster agree and one has $\sigma'(v) = \theta(v) = \tau'(v)$. A similar argumentation can be used to show that 1-clusters in $B(\sigma)$ that change their spin to 2 do not violate $\sigma' \leq \tau'$. See Figure 4.1 for an illustration.
The above argumentation moreover holds if we introduce spin boundary conditions ψ, i.e. assign to some set of vertices a fixed spin. Therefore, this Markov chain would also allow us to prove Theorem 3.8.

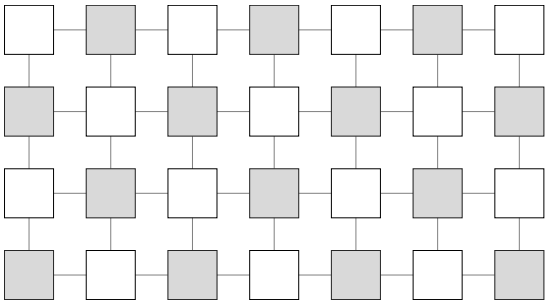

Figure 4.2: An even/odd partition of a rectangular grid graph.

4.5 Heatbath dynamics

Let us turn back to the q-state Potts model setting. We can again introduce a Glauber dynamics type of algorithm to obtain σ' from σ:

1. Choose a vertex $v \in V$ uniformly at random.
2. Draw a new spin $\sigma'(v)$ according to the conditional measure

$$\nu(\sigma' \mid \sigma(w) = \sigma'(w) \text{ for } w \neq v)\,.$$

This Markov chain, also called *heatbath dynamics*, is reversible (like in the Glauber dynamics case) and has ν as its stationary distribution. To clarify this, consider the measure ν and its interaction terms. An edge e belongs to $A(\sigma)$ if and only if both endpoints have the same spin. Therefore, we can write

$$(1-p)^{|E\setminus A(\sigma)|} = \prod_{\langle v\,w\rangle} (1-p)^{\mathbb{1}[\sigma(v)\neq\sigma(w)]} \propto \exp\left(\beta \sum_{\langle v\,w\rangle} \mathbb{1}[\sigma(v)=\sigma(w)]\right)$$

with $p = 1 - \exp(-\beta)$, as is more common for the q-state Potts model. The conditional measure at a vertex therefore only depends on the spin values at neighboring vertices, and Step 2 of the algorithm is easy to implement.
Many variants of this algorithm are used in practice, but we will mainly focus on the one described just now and the so-called *alternate scan dynamics*. Here, the vertices at Step 1 are not chosen randomly but in a fixed order, and the locality of the conditional measure allows us to do many operations in parallel. Let V_1 be a subset of V such that any $v, w \in V_1$ are not connected by an edge. Then, Step 2 of the Glauber dynamics can be performed simultaneously at all vertices in V_1. If we can partition $V = \bigcup V_i$ into such sets we get the alternate scan dynamics:

1. For $i = 1, 2, \ldots$ draw a new spin configuration σ' according to the conditional measure

$$\nu(\sigma' \mid \sigma(w) = \sigma'(w) \text{ for } w \notin V_i)\,.$$

For graphs that are subgraphs of $\mathbb{Z}^2$, one usually divides the graph into 'even' and 'odd' vertices, see Figure 4.2.

For $q = 2$, the spin Glauber dynamics and the alternate scan dynamics have a grand coupling which is monotone with respect to the partial order

$$\sigma \leq \tau \Leftrightarrow \sigma(v) \leq \tau(v) \ \forall v \in V \,,$$

which is achieved by taking the same vertices at Step 1 and taking the same random number at Step 2 to decide the new spin. This has been elaborated in Chapter 3.

4.6 Spin block dynamics

Finally let us also consider a block dynamics variant for the q-state Potts model:

1. Choose a vertex set $U \subseteq V$ according to some measure on vertex subsets.
2. Draw a new spin $\sigma'(v)$ according to the conditional measure

$$\nu(\sigma' \mid \sigma(w) = \sigma'(w) \text{ for } w \notin U) \,.$$

Reversibility follows exactly as in the block dynamics case. This algorithm is theoretically interesting due to its short mixing time in the subcritical regime, but it is not feasible in general for large 'blocks' U.

5 State of the Art Methods and Results

In this chapter, we discuss several state of the art results concerning the Markov chains introduced in Chapter 4. Most of the results will be formulated on subgraphs of $\mathbb{Z}^2$, and we begin with deriving necessary tools from Theorem 3.13, see also [Ale04] and [BS16].

The following lemma should give a good intuition for the proof techniques used in this section. Roughly it states that in the subcritical phase, the influence of edge boundary conditions on the marginal measure away from the boundary decays exponentially.

Lemma 5.1. *Let $p < p_c(q)$. Consider a lattice $E_m \subseteq \mathbb{Z}^2$ and its complement E_m^c which contains all edges in $\mathbb{Z}^2 \setminus E_m$. Let A_1^c and A_2^c be edge configurations on E_m^c. For $n < m$, consider the marginal measures on E_n, conditioned on the boundary configurations:*

$$\mu_1 = \mu\left(A = A_{|E_n} \mid A_1^c\right) \text{ and } \mu_2 = \mu\left(A = A_{|E_n} \mid A_2^c\right).$$

These probability distributions on E_n satisfy

$$\|\mu_1 - \mu_2\|_{TV} \leq C \exp(-c|m-n|),$$

where the constants $C, c > 0$ only depend on p, q.

Proof. Let us consider the extremal wired boundary condition first, i.e. $A_2^c = E_m^c$. Consider a coupling for edge configurations (A_1, A_2) on E_m that satisfies $A_1 \leq A_2$, and A_i is distributed to the random cluster measure on E_m with boundary conditions A_i^c for $i \in \{1, 2\}$. Such a coupling can be obtained from coupled Glauber dynamics, see Theorem 3.4 or [BS16]. We apply a Markov chain step to obtain a different coupling (B_1, B_2) in the following way. Let Λ be the set of edges where both endpoints are not connected to the boundary of E_m via open edges in A_2. It is easy to see that all edges with exactly one endpoint in the Λ-graph are closed in A_2, hence also closed in A_1. Therefore, we can do a resampling of the edges on Λ according to the random cluster measure on Λ with free boundary conditions to

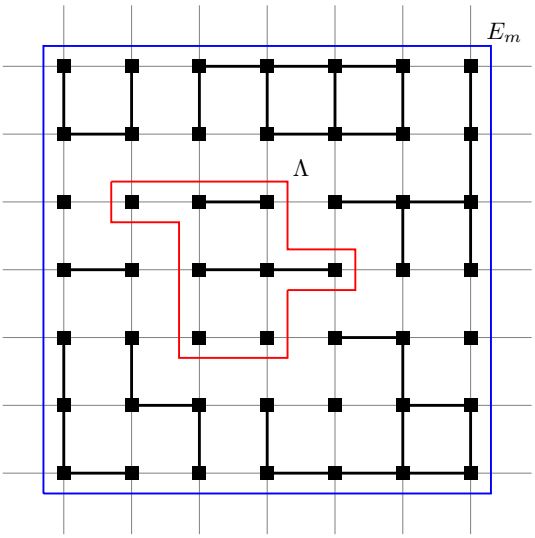

Figure 5.1: An edge configuration on E_m with $m = 3$ and wired boundary conditions. The subgraph Λ enclosed in red consists of all edges and adjacent vertices that are not connected to the boundary.

obtain (B_1, B_2) from (A_1, A_2) which now agree on edges in Λ, see Figure 5.1.
From this procedure, it follows that $B_1 \leq B_2$ and $B_1 = B_2$ on edges in Λ. Moreover B_i has the same distribution as A_i for $i \in \{1, 2\}$, because resampling according to the conditional measure on a fixed edge set is reversible (we elaborate on this fact after the proof). Using this coupling, we get that

$$\|\mu_1 - \mu_2\|_{TV} \leq \mathbb{P}[B_1 \neq B_2 \text{ on } E_n] \leq \mathbb{P}[E_n \not\subseteq \Lambda] = \mathbb{P}[V_n \leftrightarrow V_m^c \text{ via } A_2]\,.$$

We can bound this probability using the exponential decay of connectivities from Theorem 3.13, using a crude estimate over all boundary nodes of V_n and V_m:

$$\begin{aligned}
&\mathbb{P}[V_n \leftrightarrow V_m^c \text{ via } A_2] \\
\leq &\sum_{v \in V_m \setminus V_{m-1}} \mathbb{P}[V_n \leftrightarrow v \text{ via } A_2] \\
\leq &\sum_{w \in V_n \setminus V_{n-1}} \sum_{v \in V_m \setminus V_{m-1}} \mathbb{P}[w \leftrightarrow v \text{ via } A_2] \\
\leq &(4n)(4m)C(p,q)\exp(-c(p,q)|m-n|)\,.
\end{aligned}$$

Now, with a slight adjustment of the constants in this estimate one can neglect the polynomial dependence on n and m to obtain the result.
Finally, for arbitrary boundary conditions A_3^c, we get that

$$\|\mu_1 - \mu_3\|_{TV} \leq \|\mu_1 - \mu_2\|_{TV} + \|\mu_2 - \mu_3\|_{TV}\,,$$

completing the proof with an adjusted constant (multiplied by 2). □

The resampling procedure used in the proof is a powerful tool and we will use it often. The set Λ depends on the state A_2 and it is not a-priori clear that B_2 therefore

has the same law as A_2, let alone B_1 and A_1. We make this step rigorous with the following argument: Let A be distributed according to the random cluster measure μ on some graph (V, E), according to some boundary conditions. Let $\omega(\Lambda \mid A)$ be a law for subsets $\Lambda \subseteq E$ that depends on A. The procedure that takes A, randomly draws Λ according to $\omega(\Lambda \mid A)$ and draws A' according to the law $\mu(A' \mid A' = A \text{ on } \Lambda^c)$ can be described with the transition probability

$$\mathbb{P}[A \to A'] = \sum_{\Lambda} \omega(\Lambda \mid A)\mu(A' \mid A' = A \text{ on } \Lambda^c)\,.$$

If this procedure is reversible with respect to μ, we get that A' has the same law as A. First off, for any fixed Λ the resampling is clearly reversible as already seen in Chapter 4:

$$\mu(A)\mu(A' \mid A' = A \text{ on } \Lambda^c) = \mu(A')\mu(A \mid A = A' \text{ on } \Lambda^c)\,.$$

Therefore, if for a pair A, A' with $\mathbb{P}[A \to A'] \neq 0$ we have that $\omega(\Lambda \mid A) = \omega(\Lambda \mid A')$, we get reversibility of P. This situation applies to (A_2, B_2) used in the proof: $\omega(\Lambda|A_2)$ is given by

$$\mathbb{1}[\Lambda = \{e \in E\colon \text{both endpoints are not connected to the boundary via } A_2\}]$$

and this Λ does not change during $A_2 \to B_2$. We write $\Lambda = \Lambda(A_2) = \Lambda(B_2)$. For the procedure $A_1 \to B_1$, Λ is given implicitly by the coupling (A_1, A_2). Let ρ be the law of this coupling with marginals $\rho_1(A_1)$ and $\rho_2(A_2)$ (random cluster measures with corresponding boundary conditions), then

$$\begin{aligned}\mathbb{P}[B_1] &= \sum_{A_1,A_2} \rho(A_1, A_2)\rho_1(B_1 \mid B_1 = A_1 \text{ on } \Lambda(A_2)^c) \\ &= \sum_{A_1,A_2} \rho(A_1 \mid A_2)\rho_2(A_2)\rho_1(B_1 \mid B_1 = A_1 \text{ on } \Lambda(A_2)^c)\,.\end{aligned}$$

Now, it is important to note that the rightmost term does not really depend on A_1, it only depends on the fact that A_2 induces free boundary conditions on Λ. The measures ρ_1 and ρ_2 also are equal for these boundary conditions. Therefore, we can use the reversibility property to obtain

$$\begin{aligned}&\sum_{A_1,A_2} \rho(A_1 \mid A_2)\rho_2(A_2)\rho_1(B_1 \mid B_1 = A_1 \text{ on } \Lambda(A_2)^c) \\ =& \sum_{A_1,A_2} \rho(A_1 \mid A_2)\rho_1(B_1)\rho_2[A = A_2 \text{ on } \Lambda(A_2) \mid A = A_2 \text{ on } \Lambda(A_2)^c] \\ =& \rho_1(B_1) \sum_{A_2} \rho_2[A = A_2 \text{ on } \Lambda(A_2) \mid A = A_2 \text{ on } \Lambda(A_2)^c] \\ =& \rho_1(B_1)\,,\end{aligned}$$

so B_1 is distributed as A_1.
We will now briefly discuss the consequences for the q-state Potts model with $q \in \mathbb{N}$, using the Edwards-Sokal coupling. This procedure is well-defined for the $\mathbb{Z}^2$-lattice,

and we denote with $\nu^0 = \nu^0_{\mathbb{Z}^2,p,q}$ the infinite-volume measure obtained from $\mu^0_{\mathbb{Z}^2,p,q}$, as well as $\nu^1 = \nu^1_{\mathbb{Z}^2,p,q}$ obtained from $\mu^1_{\mathbb{Z}^2,p,q}$. It is easy to see that the event in the spin system setting corresponding to '$v \leftrightarrow w$ via an open path' is given by the event '$\sigma(v) = \sigma(w)$'. From the Edwards-Sokal coupling one can directly compute that for any underlying graph, one has

$$\nu[\sigma(v) = \sigma(w)] = \mu[v \leftrightarrow w] + \frac{1}{q}\left(1 - \mu[v \leftrightarrow w]\right) .$$

It is therefore very easy to derive similar statements to those of the above theorems. In the subcritical phase, $\nu[\sigma(v) = \sigma(w)]$ tends to $1/q$ exponentially fast with increasing distance between v and w, while in the supercritical phase, there is $\beta = \beta(p,q)$ such that $\nu[\sigma(v) = \sigma(w)] \geq \frac{1}{q} + \beta$ for all $v, w \in \mathbb{Z}^2$. The measure $\nu = \nu^1$ is, though obtained from the unique measure $\mu = \mu^1_{\mathbb{Z}^2,p,q}$, not the only infinite volume measure in the supercritical Potts model. Indeed, one might consider what happens if in the spin system setting, one introduces a spin boundary condition where all boundary spins have a specific spin $s \in \{1, \ldots, q\}$. The resulting infinite-volume measures ν^s are all equal in the subcritical phase, but distinct in the supercritical phase. They can be obtained from μ^1 with the Edwards-Sokal coupling if one modifies the procedure such that the unique unbounded cluster is always assigned the spin s. From this, it is also clear that

$$\nu = \frac{1}{q}(\nu^{s=1} + \ldots + \nu^{s=q})$$

in the supercritical phase.
In the subcritical phase, the model has another convenient property due to its locality. For a vertex set $U \subseteq \mathbb{Z}^2$, let $\sigma(U)$ be the spin configuration σ restricted to U. The conditional measure on U depending on a spin configuration in its complement U^c only depends on the vertices adjacent to U:

$$\nu(\sigma(U) \mid \sigma(U^c)) = \nu(\sigma(U) \mid \sigma(\partial U)) ,$$

where ∂U is the set of vertices in U^c that share an edge with a vertex in U. Consider the vertex sets V_n and V_m for $n \leq m$. Then the marginal measure on V_n conditioned on spins outside of V_m

$$\nu(\sigma(V_n) \mid \sigma(V_m^c))$$

behaves nicely in the subcritical regime: For two spin configurations σ and σ' on the boundary, one has

$$\left\|\nu(\sigma(V_n) \mid \sigma(V_m^c)) - \nu(\sigma'(V_n) \mid \sigma'(V_m^c))\right\|_{TV} \leq C \exp(-c|m-n|)$$

for some constants $C, c > 0$. This property is called *strong spatial mixing* (SSM) and follows from a similar argument as in Lemma 5.1. To see this, we need to introduce a modified Edwards-Sokal coupling for fixed spin boundary conditions. Let σ be a spin configuration on $\mathbb{Z}^2$.

1. Obtain the edge configuration B via the rule

$$\mathbb{P}[e \in B] = \begin{cases} p & \text{both endpoints of } e \text{ have the same spin,} \\ 0 & \text{else,} \end{cases}$$

considered independently for each $e \in E$.

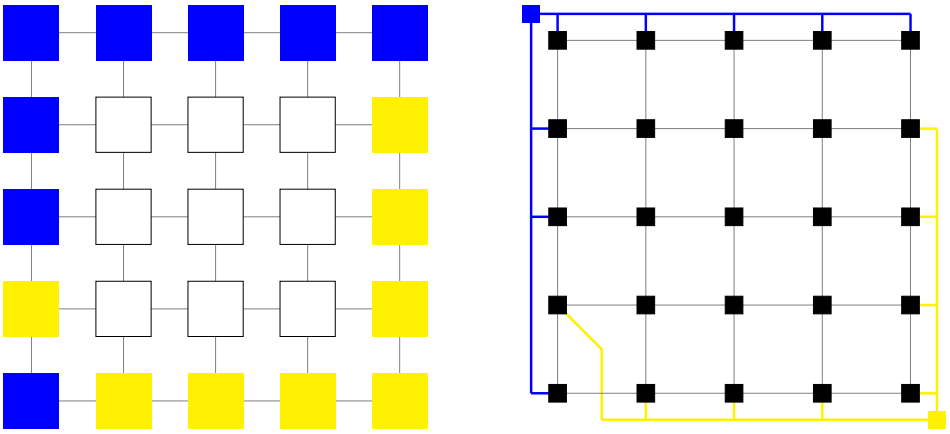

Figure 5.2: A spin boundary condition with $q = 2$ different spins (left) and the resulting modified graph E^{ψ}_{m+1} (right). The colored edges are fixed to be open. In this example, E^{ψ}_{m+1} cannot be embedded in $\mathbb{Z}^2$.

2. For each open cluster C that is completely contained in V_m, assign a new spin s, chosen uniformly at random, to each vertex of the cluster C.

It is clear that this Markov chain does not change the spin boundary conditions. For $i \in \{1, \dots, q\}$ let V_i be the collection of vertices adjacent to V_m with $\sigma(v) = i$. The collection $\psi = \{V_1 \dots, V_q\}$ completely encodes the spin boundary condition induced by σ, and we write ν^{ψ} for $\nu(\cdot \mid \sigma(V_m^c))$. We want to specify the distribution for B on E_{m+1} after applying Step 1 of the given Markov chain, therefore we define a modified graph structure. We add special vertices $v_1, \dots, v_q$ to the graph E_{m+1} and add artificial edges which connect every vertex in V_i to v_i for $i \in \{1, \dots, q\}$, see Figure 5.2. On the resulting graph E^{ψ}_{m+1}, we restrict the added edges to always be open, such that an edge configuration $B \subseteq E^{\psi}_{m+1}$ has q distinct boundary clusters connecting to the vertices $v_1, \dots, v_q$. We can reformulate the Markov chain as follows:

1. Obtain the edge configuration $B \subseteq E^{\psi}_{m+1}$ via the rule

$$\mathbb{P}[e \in B] = \begin{cases} 1 & \text{one endpoint of } e \text{ is in } \{v_1, \dots, v_q\}, \\ p & \text{both endpoints of } e \in E_{m+1} \text{ have the same spin,} \\ 0 & \text{else,} \end{cases}$$

   considered independently for each $e \in E^{\psi}_{m+1}$.

2. For $i \in \{1, \dots, q\}$ assign the spin i to the cluster containing the vertex v_i. For the remaining clusters, choose a spin uniformly at random and assign it to every vertex of the given cluster.

One readily verifies that this Markov chain has the same transition probability as the one given earlier. Let μ^{ψ}_{*} be the distribution of $B \cap E_{m+1}$ after applying Step 1 of this Markov chain, and let $\nu^{\psi}(\tau) \propto (1-p)^{|E_{m+1} \setminus A(\tau)|}$ be the q-state Potts model measure for spin configurations $\tau \colon V_{m+1} \to \{1, \dots, q\}$ that align with ψ on

the outermost vertices. Then we have

$$\begin{aligned}\mu_*^{\psi}(B) &= \sum_{\tau} \nu^{\psi}(\tau) p^{|B|}(1-p)^{|A(\tau)\setminus B|} \mathbb{1}[B \subseteq A(\tau)] \\ &\propto \sum_{\tau} (1-p)^{|E_{m+1}\setminus A(\tau)|} p^{|B|}(1-p)^{|A(\tau)\setminus B|} \mathbb{1}[B \subseteq A(\tau)] \\ &= \sum_{\tau} p^{|B|}(1-p)^{|E_{m+1}\setminus B|} \mathbb{1}[B \subseteq A(\tau)]\,.\end{aligned}$$

The boundary conditions for ψ and the term $\mathbb{1}[B \subseteq A(\tau)]$ result exactly in the event

$$\mathbb{1}[\forall i, j \colon V_i \not\leftrightarrow V_j \text{ via } B]\,,$$

and we therefore obtain

$$\begin{aligned}\mu_*^{\psi}(B) &\propto p^{|B|}(1-p)^{|E_{m+1}\setminus B|}\, q^{c(B)}\, \mathbb{1}[\forall i, j \colon V_i \not\leftrightarrow V_j \text{ via } B] \\ &\propto \mu^{\psi}(B \mid \forall i, j \colon V_i \not\leftrightarrow V_j \text{ via } B)\,,\end{aligned}$$

with μ^{ψ} actually being the random cluster measure on the graph E_{m+1}^{ψ} with the boundary condition that all artificial edges are always open. It is easy to see that applying Step 2 of the Markov chain to this measure recovers ν^{ψ}. We are now ready to prove the (SSM) property.

Lemma 5.2 (SSM)**.** *Let $p < p_c(q)$. Let $\psi = \{V_1, \dots, V_q\}$ and $\chi = \{W_1, \dots, W_q\}$ be spin boundary conditions on the vertices adjacent to V_m. Let ν^{ψ} be the q-state Potts model measure for spin configurations $\tau \colon V_{m+1} \to \{1, \dots, q\}$ that align with ψ on the outermost vertices, analogously ν^{χ}. Let $\nu_n^{\psi}, \nu_n^{\chi}$ be the corresponding marginal measures on V_n for $n < m$. Then one has*

$$\|\nu_n^{\psi} - \nu_n^{\chi}\|_{TV} \leq C \exp(-c|m-n|)$$

with constants $C, c > 0$ not depending on ψ, χ, m, n.

Proof. Let μ^{ψ} and μ^{χ} be the random cluster measures on the graphs E_{m+1}^{ψ} and E_{m+1}^{χ} where edges not in E_{m+1} are always open. It is important to note that in general, it is not possible to realize the underlying boundary conditions with a subgraph of $\mathbb{Z}^2$. However, adding open edges between the vertices $v_1, \dots, v_q$ results in a graph structure comparable to the E_{m+1}-graph with wired boundary conditions, resulting in

$$\mu^{\psi} \leq \mu_m^1\,.$$

Moreover, the event '$\forall i, j \colon V_i \not\leftrightarrow V_j$ via B' which we call $\mathcal{A}^{\psi}$ is decreasing, therefore we can use a Glauber dynamics argument similar to that in Lemma 3.1 to obtain

$$\mu^{\psi}(\cdot \mid \mathcal{A}^{\psi}) \leq \mu^{\psi} \leq \mu_m^1\,.$$

Now, as in Lemma 3.2, we can define a coupling (A, B) such that $A \sim \mu^{\psi}(\cdot \mid \mathcal{A}^{\psi})$, $B \sim \mu_m^1$, $A \leq B$ and $A = B$ on the Λ-subgraph of edges and vertices that are

not connected to the boundary by open edges in B. To this coupling, we now apply Step 2 of the modified Edwards-Sokal coupling, where we choose the same spin for clusters contained in Λ, where A and B coincide. Call the resulting spin configuration coupling (σ, τ). The configuration σ has law ν^ψ, and coincides with τ on the subgraph Λ. Therefore, if we denote by ν_n^1 the marginal law of τ on V_n, we have

$$\|\nu_n^\psi - \nu_n^1\|_{TV} \leq \mathbb{P}[E_n \not\subseteq \Lambda]\,.$$

The rest of the proof is exactly the same as in Lemma 5.1. This probability can be bounded by an estimate over the outer vertices of V_n and inner vertices of V_m^c to obtain the exponential decay property, and a triangle inequality argument implies

$$\|\nu_n^\psi - \nu_n^\chi\|_{TV} \leq C\exp(-c|m-n|)\,.$$

□

This Lemma shows that in the subcritical phase, an arbitrary change of spin boundary conditions has little influence on the marginal measure away from the boundary. However, it is possible to prove an even stronger result, namely that a local change of the spin boundary conditions will only affect the measure in its vicinity. To show this, we have to modify the proof of Lemma 5.2.

Lemma 5.3 (local SSM)**.** *Let $p < p_c(q)$. Consider spin boundary conditions $\psi = \{V_1, \dots, V_q\}$ and $\chi = \{W_1, \dots, W_q\}$ on the vertices adjacent to V_m* ***that differ only at one vertex*** *u. For $r > 0$, let Λ_r be the subgraph of $G = (V_{m+1}, E_{m+1})$ containing all vertices v with $\|u - v\|_2 \geq r$ and all edges between such vertices. Let ν^ψ be the q-state Potts model measure for spin configurations $\tau\colon V_{m+1} \to \{1, \dots, q\}$ that align with ψ on the outermost vertices, analogously ν^χ. Let ν_r^ψ, ν_r^χ be the corresponding marginal measures on Λ_r for $r > 0$. Then one has*

$$\|\nu_r^\psi - \nu_r^\chi\|_{TV} \leq C\exp(-cr)$$

with constants $C, c > 0$ not depending on ψ, χ, r, m.

Proof. Let μ^ψ and μ^χ be the random cluster measures on the graphs E_{m+1}^ψ and E_{m+1}^χ where the artificial edges are always open. The graphs only differ in the edge connecting the vertex u to one of the vertices $v_1, \dots, v_q$. Let $\mathcal{A}^\psi$ be the event that v_i is not connected to v_j via a configuration A on E_{m+1}^ψ, and define $\mathcal{A}^\chi$ analogously. As in the previous proof, we consider the random cluster measure μ_m^1 on E_{m+1} with wired boundary conditions and the coupling

$$(A^\psi, A^\chi, A^1) \text{ with } A^\psi \sim \mu^\psi(\cdot \mid \mathcal{A}^\psi),\ A^\chi \sim \mu^\chi(\cdot \mid \mathcal{A}^\chi),\ A^1 \sim \mu^1$$

as well as $A^\psi, A^\chi \leq A^1$. Now, let $\Lambda = \Lambda(A^1)$ be the subgraph induced by the set of edges in E_{m+1} where both endpoints are not connected to the vertex u via open edges in A^1. The boundary conditions on Λ induced by ψ are the same as those induced by χ, therefore we can do a coupled resampling of A^ψ and A^χ on Λ according to this law, conditioned on $\mathcal{A}^\psi$ to obtain $B^\psi \sim \mu^\psi(\cdot \mid \mathcal{A}^\psi)$, $B^\chi \mu^\chi(\cdot \mid \mathcal{A}^\chi)$ that agree on Λ.

Doing the coupled spin decision step of the Edwards-Sokal coupling guarantees that the resulting spin configurations σ^ψ and σ^χ agree on all vertices $v \in \Lambda$. Therefore, we have that

$$\|\nu_r^\psi - \nu_r^\chi\|_{TV} \leq \mathbb{P}[\Lambda_r \not\subseteq \Lambda]\,,$$

and this probability decreases exponentially in r. □

Note that the choice to work with the 2-norm on the lattice is arbitrary. Because we are working on $\mathbb{Z}^2$, all vector norms are equivalent with constants only depending on the dimension $d = 2$.

5.1 Spin block dynamics

Using the local SSM property, we can derive a good mixing time bound for the spin block dynamics. The following presentation follows [BCSV19]. We consider the graph $G = (V_{m+1}, E_{m+1})$ with a spin boundary condition $\psi = \{V_1, \ldots, V_q\}$ on vertices $v \in V_{m+1} \setminus V_m$. The parameter m is considered to be very large, and the particular choice of ψ will not have any influence on the arguments to follow. We wish to approximate the q-state Potts model measure ν^ψ using the spin block dynamics Markov chain:

1. For a given $\sigma\colon V_m \to \{1, \ldots q\}$, choose a vertex $u \in V_m$ uniformly at random.
2. Let $B_R(u)$ be the set of vertices $v \in V_m$ that satisfy $\|u - v\|_\infty \leq R$, and let $\nu_{u,R}^{\sigma,\psi}$ be the spin measure on $B_R(u)$ with boundary conditions induced by σ and ψ. Draw a sample τ from this measure and update σ to align with τ on $B_R(u)$.

The sets $B_R(u)$ might be rectangles when u is near the boundary of V_m, though it is easy to see that the local SSM property also holds in this case.

Theorem 5.4 (Mixing time of spin block dynamics)**.** *Let $p < p_c(q)$ and $\epsilon > 0$. The spin block dynamics Markov chain with parameter R satisfies*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((m/R)^2 \log m\right)$$

for some $R = R(p, q)$.

Proof. We will use a path coupling argument to derive the result. On the spin configuration space, we define a metric

$$d(\sigma, \sigma') = |\{v \in V_m \mid \sigma(v) \neq \sigma'(v)\}|\,,$$

i.e. the number of differing spins. With this metric, the maximal distance between two configurations is $|V_m| = (2m+1)^2 \in \mathcal{O}(m^2)$. To use the path coupling argument from Theorem 2.10, we need to define a coupled version of the spin block dynamics on adjacent states, i.e. spin configurations σ, σ' which differ at one specific spin $w \in V_m$. We do this in the following way: In Step 1, the same vertex $u \in V_m$ is

chosen. In Step 2, we consider the three cases $w \in B_R(u)$, $w \in B_{R+1}(u) \setminus B_R(u)$ and $w \notin B_{R+1}(u)$. In the first and third case, the measures $\nu_{u,R}^{\sigma,\psi}$ and $\nu_{u,R}^{\sigma',\psi}$ are exactly the same due to the locality of the model, and we can therefore sample (τ, τ') to be identical. In the second case, w lies on the boundary of $B_R(u)$ and influences the corresponding measures. Let $r > 0$ and $\delta = \delta(r) > 0$ such that the local SSM property holds with

$$\|\nu_r^{\psi} - \nu_r^{\chi}\|_{TV} \leq C \exp(-cr) = \delta \,.$$

This estimate holds independently of m. Thus, we can choose a coupled resampling (τ, τ') such that

$$\mathbb{P}[\tau(v) \neq \tau'(v) \text{ for some } v \text{ with } \|v - w\|_2 > r] \leq \delta \,.$$

Denote with (ω, ω') the result of this procedure applied to (σ, σ'). We proceed to estimate the expected distance. The first case $w \in B_R(u)$ happens with probability smaller or equal to $(2R+1)^2/(2m+1)^2$ and results in $d(\omega, \omega') = 0$, not contributing to our estimation. The third case $w \notin B_{R+1}(u)$ happens with probability smaller or equal to $((2m+1)^2 - R^2)/(2m+1)^2$ and results in $d(\omega, \omega') = 1$. The second case happens with probability smaller or equal to $(8R+4)/(2m+1)^2$ and results in $d(\omega, \omega') \leq \delta(2R+1)^2 + \frac{\pi}{2}(r+1)^2$. Combined we get

$$\mathbb{E}\left[d(\omega, \omega') \mid (\sigma, \sigma')\right] \leq \frac{(2m+1)^2 - R^2}{(2m+1)^2} + \frac{8R+4}{(2m+1)^2}\left(\delta(2R+1)^2 + \frac{\pi}{2}(r+1)^2\right) .$$

Due to the fact that $\delta = \delta(r)$ is decreasing exponentially in r, we can choose R and r such that the second term diminishes, leading to the estimate

$$\mathbb{E}\left[d(\omega, \omega') \mid (\sigma, \sigma')\right] \leq 1 - \alpha \frac{R^2}{(2m+1)^2}$$

for all m big enough and for some α slightly less than 1. This is due to the fact that the probability to erase the distance scales with the area of $B_R(u)$, while the probability to increase the distance by a minor r-dependent term scales with the perimeter of $B_R(u)$. Applying the path coupling theorem gives

$$\tau_{mix}(\epsilon) \leq \log((2m-1)^2) \frac{|\log(\epsilon)|}{|\log\left(1 - \alpha \frac{R^2}{(2m+1)^2}\right)|} \in \mathcal{O}\left((m/R)^2 \log m\right) .$$

□

The R in the asymptotic statement is not necessary, though it aligns intuitively with a heatbath-type estimate if one thinks of a block dynamics step to have constant cost. The spin block dynamics would be realizable using a heatbath algorithm on $B_R(u)$ with a stopping time such that the resulting configuration is almost as drawn from the true measure. This means that using a heatbath algorithm which has a more elaborate strategy of choosing vertices one can achieve a $\mathcal{O}\left(m^2 \log m\right)$ bound for the mixing time, though to show this for the general heatbath algorithm is not so easy.

5.2 The Ising model case

In the Ising model case with $q = 2$, we can use the additional property of spin monotonicity to achieve mixing time bounds. We begin with the following variant of the heatbath dynamics:

1. Let V be the set of eligible vertices at Step 1 of the heatbath dynamics. Choose a random enumeration $v_1, \dots v_M$ of V, with $M = |V|$.

2. Apply a total of M heatbath steps, where at time t, the vertex v_t is chosen at Step 1.

We call this variant *heatbath sweep dynamics* and think of it as a global Markov chain. The grand monotone coupling is evident: Choose the random enumeration at Step 1 independently of the state of the chain, and do monotone coupled heatbath steps. The following proof roughly follows the argumentation in Theorem 5.1 of [BS16]. Also note that the statements derived throughout the rest of this section are covered by the censoring framework results given in [BCV18].

Theorem 5.5 (Heatbath sweep dynamics for the Ising model)**.** *Let $q = 2$, $p < p_c(q)$ and $\epsilon > 0$. The heatbath sweep dynamics Markov chain on V_m with an arbitrary boundary condition ψ on $V_{m+1} \setminus V_m$ satisfies*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

Proof. For a fixed boundary condition ψ, we have two extremal states σ^0 and τ^0 with

$$\forall v \in V_m \colon \sigma^0(v) = 1,\ \tau^0(v) = 2.$$

The heatbath sweep dynamics can be coupled to conserve monotonicity such that we obtain Markov chains $(\sigma^t)_{t\geq 0}$ and $(\tau^t)_{t\geq 0}$ with

$$\forall t \geq 0 \colon \sigma^t \leq \tau^t .$$

Let $u \in V_m$ be fixed and consider the domain $B_R(u)$ for some $R > 0$. We introduce auxiliary Markov chains $(\sigma_u^t)_{t\geq 0}$ and $(\tau_u^t)_{t\geq 0}$ such that $\sigma_u^0 = \sigma^0$, $\tau_u^0 = \tau^0$ and heathbath step updates for these chains only happen if the chosen vertex at Step 1 of the heatbath dynamics is inside $B_R(u)$. This means that while σ^t and τ^t are doing heatbath sweep steps on V_m, the Markov chains σ_u^t and τ_u^t are doing heatbath sweep steps on $B_R(u)$. If we couple all chains together, we achieve

$$\sigma_u^t \leq \sigma^t \leq \tau^t \leq \tau_u^t$$

for all $t \geq 0$. Therefore, $\sigma_u^t(u) = \tau_u^t(u)$ implies $\sigma^t(u) = \tau^t(u)$. The Markov chains σ_u^t and τ_u^t converge to the q-state Potts model measure on $B_R(u)$ with corresponding boundary conditions, and the influence of these boundary conditions will reach the spin at u with a very low probability. Call these measures $\nu_{u,R}^{\psi,1}$ and $\nu_{u,R}^{\psi,2}$. Let

$T_{u,R} = T_{u,R}(\delta)$ be the mixing time with respect to $\delta = \frac{1}{4}$ of the heatbath sweep dynamics on $B_R(u)$. For fixed R, the number of shapes and boundary conditions $B_R(u)$ can have is finite and independent of m, so we define

$$T_R = \max_u \{T_{u,R}\}\,.$$

After $t = T_R$ heatbath sweep steps, we have that

$$\begin{aligned}
\mathbb{P}[\sigma^t(u) \neq \tau^t(u)] \leq\ & \mathbb{P}[\sigma_u^t(u) \neq \tau_u^t(u)] \\
=\ & \mathbb{P}[\sigma_u^t(u) = 1 \text{ and } \tau_u^t(u) = 2] \\
=\ & \mathbb{P}[\tau_u^t(u) = 2] - \mathbb{P}[\sigma_u^t(u) = 2] \\
\leq\ & \left|\mathbb{P}[\tau_u^t(u) = 2] - \nu_{u,R}^{\psi,2}\left[\text{the spin at } u \text{ is } 2\right]\right| \\
& + \left|\nu_{u,R}^{\psi,2}\left[\text{the spin at } u \text{ is } 2\right] - \nu_{u,R}^{\psi,1}\left[\text{the spin at } u \text{ is } 2\right]\right| \\
& + \left|\nu_{u,R}^{\psi,1}\left[\text{the spin at } u \text{ is } 2\right] - \mathbb{P}[\sigma_u^t(u) = 2]\right|\,.
\end{aligned}$$

The first and third term in the final estimate describe how close the auxiliary distributions are to equilibrium, and per definition of the mixing time T_R they are bounded by $\delta = \frac{1}{4}$. The second term describes the difference of the boundary distributions at u. For any u, differing boundary vertices have distance at least R to u, and the number of differing boundary vertices is at most $(2R+1)^2$, therefore the local SSM property together with a crude triangle inequality estimation gives

$$\left|\nu_{u,R}^{\psi,2}\left[\text{the spin at } u \text{ is } 2\right] - \nu_{u,R}^{\psi,1}\left[\text{the spin at } u \text{ is } 2\right]\right| \leq C(2R+1)^2 \exp(-cR)\,.$$

Now, let $M = |V_m| = (2m+1)^2$ and $R = b\log(M) \in \mathbb{N}$ with a constant $b > \frac{1}{c}$, for a fixed but large m. Then the above bound reads

$$C(2b\log(M)+1)^2 M^{-bc} < \frac{1}{6}M^{-1}$$

for m big enough. Therefore after $t = T_R$ steps, we have

$$\mathbb{P}[\sigma^t(u) \neq \tau^t(u)] \leq 2\delta + \frac{1}{6}M^{-1}\,.$$

Moreover, we can use Lemma 2.7 and get that the mixing time $T_R(\frac{1}{6}M^{-1})$ is smaller or equal to $\lceil \log_2(6M) \rceil T_R$. So for $t = \lceil \log_2(6M) \rceil T_R$, we get

$$\mathbb{P}[\sigma^t(u) \neq \tau^t(u)] < \frac{1}{2M}\,.$$

Finally, a union bound over the vertices gives

$$\mathbb{E}[\#\text{differing vertex spins of } \sigma^t \text{ and } \tau^t] \leq \sum_{u \in V_m} \mathbb{P}[\sigma^t(u) \neq \tau^t(u)] < \frac{1}{2}\,.$$

Due to the fact that $\sigma^s = \tau^s$ implies $\sigma^t = \tau^t$ for $t \geq s$, we get that

$$\mathbb{P}[\sigma^{nt} \neq \tau^{nt}] \leq \frac{1}{2^n} \leq \epsilon$$

for n big enough, independently of m. Therefore the mixing time of the heatbath sweep algorithm on V_m is bounded from above by

$$n\lceil \log_2(6M)\rceil T_R = n\lceil 2\log_2(M) + \log_2(6)\rceil T_R\,.$$

Now, we can use an induction argument to prove the final estimate. Let $\phi(R)$ be any strictly increasing function in R. We can find a constant $a > 0$ such that for all $R' \leq R$, we have have that $T_{R'} \leq a\phi(R')$. With $R = b\log(M)$ we get that our mixing time bound becomes

$$n\lceil 2\log_2(M) + \log_2(6)\rceil a\phi(b\log(M))\,.$$

If it were true that this expression is dominated by $a\phi(M)$, an induction argument implies that the mixing time of the heatbath sweep dynamics is $\mathcal{O}(\phi(M))$. One readily verifies that for m big enough this is true for $\phi(M) = \log(M)^2$. The result follows from

$$\log(M) = 2\log(2m+1) \leq 8\log m\,.$$

□

The main tool of this proof is the monotone coupling of the global algorithm to local versions of itself, and this is also possible for the standard heathbath algorithm. We have to modify the proof to account for the fact that heatbath dynamics modifies the configuration in a non-homogeneous way due to its random behavior at Step 1.

Corollary 5.6 (Heatbath dynamics for the Ising model)**.** *Let $q = 2$, $p < p_c(q)$ and $\epsilon > 0$. The heatbath dynamics Markov chain on V_m with an arbitrary boundary condition ψ on $V_{m+1} \setminus V_m$ satisfies*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left(m^2(\log m)^2\right)\,.$$

Proof. The proof is mostly the same as in the previous theorem. We define $(\sigma^t)_{t\geq 0}$ to be the heatbath dynamics Markov chain starting in the minimal state σ^0 with ψ boundary conditions, and $(\tau^t)_{t\geq 0}$ to be the heatbath dynamics Markov chain starting in the maximal state τ^0 with ψ boundary conditions. The auxiliary chains $(\sigma_u^t)_{t\geq 0}$ and $(\tau_u^t)_{t\geq 0}$ are defined analogously on $B_R(u)$, only realizing an update if the chosen vertex of the global heatbath dynamics is in $B_R(u)$. These chains can again be coupled together to achieve

$$\sigma_u^t \leq \sigma^t \leq \tau^t \leq \tau_u^t$$

for all $t \geq 0$. As above, we can define T_R to be the maximal mixing time with respect to $\delta = \frac{1}{4}$ of the heatbath dynamics on domains $B_R(u)$ with arbitrary boundary conditions. Let $U(u, R, t)$ be the number of times the heatbath dynamics chooses a vertex in $B_R(u)$ out of t times. Then $U(u, R, t)$ has a binomial distribution $B(n, p)$ with parameters $n = t$ and $p \geq (R+1)^2/M$. If we take

$$t = t(R, m) = a\log(6M)T_R M/(R+1)^2 \in \mathbb{N}$$

with $a > 1$, we get

$$\mathbb{E}[U(u, R, t)] \geq a \log(6M) T_R \,.$$

Therefore a Chernoff bound gives

$$\mathbb{P}[U(u, R, t) < \log(6M) T_R] \leq \frac{1}{6M} \,.$$

We again use the estimate

$$\begin{aligned} \mathbb{P}[\sigma^t(u) \neq \tau^t(u)] \leq \quad & \left| \mathbb{P}[\tau_u^t(u) = 2] - \nu_{u,R}^{\psi,2} [\text{the spin at } u \text{ is } 2] \right| \\ & + \left| \nu_{u,R}^{\psi,2} [\text{the spin at } u \text{ is } 2] - \nu_{u,R}^{\psi,1} [\text{the spin at } u \text{ is } 2] \right| \\ & + \left| \nu_{u,R}^{\psi,1} [\text{the spin at } u \text{ is } 2] - \mathbb{P}[\sigma_u^t(u) = 2] \right| \end{aligned}$$

and see that for $t = t(R, m)$ the first and third summand are bound by

$$\frac{1}{6M} + \delta' = \frac{1}{3M} \,,$$

where we obtain $\delta' = \frac{1}{6M}$ by using Lemma 2.7 with $t = t(R, m)$. Now, same as in the above theorem, we take $R = b \log(M)$ with $b > \frac{1}{c}$ to bound the second term by $\frac{1}{6M}$ for M big enough, and obtain the overall bound

$$\mathbb{P}[\sigma^t(u) \neq \tau^t(u)] < \frac{5}{6M} \,,$$

and a union bound over the vertices gives

$$\mathbb{E}[\#\text{differing vertex spins of } \sigma^t \text{ and } \tau^t] \leq \frac{5}{6} \,.$$

This gives

$$\mathbb{P}[\sigma^{nt} \neq \tau^{nt}] < \epsilon$$

for some n big enough, independently of m. Therefore we get the mixing time bound

$$\tau_{mix}(\epsilon) \leq na \log(6M) T_R M / (R+1)^2 \,.$$

Using the same induction argument as above, we consider an increasing function $\phi(R)$ and have that there exists a constant A such that for all $R' \leq R$, we get $T_{R'} \leq A\phi(R')$. Plugging in $R = b \log(M)$, we see that our mixing time is $\mathcal{O}(\phi(M))$ if

$$na \log(6M) A \phi(b \log(M)) \frac{M}{b^2 \log(M)^2} \leq A \phi(M)$$

holds. This is true for $\phi(M) = M \log(M)^2$ and m big enough, resulting in a mixing time

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left(m^2 (\log m)^2\right) \,.$$

□

The proof is essentially the same, with an additional use of a Chernoff bound to ensure that enough heatbath updates happen in the subareas $B_R(u)$. Let us consider the heatbath sweep dynamics again. For the proofs of Theorem 5.5 and Corollary 5.6 we used the fact that the global algorithm reduces to local versions of itself on subdomains. This is also true for the alternate scan dynamics with an even-odd partition of the vertices, as well as for the systematic scan dynamics where the vertices are traversed in a fixed order from top-left to bottom-right. We obtain

Corollary 5.7 (Alternate scan and systematic scan dynamics for the Ising model)**.** *Let $q = 2$, $p < p_c(q)$ and $\epsilon > 0$. The alternate scan dynamics Markov chain on V_m with an arbitrary boundary condition ψ on $V_{m+1} \setminus V_m$ satisfies*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

The same holds for the systematic scan dynamics Markov chain.

We omit the proof here because it is exactly the same as in Theorem 5.5. Last but not least, the monotone Swendsen-Wang dynamics also allows for a local-global coupling and the induction argument to obtain

Corollary 5.8 (Monotone Swendsen-Wang for the Ising model)**.** *Let $q = 2$, $p < p_c(q)$ and $\epsilon > 0$. The monotone Swendsen-Wang Markov chain on V_m with an arbitrary boundary condition ψ on $V_{m+1} \setminus V_m$ satisfies*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

As we have seen, the Ising model at subcritical temperature allows for a thorough treatment of Markov chains, due to the spin monotonicity property and the local SSM property. Let us mention that in the above mixing time bounds, the logarithmic powers are not optimal. Of course one can find more precise asymptotics $\phi(M)$ in the last argument of the proofs, we will see however that the stated bounds are good enough to use a bootstrap argument, obtaining an even more precise bound. This argument will not depend on the spin monotonicity property, therefore we introduce it at a later point.

5.3 The Fortuin-Kasteleyn random cluster setting

We see from the proof for the local SSM property that a change of the spin boundary condition at one vertex induces an edge boundary condition change which does not have a long distance effect on the resulting random cluster measure. This is not always the case. For instance, adding an edge between two special vertices v_i and v_j can have an effect at different regions of the boundary. We can make a similar statement for edge boundary conditions that are realized in $\mathbb{Z}^2$. Consider $C, D \subseteq E_m^c$ that differ only at one edge $e \in E_m^c$. Then without loss of generality we can say

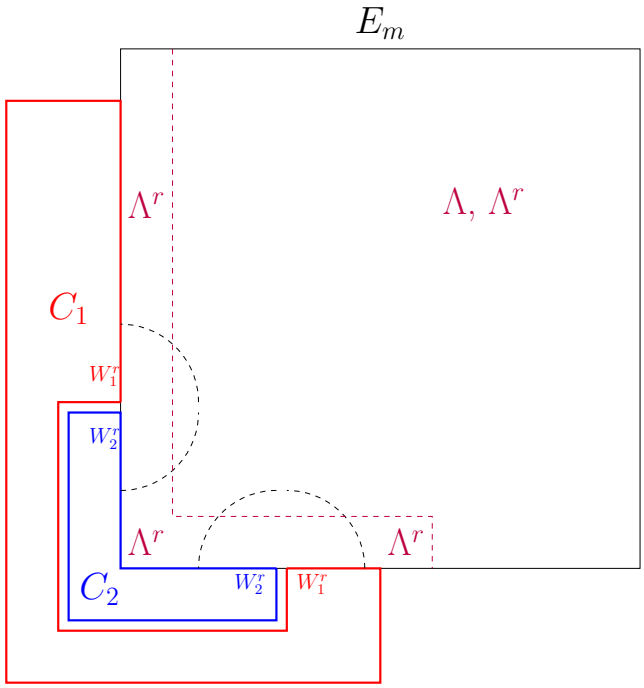

Figure 5.3: The boundary situation for boundary conditions differing at one pivotal edge. The straight dashed line roughly describes the influence of ∂C_1 and ∂C_2 on the distribution, or equivalently the approximate boundary of Λ. The curved dashed lines describe the approximate boundary of Λ^r. The set $\Lambda^r \setminus \Lambda$ decomposes into disconnected regions with high probability, and these regions connect to either ∂C_1 or ∂C_2.

$C \leq D$ and define a coupling (A, B) with $A \sim \mu^{E_m^c, C}$, $B \sim \mu^{E_m^c, D}$ and $A \leq B$. We think of A, B as configurations on E_m. As above, we can define Λ to be the subgraph of E_m of all edges and vertices that do not connect to e via an open path in B, and resample A and B on this graph to be equal. However, this will only give us an estimate as in Lemma 5.2. We need to make an effort to precisely localize the effect of the differing boundary conditions. First off, $\mu^{E_m^c, C}$ and $\mu^{E_m^c, D}$ only differ if e is pivotal to D and the open clusters in C corresponding to both endpoints of e connect to at least one vertex on the boundary of E_m. Call these clusters C_1 and C_2, and define $\partial C_1 = C_1 \cap E_m$, $\partial C_2 = C_2 \cap E_m$ to be the vertices of E_m that belong to these clusters. For a given $r > 0$, let W_1^r be the set of vertices w in ∂C_1 such that a vertex v in ∂C_2 with $\|w - v\|_2 \leq r$ exists and define W_2^r analogously. Due to the fact that C_1 and C_2 are embedded in $\mathbb{Z}^2$ and do not intersect each other, the cardinality of these sets can be bound independently of m for fixed r, and these vertices will be around at most two regions of the boundary of E_m, see Figure 5.3. Now, let Λ^r be the graph of edges and vertices in E_m that do not connect to a vertex in W_1^r or W_2^r via an open path in B. The distribution on Λ^r is not necessarily independent of A and B, therefore we cannot do a resampling on Λ^r. We can however resample on Λ first, and consider the conditional measures on $\Lambda^r \setminus \Lambda$ after this resampling. For $p < p_c(q)$, this set should consist of disconnected regions around the boundary

of E_m which are in the vicinity of at most one of the sets W_1^r or W_2^r, and therefore the distributions on these regions can also be resampled. We arrive at

Lemma 5.9 (Local change for realizable boundary conditions in $\mathbb{Z}^2$)**.** *Let $p < p_c(q)$. Let $C, D \subseteq E_m^c \subseteq \mathbb{Z}^2$ be edge boundary conditions that differ at only one edge $e \in E_m^c$. Define W_1^r, W_2^r, Λ and Λ^r as above. Let $L^r \subseteq E_m$ be the set edges where both endpoints have distance at least r to all vertices in W_1^r and W_2^r. Let μ_r^C be the marginal measure of μ_m^{C,E_m^c} on L^r, analogously μ_r^D. Then we have*

$$\|\mu_r^C - \mu_r^D\|_{TV} \leq c_1 \exp(-cr)$$

with constants $c_1, c > 0$ independent of r, C, D.

Proof. As already seen, we can define a coupling (A, B) with $A \sim \mu_m^{C,E_m^c}$, $B \sim \mu_m^{D,E_m^c}$ and $A \leq B$. We can furthermore modify this coupling such that A and B align on Λ. Now, consider a connected component of $\Lambda^r \setminus \Lambda$, and let K be the subgraph of this component after removing all edges that are closed in B and have one endpoint in Λ. On the region K we can do a joint resampling of A and B if it connects to at most one of the sets ∂C_1 and ∂C_2. Because K does not connect to W_1^r or W_2^r, the distance between $\partial C_1 \setminus W_1^r$ and $\partial C_2 \setminus W_2^r$ is at least r though, so this happens with a probability that decays exponentially in r due to Theorem 3.13. The number of such components is at most three, see Figure 5.3. Finally, the probability that L^r is not contained in Λ^r decays also exponentially in r, so we get a coupling (A, B) that differs on L^r with a probability that decays exponentially in r:

$$\begin{aligned}
\|\mu_r^C - \mu_r^D\|_{TV} \leq& \mathbb{P}[A \neq B \text{ on } L^r] \\
\leq& \mathbb{P}[L^r \not\subseteq \Lambda^r \text{ or } (\partial C_1 \setminus W_1^r) \leftrightarrow (\partial C_2 \setminus W_2^r) \text{ in } B] \\
\leq& \mathbb{P}[L^r \not\subseteq \Lambda^r] + \mathbb{P}[(\partial C_1 \setminus W_1^r) \leftrightarrow (\partial C_2 \setminus W_2^r) \text{ in } B] \\
=& \mathbb{P}[L^r \leftrightarrow (W_1^r \cup W_2^r)] + \mathbb{P}[(\partial C_1 \setminus W_1^r) \leftrightarrow (\partial C_2 \setminus W_2^r) \text{ in } B]
\end{aligned}$$

and as in the proof of Lemma 5.1 we get that both summands are bound by $c_1 \exp(-cr)$ for some constants $c_1, c > 0$ independent of A, B, C, D, r. □

This result can be compared to the local SSM property in the spin system setting, with the difference that any differing edge $e \in E_m^c$ can have an influence on the resulting marginal measures on E_m. We will use this property to prove a mixing time bound for the block dynamics in the Fortuin-Kasteleyn random cluster setting. Let us reconfirm the setting: We consider edge configurations $A \subseteq E_m$ and some boundary condition $C \subseteq E_m^c$, together with the model $\mu = \mu_m^{C,E_m^c}$. We introduce a specific instance of the block dynamics on this model:

1. Choose $u \in V_m$ uniformly at random.
2. Let $E_R(u)$ be the set of edges $e \in E_m$ such that the endpoints x, y satisfy $\|u - x\|_\infty \leq R$, $\|u - y\|_\infty \leq R$. Resample the given configuration A on the set $E_R(u)$ with respect to the conditional measure

$$\mu_{u,R} = \mu^{A \cup C, E_R(u)^c}$$

that is given by the boundary conditions on $E_R(u)^c$ induced by A and C. The resulting configuration A' aligns with A on $E_m \setminus E_R(u)$.

We wish to apply the path coupling theorem, therefore we define the distance function

$$d(A, B) = |A \cup B| - |A \cap B|$$

on edge configurations, which is the number of differing edges between A and B. Let $A \leq B$ be edge configurations that differ at exactly one edge $e \in E_m$. We define the following coupled version of the block dynamics:

1. Choose $u \in V_m$ uniformly at random.

2. Let $E_R(u)$ be the set of edges $e \in E_m$ such that the endpoints x, y satisfy $\|u - x\|_\infty \leq R$, $\|u - y\|_\infty \leq R$. If the conditional measures $\mu_{u,R}$ with respect to A and B are equal, then do a coupled resampling on $E_R(u)$ such that $A' = B'$ on $E_R(u)$. If the conditional measures are different, do a coupled resampling such that

$$\mathbb{P}[A \neq B \text{ on } L^r = L^r(E_R(u), A, B, C)] \leq C \exp(-cr)$$

and $A' \leq B'$ holds as is possible due to Lemma 5.9.

This coupling for adjacent states A, B is similar to the spin block dynamics coupling as it restricts the influence of the differing edge in A and B on the resulting states A' and B'. To apply the path coupling theorem, we need to be in the situation that the differing edge e needs to be close to $E_R(u)$ to have an influence on the conditional measures. If we assume that $p < p_c(q)$ and B has law μ^{C,E_m^c}, the exponential decay of correlations property given by Theorem 3.13 allows us to bound the number of edges that can have an influence on the conditional measures on $E_R(u)$ with high probability. Note that however, the exponential decay property holds for B' on the subset $E_R(u)$ without putting any restrictions on B, due to the fact that B' restricted to $E_R(u)$ follows a random cluster model distribution on $E_R(u)$ with some arbitrary boundary condition. Therefore, after enough block dynamics steps, the exponential decay of correlations property should hold on E_m.

Lemma 5.10. *Let $p > p_c(q)$. Let $B^0 \subseteq E_m$ be an initial state to the block dynamics Markov chain $(B^t)_{t\geq 0}$ with parameter R. Assume that at time T, every vertex pair $v, w \in V_m$ with $\|v - w\|_2 \leq R$ has been contained in $B_R(u)$ at least once for some $u \in V_m$. Then for all $v, w \in V_m$ with $\|v - w\|_2 \leq R$ one has*

$$\mathbb{P}\left[v \leftrightarrow w \text{ via an open path in } B^T\right] \leq C(p, q) \exp(-c(p, q)\|v - w\|_2)\,.$$

Proof. Let u^t be the vertex chosen at time t of the block dynamics, and consider the marginal distribution of B^t on $E_R(u^t)$. Then for all $v, w \in B_R(u^t)$ it holds

$$\mathbb{P}\left[v \leftrightarrow w \text{ via an open path in } B^t\right] \leq C(p, q) \exp(-c(p, q)\|v - w\|_2)\,,$$

for arbitrary boundary conditions induced by B^{t-1}. Let $D = E_m$ be the maximal state and let D' be the resulting state after applying the block dynamics Markov chain. Due to $B^{t-1} \leq D$ we can couple the block dynamics to ensure $B^t \leq D'$ by choosing the same vertex u^t at Step 1 and doing a coupled resampling. Let $t_1 > t$ be the first successive timestep where $B_R(u^t) \cap B_R(u^{t_1}) \neq \emptyset$. We couple D'' to B^{t_1} in the following way: For $t < s < t_1$, apply the block dynamics step to B^{s-1} and do nothing with D'. For $s = t_1$, first resample B^{t_1-1} on $B_R(u^{t_1}) \setminus B_R(u^t)$ with respect to the marginal measure with B^{t_1-1}-boundary conditions on $B_R(u^{t_1})$. Then, do a coupled resampling on $E_R(u^t) \cap E_R(u^{t-1})$ such that $B^{t_1} \leq D''$. One readily verifies that this procedure is a valid coupling that reproduces the distribution of B^{t_1} correctly. Moreover, D'' has the same distribution as D' and therefore has exponential decay of correlations on $B_R(u^t)$. Hence B^{t_1} also has this property, because $B^{t_1} \leq D''$. We have shown that exponential decay of correlations between two vertices v, w, once established by the block dynamics, is retained for successive iterations. If all vertex pairs v, w with $\|v - w\|_2 \leq R$ fall into $B_R(u^t)$ for some $t \leq T$, exponential decay of correlations holds up to distance R under the law of B^T. □

We can now show a mixing time bound for the block dynamics, in case of free or wired boundary conditions.

Theorem 5.11. *Consider the random cluster model on E_m with free or wired boundary conditions, and let $p < p_c(q)$. The block dynamics with parameter R has a mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((m^2/R^2)\log m\right)$$

for R big enough.

Proof. Let $A^0 = \emptyset$ and $B^0 = E_m$ be the extremal states and consider the coupled block dynamics Markov chain (A^t, B^t) such that $A^t \leq B^t$ for all $t \geq 0$. Let u^t be the vertex chosen in Step 1 of the block dynamics at time t. Let $v, w \in V_m$ with $\|v - w\|_2 \leq R$. Then the probability that both v, w are in $B_R(u^t)$ during a block dynamics step is greater or equal to R^2/M, with $M = |V_m| = (2m+1)^2$. After $T = a\log(M)M/R^2 \in \mathbb{N}$ block dynamics steps with $a > 2$, a Chernoff bound implies that

$$\mathbb{P}[v, w \text{ are not in } B_R(u^t) \text{ for any } t \leq T] \leq \frac{1}{M^2}\,.$$

The number of such pairs scales linearly in M and R^2, therefore we get for M big enough that after T steps, the law of B^T has exponential decay of correlations up to distance R due to the previous Lemma, with a probability tending to 1 for increasing M. This property is conserved for $t \geq T$, so we may assume that in the following, we have

$$\mathbb{P}\left[v \leftrightarrow w \text{ via an open path in } B^t\right] \leq C(p,q)\exp(-c(p,q)\|v - w\|_2)\,.$$

Now, assume that $C \leq D \leq B^t$ such that C and D differ at one edge $e \in E_m$. We apply the special coupling introduced above to (C, D) and want to show that the expected distance is smaller than 1. Similar to the spin block dynamics we have to

consider how the chosen vertex u and e may relate to each other. For wired and free boundary conditions, a differing edge can only have an influence on the conditional measures at $E_R(u)$ if it is connected to $E_R(u)$ via a path in E_m. With probability at most $(2R+1)^2/M$, we have $e \in E_R(u)$ and $d(C', D') = 0$ follows. With probability at most

$$\frac{M - R^2 - 2(r-1)R}{M},$$

the distance of both endpoints of e to all vertices in $B_R(u)$ is bigger or equal to r. In this case, the probability that one endpoint of e connects to $B_R(u)$ via an open path in B^t is smaller than $C\exp(-cr)$, and therefore the differing edge will influence the conditional measures on $E_R(u)$ with very low probability. With probability at most $4(r+1)(2R+1+r)/M$ one endpoint of e has distance less or equal to r to the set $E_R(u)$. In this case, we get a potential influence of e on the conditional measures on $E_R(u)$, however this influence is bounded due to our choice of the coupling and Lemma 5.9. In the second or third case, if the edge has influence on the conditional measures, it is bounded to $E_R(u) \setminus L^r$ with high probability, and we roughly bound the number of edges in this area by $8r^2$. It follows

$$\begin{aligned} & \mathbb{E}\left[d(C', D') \mid C, D\right] \\ \leq \quad & \frac{M - R^2 - 2(r-1)R}{M}\left(1 + C\exp(-cr)\left(8r^2 + (2R+1)^2 C\exp(-cr)\right)\right) \\ & + \frac{4(r+1)(2R+1+r)}{M}\left(8r^2 + (2R+1)^2 C\exp(-cr)\right). \end{aligned}$$

As in Lemma 5.4, we see that for R big enough and r chosen appropriately (for instance $r = \log(6R)/c$), we get that

$$\mathbb{E}\left[d(C', D') \mid C, D\right] \leq \frac{M - \alpha R^2}{M}$$

with some $\alpha < 1$ and R big enough. This means that the path coupling theorem applies after a burn-in time of $T = a\log(M)M/R^2$ steps, resulting in a mixing time of

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((M/R^2)\log(M)\right)$$

overall, and translating to m gives

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((m^2/R^2)\log m\right).$$

□

Note that it is crucial for this proof that the boundary conditions do not allow connectivity information to be distributed by the boundary. This kind of proof also works for side-homogeneous boundary conditions introduced in [BS16], as well as for boundary conditions drawn according to the infinite-volume measure μ in the subcritical regime. We proceed to prove a mixing time bound for the Glauber dynamics, in the case of free or wired boundary conditions. It has been introduced in [BS16] and we have already seen the proof techniques in Theorem 5.6. We will however need a small modification to Lemma 5.1 in the case of side-homogeneous boundary conditions.

Lemma 5.12. *Let $p < p_c(q)$ and $G = (V, E)$ be a rectangular subgraph of $\mathbb{Z}^2$, and let $C \subseteq D \subseteq \mathbb{Z}^2 \setminus E$ be two boundary conditions that satisfy the following property: There is exactly one open cluster that contains more than one vertex, and the set of vertices in V belonging to this cluster constitute a connected component of the boundary of G, see Figure 5.4. Let $\Delta \subseteq V$ be the set of vertices in V that belong to the unique open cluster in D, but not in C. Let L_r be the subset of edges in E where both endpoints v_1, v_2 satisfy*

$$\|w - v_i\|_2 > r \text{ for all } w \in \Delta\,.$$

For μ_r^C, μ_r^D, which are the marginal random cluster measures on L^r with boundary condition C or D respectively, one has

$$\|\mu_r^C - \mu_r^D\|_{TV} \leq C_1 \exp(-cr)$$

with constants $C_1, c > 0$ that only depend on p, q.

Proof. We consider a coupling (A, B) of configurations on G that satisfies

$$A \sim \mu^C,\ B \sim \mu^D,\ A \leq B$$

as is possible due to Theorem 3.4. Let $\Lambda = \Lambda(B)$ be the set of edges where both endpoints are not connected to any vertex in Δ via B. We can do a coupled resampling on Λ because (A, C) and (B, D) induce the same boundary conditions on Λ. We get a coupling (A', B') with

$$A' \sim \mu^C,\ B' \sim \mu^D,\ A' \leq B',\ A' = B' \text{ on } \Lambda\,,$$

and from this coupling it follows

$$\|\mu_r^C - \mu_r^D\|_{TV} \leq \mathbb{P}[L^r \not\subseteq \Lambda]\,.$$

This probability decays exponentially in r. □

Theorem 5.13 (Mixing time of the Glauber dynamics for free and wired boundary conditions)**.** *Let $p < p_c(q)$. For the random cluster model on E_m with free or wired boundary conditions, the Glauber dynamics Markov chain has mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left(m^2(\log m)^2\right)\,.$$

Proof. Let $A^0 = \emptyset$ and $B^0 = E_m$ be the extremal states and let $(A^t, B^t)_{t \geq 0}$ be the coupled monotone Glauber dynamics process. Let $u \in V_m$ be a vertex and consider the box $E_R(u)$ of all edges $e \in E_m$ with endpoints v that satisfy $\|u - v\|_\infty \leq R$. We introduce auxiliary Markov chains A_u^t and B_u^t with initial conditions $A_u^0 = A^0$ and $B_u^0 = B^0$, coupled to (A^t, B^t) as follows: If the edge chosen in Step 1 of

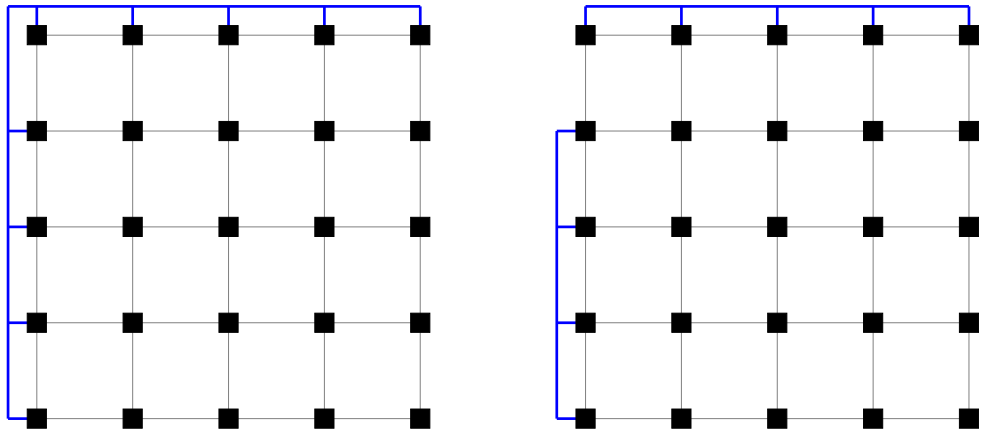

Figure 5.4: Realizable boundary conditions in the Fortuin-Kasteleyn random cluster model. On the left, the prerequisites of Lemma 5.12 are satisfied, but not on the right.

the Glauber dynamics is in $E_R(u)$, perform a coupled Glauber dynamics update, otherwise do nothing. The auxiliary chains are Glauber dynamics processes on $E_R(u)$ with side-homogeneous boundary conditions, i.e. connectivity information cannot travel across the boundary of $E_R(u)$. Moreover, the coupling implies

$$A_u^t \leq A^t \leq B^t \leq B_u^t$$

for all $t \geq 0$. Let $\mu_{u,R}^0$ be the random cluster measure on $E_R(u)$ induced by A^0 and $\mu_{u,R}^1$ the random cluster measure on $E_R(u)$ induced by B^0. After t steps, one has for any edge $e \in E_m$

$$\begin{aligned} & \mathbb{P}[e \in B^t \text{ and } e \notin A^t] \\ \leq\ & \mathbb{P}[e \in B_u^t \text{ and } e \notin A_u^t] \\ =\ & \mathbb{P}[e \in B_u^t] - \mathbb{P}[e \in A_u^t] \\ \leq\ & \left|\mathbb{P}[e \in B_u^t] - \mu_{u,R}^1[e \text{ is open}]\right| \\ & + \left|\mu_{u,R}^1[e \text{ is open}] - \mu_{u,R}^0[e \text{ is open}]\right| \\ & + \left|\mu_{u,R}^0[e \text{ is open}] - \mathbb{P}[e \in A_u^t]\right| . \end{aligned}$$

For a given R, let T_R be the maximal mixing time (with respect to $\delta = \frac{1}{4}$) of the Glauber dynamics on any $E_R(u)$ with side-homogeneous boundary conditions. The box $E_R(u)$ contains at least $2(R-1)^2$ edges, therefore a Chernoff bound argument implies that after $t = a\log(6M)T_R M/(2(R-1)^2) \in \mathbb{N}$ global updates with $M = |E_m|$ and some $a > 2$, the number of updates in $E_R(u)$ is at least $\log(6M)T_R$ with probability at least $\frac{1}{6M}$. Using this together with Lemma 2.7, we can bound the first and third term in the above estimate by $\frac{1}{3M}$. For the second term, we notice that $\mu_{u,R}^0$ and $\mu_{u,R}^1$ satisfy the prerequisites of Lemma 5.12. Set $R = b\log(M)$ with $b > 1/c$, where $c > 0$ is the exponential decay constant in Lemma 5.12. Edges e with u as an endpoint have distance $R-1$ to the set Δ given by the difference of

the boundary conditions for $\mu^0_{u,R}$ and $\mu^1_{u,R}$. Therefore we get

$$\left|\mu^1_{u,R}[e \text{ is open}] - \mu^0_{u,R}[e \text{ is open}]\right| \leq C\exp(-c(R-1)) \leq CM^{-bc} \leq \frac{1}{6}M^{-1}$$

for m big enough. Overall, we get that for edges e adjacent to u, after $t = a\log(6M)T_R M/(2(R-1)^2)$ Glauber dynamics steps, one has

$$\mathbb{P}[e \in B^t \text{ and } e \notin A^t] \leq \frac{1}{2M},$$

and therefore a union bound over the edges gives

$$\mathbb{E}\left[|B^t \setminus A^t|\right] \leq \frac{1}{2}.$$

This implies

$$\mathbb{P}[A^t \neq B^t] \leq \frac{1}{2},$$

and the fact that the coupling (A^t, B^t) is nondiverging implies

$$\mathbb{P}[A^{nt} \neq B^{nt}] \leq \frac{1}{2^n} \leq \epsilon$$

for some n independent of m. Therefore we have $\tau_{mix}(\epsilon) \leq nt$. The rest follows from an induction argument. Let $\phi(R)$ be a strictly increasing function in R and let a' be a constant such that for all $R' \leq R$, one has $T_{R'} \leq a'\phi(R')$. With $R = b\log(M)$ we get a mixing time bound

$$a\log(6M)M/(2(b\log(M)-1)^2)a'\phi(b\log(M))$$

and if this expression is dominated by $a'\phi(M)$, we get that $\tau_{mix}(\epsilon) \in \mathcal{O}(\phi(M))$. This is true for $\phi(M) = M\log(M)^2$, and translating this to m we get

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((m^2/R^2)(\log m)^2\right).$$

□

Note that this result also holds for $p > p_c(q)$ because the Glauber dynamics Markov chain can also be interpreted as Glauber dynamics on the dual model, see [BS16] for the details.

5.4 Speed of disagreement percolation

Consider the Ising model on some graph $G = (V, E)$ and two configurations $\sigma^0 \leq \tau^0$ that only differ in one vertex $v \in V$. Using a monotone coupled Markov chain, one can ask for upper bounds on $\left|\{v\colon \sigma^t(v) \neq \tau^t(v)\}\right|$ that only depend on G and t. Such bounds are useful as they allow for sharp path coupling arguments. We will improve some mixing time results of the previous subsections. Note that the final mixing time results given here are known, see for instance [BCV18]. See also [DSVW02], which uses the speed of disagreement percolation idea in a very similar way.

Lemma 5.14 (Speed of disagreement for alternate scan dynamics)**.** *Let $G = (V, E)$ be a graph and $d(v, w) = \min\{n - 1 \colon e_1, \dots, e_n$ is a path from v to $w\}$. Let σ^0 and τ^0 be Ising spin configurations on V that differ at one vertex $u \in V$. Let $V_1, \dots, V_m$ be a partition of V such that the endpoints of any edge $e \in E$ are in different sets V_i. With this partition, the coupled alternate scan dynamics Markov chain (σ^t, τ^t) satisfies*

$$\sigma^t(w) = \tau^t(w) \text{ for all } w \in V \text{ with } \|v - w\|_1 > t\,.$$

Proof. It is easy to see that the coupled alternate scan dynamics can propagate disagreement only to vertices adjacent to disagreeing spins in one timestep. The result follows immediately. □

We see that for alternate scan dynamics, the reach of disagreement is at most linear in the number of timesteps. This result obviously extends to the q-state Potts model case.

Theorem 5.15 (Mixing time of the alternate scan dynamics II)**.** *Let $q = 2$, $p < p_c(q)$ and consider the graph $G = (V_m, E_m)$ with spin boundary conditions ψ on $V_{m+1} \setminus V_m$. The alternate scan dynamics with even/odd partition in the Ising model case has mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}(\log m)\,.$$

Proof. We already know that $\tau_{mix}(\epsilon) \in \mathcal{O}((\log m)^2)$, but for the upcoming argument it suffices to know that $\tau_{mix}(\epsilon)$ is polynomial in $M = (2m + 1)^2$. Let $\sigma^0 \leq \tau^0$ differ at only one vertex $u \in V_m$, and let (σ^t, τ^t) be the coupled monotone alternate scan dynamics Markov chain. Let $R > 0$ and for $v \in V_m$, consider auxiliary Markov chains (σ_v^t, τ_v^t) given by the alternate scan dynamics on $B_R(v)$, starting in in the extremal states and coupled to (σ^t, τ^t) such that at each timestep t, one has

$$\sigma_v^t \leq \sigma^t \leq \tau^t \leq \tau_v^t\,.$$

For a given R, let $T_R = T_R(1/4)$ be the maximal mixing time of the alternate scan dynamics on $B_R(v)$, which is at most polynomial in R by assumption. Let $\Gamma(u, T_R)$ be the set of vertices $w \in V_m$ that satisfy $\|v - w\|_1 \leq T_R$. Due to Lemma 5.14 we know that $|\Gamma(u, T_R)|$ grows like T_R^2, and therefore also grows polynomially in R. After $t = T_R$ timesteps, the disagreement at u can only travel to vertices in $\Gamma(u, T_R)$, and we therefore obtain

$$\mathbb{E}\left[|\{v \in V_m \colon \sigma^t(v) \neq \tau^t(v)\}|\right] \leq \sum_{w \in \Gamma(u, T_R)} \mathbb{P}[\sigma^t(w) \neq \tau^t(w)]\,.$$

Furthermore, we can bound

$$\mathbb{P}[\sigma^{nt}(w) \neq \tau^{nt}(w)] \leq \mathbb{P}[\sigma_w^{nt}(w) \neq \tau_w^{nt}(w)] \leq \frac{2}{4^n} + C(2R + 1)^2 \exp(-cR)\,,$$

where the latter term comes from the local SSM property, see the proof of Theorem 5.5 for details. Therefore we get an overall bound

$$\begin{aligned}&\mathbb{E}\left[|\{v \in V_m \colon \sigma^{nt}(v) \neq \tau^{nt}(v)\}|\right] \\ &\leq |\Gamma(u, nT_R)| \left(\frac{2}{4^n} + C(2R+1)^2 \exp(-cR)\right) .\end{aligned}$$

Choosing $n = R$, we see that this expression is asymptotically

$$\mathcal{O}(\exp(-c'R))$$

for some $c' > 0$. Therefore there exists R big enough such that after $nt = RT_R$ steps, the expected disagreement between σ^{nt} and τ^{nt} is bounded by $\frac{1}{2}$. It is important to note here that R does not depend on m. Interpreting the alterate scan dynamics applied nt times as a contractive Markov chain, we get that

$$\tau_{mix}(\epsilon) \in \mathcal{O}(\log(M) R T_R) = \mathcal{O}(\log m)$$

using the path coupling theorem. □

We see that this proof crucially depends on the fact that $|\Gamma(u, nT_R)|$ grows polynomially in nT_R. For the alternate scan dynamics with even/odd partitioning this is easy to see. For the upcoming algorithms, we need to bound the reach of disagreement in a probabilistic manner. We continue with the heatbath sweep Markov chain.

Lemma 5.16 (Speed of disagreement for heatbath sweep dynamics)**.** *Let $\sigma^0 \leq \tau^0$ be Ising spin configurations on $\mathbb{Z}^2$ that differ in only one vertex $u \in V_m$. Consider the monotone coupling (σ^t, τ^t) of the heatbath sweep dynamics Markov chain on V_m, and let*

$$\Gamma(u, t) = \bigcup_{s=0}^{t} \{v \in V_m \colon \sigma^s(v) \neq \tau^s(v)\}$$

be the set of vertices which are or were disagreeing at some time $s \leq t$. Then we have

$$\mathbb{E}[|\Gamma(u, t)|] \in \mathcal{O}(t^b)$$

for all $b > 2$.

Proof. At time $t - 1$, consider the states

$$\sigma'(v) = \begin{cases} \sigma^{t-1}(v) & v \notin \Gamma(u, t-1) \\ 1 & v \in \Gamma(u, t-1) \end{cases}, \quad \tau'(v) = \begin{cases} \tau^{t-1}(v) & v \notin \Gamma(u, t-1) \\ 2 & v \in \Gamma(u, t-1) \end{cases}.$$

Then we have $\sigma' \leq \sigma^{t-1} \leq \tau^{t-1} \leq \tau'$, and if we apply the heatbath sweep coupling to (σ', τ') to obtain (σ, τ) in a monotone way, we can guarantee

$$\sigma \leq \sigma^t \leq \tau^t \leq \tau .$$

Therefore we have

$$\Gamma(u,t) \subseteq \Gamma \coloneqq \Gamma(u,t-1) \cup \{v \in V_m \colon \sigma(v) \neq \tau(v)\},$$

and this implies

$$\mathbb{P}[v \in \Gamma(u,t) \mid \Gamma(u,t-1)] \leq \mathbb{P}[v \in \Gamma \mid \Gamma(u,t-1)].$$

Let $v_1, \ldots, v_M \in V_m$ be the enumeration of vertices at the t-th timestep that is generated at Step 1 of the heatbath sweep dynamics. If $v \in \Gamma$, there exists a (self-avoiding) path $v_{i(1)}, \ldots, v_{i(n)}$ of adjacent vertices with

$$v_{i(1)} \text{ is adjacent to } \Gamma(u,t-1), v_{i(n)} = v \text{ and } i(1) < \ldots < i(n)$$

for the disagreement to 'travel' from $\Gamma(u,t-1)$ to v. For any such path, the probability that $v_{i(1)}$ to $v_{i(n)}$ are chosen in increasing order is given by

$$\frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{2} = \frac{1}{n!}.$$

Therefore we can bound

$$\begin{aligned} &\mathbb{P}[v \in \Gamma \mid \Gamma(u,t-1)] \\ \leq &\mathbb{P}\left[\exists v_{i(1)}, \ldots, v_{i(n)} = v \text{ path}\right] \\ \leq &\sum_{n \geq d} \sum_{\text{paths of length } n} \frac{1}{n!} \\ \leq &\sum_{n \geq d} \frac{(2n)^2 3^{n-1}}{n!}, \end{aligned}$$

where $(2n)^2$ is a bound for the possible starting positions of paths ending in v, 3^{n-1} is a crude bound for the number of self-avoiding walks of length n starting in a fixed vertex, and

$$d = d(v, \Gamma(u,t-1)) = \min_{w \in \Gamma(u,t-1)} \|v - w\|_1$$

is the path distance of v to $\Gamma(u,t-1)$. We see that for increasing d, $\mathbb{P}[v \in \Gamma \mid \Gamma(u,t-1)]$ decays faster than any exponential (in the asymptotical sense). Therefore we can bound

$$\mathbb{P}[v \in \Gamma \mid \Gamma(u,t-1)] \leq \exp(-cd)$$

with $c > 0$ small enough. Now, let $\Delta(u,t-1)$ be the smallest l^1-ball around u containing $\Gamma(u,t-1)$, i.e.

$$R(t-1) = \max_{v \in \Gamma(u,t-1)} \|u - v\|_1, \ \Delta(u,t-1) = \{v \in V_m \colon \|u - v\|_1 \leq R(t-1)\}.$$

If we assume that $\Gamma(u,t-1) = \Delta(u,t-1)$ holds (which is the maximal set under the condition that $R(t-1)$ stays fixed), we can bound

$$\mathbb{P}[R(t) - R(t-1) \geq r \mid \Delta(u,t-1)] \leq \min\{1, 4(R(t-1)+r)\exp(-cr)\},$$

using the estimate from above and summing over all vertices with distance r from $\Delta(u, t-1)$. Then we consider

$$\begin{aligned}
&\mathbb{E}\left[R(t) - R(t-1) \mid R(t-1)\right] \\
\leq &\mathbb{E}\left[R(t) - R(t-1) \mid \Gamma(u,t-1) = \Delta(u,t-1)\right] \\
\leq &\mathbb{E}\left[\sum_{r\geq 1} \mathbb{1}[\exists v \in \Gamma(u,t) \text{ with } d(v,\Gamma(u,t-1)) = r \,\Big|\, \Gamma(u,t-1) = \Delta(u,t-1)]\right] \\
= &\sum_{r\geq 1} \mathbb{P}[\exists v \in \Gamma(u,t) \text{ with } d(v,\Gamma(u,t-1)) = r \mid \Gamma(u,t-1) = \Delta(u,t-1)] \\
= &\sum_{r\geq 1} \mathbb{P}[R(t) - R(t-1) \geq r \mid \Gamma(u,t-1) = \Delta(u,t-1)] \\
\leq &\sum_{r\geq 1} \min\left\{1, 4(R(t-1)+r)\exp(-cr)\right\} .
\end{aligned}$$

Simplifying this sum further, we see that

$$\begin{aligned}
&\sum_{r\geq 1} \min\left\{1, 4(R(t-1)+r)\exp(-cr)\right\} \\
\leq &\log(4R(t-1))/c + \sum_{s\geq 1} \min\left\{1, \frac{4(R(t-1)+\log(4R(t-1))+s)}{4R(t-1)}\exp(-cs)\right\} \\
\leq &\log(4R(t-1))/c + \sum_{s\geq 1} \min\left\{1, (1+1/c+s)\exp(-cs)\right\} \\
\leq &\log(4R(t-1))/c + A
\end{aligned}$$

with some constant A independent of $R(t-1)$. We have shown that

$$\mathbb{E}[R(t) \mid R(t-1)] \leq R(t-1) + \log(4R(t-1))/c + A\,,$$

which shows that $\mathbb{E}[R(t)] \in \mathcal{O}(t^{b'})$ for all $b' > 1$. The estimate $|\Gamma(u,t)| \leq (2R(t)+1)^2$ gives the result. □

Theorem 5.17 (Mixing time of the heatbath sweep dynamics II)**.** *Let $q = 2$, $p < p_c(q)$ and consider the graph $G = (V_m, E_m)$ with spin boundary conditions ψ on $V_{m+1} \setminus V_m$. The heatbath sweep dynamics in the Ising model case has mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}(\log m)\,.$$

Proof. Let $\sigma^0 \leq \tau^0$ differ at only one vertex $u \in V_m$, and let (σ^t, τ^t) be the coupled monotone heatbath sweep dynamics Markov chain. Let $R > 0$ and for $v \in V_m$, consider auxiliary Markov chains (σ_v^t, τ_v^t) given by the heatbath sweep dynamics on $B_R(v)$, starting in in the extremal states and coupled to (σ^t, τ^t) such that at each timestep t, one has

$$\sigma_v^t \leq \sigma^t \leq \tau^t \leq \tau_v^t\,.$$

For a given R, let $T_R = T_R(1/4)$ be the maximal mixing time of the heatbath sweep dynamics on $B_R(v)$, which is at most polynomial in R by Theorem 5.5. Let $\Gamma(u,t)$ be defined as in Lemma 5.16. We have

$$\begin{aligned}
& \mathbb{E}[|\{v \in V_m \colon \sigma^{nt}(v) \neq \tau^{nt}(v)\}|] \\
\leq & \sum_{v \in V_m} \mathbb{P}[\sigma^{nt}(v) \neq \tau^{nt}(v) \mid v \in \Gamma(u,nt)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
\leq & \sum_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v) \mid v \in \Gamma(u,nt)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
= & \sum_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
\leq & \mathbb{E}[|\Gamma(u,nt)|] \max_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v)]\,.
\end{aligned}$$

The first factor grows polynomially in nt, the second factor can be bounded by $\frac{2}{4^n} + C(2R+1)^2 \exp(-cR)$ for $t = T_R$, see the proof of Theorem 5.5 for details. Choosing $n = R$ and R big enough gives an expected difference less than 1, hence the path coupling theorem gives the result. □

We continue with the heatbath dynamics.

Lemma 5.18 (Speed of disagreement for heatbath dynamics)**.** *Let $\sigma^0 \leq \tau^0$ be Ising spin configurations on $\mathbb{Z}^2$ that differ in only one vertex $u \in V_m$. Consider the monotone coupling (σ^t, τ^t) of the heatbath sweep dynamics Markov chain on V_m, and let*

$$\Gamma(u,t) = \bigcup_{s=0}^{t} \{v \in V_m \colon \sigma^s(v) \neq \tau^s(v)\}$$

be the set of vertices which are or were disagreeing at some time $s \leq t$. With $M = |V_m| = (2m+1)^2$, we have

$$\mathbb{E}[|\Gamma(u,Mt)|] \in \mathcal{O}(t^b)$$

for all $b > 2$.

Proof. As in the previous lemma, we consider the probability that $v \in \Gamma(u,Mt)$, conditioned on $\Gamma(u,M(t-1))$. Let $v_1, \dots v_M$ be the chosen vertices of the upcoming heatbath dynamics. For the disagreement to travel from $\Gamma(u,t-1)$ to v, there needs to be a path $v_{i(1)}, \dots, v_{i(n)}$ of adjacent vertices with $v_{i(1)}$ adjacent to $\Gamma(u,M(t-1))$ and $v_{i(n)} = v$, as well as

$$i(1) < \ldots < i(n)\,.$$

Consider this path to be fixed and let $S = \{v_{i(1)}, \dots, v_{i(n)}\}$ be the set of path vertices. Then the probability that disagreement travels along this path can be bounded by

$$\sum_{k=n}^{M} \mathbb{P}[k \text{ draws from } S]\mathbb{P}[\exists \text{increasing complete subsequence} \mid k \text{ draws from } S]\,.$$

The first factor in this sum comes from the binomial distribution, while the second factor describes the probability that after k ordered draws from S, there is the subsequence $(v_{i(1)}, \ldots, v_{i(n)})$. The second term can be bounded by

$$\binom{k}{n} n^{k-n} n^{-k} = \binom{k}{n} n^{-n} ,$$

where n^{-k} is the total number of ordered k-draws from S and $\binom{k}{n} n^{k-n}$ describes the procedure of taking any n out of k positions, assigning to them $v_{i(1)}, \ldots, v_{i(n)}$ in increasing order and filling the rest up arbitrarily. We arrive at

$$\sum_{k=n}^{M} B(M, n/M, k) \binom{k}{n} n^{-n} ,$$

with $B(M, n/M, k)$ the binomial distribution with parameters M and $p = n/M$. Because M is very large, we approximate with the Poisson distribution $P(n, k) = \exp(-n)n^k/k!$

$$\begin{aligned} &\sum_{k=n}^{M} B(M, n/M, k) \binom{k}{n} n^{-n} \\ \leq &\sum_{k=n}^{M} |B(M, n/M, k) - P(n, k)| \binom{k}{n} n^{-n} + \sum_{k=n}^{M} P(n, k) \binom{k}{n} n^{-n} . \end{aligned}$$

For the second sum, we can take M to infinity and obtain

$$\sum_{k \geq n} P(n, k) \binom{k}{n} n^{-n} = \frac{1}{n!} \exp(-n) \sum_{k \geq n} \frac{n^{k-n}}{(k-n)!} = \frac{1}{n!} .$$

The first sum behaves similar:

$$\begin{aligned} &\sum_{k=n}^{M} |B(M, n/M, k) - P(n, k)| \binom{k}{n} n^{-n} \\ = &\sum_{k=n}^{M} \frac{n^k}{k!} \binom{k}{n} n^{-n} \left| \frac{M!}{(M-k)! M^k} \left(1 - \frac{n}{M}\right)^{M-k} - \exp(-n) \right| \\ \leq &\sum_{k \geq n} \frac{n^k}{k!} \binom{k}{n} n^{-n} (2e^{-n}) \\ = &\frac{2}{n!} , \end{aligned}$$

where the estimate is true for M big enough independently of n. We have shown that after M heatbath steps, the probability of disagreement traveling along a fixed path of length n is bounded by $\frac{3}{n!}$. The rest of the proof is exactly the same as in Lemma 5.16. □

Corollary 5.19 (Mixing time for the heatbath dynamics II)**.** *Let $q = 2$, $p < p_c(q)$ and consider the graph $G = (V_m, E_m)$ with spin boundary conditions ψ on $V_{m+1} \setminus V_m$. The heatbath dynamics in the Ising model case has mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}(M \log(M)) = \mathcal{O}(m^2 \log m)$$

with $M = |V_m| = (2m+1)^2$.

Proof. The proof is almost identical to the case of heatbath sweep dynamics. Let $\sigma^0 \leq \tau^0$ differ at only one vertex $u \in V_m$, and let (σ^t, τ^t) be the coupled monotone heatbath dynamics Markov chain. Let $R > 0$ and for $v \in V_m$, consider auxiliary Markov chains (σ_v^t, τ_v^t) given by the heatbath dynamics on $B_R(v)$, starting in in the extremal states and coupled to (σ^t, τ^t) such that at each timestep t, one has

$$\sigma_v^t \leq \sigma^t \leq \tau^t \leq \tau_v^t \,.$$

For a given R, let $T_R = T_R(1/4)$ be the maximal mixing time of the heatbath dynamics on $B_R(v)$, which is at most polynomial in R by Theorem 5.6. Let $\Gamma(u,t)$ be defined as in Lemma 5.18. We have

$$\begin{aligned}
&\mathbb{E}[|\{v \in V_m \colon \sigma^{nt}(v) \neq \tau^{nt}(v)\}|] \\
\leq &\sum_{v \in V_m} \mathbb{P}[\sigma^{nt}(v) \neq \tau^{nt}(v) \mid v \in \Gamma(u,nt)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
\leq &\sum_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v) \mid v \in \Gamma(u,nt)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
= &\sum_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v)]\, \mathbb{P}[v \in \Gamma(u,nt)] \\
\leq &\mathbb{E}[|\Gamma(u,nt)|] \max_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v)] \,.
\end{aligned}$$

If we set $n = 2RM$, we get that the first factor grows polynomially in R. For the second factor, we see that the expected number of updates in $B_R(v)$ at time nt is given by

$$2RMT_R|B_R(v)|/M = 2RT_R|B_R(v)| \,.$$

Therefore a Chernoff bound gives us that the probability of less than RT_R updates happening in $|B_R(v)|$ decays exponentially in R. If at least RT_R updates do happen in $B_R(v)$ (call this event $U_{R,v}$), we get the standard bound

$$\mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v) | U_{R,v}] \leq \frac{2}{4^R} + C(2R+1)^2 \exp(-cR) \,.$$

Overall, we get that

$$\mathbb{E}[|\Gamma(u,nt)|] \max_{v \in V_m} \mathbb{P}[\sigma_v^{nt}(v) \neq \tau_v^{nt}(v)]$$

decays exponentially in R, implying that for R big enough, the path coupling theorem applies for the heatbath dynamics applied $M2RT_R$ times. Here, R is independent of M, therefore this results in a mixing time

$$\tau_{mix}(\epsilon) \in \mathcal{O}(M \log(M)) = \mathcal{O}(m^2 \log m)\,.$$

□

Finally, we turn to the monotone Swendsen-Wang dynamics Markov chain.

Lemma 5.20 (Speed of disagreement for monotone Swendsen-Wang dynamics)**.** *Let $\sigma^0 \leq \tau^0$ be Ising spin configurations on $\mathbb{Z}^2$ that differ in only one vertex $u \in V_m$. Consider the monotone coupling (σ^t, τ^t) of the monotone Swendsen-Wang dynamics Markov chain on V_m, and let*

$$\Gamma(u,t) = \bigcup_{s=0}^{t} \{v \in V_m \colon \sigma^s(v) \neq \tau^s(v)\}$$

be the set of vertices which are or were disagreeing at some time $s \leq t$. Then we have

$$\mathbb{E}[|\Gamma(u,t)|] \in \mathcal{O}(t^b)$$

for all $b > 2$.

Proof. We consider the probability that $v \in \Gamma(u,t)$ conditioned on $\Gamma(u,t-1)$. This can only happen if the connected component $C(v)$ of v after Step 1 of the monotone Swendsen-Wang dynamics intersects $\Gamma(u,t-1)$, and all proposal spins in $C(v) \setminus \Gamma(u,t-1)$ choose the same spin. The probability for this event is less or equal to $\frac{1}{2^d}$ with $d = \min_{w \in \Gamma(u,t-1)} \|v - w\|_1$. The rest of the proof follows Lemma 5.16. □

We state the implied mixing time result without proof.

Corollary 5.21 (Mixing time for monotone Swendsen-Wang dynamics II)**.** *Let $q = 2$, $p < p_c(q)$ and consider the graph $G = (V_m, E_m)$ with spin boundary conditions ψ on $V_{m+1} \setminus V_m$. The monotone Swendsen-Wang dynamics in the Ising model case has mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}(\log m)\,.$$

5.5 The supercritical phase

Many of the introduced techniques apply in the supercritical phase as well, though we have to adjust our understanding of the mixing time a little bit. As an example, consider the q-state Potts model on $G = (V_m, E_m)$ without boundary conditions, i.e. the edges connecting V_m to $\mathbb{Z}^2 \setminus V_m$ are removed. For $p > p_c(q)$, the random cluster model is in the supercritical phase and therefore tends to states where the majority

of spins have the same color, with a few excitations. Due to the absence of boundary conditions, every spin color $s \in \{1, \dots, q\}$ has equal probability to be the dominating one. This results in a probability distribution with mass concentrated around the q monochromatic states, and local Markov chains like the heatbath dynamics have a hard time transitioning between these regions. We will show in this section that this in not the complete picture. In fact, if we restrict ourselves to studying the dynamics in the vicinity of a single dominating color, mixing takes place at an optimal rate. Using the techniques of this chapter, we will show that for the Ising model, positive mixing time results are possible.

We begin with a reformulation of the exponential decay property, suited for the supercritical case. Let $G = (V_m, E_m)$ be the usual rectangular subgraph of $\mathbb{Z}^2$, with a realizable boundary condition. We consider the random cluster model μ_m on this graph, with $p > p_c(q)$. Then we know that this model is equivalent to a random cluster model μ'_m on the planar dual graph G', with the same q and parameter $p' < p_c(q)$. For a configuration A on the primal graph, the dual configuration A' is given by

$$e \in A \Leftrightarrow e' \notin A' \,,$$

where e' is the dual edge of e. For a dual vertex w', let $C'(w')$ be the set of dual open edges connected to w' via the dual configuration A'. If A has distribution μ_m, then A' has distribution μ'_m and therefore $C'(w')$ should not contain many edges. We know that for dual vertices u', v' one has

$$\mathbb{P}[u' \leftrightarrow v' \text{ via } A'] \leq C \exp(-c\|u' - v'\|_2) \,,$$

for some constants $C, c > 0$ and it follows

$$\mathbb{P}[\exists u' \leftrightarrow w' \text{ with } \|u' - w'\|_2 \geq r] \leq D' r \exp(-cr)$$

with an adjusted constant D'. To obtain this bound, it suffices to sum over the connection probabilities of vertices u' that have distance exactly r to w', which are $\mathcal{O}(r)$ in number. Now, we can sum over all dual vertices to obtain

$$\mathbb{P}[\exists w' \text{ s.t. } \exists u' \leftrightarrow w' \text{ with } \|u' - w'\|_2 \geq r] \leq |V'_m| D' r \exp(-cr)$$

with $|V'_m| \in \mathcal{O}(m^2)$ being the number of dual vertices. Choosing $r = b \log m$ with b big enough, we see that this probability is diminishing for increasing m.

Lemma 5.22. *In the supercritical phase, dual clusters of the form $C'(w')$ have maximal radius $r = b \log m$ with probability*

$$\mathbb{P}[\exists w' \text{ s.t. } \exists u' \leftrightarrow w' \text{ with } \|u' - w'\|_2 \geq r] \in \mathcal{O}((\log m) m^{2-bc}) \,.$$

This implies the following for the primal configuration A: with a probability tending to 1 for increasing m, the configuration A has a unique biggest open cluster, the so called global cluster. This global cluster contains all open edges which comprise the boundaries of dual clusters, which only have a certain radius. Therefore, the

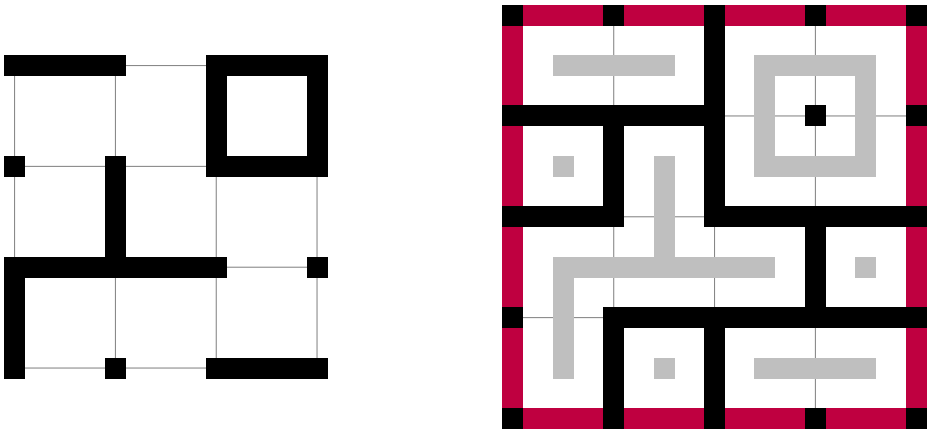

Figure 5.5: A dual configuration on a 4×4 grid with free boundary conditions (left), as well as the resulting primal configuration (right). The purple edges are always open, imposing a wired-like boundary condition. The global cluster encloses all dual clusters.

global cluster permeates the whole configuration A with increasing probability, see Figure 5.5 for a clarification. Lemma 5.22 is a global property in the sense that it holds for the whole lattice and not only for a finite number of vertices. An addition to this law can be given with the following local property. From here on out we use wired boundary conditions or the dual boundary conditions to free boundary conditions which are given by wired boundary conditions plus the condition that all edges between boundary vertices in $V_m \setminus V_{m-1}$ are also always open. In both cases, we define the global cluster to be the unique cluster connecting to the boundary.

Lemma 5.23 (Exponential decay of non-global connectivity)**.** *Let $G = (V_m, E_m)$ be the rectangular subgraph of $\mathbb{Z}^2$ and consider the associated random cluster measure μ_m^1 with wired-like boundary conditions, as well as $p > p_c(q)$. Let $\mathcal{N}(A, u, v)$ be the following event: There exists a path of dual vertices $w_1', \dots, w_n'$ with $w_n' = w_1'$ such that the connecting edges are open in A' and both u and v are on the inside of this loop. In the supercritical phase with $A \sim \mu_m^1$, there exist $C, c > 0$ such that*

$$\mathbb{P}[\mathcal{N}(A, u, v)] \leq C \exp(-c\|u - v\|_2)$$

holds for all m and $u, v \in V_m$.

Note that the event $\mathcal{N}(A, u, v)$ encompasses the event 'u and v are connected via a non-global cluster in A'.

Proof. Let $u, v \in V_m$ be non-boundary vertices, and assume without loss of generality that $u_1 \leq v_1$ and $u_2 \leq v_2$ (we can rotate the graph otherwise). Let $u' = (u_1 - 0.5, u_2 - 0.5)$ be the dual vertex that is located at the lower left of u, and let $v' = (v_1 + 0.5, v_2 + 0.5)$ be the dual vertex that is connected at the upper right of v. We consider the dual vertices $u_i' = u' - (0, i)$ and $v_j' = v' + (0, j)$ that have $\|\cdot\|_\infty$-norm smaller than m. Notice that the event $\mathcal{N}(A, u, v)$ implies that there

must exist a dual path from a vertex in $\{u'_0, u'_1, \ldots\}$ to a vertex in $\{v'_0, v'_1, \ldots\}$. We can bound the corresponding probability by

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\mathbb{P}[u'_i \leftrightarrow v'_j \text{ in the dual configuration}] \leq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} C\exp(-c\|u'_i - v'_j\|_2)$$

because Theorem 3.13 holds for the dual configuration. Moreover, we have

$$\|u'_i - v'_j\|_2 = \|v - u + (1, 1+i+j)\|_2 \geq \frac{1}{\sqrt{2}}(\|u-v\|_2 + \|(1,1+i+j)\|_1)$$

and thus we see

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} C\exp(-c\|u'_i - v'_j\|_2) \leq C\exp(-c'\|u-v\|_2)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\exp(-c'(i+j+2))$$

with $c' = c/\sqrt{2}$. The double sum converges and therefore we get the result. □

For the rest of this section, we consider the supercritical Ising model with $p > p_c(2)$ on $G = (V_m, E_m)$ with monochromatic boundary conditions, i.e. all vertices in $V_{m+1} \setminus V_m$ have color $s = 2$. We denote the corresponding measure with ν_m. In this setting, $s = 2$ is the dominating color, and we will analyze Markov chain dynamics P that are close to the monochromatic $s = 2$ state σ^*. To do this, consider

$$d^*(\nu_m, P^s) = \|\nu_m - P^s(\sigma^*, \cdot)\|_{TV}$$

and the *monochromatic mixing time*

$$\tau^*_{mix}(P, \epsilon) = \min\{s\colon d^*(\nu_m, P^s) \leq \epsilon\}\,.$$

It is easy to see that this is a weaker concept than the original mixing time, and it holds

$$\tau^*_{mix}(P, \epsilon) \leq \tau_{mix}(P, \epsilon)\,.$$

But it is still possible to derive constructive Markov chain Monte-Carlo integration results from here, because the initial distribution δ_{σ^*} is realizable. Therefore, error bounds for Markov chain Monte-Carlo simulations that start in σ^* can be derived like in Chapter 2. It is important to note that Lemma 2.7 holds also for the monochromatic mixing time.
The next step of our preparations consists of a local to global coupling result. For a vertex $u \in V_m$, we consider $B_R(u) = \{v \in V_m\colon \|u - v\|_\infty \leq R\}$ and the Ising model measure ν_u on $B_R(u)$ with monochromatic $s = 2$ boundary conditions, i.e. all vertices adjacent to $B_R(u)$ have color $s = 2$.

Lemma 5.24 (Local to global coupling)**.** *Let $p > p_c(2)$ and ν_m be the Ising model measure on $G = (V_m, E_m)$ with monochromatic $s = 2$ boundary conditions. Let $0 \leq r \leq R$ and let ν_u be the Ising model measure on $B_R(u)$ with monochromatic $s = 2$ boundary conditions. Let ν^r_m and ν^r_u be the corresponding marginal measures*

on the set W^r of vertices that have at least distance $r > 0$ to vertices in $V_m \setminus B_R(u)$. Then there exist constants $C, c > 0$ such that

$$\|\nu_m^r - \nu_u^r\|_{TV} \leq C \exp(-cr)$$

for all m, r, R and $u \in V_m$.

Proof. As usual, we construct a coupling of ν_m and ν_u that agrees with high probability on W^r. Let μ_m be the random cluster model measure on E_{m+1} with wired boundary conditions, and let μ_u be the random cluster model measure on edges with both endpoints in $B_{R+1}(u)$ together with wired boundary conditions, where we call the edge set $E_{u,R}$. The partial order in this setting gives us $\mu_m \leq \mu_u$, and we can define a coupling with

$$A \sim \mu_m, \quad A_u \sim \mu_u, \quad A \leq A_u$$

(using for instance the Glauber dynamics Markov chain with a CFTP procedure). Now, we do a resampling to obtain a different coupling. For a given realization (A, A_u), let V' be the set of dual vertices enclosed by $B_{R+1}(u)$ that are connected to a dual vertex outside of $B_{R+1}(u)$ via open dual edges in A'. This set gives us a connected boundary given by the edge set F defined via

$$\begin{aligned} \tilde{F} &= \{e \in E_{u,R} \colon \text{the dual edge } e' \text{ has exactly one endpoint in } V'\} \\ &\quad \cup \{e \in E_{u,R} \colon e \text{ is an outermost edge and open in } A\}\,, \\ F &= \tilde{F} \setminus \{e \text{ is an outermost edge in } E_{m+1}\} \end{aligned}$$

that encompasses V', see Figure 5.6.
Per definition, all edges $e \in F$ are open in A, and these open edges form a set of disjoint closed loops. Due to monotonicity, this is true for A_u as well, and we see that the conditional measures on the set enclosed by F coincide for both A and A_u. We can do a coupled resampling on

$$\Lambda = \{e \in E_{u,R} \colon e' \text{ has no endpoint in } V'\}$$

to obtain a coupling

$$B \sim \mu_m, \quad B_u \sim \mu_u, \quad B \leq B_u, \quad B = B_u \text{ on } \Lambda\,.$$

Let W be the set of vertices that are adjacent to an edge in Λ. Using the Edwards-Sokal coupling step that always assigns $s = 2$ to the global cluster, we obtain a coupling of the corresponding spin models

$$\begin{gathered} \sigma \sim \nu_m, \quad \sigma_u \sim \nu_u, \\ \sigma = \sigma_u \text{ on vertices in } W \text{ belonging to non-global clusters of } B_u\,, \\ \sigma = \sigma_u \text{ on vertices belonging to the global cluster of } B\,. \end{gathered}$$

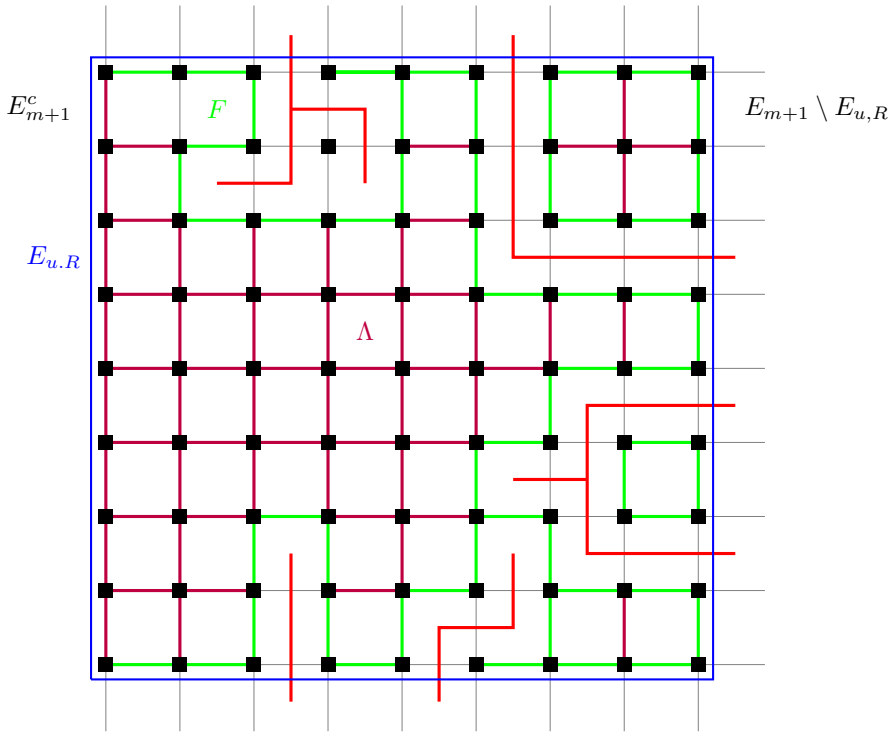

Figure 5.6: The edge sets involved in the proof of Lemma 5.24, here for a set $E_{u,R}$ that is aligning with the boundary of E_{m+1} on the left boundary. The set F decomposes in components that are loop-like. The red edges mark the connection of dual vertices in V' to the outside of $B_{R+1}(u)$. The edges in F are open in A and A_u, therefore a resampling on Λ is possible.

This is true because the non-global clusters of W align in B and B_u, and the global cluster in B automatically is contained in the global cluster of B_u. From this coupling, we get that

$$\begin{aligned}\|\nu_m^r - \nu_u^r\|_{TV} &\leq \mathbb{P}[\sigma \neq \sigma_u \text{ on } W^r] \\ &\leq \mathbb{P}[W^r \not\subseteq W] + \mathbb{P}[W^r \subseteq W]\mathbb{P}[\sigma \neq \sigma_u \text{ on } W^r \mid W^r \subseteq W]\,.\end{aligned}$$

However, the construction of W gives us that $W^r \not\subseteq W$ implies that there exists a vertex in W^r that is enclosed by a dual path connecting to a dual vertex outside of $B_{R+1}(u)$, and this probability decays exponentially with increasing distance from W^r to $V_m \setminus B_R(u)$, which is r. For the second summand, we note that for a vertex $v \in W^r \subseteq W$ we have that

$$\sigma(v) \neq \sigma_u(v) \Rightarrow v \text{ belongs to the global cluster in } B_u, \text{ but not in } B.$$

In this case, v is connected to a unique component of F via open edges in Λ which does not belong to the global cluster of B. Therefore it is enclosed by an open dual component C', which connects to a dual vertex outside of $B_{R+1}(u)$ by construction. Such a dual component contains vertices which have distance bigger than R from each other, therefore this probability decays exponentially in R. Overall, we get

$$\begin{aligned}\|\nu_m^r - \nu_u^r\|_{TV} &\leq \mathbb{P}[\sigma \neq \sigma_u \text{ on } W^r] \\ &\leq \mathbb{P}[W^r \not\subseteq W] + \mathbb{P}[W^r \subseteq W]\mathbb{P}[\sigma \neq \sigma_u \text{ on } W^r \mid W^r \subseteq W] \\ &\leq C\exp(-cr) + D\exp(-dR) \\ &\leq \tilde{C}\exp(-\tilde{c}r)\end{aligned}$$

for some constants $C, \tilde{c} > 0$.

□

From the proof it is clear that a generalization to the q-state Potts model with monochromatic boundary conditions is possible, as we did not use any monotonicity in the spin setting. Using this vital property, we proceed with the central theorems of this section.

Theorem 5.25 (Monochromatic mixing time bound for monotone dynamics with local/global coupling)**.** *Let $p > p_c(2)$ and ν_m be the Ising model measure on $G = (V_m, E_m)$ with monochromatic $s = 2$ boundary conditions. The alternate scan dynamics, heatbath sweep dynamics and monotone Swendsen-Wang dynamics have monochromatic mixing time*

$$\tau_{mix}^*(\epsilon) \in \mathcal{O}(\log m)\,.$$

Proof. Let $(\sigma^t)_{t\geq 0}$ be the alternate scan dynamics Markov chain with transition matrix P and initial state σ^0 given by

$$\sigma^0(v) = 2 \quad \forall v \in V_m\,,$$

i.e. the monochromatic $s = 2$ state. Moreover, let $\tau^0 \sim \nu_m$ and consider the Markov chain $(\tau^t)_{t\geq 0}$ coupled to $(\sigma^t)_{t\geq 0}$ using the monotone grand coupling of the alternate scan dynamics to obtain

$$\tau^t \leq \sigma^t$$

for $t \geq 0$. The monochromatic mixing time is then bounded by

$$\tau^*_{mix}(P, \epsilon) = \min\{t\colon \|\nu_m - P^t(\sigma^0, \cdot)\|_{TV} \leq \epsilon\} \leq \min\{t\colon \mathbb{P}[\tau^t \neq \sigma^t] \leq \epsilon\}\,.$$

Let $u \in V_m$ and $R > 0$ and consider the local version $(\sigma^t_{u,R})_{t\geq 0}$ of the alternate scan dynamics on $B_R(u)$, coupled to $(\tau^t)_{t\geq 0}$ and $(\sigma^t)_{t\geq 0}$ such that

$$\tau^t \leq \sigma^t \leq \sigma^t_{u,R}$$

for $t \geq 0$ (also starting in the monochromatic $s = 2$ state). Denote with $\nu_{u,R}$ the stationary distribution of this chain, which is given by the Ising model measure on $B_R(u)$ with $s = 2$ monochromatic boundary conditions. For any $v \in B_R(u)$ one has

$$\mathbb{P}[\tau^t(v) \neq \sigma^t(v)] \leq \mathbb{P}[\tau^t(v) \neq \sigma^t_{u,R}(v)] = \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\tau^t(v) = 2]$$

due to monotonicity. The Markov chain τ^t has been initialized with $\tau^0 \sim \nu_m$, therefore $\mathbb{P}[\tau^t(v) = 2] = \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m]$. We continue with

$$\begin{aligned} & \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\tau^t(v) = 2] \\ = \quad & \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] \\ & + \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m] \end{aligned}$$

and see that the first difference will be small for sufficiently high t, while the second difference will be small if v is close to u and R is big. More precisely, let $C, c > 0$ be constants such that Lemma 5.24 holds, and choose $R = b \log m$ with $b > \frac{2}{c}$. For any $v \in V_m$, we can find a u such that $B_R(u)$ is quadratic and the distance of v to $V_m \setminus B_R(u)$ is at least $R+1$. For such a pair (u, v), Lemma 5.24 applies with $r = R$ and we get

$$\begin{aligned} & \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m] \\ \leq \quad & \|\nu^0_{u,R} - \nu^0_m\|_{TV} \leq C \exp(-cR) \\ = \quad & Cm^{-bc} \end{aligned}$$

with $bc > 2$. If we choose $t = T_R = T_R(1/4)$ to be the monochromatic mixing time of the alternate scan dynamics on $B_R(u)$ (for quadratic $B_R(u)$), we get

$$\mathbb{P}[\sigma^{nt}_{u,R}(v) = 2] - \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] \leq \frac{1}{4^n}$$

for $n \in \mathbb{N}$. Choosing $n = bc \log_4(m)$ results in an overall bound

$$\mathbb{P}[\tau^{nt}(v) \neq \sigma^{nt}(v)] \leq (C + 1)m^{-bc}$$

with $bc > 2$. For m big enough, we therefore get that

$$\mathbb{E}[|\{v\colon \sigma^{nt}(v) \neq \tau^{nt}(v)\}|] \leq |V_m|(C + 1)m^{-bc} \leq \frac{1}{4}$$

and it immediately follows

$$\mathbb{P}[\sigma^{nt} \neq \tau^{nt}] \leq \frac{1}{4}\,.$$

Therefore we get for some $m = m_0$ the monochromatic mixing time bound

$$\tau^*_{mix}(P, 1/4) \leq nt = bc \log_4(m) T_R\,,$$

where T_R however still depends on $R = b \log m$. We now use an induction argument to obtain a concrete monochromatic mixing time bound. With our notation, we see that $\tau^*_{mix}(P, 1/4) = T_m$ because the graph $G = (V_m, E_m)$ with monochromatic boundary conditions is equivalent to $B_m((0,0))$ with the same boundary conditions. Let $a > 0$ and $\phi(R)$ be an increasing function in R such that $T_r \leq a\phi(r)$ for all $r \leq R$. If

$$T_m \leq bc \log_4(m)\phi(R) \leq a\phi(m)$$

holds, we get $\tau^*_{mix}(P, 1/4) \in \mathcal{O}(\phi(m))$. One readily verifies that the above is true for $\phi(m) = (\log m)^2$.
As a final step, we use the speed of disagreement percolation argument to show the final mixing time bound. We consider the Markov chains $(\sigma^t)_{t\geq 0}$ and $(\tau^t)_{t\geq 0}$, and using the grand coupling we can initialize Markov chains $(\pi^t)_{t\geq 0}$ for any $\pi^0 \geq \tau^0$ such that

$$\tau^t \leq \pi^t \leq \sigma^t$$

holds for all t. We take such chains $(\pi_1^t)_{t\geq 0}$ and $(\pi_2^t)_{t\geq 0}$ that at time $t = 0$ differ exactly at one vertex $u \in V_m$, i.e. $\pi_1^0(u) = 1$ and $\pi_2^0(u) = 2$ and elsewhere the initial states agree. For a given R, we consider the expected difference at time $t = RT_R$:

$$\mathbb{E}[|\{v\colon \pi_1^t(v) \neq \pi_2^t(v)\}|] \leq \sum_{v\in V_m} \mathbb{P}[v \in \Gamma(u,t)]\mathbb{P}[\sigma^t(v) \neq \tau^t(v)]\,.$$

The summands here can be bounded using the local chains as above to obtain

$$\mathbb{E}[|\{v\colon \pi_1^t(v) \neq \pi_2^t(v)\}|] \leq \mathbb{E}[|\Gamma(u,t)|]\left(\frac{1}{4^R} + C\exp(-c(R-1))\right)\,,$$

and we see that this expression is decaying in R due to the polynomial growth of $\mathbb{E}[|\Gamma(u,t)|]$. Therefore, for R big enough we have that the Markov chain applied $t = RT_R$ times is contractive:

$$\mathbb{P}[\pi_1^t \neq \pi_2^t] \leq \frac{1}{2}\,.$$

The path coupling theorem implies the monochromatic mixing time

$$\tau^*_{mix}(P, 1/4) \in \mathcal{O}(\log m)$$

because $t = RT_R$ is independent of m.
The properties of the alternate scan dynamics we used in this proof are mainly the local/global coupling, monotonicity and the polynomial speed of disagreement percolation bound. All of these properties also hold for the monotone Swendsen-Wang dynamics and the heatbath sweep dynamics, implying the same monochromatic mixing time bound. □

Aside from the speed of disagreement percolation bound, we get that this proof also works for systematic scan dynamics.

Corollary 5.26. *Let $p > p_c(2)$ and ν_m be the Ising model measure on $G = (V_m, E_m)$ with monochromatic $s = 2$ boundary conditions. The systematic scan dynamics has monochromatic mixing time*

$$\tau^*_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

The proof for the heatbath dynamics needs the same modification like in the subcritical case, keeping track of the expected number of updates in small boxes.

Corollary 5.27. *Let $p > p_c(2)$ and ν_m be the Ising model measure on $G = (V_m, E_m)$ with monochromatic $s = 2$ boundary conditions. The heatbath dynamics has monochromatic mixing time*

$$\tau^*_{mix}(\epsilon) \in \mathcal{O}\left(m^2 \log m\right) .$$

Proof. Let $(\sigma^t)_{t\geq 0}$ be the heatbath dynamics Markov chain with transition matrix P and initial state σ^0 given by the monochromatic $s = 2$ state. Moreover, let $\tau^0 \sim \nu_m$ and consider the Markov chain $(\tau^t)_{t\geq 0}$ coupled to $(\sigma^t)_{t\geq 0}$ using the monotone grand coupling of the heatbath dynamics to obtain

$$\tau^t \leq \sigma^t$$

for $t \geq 0$. The monochromatic mixing time is then bounded by

$$\tau^*_{mix}(P, \epsilon) = \min\{t \colon \|\nu_m - P^t(\sigma^0, \cdot)\|_{TV} \leq \epsilon\} \leq \min\{t \colon \mathbb{P}[\tau^t \neq \sigma^t] \leq \epsilon\} .$$

Let $u \in V_m$ and $R > 0$ and consider the local version $(\sigma^t_{u,R})_{t\geq 0}$ of the heatbath dynamics on $B_R(u)$, coupled to $(\tau^t)_{t\geq 0}$ and $(\sigma^t)_{t\geq 0}$ such that

$$\tau^t \leq \sigma^t \leq \sigma^t_{u,R}$$

for $t \geq 0$ (also starting in the monochromatic $s = 2$ state). Denote with $\nu_{u,R}$ the stationary distribution of this chain, which is given by the Ising model measure on $B_R(u)$ with $s = 2$ monochromatic boundary conditions. For any $v \in B_R(u)$ one has

$$\mathbb{P}[\tau^t(v) \neq \sigma^t(v)] \leq \mathbb{P}[\tau^t(v) \neq \sigma^t_{u,R}(v)] = \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\tau^t(v) = 2]$$

due to monotonicity. The Markov chain τ^t has been initialized with $\tau^0 \sim \nu_m$, therefore $\mathbb{P}[\tau^t(v) = 2] = \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m]$. We continue with

$$\begin{aligned} & \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\tau^t(v) = 2] \\ = \quad & \mathbb{P}[\sigma^t_{u,R}(v) = 2] - \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] \\ & + \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m] \end{aligned}$$

and see that the first difference will be small for sufficiently high t, while the second difference will be small if v is close to u and R is big. More precisely, let $C, c > 0$ be constants such that Lemma 5.24 holds, and choose $R = b \log m$ with $b > \frac{2}{c}$. For any $v \in V_m$, we can find a u such that $B_R(u)$ is quadratic and the distance of v to $V_m \setminus B_R(u)$ is at least $R+1$. For such a pair (u, v), Lemma 5.24 applies with $r = R$ and we get

$$\begin{aligned} & \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\tau(v) = 2 \mid \tau \sim \nu_m] \\ \leq \; & \|\nu^0_{u,R} - \nu^0_m\|_{TV} \leq C \exp(-cR) \\ = \; & Cm^{-bc} \end{aligned}$$

with $bc > 2$. Let $T_R = T_R(1/4)$ be the monochromatic mixing time of the heatbath dynamics on quadratic sets $B_R(u)$, and choose

$$t = aT_R \log(3M)M/(2R+1)^2 \in \mathbb{N}$$

with $1 \leq a < 2$. After t heatbath steps, the expected number of updates in $B_R(u)$ is given by $a \log(3M)T_R$, and a Chernoff bound then implies that

$$\mathbb{P}[\text{less than } \log(3M)T_R \text{ updates happen in } B_R(u)] \leq \frac{1}{3M}\,.$$

If at least $\log(3M)T_R$ updates do happen in $B_R(u)$ (call this event U), we get

$$\mathbb{P}[\sigma^t_{u,R}(v) = 2 \mid U] - \mathbb{P}[\sigma_{u,R}(v) = 2 \mid \sigma_{u,R} \sim \nu_{u,R}] \leq 4^{-\log(3M)} \leq \frac{1}{3M}\,.$$

Collecting all estimates, we get

$$\mathbb{P}[\tau^t(v) \neq \sigma^t(v)] \leq \frac{2}{3M} + Cm^{-bc} \leq \frac{3}{4M}$$

with $bc > 2$ and m big enough. Summing over all vertices, we get

$$\mathbb{E}\left[|\{v \in V_m \colon \tau^t(v) \neq \sigma^t(v)\}|\right] \leq \frac{3}{4}$$

directly implying

$$\mathbb{P}[\tau^t \neq \sigma^t] \leq \frac{3}{4}\,.$$

From Lemma 2.7 then it follows that after nt steps with n big enough (independent of m), we get

$$\mathbb{P}[\tau^{nt} \neq \sigma^{nt}] \leq \frac{3^n}{4^n} \leq \epsilon\,,$$

such that the monochromatic mixing time is bounded by

$$nt = naT_R \log(3M)M/(2R+1)^2\,,$$

however in this bound R is depending on m. The standard induction argument gives a monochromatic mixing time

$$\tau^*_{mix}(\epsilon) \in \mathcal{O}\left(m^2(\log m)^2\right)\,,$$

and again we can use the speed of disagreement percolation argument to obtain

$$\tau^*_{mix}(\epsilon) \in \mathcal{O}\left(m^2 \log m\right)\,.$$

□

6 The Down-Up Model

In this chapter, we introduce a new model closely related to the random cluster model and apply the ideas gained throughout this thesis. With this model, we generalize the Ising model to non-integer $q \geq 2$ and discover a rich monotone structure. Moreover, we generalize the monotone Swendsen-Wang dynamics to this model and prove mixing time results for the subcritical and for the supercritical phase. The idea for this model is inspired by the splitting procedure for variants of the Chayes-Machta algorithm, see [GOPS11].

Let $G = (V, E)$ be a finite graph and

$$\mu(A) = \mu_{G,p,q}(A) = Z^{-1} p^{|A|} (1-p)^{|E \setminus A|} q^{c(A)}$$

be the standard random cluster model measure on G without boundary conditions, with parameters $p \in (0,1)$ and $q \in [2, \infty)$. We introduce two spins $\{\nabla, \Delta\}$ ("down" and "up") and perform the following assignment:

- For a given random cluster configuration A, assign to all clusters of A a spin $s \in \{\nabla, \Delta\}$ with equal probability. This defines a spin configuration

$$\omega \colon V \to \{\nabla, \Delta\}\,.$$

It is clear that for $q = 2$, this is just the usual Edwards-Sokal coupling and ω will be distributed according to the Ising model measure. In the general case, we see that with $r = q/2$ the joint model measure

$$\begin{aligned} \rho(A, \omega) &= Z^{-1} p^{|A|} (1-p)^{|E \setminus A|} q^{c(A)} \mathbb{1}[A \subseteq A(\omega)] 2^{-c(A)} \\ &= Z^{-1} p^{|A|} (1-p)^{|A(\omega) \setminus A|} r^{c(A)} \mathbb{1}[A \subseteq A(\omega)] (1-p)^{|E \setminus A(\omega)|} \end{aligned}$$

decomposes into a local interaction part $(1-p)^{|E \setminus A(\omega)|}$ and a random cluster model part on the subgraph $G_\omega = (V, A(\omega))$. The down-up model is then obtained if we forget the underlying cluster configuration A and only care for ω:

$$\nu(\omega) = \sum_A \rho(A, \omega) = Z^{-1} (1-p)^{|E \setminus A(\omega)|} Z(\omega)$$

with $Z(\omega) = Z_{G_\omega, p, r}$ being the partition function of the random cluster model on G_ω with parameters p and r. The generalized Edwards-Sokal coupling or Swendsen-Wang dynamics between these models is then given by

1. For a given random cluster configuration A, assign to all clusters of A a spin $s \in \{\nabla, \Delta\}$ with equal probability. This defines a spin configuration
$$\omega \colon V \to \{\nabla, \Delta\}\,.$$
2. For a given spin configuration ω, generate B according to the rule
$$B \sim \mu_{G_\omega, p, r}\,.$$

Both steps can be viewed as a Gibbs sampler on the joint space, holding one of the variables fixed.

6.1 Monotonicity with respect to boundary conditions

We define the usual partial order induced by $\nabla \leq \Delta$ on spin configurations:

$$\omega \leq \pi \quad \Leftrightarrow \quad \omega(v) \leq \pi(v) \text{ for all } v \in V\,.$$

We will show the following

Theorem 6.1 (Monotonicity with respect to boundary conditions)**.** *Let ν be the down-up model measure on $G = (V, E)$ with parameters $p \in (0, 1)$ and $q \geq 2$. Let $W^1_\nabla, W^2_\nabla, W^1_\Delta, W^2_\Delta$ be subsets of V satisfying*

$$W^1_\nabla \cap W^1_\Delta = \emptyset, \quad W^2_\nabla \cap W^2_\Delta = \emptyset$$

and

$$W^2_\nabla \subseteq W^1_\nabla, \quad W^1_\Delta \subseteq W^2_\Delta$$

such that they are compatible boundary conditions. Then the measures

$$\nu^i(\omega) = \nu(\omega \mid \omega \equiv \nabla \text{ on } W^i_\nabla \text{ and } \omega \equiv \Delta \text{ on } W^i_\Delta)$$

for $i \in \{1, 2\}$ satisfy

$$\nu^1 \leq \nu^2\,.$$

But we do not stop at that. The monotonicity in the spin setting actually carries over to the joint model. For a spin configuration ω, define

$$A_\nabla(\omega) = \{e \in E\colon \text{ both endpoints } v, w \text{ of } e \text{ have spin } \omega(v) = \omega(w) = \nabla\}$$
$$A_\Delta(\omega) = \{e \in E\colon \text{ both endpoints } v, w \text{ of } e \text{ have spin } \omega(v) = \omega(w) = \Delta\}$$

such that $A(\omega)$ decomposes into $A_\nabla(\omega)$ and $A_\Delta(\omega)$. Moreover, we have that $\omega \leq \pi$ implies $A_\nabla(\pi) \subseteq A_\nabla(\omega)$ nad $A_\Delta(\omega) \subseteq A_\Delta(\pi)$. We therefore define the following partial order on the joint space:

$$\begin{aligned}(A, \omega) \leq (B, \pi) \quad \Leftrightarrow \quad & \omega \leq \pi \text{ and} \\ & A \cap A_\Delta(\omega) \subseteq B \cap A_\Delta(\omega) \text{ and} \\ & B \cap A_\nabla(\pi) \subseteq A \cap A_\nabla(\pi)\,.\end{aligned}$$

In other words, we have $(A,\omega) \leq (B,\pi)$ if $\omega \leq \pi$ and A dominates B on the ∇-parts, as well as B dominates A on the Δ-parts.

Theorem 6.2 (Monotonicity with respect to boundary conditions in the joint model)**.** *Let ρ be the joint model measure on $G=(V,E)$ with parameters $p \in (0,1)$ and $q \geq 2$. Let $W^1_\nabla, W^2_\nabla, W^1_\Delta, W^2_\Delta$ be subsets of V satisfying*

$$W^1_\nabla \cap W^1_\Delta = \emptyset, \quad W^2_\nabla \cap W^2_\Delta = \emptyset$$

and

$$W^2_\nabla \subseteq W^1_\nabla, \quad W^1_\Delta \subseteq W^2_\Delta$$

such that they are compatible spin boundary conditions. Moreover, let F^i_∇ be subsets of edges with both endpoints in W^i_∇, and let F^i_Δ be subsets of edges with both endpoints in W^i_Δ for $i \in \{1,2\}$. Consider edge boundary conditions $D^i_\nabla \subseteq F^i_\nabla$ and $D^i_\Delta \subseteq F^i_\Delta$ for $i \in \{1,2\}$ that additionally satisfy

$$\begin{aligned} D^2_\nabla \subseteq D^1_\nabla\,, \qquad & D^1_\Delta \subseteq D^2_\Delta\,, \\ F^1_\nabla \setminus F^2_\nabla \subseteq D^1_\nabla\,, \qquad & F^2_\Delta \setminus F^1_\Delta \subseteq D^2_\Delta\,, \end{aligned}$$

see also Figure 6.1. Then the measures

$$\rho^i(A,\omega) = \rho\left(A,\omega \mid \begin{cases} \omega \equiv \nabla \text{ on } W^i_\nabla \\ \omega \equiv \Delta \text{ on } W^i_\Delta \\ A \cap F^i_\nabla = D^i_\nabla \\ A \cap F^i_\Delta = D^i_\Delta \end{cases}\right)$$

for $i \in \{1,2\}$ satisfy

$$\rho^1 \leq \rho^2$$

with respect to the partial order on the joint space. Moreover, the marginal measures $\nu^i(\omega) = \sum_A \rho^i(A,\omega)$ for $i \in \{1,2\}$ satisfy

$$\nu^1 \leq \nu^2$$

with respect to the partial order for spin configurations.

Before proving these theorems, we elaborate on the spin marginal measures introduced just now. For given compatible spin and edge boundary conditions $\phi = (W_\nabla, W_\Delta, F_\nabla, F_\Delta, D_\nabla, D_\Delta)$, we get the joint measure

$$\rho^\phi(A,\omega) = \rho\left(A,\omega \mid \begin{cases} \omega \equiv \nabla \text{ on } W_\nabla \\ \omega \equiv \Delta \text{ on } W_\Delta \\ A \cap F_\nabla = D_\nabla \\ A \cap F_\Delta = D_\Delta \end{cases}\right)$$

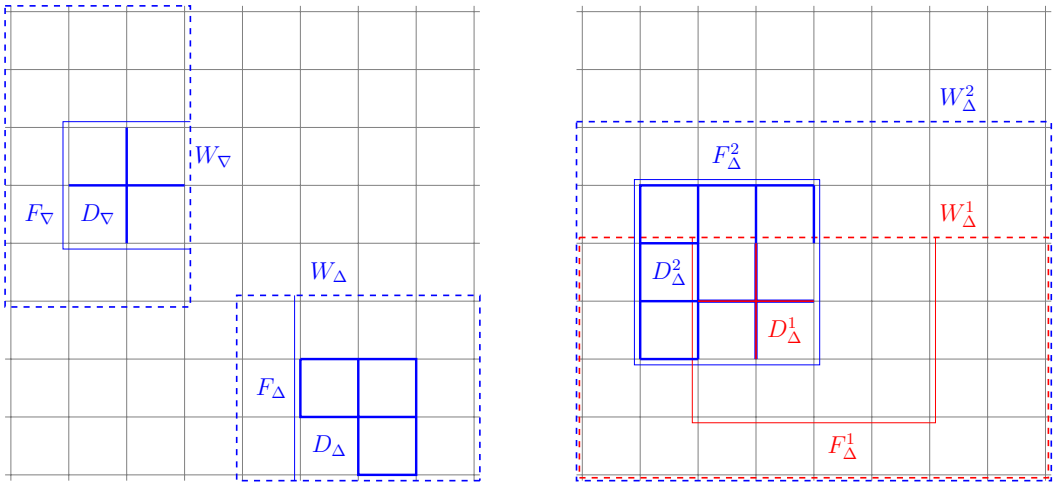

Figure 6.1: Compatible boundary conditions. On the left, we see a boundary condition ϕ on a subgraph of $\mathbb{Z}^2$. On the right, we see comparable boundary conditions ϕ^1 and ϕ^2 from the setting in Theorem 6.2, to simplify only for the Δ-parts.

and we write $(A,\omega) \in \phi$ if $\rho^\phi(A,\omega) > 0$. From here, we see that

$$\begin{aligned}\nu^\phi(\omega) &= \sum_{A\subseteq E} \rho^\phi(A,\omega) \\ &\propto \mathbb{1}[\omega \equiv \nabla \text{ on } W_\nabla]\mathbb{1}[\omega \equiv \Delta \text{ on } W_\Delta] \\ &\quad \cdot \sum_{A\subseteq A(\omega)} \mathbb{1}[A\cap F_\Delta = D_\Delta]\mathbb{1}[A\cap F_\nabla = D_\nabla] p^{|A|}(1-p)^{|E\setminus A|} r^{c(A)}\,.\end{aligned}$$

Furthermore, any $A \subseteq A(\omega)$ uniquely decomposes into $A_\nabla \subseteq A_\nabla(\omega)$ and $A_\Delta \subseteq A_\Delta(\omega)$ such that we can write

$$\nu^\phi(\omega) \propto \mathbb{1}[\omega \in \phi](1-p)^{|E\setminus A(\omega)|} Z^\phi_\nabla(\omega) Z^\phi_\Delta(\omega)\,,$$

where we abbreviated

$$\omega \in \phi \quad \Leftrightarrow \quad \begin{cases} \omega \equiv \nabla \text{ on } W_\nabla \\ \omega \equiv \Delta \text{ on } W_\Delta \end{cases}$$

and the partition function terms are given by

$$\begin{aligned} Z^\phi_\nabla(\omega) &= Z^{D_\nabla, F_\nabla}_{A_\nabla(\omega),p,r} \quad \text{and} \\ Z^\phi_\Delta(\omega) &= Z^{D_\Delta, F_\Delta}_{A_\Delta(\omega),p,r}\,. \end{aligned}$$

For $F_\nabla = F_\Delta = \emptyset$, we see that

$$\mathbb{1}[\omega \in \phi] Z^\phi_\nabla(\omega) Z^\phi_\Delta(\omega) \propto \mathbb{1}[\omega \in \phi] Z(\omega)$$

such that Theorem 6.1 follows from Theorem 6.2. Moreover, we see that the down-up model measure $\nu^\phi(\omega)$ does in fact depend on the edge boundary conditions for

$q > 2$ because of the partition function terms.
On the other hand, we can consider the marginal measure with respect to edge configurations

$$\mu^\phi(A) = \sum_\omega \rho^\phi(A,\omega) \propto \sum_\omega \mathbb{1}[(A,\omega) \in \phi]\rho(A,\omega)\,.$$

We calculate

$$\begin{aligned} & \sum_\omega \mathbb{1}[(A,\omega) \in \phi]\rho(A,\omega) \\ = & \sum_\omega \mathbb{1}[(A,\omega) \in \phi]Z^{-1}p^{|A|}(1-p)^{|E\setminus A|}r^{c(A)} \\ \propto & \sum_\omega \mathbb{1}[(A,\omega) \in \phi]p^{|A|}(1-p)^{|E\setminus A|}r^{c^\phi(A)}r^{c(A)-c^\phi(A)} \end{aligned}$$

with $c^\phi(A)$ being the number of open clusters in A that do not connect to $W_\nabla \cup W_\Delta$. This sum goes over the possible colorings of such clusters, and therefore we get

$$\sum_\omega \mathbb{1}[(A,\omega) \in \phi]p^{|A|}(1-p)^{|E\setminus A|}r^{c(A)} = \mathbb{1}[A \in \phi]p^{|A|}(1-p)^{|E\setminus A|}q^{c^\phi(A)}r^{c(A)-c^\phi(A)}$$

where we abbreviated

$$A \in \phi \quad \Leftrightarrow \quad \begin{cases} A \cap (F_\nabla \cup F_\Delta) = D_\nabla \cup D_\Delta \\ W_\nabla \not\leftrightarrow W_\Delta \text{ via } A\,. \end{cases}$$

We see that μ^ϕ corresponds to a random cluster model distribution where clusters connected to $W = W_\nabla \cup W_\Delta$ only get the cluster weight $r = q/2$, with boundary condition $D = D_\nabla \cup D_\Delta \subseteq F_\nabla \cup F_\Delta = F$ and conditioned on the event $\mathbb{1}[W_\nabla \not\leftrightarrow W_\Delta \text{ via } A]$. As the next step, we introduce a generalization of the monotone Swendsen-Wang dynamics for the down-up model. Consider the following procedure:

1. For $\omega \in \phi$, generate the joint configuration $(A,\omega) \in \phi$ according to the conditional measure

$$\rho^{\phi,\omega'}(B) = \rho^\phi(B, \pi \mid \pi = \omega) = \rho\left(B, \pi \mid \begin{cases} \pi = \omega \\ B \cap F_\nabla = D_\nabla \\ B \cap F_\Delta = D_\Delta \end{cases}\right).$$

2. For the given joint configuration $(A,\omega) \in \phi$, assign to all vertices of V a proposal spin in $\{\nabla, \Delta\}$ with the following probabilities:

$$\mathbb{P}[\theta(v) = \Delta] = \begin{cases} 0 & v \in W_\nabla \\ 1 & v \in W_\Delta \\ \frac{1}{2} & \text{else}\,, \end{cases}$$

creating a proposal spin configuration $\theta\colon V \to \{\nabla, \Delta\}$.

3. For all open clusters of A, check if all proposal spins on this cluster align. If yes, assign to all vertices of the given cluster the proposed new spin. If not, keep the old spins. This creates a new spin configuration $\omega' \in \phi$.

For $q = 2$, the conditional measure for A is given by simple percolation on $A(\omega)$ and we obtain the monotone Swendsen-Wang dynamics. For general $q \geq 2$ we call this procedure monotone down-up dynamics.

Lemma 6.3. *Let ϕ be a compatible boundary condition as in Theorem 6.2. Then the monotone down-up dynamics P^ϕ are ergodic and reversible with respect to the down-up model measure ν^ϕ, and therefore satisfy*

$$\nu^\phi P^\phi = \nu^\phi \,.$$

Proof. For any $\omega \in \phi$, we get that the event $A = D_\nabla \cup D_\Delta$ at Step 1 has positive probability. In this case, at Step 2 any possible θ will be accepted such that the outcome of Step 3 will be $\omega' = \theta$. But the generation rule for θ implies that $\omega' \in \phi$, as well as all possible ω' can be reached. Therefore, the monotone down-up dynamics are ergodic.
We proceed to show that the monotone down-up dynamics with transition matrix P^ϕ is reversible with respect to ν^ϕ. Consider the probability to go from ω to ω'

$$P^\phi(\omega, \omega') = \sum_A \rho^{\phi,\omega}(A) \sum_\theta \mathbb{P}[\theta \mid \phi]\mathbb{P}[\omega' \mid A, \omega, \theta, \phi]$$

with the θ-generation probability

$$\mathbb{P}[\theta \mid \phi] = 2^{-|V\setminus(W_\nabla \cup W_\Delta)|}\mathbb{1}[\theta \equiv \nabla \text{ on } W_\nabla]\mathbb{1}[\theta \equiv \Delta \text{ on } W_\Delta]$$

and the transition probability of Step 3

$$\begin{aligned}\mathbb{P}[\omega' \mid A, \omega, \theta, \phi] = {} & \mathbb{1}[A \subseteq A(\omega')] \prod_{C \subseteq V} \mathbb{1}[C \text{ is a cluster of } A] \\ & \cdot \Big(\mathbb{1}[\omega \equiv \omega' \text{ on } C](1 - 2^{-|C\setminus(W_\nabla \cup W_\Delta)|}) + \mathbb{1}[\omega \not\equiv \omega' \text{ on } C]2^{-|C\setminus(W_\nabla \cup W_\Delta)|}\Big) \,.\end{aligned}$$

For reversibility, we need to show that

$$\nu^\phi(\omega)P^\phi(\omega, \omega')$$

is symmetric in ω and ω'. Due to ν^ϕ being the marginal of ρ^ϕ with respect to the spin variable, we get

$$\nu^\phi(\omega)\rho^{\phi,\omega}(A) = \rho^\phi(A, \omega)$$

and therefore

$$\nu^\phi(\omega)P^\phi(\omega, \omega') = \sum_A \rho^\phi(A, \omega) \sum_\theta \mathbb{P}[\theta \mid \phi]\mathbb{P}[\omega' \mid A, \omega, \theta, \phi] \,.$$

Because of $\rho^\phi(A,\omega) = \mathbb{1}[A \subseteq A(\omega)]\rho^\phi(A,\omega)$, we see that the whole term is symmetric if for all A with $A \subseteq A(\omega)$ and $A \subseteq A(\omega')$ one has

$$\rho^\phi(A,\omega) = \rho^\phi(A,\omega')\,.$$

But this is simply true because for fixed A, one has that

$$\rho^\phi(A,\omega) = \mu^\phi(A)\rho^{\phi,A}(\omega) = \mu^\phi(A)\mathbb{1}[A \subseteq A(\omega)]2^{-c^\phi(A)}\,,$$

with $c^\phi(A)$ being the number of open clusters of A that do not connect to $W_\nabla \cup W_\Delta$ and

$$\mu^\phi(A) = \sum_\omega \rho^\phi(A,\omega)$$

being the marginal measure with respect to the edge configuration. □

Now we prove one of the central statements of this chapter.

Theorem 6.4 (Monotonicity of the dynamics)**.** *Consider the setting of Theorem 6.2 and call the respective boundary conditions ϕ^1 and ϕ^2. Let P^1 and P^2 be the monotone down-up dynamics with respect to the models ν^1 and ν^2. There exists a coupling Q of the Markov chains P^1 and P^2 mapping (ω^1,ω^2) to (π^1,π^2) such that*

$$\omega^1 \in \phi^1,\, \omega^2 \in \phi^2,\, \omega^1 \le \omega^2 \Rightarrow \pi^1 \in \phi^1,\, \pi^2 \in \phi^2,\, \pi^1 \le \pi^2\,.$$

Proof. Consider the conditional measures $\rho^{i,\omega^i}(A) = \rho^i(A, \omega \mid \omega = \omega^i)$ for $i \in \{1,2\}$. We calculate using the unique splitting $A = (A_\nabla, A_\Delta)$:

$$\begin{aligned}
\rho^{i,\omega^i}(A) &\propto \mathbb{1}[A \subseteq A(\omega^i)]\mathbb{1}[A \in \phi^i]p^{|A|}(1-p)^{|E\setminus A|}r^{c(A)} \\
&= (1-p)^{|E\setminus A(\omega^i)|}\mathbb{1}\begin{bmatrix} A_\nabla \subseteq A_\nabla(\omega^i) \\ A_\nabla \cap F^i_\nabla = D^i_\nabla \\ A_\Delta \subseteq A_\Delta(\omega^i) \\ A_\Delta \cap F^i_\Delta = D^i_\Delta \end{bmatrix} p^{|A|}(1-p)^{|A(\omega^i)\setminus A|}r^{c(A)} \\
&\propto \mu^{i,\omega^i}_\nabla(A_\nabla)\mu^{i,\omega^i}_\Delta(A_\Delta)
\end{aligned}$$

with μ^{i,ω^i}_∇ being the random cluster measure on $A_\nabla(\omega^i)$ with boundary condition $D^i_\nabla \subseteq F^i_\nabla$ and parameters p and $r = q/2$, analogously μ^{i,ω^i}_Δ. We see that $\rho^{i,\omega^i}(A)$ actually is a product of random cluster measures in the variables (A_∇, A_Δ). Moreover, for $\omega^1 \le \omega^2$ we get that

$$A_\nabla(\omega^2) \subseteq A_\nabla(\omega^1) \text{ and } A_\Delta(\omega^1) \subseteq A_\Delta(\omega^2)\,.$$

Due to the compatible edge boundary conditions it follows

$$\mu^{2,\omega^2}_\nabla \le \mu^{1,\omega^1}_\nabla \text{ and } \mu^{1,\omega^1}_\Delta \le \mu^{2,\omega^2}_\Delta$$

with respect to the inclusion partial order on edges. This means that there exists a coupling (A^1, A^2) such that

$$A^i = (A^i_\nabla, A^i_\Delta) \sim \rho^{i,\omega^i} \text{ for } i \in \{1,2\}, \quad A^2_\nabla \subseteq A^1_\nabla, \quad A^1_\Delta \subseteq A^2_\Delta,$$

which just means that $(\omega^1, A^1) \leq (\omega^2, A^2)$ in the joint setting. This coupling covers Step 1 of the monotone down-up dynamics, transferring spin monotonicity to the joint model.
For Step 2, we couple both dynamics to generate (θ_1, θ_2) to be maximally aligned

$$\theta_1(v) = \theta_2(v) \text{ for all } v \in V \setminus (W^1_\nabla \cup W^2_\nabla \cup W^1_\Delta \cup W^2_\Delta),$$

such that $\theta_1 \leq \theta_2$ in the spin sense follows.
For Step 3, we see that $\pi^1(v) > \pi^2(v)$ is not possible at vertices v with $\omega^1(v) \neq \omega^2(v)$. This comes from the fact that $\pi^i(v)$ either keeps the old spin or changes to the proposal $\theta^i(v)$ for $i \in \{1,2\}$. We therefore consider vertices $v \in V$ with $\omega^1(v) = \omega^2(v)$.

- Assume that $\omega^1(v) = \omega^2(v) = \nabla$ and $\pi^1(v) = \Delta$. Let C^i be the A^i_∇-Cluster that contains v for $i \in \{1,2\}$. Then we have $C^2 \subseteq C^1$ due to the Step 1 coupling. Moreover, v changed spin in the first configuration, therefore $\theta_1(w) = \Delta$ for all $w \in C^1$. Together with $\theta^1 \leq \theta^2$ we get that $\theta_2(w) = \Delta$ for all $w \in C^2$, which implies $\pi^2(v) = \Delta$.
- Assume that $\omega^1(v) = \omega^2(v) = \Delta$ and $\pi^2(v) = \nabla$. Let C^i be the A^i_Δ-Cluster that contains v for $i \in \{1,2\}$. Then we have $C^1 \subseteq C^2$ due to the Step 1 coupling. Moreover, v changed spin in the second configuration, therefore $\theta_2(w) = \nabla$ for all $w \in C^2$. Together with $\theta^1 \leq \theta^2$ we get that $\theta_1(w) = \nabla$ for all $w \in C^1$, which implies $\pi^1(v) = \nabla$.

□

It is easy to see that this property implies Theorem 6.2.

Proof of Theorem 6.2. The statement $\nu^1 \leq \nu^2$ immediately follows from the existence of a monotone coupling of Markov chains, as seen for instance in the proofs of Lemma 3.1 to 3.3. For the statement $\rho^1 \leq \rho^2$, consider the monotone down-up dynamics starting in a joint configuration and apply Steps 2, 3, and 1 to obtain a new joint configuration. The considerations of the previous theorem show that such Markov chains R^1 and R^2 also have a monotone coupling, and it is clear that their stationary distributions are given by ρ^1 and ρ^2. □

We conclude this section with showing that for the monotone down-up dynamics there exists a monotone grand coupling. Consider the class of compatible boundary conditions

$$\Phi = \left\{ \phi = (W_\nabla, W_\Delta, F_\nabla, F_\Delta, D_\nabla, D_\Delta) \mid \begin{cases} W_\nabla \cap W_\Delta = \emptyset \\ F_\nabla \subseteq \{e\text{: both endpoints in } W_\nabla\} \\ F_\Delta \subseteq \{e\text{: both endpoints in } W_\Delta\} \\ D_\nabla \subseteq F_\nabla \\ D_\Delta \subseteq F_\Delta \end{cases} \right\}$$

and define the partial order on compatible boundary conditions $\phi^1, \phi^2 \in \Phi$ via

$$\phi^1 \leq \phi^2 \Leftrightarrow \begin{cases} W^2_\nabla \subseteq W^1_\nabla & D^2_\nabla \subseteq D^1_\nabla \quad F^1_\nabla \setminus F^2_\nabla \subseteq D^1_\nabla \\ W^1_\Delta \subseteq W^2_\Delta & D^1_\Delta \subseteq D^2_\Delta \quad F^2_\Delta \setminus F^1_\Delta \subseteq D^2_\Delta \end{cases}$$

as in Theorem 6.2. Let $\mathbb{X} = \{(\omega, \phi) \mid \omega \in \phi\}$ be the set of all compatible boundary conditions ϕ and ω that align with ϕ, and let $\mathbb{V} = \{\omega \colon V \to \{\nabla, \Delta\}\}$ be the set of all spin configurations. We now describe a procedure to create a rule

$$g \colon \mathbb{X} \to \mathbb{V}$$

with law $\mathcal{G}$ satisfying

$$g \sim \mathcal{G} \Rightarrow \begin{cases} g(\omega, \phi) \in \phi \\ g(\omega, \phi) \sim P^\phi(\omega, \cdot) \end{cases}$$

as well as

$$\begin{cases} g \sim \mathcal{G} \\ \omega^1 \in \phi^1 \\ \omega^2 \in \phi^2 \\ \phi^1 \leq \phi^2 \\ \omega^1 \leq \omega^2 \end{cases} \quad \Rightarrow \quad g(\omega^1, \phi^1) \leq g(\omega^2, \phi^2)\,.$$

This rule generation has to cover all steps of the monotone down-up dynamics, and we start with the vital Step 1. Reformulate it as follows:

1. Generate Glauber dynamics rules $g_0, g_{-1}, \ldots$ on $G = (V, E)$ with parameters p and $r = q/2$. Each of such rules consists of an edge $e \in E$ chosen uniformly at random, and a parameter $u \in [0, 1]$ distributed uniformly. For a given $(\omega, \phi) \in \mathbb{X}$, these rules transfer to Glauber dynamics rules on $A_\nabla(\omega)$ and $A_\Delta(\omega)$, with boundary conditions $(F_\nabla, F_\Delta, D_\nabla, D_\Delta)$. Obtain the joint configuration (A, ω) with A being the result of the (CFTP)-procedure with the generated rules.

Due to the monotonicity of the Glauber dynamics for the random cluster model, we see that the assignment $(\omega, \phi) \to (A, \omega, \phi)$ is monotone in the sense that

$$\begin{cases} \omega^1 \in \phi^1 \\ \omega^2 \in \phi^2 \\ \phi^1 \leq \phi^2 \\ \omega^1 \leq \omega^2 \end{cases} \quad \Rightarrow \quad (A^1, \omega^1) \leq (A^2, \omega^2)\,.$$

Continuing from here, we reformulate Step 2 as follows:

2. Generate $\theta \colon V \to \{\nabla, \Delta\}$ with

$$\mathbb{P}[\theta(v) = \nabla] = \mathbb{P}[\theta(v) = \Delta] = 1/2$$

for all vertices $v \in V$. For a given boundary condition ϕ, we define

$$\theta^\phi(v) = \begin{cases} \nabla & v \in W_\nabla, \\ \Delta & v \in W_\Delta, \\ \theta(v) & \text{else} \end{cases}$$

and use this proposal spin configuration to proceed.

For $\phi^1 \leq \phi^2$ we get $\theta^1 \leq \theta^2$, such that in Step 3 we have the situation described in Step 3 of the proof of Lemma 6.4. Step 3 itself does not need a reformulation, as it only uses information of the state in question. Overall, we get

Theorem 6.5 (Grand coupling of the monotone down-up dynamics)**.** *There exists a monotone grand coupling for the monotone down-up dynamics, with respect to pairs* $(\omega, \phi) \in \mathbb{X}$.
Likewise, the monotone down-up dynamics starting in the joint configuration space has a monotone grand coupling with respect to tuples (A, ω, ϕ) *with* $(A, \omega) \in \phi$.

6.2 Monotonicity with respect to increasing events

We turn to increasing events in addition to boundary conditions. In order to do this, we first elaborate on the partial order in the joint setting. For states $(A, \omega) = (A_\nabla, A_\Delta, \omega)$ and $(B, \pi) = (B_\nabla, B_\Delta, \pi)$ we introduce join and meet operations

$$\begin{aligned}(\omega \vee \pi)(v) &= \max\{\omega(v), \pi(v)\}\\ (\omega \wedge \pi)(v) &= \min\{\omega(v), \pi(v)\}\\ (A, \omega) \vee (B, \pi) &= (A_\nabla \cap B_\nabla, A_\Delta \cup B_\Delta, \omega \vee \pi)\\ (A, \omega) \wedge (B, \pi) &= (A_\nabla \cup B_\nabla, A_\Delta \cap B_\Delta, \omega \wedge \pi)\,.\end{aligned}$$

It is easy to verify that the joint configuration space is a distributive lattice with these relations, already indicating that FKG-type inequalities hold for the joint model.
While the monotone down-up dynamics are a fairly fast algorithm in terms of possible updates happening in one step, to understand increasing event properties we introduce Glauber dynamics on the joint model.

1. Decide whether to do a spin or an edge update with probability 1/2.

2a. For spin updates, choose a vertex $v \in V$ uniformly at random and update the spin according to the conditional measure, while keeping other spins and all edges fixed.

2b. For edge updates, choose an edge $e \in E$ uniformly at random and update the edge according to the conditional measure, while keeping other edges and all spins fixed.

It is easy to see that this procedure has a grand coupling by introducing a parameter at Steps 2a and 2b that decides the outcome of the resampling. Moreover, edge resampling is just Glauber dynamics on a random cluster model and therefore monotone in the boundary conditions. Spin resampling is only possible if the chosen spin is not connected to any other vertex, and therefore also easily seen to be monotone in the boundary conditions. Indeed, both steps can be viewed as parts of the monotone down-up dynamics with respect to boundary conditions on all other edges and spins. If we condition this process to additionally not leave a given nonempty increasing event $\mathcal{I}$, we get

Theorem 6.6 (FKG-type inequality)**.** *Consider the setting of Theorem 6.2 and call the respective boundary conditions ϕ^1 and ϕ^2. Let $\mathcal{I}$ be an increasing event for the joint model such that*

$$\phi^2 \cap \mathcal{I} = \{(A,\omega)\colon (A,\omega) \in \phi^2\} \cap \mathcal{I}$$

is nonempty. Then one has

$$\rho^1 \leq \rho^2(\cdot \mid \mathcal{I})\,.$$

Moreover, it holds

$$\rho^2[\mathcal{I} \cap \mathcal{J}] \geq \rho^1[\mathcal{J}]\rho^2[\mathcal{I}]$$

for all increasing events $\mathcal{I}$, $\mathcal{J}$, reprocducing the FKG inequality.

Proof. Let (A,ω) be an element of the set $\phi^2 \cap \mathcal{I}$. Due to $\phi^1 \leq \phi^2$ we have $(A,\omega) \in \phi^1$, such that we can start a coupled monotone Glauber dynamics process with respect to the measures ϕ^1 and $\phi^2(\cdot \mid \mathcal{I})$. The standard Markov chain argument gives the first statement, and the second statement follows immediately, see the proof of Lemma 3.2. □

6.3 Phase properties in the $\mathbb{Z}^2$-planar case

In this section, we transfer phase properties from the random cluster model to the down-up model. For a given boundary condition ϕ, we have already seen that the marginal measure with respect to the edge part is given by

$$\mu^\phi(A) \propto \begin{array}{l} \mathbb{1}[A \cap (F_\nabla \cup F_\Delta) = D_\nabla \cup D_\Delta]\mathbb{1}[W_\nabla \not\leftrightarrow W_\Delta \text{ via } A] \\ \cdot\ p^{|A|}(1-p)^{|E\setminus A|} q^{c^\phi(A)} r^{c(A)-c^\phi(A)} \end{array}$$

where the first factor describes an edge boundary condition and the second factor describes a decreasing event. The exponent $c^\phi(A)$ is given by the number of clusters of A that do not connect to W_∇ or W_Δ. If we consider boundary conditions of the form

$$D_\nabla = F_\nabla = \{\text{edges with endpoints in } W_\nabla\}$$
$$D_\Delta = F_\Delta = \{\text{edges with endpoints in } W_\Delta\}$$

we see that

$$c(A) - c^\phi(A) = |\{\text{clusters of } A \text{ connecting to } W_\nabla \text{ or } W_\Delta\}| = |\{\text{clusters of } F_\nabla \cup F_\Delta\}|$$

is not depending on A. This gives

$$\mu^\phi(A) \propto \begin{array}{l} \mathbb{1}[F_\nabla \cup F_\Delta \subseteq A]\mathbb{1}[W_\nabla \not\leftrightarrow W_\Delta \text{ via } A] \\ \cdot\ p^{|A|}(1-p)^{|E\setminus A|} q^{c(A)} \end{array}$$

and we recover a proper random cluster model, with boundary conditions and the decreasing event condition.

The concrete setting we want to work with now is as follows: Let $G = (V, E)$ be the $\mathbb{Z}^2$-graph, with V_m and E_m the usual subsets describing the quadratic subgraph around the origin. We consider realizable boundary conditions

$$\Phi_m = \left\{ \phi \in \Phi \mid \begin{cases} W_\nabla \cup W_\Delta = V \setminus V_m \\ F_\nabla = A(W_\nabla) = \{\text{edges with endpoints in } W_\nabla\} \\ F_\Delta = A(W_\Delta) = \{\text{edges with endpoints in } W_\Delta\} \end{cases} \right\}$$

such that only vertices in V_m and edges in E_m are not fixed. Even though the underlying graph is infinite, the marginal models ν_m^ϕ on V_m (and ρ_m^ϕ on (V_m, E_m)) are well-defined because every realizable boundary condition can also be formulated using a finite graph. In this setting,

$$\phi_m^{min} = \begin{cases} W_\nabla = V \setminus V_m \\ D_\nabla = F_\nabla \end{cases} \quad \text{and} \quad \phi_m^{max} = \begin{cases} W_\Delta = V \setminus V_m \\ D_\Delta = F_\Delta \end{cases}$$

are the extremal boundary conditions. The induced measures on the edge side are both given by the pure random cluster model on E_m with wired boundary conditions, such that phase properties can be translated to the down-up model. As an analogue to Theorem 3.13, in the subcritical phase we get

Theorem 6.7 (Exponential decay of connectivities in the subcritical phase)**.** *Let $\phi \in \Phi_m$ and $p < p_c(q)$. Then there exist constants $C, c > 0$ such that*

$$\rho^\phi [v \longleftrightarrow w \text{ via } A_\nabla] \leq C \exp(-c\|v - w\|)$$

and

$$\rho^\phi [v \longleftrightarrow w \text{ via } A_\Delta] \leq C \exp(-c\|v - w\|)$$

for all $v, w \in V_m$.

Proof. The event '$v \longleftrightarrow w$ via A_∇' is decreasing, therefore we have

$$\rho^\phi [v \longleftrightarrow w \text{ via } A_\nabla] \leq \rho^{\phi_m^{min}} [v \longleftrightarrow w \text{ via } A_\nabla] \, .$$

Using the connection to the random cluster model measure μ with wired-like boundary conditions we immediately see that

$$\rho^{\phi_m^{min}} [v \longleftrightarrow w \text{ via } A_\nabla] \leq \mu[v \longleftrightarrow w \text{ via } A] \, ,$$

and this probability decays exponentially due to Theorem 3.13. The second statement is obtained analogously. □

In the supercritical phase, consider $\phi = \phi_m^{max}$. For a configuration $(A, \omega) \in \phi$, there exists a unique A_Δ-cluster that connects to the boundary, and we call it the global cluster.

Theorem 6.8 (Exponential decay of connectivities in the supercritical phase)**.** *Let $\phi = \phi_m^{max}$ and $p > p_c(q)$. Then there exist constants $C, c > 0$ such that*

$$\rho^{\phi}\left[v \text{ and } w \text{ are enclosed by an open loop in the dual of } A\right] \leq C \exp(-c\|v-w\|)$$

for all m and $v, w \in V_m$.

Proof. This directly follows from Lemma 5.23. □

Note that Theorem 6.8 implies that

$$\rho^{\phi}[v \longleftrightarrow w \text{ via a non-global cluster of } A]$$

also decays exponentially in $\|v-w\|$ due to being a stronger restriction.

6.3.1 Monotone down-up dynamics in the subcritical phase

We first derive an extremal local SSM property for the down-up model similar to Lemma 5.3. Afterwards, we use the techniques from Chapter 5 to obtain a $\mathcal{O}\left((\log m)^2\right)$ bound for the mixing time of the monotone down-up dynamics with extremal boundary conditions.
Let $R > 0$ and $u \in V_m$ such that $B_R(u) = \{v \in V \colon \|u-v\|_\infty \leq R\}$ is completely contained in V_m, and let $\phi = \phi_m^{max}$ be fixed. We consider expansions of ϕ on $V_m \setminus B_R(u)$ given by

$$\phi^1_{u,R} = \begin{cases} W^1_\nabla = V_m \setminus B_R(u) \\ W^1_\Delta = V \setminus V_m \\ D^1_\nabla = F^1_\nabla = A(W^1_\nabla) \\ D^1_\Delta = F^1_\Delta = A(W^1_\Delta) \end{cases} \quad \text{and} \quad \phi^2_{u,R} = \begin{cases} W^2_\nabla = \emptyset \\ W^2_\Delta = V \setminus B_R(u) \\ D^2_\Delta = F^2_\Delta = A(W^2_\Delta) \end{cases}$$

such that $\phi^1_{u,R} \leq \phi \leq \phi^2_{u,R}$ holds. These boundary conditions naturally come from the usual local/global procedure we apply to monotone Markov chains. We call ν_m the marginal measure on V_m with respect to the boundary conditions ϕ, and we call $\nu^1_{u,R}$ and $\nu^2_{u,R}$ the marginal models on $B_R(u)$ with respect to $\phi^1_{u,R}$ and $\phi^2_{u,R}$.

Lemma 6.9 (Local SSM for extremal boundary conditions)**.** *Let $p < p_c(q)$, and $\phi = \phi_m^{max}$. Let $R > 0$ and $u \in V_m$ such that $B_R(u) = \{v \in V \colon \|u-v\|_\infty \leq R\}$ is completely contained in V_m. Let Λ_r be the set of vertices $v \in B_R(u)$ satisfying*

$$\forall w \in V_m \setminus B_R(u) \colon \|v-w\|_\infty \geq r\,.$$

For the marginals $\nu_{m,r}$, $\nu^1_{u,R,r}$ and $\nu^2_{u,R,r}$ on Λ_r of their respective counterparts ν_m, $\nu^1_{u,R}$ and $\nu^2_{u,R}$ one has

$$\nu^1_{u,R,r} \leq \nu_{m,r} \leq \nu^2_{u,R,r}$$

as well as

$$\|\nu^1_{u,R,r} - \nu^2_{u,R,r}\|_{TV} \leq C \exp(-cr)$$

with constants $C, c > 0$ independent of R, u, r.

Proof. The monotonicity of the marginal models directly follows from Theorem 6.1. For the second statement, consider a coupling (ω_1, ω_2) satisfying

$$\omega_1 \sim \nu^1_{u,R}, \quad \omega_2 \sim \nu^2_{u,R}\,.$$

For such a coupling one has

$$\|\nu^1_{u,R,r} - \nu^2_{u,R,r}\|_{TV} \leq \mathbb{P}[\omega_1 \neq \omega_2 \text{ on } \Lambda_r],$$

and we construct a coupling with decay properties using the connection to the random cluster model. Let $\phi^1 = \phi^1_{u,R}$ and consider the random cluster measure

$$\mu^1 = \mu^{\phi^1} \leq \mu = \mu^{\phi^2}$$

where $\mu = \mu^{F,D}$ is the random cluster measure with wired-like boundary conditions

$$D = F = \{\text{edges with at least one endpoint in } V \setminus B_R(u)\}\,.$$

For a coupling (A_1, A) with $A_1 \sim \mu^1$, $A \sim \mu$ and $A_1 \leq A$ we consider the set of edges

$$\Lambda = \Lambda(A) = \{\text{edges with both endpoints not connected to } V_m \setminus B_R(u) \text{ via } A\}\,.$$

Leaving edges in $E \setminus \Lambda$ fixed, we see that we can do an identical resampling of (A_1, A) on Λ because both conditional measures are the same here. Indeed, the condition $\mathbb{1}[W^1_\nabla \not\leftrightarrow W^1_\Delta \text{ via } A_1]$ is always satisfied for any resampling on Λ, such that we obtain a coupling (B_1, B_2) with

$$B_1 \sim \mu^1, \quad B_2 \sim \mu, \quad B_1 \leq B_2, \quad B_1 = B_2 \text{ on } \Lambda\,.$$

From here, we do a coupled generalized Edwards-Sokal coupling step in the following way:

- For $i \in \{1, 2\}$ obtain ω_i from B_i by assigning to each open cluster of B_i a spin from $\{\nabla, \Delta\}$, and assigning to each vertex of this cluster the given spin.
- This assignment has the following restrictions. If the cluster in question is connected to W^i_∇ of ϕ^i, the chosen spin is ∇. Likewise, if it is connected to W^i_Δ of ϕ^i, the chosen spin is Δ. If the given cluster is not connected to either W_∇ or W_Δ, assign the spin with equal probability $1/2$, and couple this assignment for clusters that align in B_1 and B_2 such that they get the same spin.

This procedure gives us a coupling (ω_1, ω_2) with the desired distribution and additionally

$$\mathbb{P}[\omega_1 \neq \omega_2 \text{ on } \Lambda_r] \leq \mathbb{P}[\Lambda_r \not\subseteq \Lambda]\,.$$

This probability decays exponentially in r because of $p < p_c(q)$, as already seen in the proof of Lemma 5.1. □

Due to the down-up symmetry of the model, a similar statement holds for ϕ_m^{min}. For this proof, it was crucial that boundary information on the common boundary of V_m and $B_R(u)$ is not changed via the expansion of boundary conditions. A similar statement can be derived for boundary conditions with bounded connectivities, which would be more natural in the subcritical phase. For simplicity, we stick to extremal boundary conditions.

Theorem 6.10 (Mixing time of the monotone down-up dynamics in the subcritical phase)**.** *We consider the down-up model ν^ϕ on (V_m, E_m) with the maximal boundary condition $\phi = \phi_m^{max}$. For $p > p_c(q)$ and any $\epsilon > 0$, the monotone down-up dynamics for this model has a mixing time*

$$\tau_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

Proof. The proof proceeds exactly the same as in Chapter 5. Let $\omega^0 \equiv \nabla$ on V_m be the minimal state and $\pi^0 \equiv \Delta$ on V_m be the maximal state. For a given $R > 0$ let $u, v \in V_m$ be vertices with v arbitrary and u chosen such that $B_R(u) \subseteq V_m$ and $\|v - w\|_\infty > R$ for all $w \in V_m \setminus B_R(u)$. We consider the expanded boundary conditions $\phi^1 = \phi^1_{u,R}$ and $\phi^2 = \phi^2_{u,R}$ such that

$$\phi^1 \leq \phi \leq \phi^2 .$$

We now use the monotone grand coupling for monotone down-up dynamics to obtain Markov chains

$$\omega^t_{u,R} \leq \omega^t \leq \pi^t \leq \pi^t_{u,R}$$

where $(\omega^t_{u,R})_{t\geq 0}$ and $(\pi^t_{u,R})_{t\geq 0}$ come from applying the grand coupling to ω^0 with respect to ϕ^1 or to π^0 with respect to ϕ^2 boundary conditions. Therefore we have

$$\mathbb{P}[\omega^t(v) \neq \pi^t(v)] \leq \mathbb{P}[\omega^t_{u,R}(v) \neq \pi^t_{u,R}(v)] = \mathbb{P}[\pi^t_{u,R}(v) = \Delta] - \mathbb{P}[\omega^t_{u,R}(v) = \Delta]$$

and furthermore

$$\begin{aligned}\mathbb{P}[\pi^t_{u,R}(v) = \Delta] - \mathbb{P}[\omega^t_{u,R}(v) = \Delta] \quad \leq \quad & \left|\mathbb{P}[\pi^t_{u,R}(v) = \Delta] - \nu^{\phi^2}[\text{spin at } v \text{ is } \Delta]\right| \\ & + \left|\nu^{\phi^2}[\text{spin at } v \text{ is } \Delta] - \nu^{\phi^1}[\text{spin at } v \text{ is } \Delta]\right| \\ & + \left|\nu^{\phi^1}[\text{spin at } v \text{ is } \Delta] - \mathbb{P}[\omega^t_{u,R}(v) = \Delta]\right| .\end{aligned}$$

The local SSM property from Lemma 6.9 implies that the second term can be bounded by $C\exp(-cR)$. If we choose $R = b\log m$ with $bc > 2$, this bound becomes

$$\left|\nu^{\phi^2}[\text{spin at } v \text{ is } \Delta] - \nu^{\phi^1}[\text{spin at } v \text{ is } \Delta]\right| < \frac{1}{6}(2m+1)^{-2}$$

for m big enough. For the first and third term, we consider $t = T_R$ with $T_R = T_R(1/4)$ being the maximal mixing time of the monotone down-up dynamics on

boxes $B_R(u)$ with the expanded boundary conditions. For $t = bcT_R \log m$, the scaling Lemma 2.7 implies

$$\left|\mathbb{P}[\pi^t_{u,R}(v) = \Delta] - \nu^{\phi^2}[\text{spin at } v \text{ is } \Delta]\right| < \frac{1}{6}(2m+1)^{-2}$$

for m big enough. The same holds for the third term, and overall we get

$$\mathbb{P}[\omega^t(v) \neq \pi^t(v)] < \frac{1}{2}(2m+1)^{-2} = \frac{1}{2}|V_m|^{-1}\,.$$

Summing over all vertices $v \in V_m$, we get

$$\mathbb{E}\left[|\{v \in V_m \mid \omega^t(v) \neq \pi^t(v)\}|\right] \leq \frac{1}{2}$$

which implies

$$\mathbb{P}[\omega^t \neq \pi^t] \leq \frac{1}{2}\,.$$

Therefore we have the mixing time bound

$$\tau_{mix}(1/2) \leq bcT_R \log m\,,$$

where T_R still depends on m via $R = bc \log m$. The standard induction argument gives a mixing time (see for instance Theorem 5.5)

$$\tau_{mix}(1/2) \in \mathcal{O}\left((\log m)^2\right)\,,$$

and the same holds for any fixed $\epsilon > 0$ due to the scaling Lemma 2.7. □

For the optimal mixing time bound of order $\mathcal{O}(\log m)$, a speed of disagreement percolation bound would be helpful. In the down-up dynamics case though, it is not immediately clear how to prove this because the model is non-local, and Step 1 of the monotone down-up dynamics is non-local as well. The censoring framework from [BCV18] could also be helpful.

6.3.2 Monotone down-up dynamics in the supercritical phase

We first state the corresponding local to global coupling result similar to Lemma 5.24, which will be used to obtain a $\mathcal{O}\left((\log m)^2\right)$ bound on the monochromatic mixing time of the monotone down-up dynamics.

Lemma 6.11 (Local to global coupling)**.** *Let $p > p_c(q)$ and $\phi^1 = \phi^{max}_m$. For $u \in V_m$ and $R > 0$ such that $B_R(u) \subseteq V_m$, let $\phi^2 = \phi^2_{u,R}$ such that*

$$\phi^1 \leq \phi^2 \text{ and } \nu^1 \leq \nu^2$$

holds for the models $\nu^1 = \nu^{\phi^1}$ and $\nu^2 = \nu^{\phi^2}$. Let W_r be the set of vertices $v \in V_m$ that have at least distance r to $V_m \setminus B_R(u)$. Then the marginal models ν^1_r and ν^2_r on W_r satisfy

$$\|\nu^1_r - \nu^2_r\|_{TV} \leq C \exp(-cr)$$

with constants $C, c > 0$ that are independent from u, R, r, m.

Proof. Let μ^1 and μ^2 be the corresponding random cluster model measures, which are pure models with wired-like boundary conditions on (V_m, E_m) and $(B_R(u), E_{u,R})$ respectively. Here, $E_{u,R}$ is the set of edges with one endpoint in $B_R(u)$. These models satisfy $\mu^1 \leq \mu^2$ with respect to the partial order in the edge setting, and the rest of the proof is analogous to the proof of Lemma 5.24. □

We define the monochromatic mixing time for the down-up model and a Markov chain P to be

$$\tau^*_{mix}(P, \epsilon) = \min\{t \colon \|\nu^\phi - P^t(\omega^*, \cdot)\|_{TV} \leq \epsilon\}$$

with $\omega^* \equiv \Delta$ on V_m being the maximal state.

Theorem 6.12 (Monochromatic mixing time bound for the monotone down-up dynamics)**.** *Let $\phi = \phi_m^{max}$ and $p > p_c(q)$. The monotone down-up dynamics has monochromatic mixing time*

$$\tau^*_{mix}(\epsilon) \in \mathcal{O}\left((\log m)^2\right) .$$

Proof. The proof proceeds exactly the same as in Theorem 5.25, up to the point where speed of disagreement percolation results are needed. Let $(\omega^t)_{t\geq 0}$ be the monotone down-up dynamics Markov chain with transition matrix P and initial state ω^0 given by

$$\omega^0(v) = \Delta \quad \forall v \in V_m ,$$

i.e. the monochromatic Δ-state. Moreover, let $\pi^0 \sim \nu^\phi$ and consider the Markov chain $(\pi^t)_{t\geq 0}$ coupled to $(\omega^t)_{t\geq 0}$ using the monotone grand coupling of the monotone down-up dynamics to obtain

$$\pi^t \leq \omega^t$$

for $t \geq 0$. The monochromatic mixing time is then bounded by

$$\tau^*_{mix}(P, \epsilon) = \min\{t \colon \|\nu^\phi - P^t(\omega^0, \cdot)\|_{TV} \leq \epsilon\} \leq \min\{t \colon \mathbb{P}[\pi^t \neq \omega^t] \leq \epsilon\} .$$

Let $u \in V_m$, $R > 0$ and consider the local version $(\omega^t_{u,R})_{t\geq 0}$ of the monotone down-up dynamics on $B_R(u)$, coupled to $(\pi^t)_{t\geq 0}$ and $(\omega^t)_{t\geq 0}$ such that

$$\pi^t \leq \omega^t \leq \omega^t_{u,R}$$

for $t \geq 0$ (also starting in the monochromatic Δ-state). Here, we use extremal boundary conditions $\phi^2_{u,R}$ such that $\nu_{u,R} = \nu^{\phi^2_{u,R}}$ is the stationary distribution of this chain. For any $v \in B_R(u)$ one has

$$\mathbb{P}[\pi^t(v) \neq \omega^t(v)] \leq \mathbb{P}[\pi^t(v) \neq \omega^t_{u,R}(v)] = \mathbb{P}[\omega^t_{u,R}(v) = \Delta] - \mathbb{P}[\pi^t(v) = \Delta]$$

due to monotonicity. The Markov chain π^t has been initialized with $\pi^0 \sim \nu^\phi$, therefore $\mathbb{P}[\pi^t(v) = \Delta] = \mathbb{P}[\pi(v) = \Delta \mid \pi \sim \nu^\phi]$. We continue with

$$\begin{aligned} & \mathbb{P}[\omega^t_{u,R}(v) = \Delta] - \mathbb{P}[\pi^t(v) = \Delta] \\ = \quad & \mathbb{P}[\omega^t_{u,R}(v) = \Delta] - \mathbb{P}[\omega_{u,R}(v) = \Delta \mid \omega_{u,R} \sim \nu_{u,R}] \\ & + \mathbb{P}[\omega_{u,R}(v) = \Delta \mid \omega_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\pi(v) = \Delta \mid \pi \sim \nu^\phi] \end{aligned}$$

and see that the first difference will be small for sufficiently high t, while the second difference will be small if v is close to u and R is large. More precisely, let $C, c > 0$ be constants such that Lemma 5.24 holds, and choose $R = b \log m$ with $b > \frac{2}{c}$. For any $v \in V_m$, we can find a $u \in V_m$ such that $B_R(u) \subseteq V_m$ is quadratic and the distance of v to $V_m \setminus B_R(u)$ is at least $R + 1$. For such a pair (u, v), Lemma 6.11 applies with $r = R$ and we get

$$\begin{aligned} & \mathbb{P}[\omega_{u,R}(v) = \Delta \mid \omega_{u,R} \sim \nu_{u,R}] - \mathbb{P}[\pi(v) = \Delta \mid \pi \sim \nu^\phi] \\ \leq \; & \|\nu^0_{u,R} - (\nu^\phi)^0\|_{TV} \\ \leq \; & C \exp(-cR) \\ = \; & C m^{-bc} \end{aligned}$$

with $bc > 2$. If we choose $t = T_R = T_R(1/4)$ to be the monochromatic mixing time of the monotone down-up dynamics on $B_R(u)$ (for quadratic $B_R(u)$), we get

$$\mathbb{P}[\omega^{nt}_{u,R}(v) = \Delta] - \mathbb{P}[\omega_{u,R}(v) = \Delta \mid \omega_{u,R} \sim \nu_{u,R}] \leq \frac{1}{4^n}$$

for $n \in \mathbb{N}$. Choosing $n = bc \log_4(m)$ results in an overall bound

$$\mathbb{P}[\pi^{nt}(v) \neq \omega^{nt}(v)] \leq (C + 1) m^{-bc}$$

with $bc > 2$. For m big enough, we therefore get that

$$\mathbb{E}[|\{v \colon \omega^{nt}(v) \neq \pi^{nt}(v)\}|] \leq |V_m|(C + 1) m^{-bc} \leq \frac{1}{4}$$

and it immediately follows

$$\mathbb{P}[\omega^{nt} \neq \pi^{nt}] \leq \frac{1}{4}\,.$$

Therefore we get for some $m = m_0$ the monochromatic mixing time bound

$$\tau^*_{mix}(P, 1/4) \leq nt = bc \log_4(m) T_R\,,$$

where T_R however still depends on $R = b \log m$. The standard induction argument then gives a monochromatic mixing time bound

$$\tau^*_{mix}(P, 1/4) \in \mathcal{O}\left((\log m)^2\right)\,,$$

and the same holds for all $\epsilon > 0$ due to the scaling Lemma 2.7. □

7 Couplings for the Swendsen-Wang Dynamics

In this chapter we will focus on results for the mixing time of the Swendsen-Wang dynamics Markov chain in the special case of rectangular subgraphs $G = (V_m, E_m)$ of $\mathbb{Z}^2$. We then continue to provide simulation results on couplings for the Swendsen-Wang dynamics in the supercritical phase, suggesting that the existing bounds are not optimal.

7.1 Grand coupling

Let $G = (V, E)$ be a finite graph and $F \subseteq E$ a boundary set with boundary condition $D \subseteq F$. Let $\mu = \mu_{G,p,q}^{F,D}$ be the associated random cluster model with $q \in \mathbb{N}$ and $p \in (0, 1)$. We can formulate a general grand coupling $A \to A'$ for the Swendsen-Wang dynamics in the following way:

1. Generate a random variable $x \in X$, according to some fixed law.
2. For a given configuration $A \subseteq E$, assign to each open cluster C of A a spin $s = f(x, C, A) \in \{1, \dots, q\}$.
3. For $e \in F$, set $e \in A'$ if and only if $e \in D$. For $e \notin F$, draw $r_e \in [0, 1]$ uniformly at random and perform the assignment

$$\begin{cases} e \notin A' & \text{both endpoints have differing spin or } r_e > p\,, \\ e \in A' & \text{else}\,. \end{cases}$$

The function $f(x, C, A)$ has to be chosen such that for all (C, A), each spin $s = s(C)$ has a uniform distribution and is independent of the spins at other clusters. This grand coupling is in general not monotone, and therefore many of the introduced techniques do not apply here. The mechanism with x and f is usually chosen such that a local agreement of states A and B most likely is conserved. We will consider a specific version now.
Let X be the set of pairs (σ, i), where $\sigma\colon V \to \{1, \dots, q\}$ is a spin configuration and $i = (v_1, \dots, v_M)$ is a random enumeration of V. The law of $x = (\sigma, i)$ is given by

the respective uniform distributions. Then, we can define $f(x, C, A) = f(x, C)$ to be the spin $\sigma(v_j)$ with

$$j = \min\{k \colon v_k \in C\}\,.$$

In other words, the spin $s(C)$ is given by the spin at the vertex $v \in C$ which appears first in the random enumeration i. We call this coupling H. For this grand coupling ρ, the following result has been shown in [Hub03].

Theorem 7.1 (Coupling time for the Swendsen-Wang dynamics [Hub03])**.** *Let $F = \emptyset$ and Δ be a bound for the degree of each vertex. If*

$$\gamma = p\left(\frac{\Delta - 2}{1 - p(\Delta - 1)} + 2\right)$$

is less than one, it holds

$$\tau_{coup}(\rho, 1/2) \leq -\log_\gamma(2|E|)\,.$$

Note that this result holds independently of q. If we apply this result to the case where $G = (V_m, E_m)$ is the usual subgraph of $\mathbb{Z}^2$, we get $\Delta = 4$ and therefore

$$\gamma < 1 \quad \Leftrightarrow \quad p < \frac{1}{6}\,.$$

This value is far from the critical threshold $p_c = \sqrt{q}/(\sqrt{q} + 1)$, this result however is still noteworthy as it shows that coupling arguments are possible without using monotonicity properties of the model. In the recent paper [NS19], there appears a very similar result (Proposition 3.4), though it seems to be weaker in this specific case. In [Hub03] there also is a result for very low temperatures.

Theorem 7.2. *[Hub03] Let $F = \emptyset$ and $p \geq 1 - (|E|q)^{-1}$. Then*

$$\tau_{coup}(\rho, 1/2) \leq 2(|E|q)^2\,.$$

This result is rather weak in the setting of subgraphs of $\mathbb{Z}^2$, because p is growing in $|E|$. Moreover, the result is believed to be not optimal in the supercritical phase of the model. The proofs for both of these results do not use derived model properties such as exponential decay of correlations, and the question stands whether stronger results can be derived in the specific setting of $G = (V_m, E_m)$ being a rectangular subgraph of $\mathbb{Z}^2$.

7.2 Spectral gap comparison results

The next big step in understanding the Swendsen-Wang algorithm and other random cluster dynamics has been achieved in the recent works [Ull13], [Ull12a], [Ull12b].

Many spectral gap results for the spin heatbath dynamics, Glauber dynamics and Swendsen-Wang dynamics Markov chains have been elaborated using powerful comparison results.

Theorem 7.3 (Spectral gap results [Ull12a])**.** *The following results hold for the spectral gaps of random cluster dynamics.*

a) On a graph with finite maximum degree Δ, for the q-state Potts model at inverse temperate β one has

$$1 - \lambda_2(SW) \geq q^{-1} q \exp(2\beta)^{-2} (1 - \lambda_2(HB)) \,,$$

where $\lambda_2(SW)$ is the second biggest Eigenvalue of the Swendsen-Wang dynamics and $\lambda_2(HB)$ is the second biggest Eigenvalue of the spin heatbath dynamics.

b) On a graph with $m \geq 3$ edges, one has

$$1 - \lambda_2(G) \leq 1 - \lambda_2(SW) \leq 8m(\log m)(1 - \lambda_2(G))$$

with $\lambda_2(SW)$ the second biggest Eigenvalue of the Swendsen-Wang dynamics (starting from a random cluster configuration) and $\lambda_2(G)$ is the second biggest Eigenvalue of the Glauber dynamics Markov chain.

c) On the rectangular subgraph (V_m, E_m) of $\mathbb{Z}^2$, for $q = 2$ at the critical temperature $p = p_c(q)$ one has

$$1 - \lambda_2(SW) \geq C m^{-c}$$

for some constants $C, c > 0$.

Together with already known results for heatbath dynamics in the subcritical phase, these were the first results for the Swendsen-Wang algorithm that hold for all $p < p_c(q)$. Note that the rapid mixing result for $q = 2$ at the critical temperature transfers to Glauber dynamics as well.

7.3 Spectral gap results in the subcritical phase

In the very recent paper [BCV18], it has been shown that the monotone Swendsen-Wang dynamics on rectangular subgraphs of $\mathbb{Z}^2$ (in the $q = 2$ case) has mixing time $\mathcal{O}(\log m)$ throughout the subcritical phase, and therefore it has a spectral gap which is independent of m. We have seen this result in Chapter 5. Concerning the spectral gap of the Swendsen-Wang dynamics, the following result has been derived in [BCSV19].

Theorem 7.4 (Spectral gap of the Swendsen-Wang dynamics)**.** *Let $G = (V_m, E_m)$ be the usual quadratic subgraph of $\mathbb{Z}^2$ and $q \geq 2$. Throughout the subcritical regime,*

the Swendsen-Wang dynamics Markov chain satisfies

$$\frac{1}{1-\lambda_2(SW)} \in \mathcal{O}(1)\,,$$

where $\lambda_2(SW)$ is the second biggest Eigenvalue of the Swendsen-Wang dynamics Markov chain on G.

This is a substantial improvement to the previously known bounds that follow from [Ull13],[Ull12a],[Ull12b].

7.4 Recent mixing time results

Finally,the paper [BCP$^+$20] released in 2020 provides the optimal mixing time bound in the subcritical phase.

Theorem 7.5 (Mixing time of the Swendsen-Wang dynamics for $p < p_c(q)$)**.** *The Swendsen-Wang dynamics Markov chain on the subgraph (V_m, E_m) with parameter $p < p_c(q)$ has a mixing time $T_{mix}(SW) \in \mathcal{O}(\log m)$.*

This has been achieved by considering the relative entropy of the system, which can be shown to contract at a constant rate. The result generalizes to subcubes of $\mathbb{Z}^d$ in the presence of strong spatial mixing (SSM). In [BCP$^+$20], there also is a mixing time result for the supercritical phase, achieved with similar methods.

Theorem 7.6 (Mixing time of the Swendsen-Wang dynamics for $p > p_c(q)$)**.** *The Swendsen-Wang dynamics Markov chain on the subgraph (V_m, E_m) with parameter $p > p_c(q)$ has a mixing time $T_{mix}(SW) \in \mathcal{O}\left(m^2 \log m\right)$.*

This result is believed to not be the optimal rate (which should be $\log m$).
Overall, this leaves us with the following state of affairs for the Swendsen-Wang algorithm on rectangular subgraphs of $\mathbb{Z}^2$:

- In the subcritical phase $p < p_c(q)$, optimal spectral gap and mixing time results are present. The spectral gap stays constant and the mixing time increases as $\log m$ for increasing system size m.
- For $q = 2$ and $p = p_c(q)$, the spectral gap depends polynomially on m, making the Swendsen-Wang dynamics rapidly mixing.
- In the supercritical phase, there is the (suboptimal) mixing time result of $\mathcal{O}(m^2 \log m)$.

7.5 Experimental results in the supercritical phase

One of the main contributions of this work is the observation that the grand coupling introduced above (due to [Hub03]) shows a remarkably good convergence behavior throughout the supercritical phase of the random cluster model on the subgraph (V_m, E_m), for any $q \geq 2$. We provide numerical evidence to support the conjecture that the mixing time of the Swendsen-Wang dynamics is in $\mathcal{O}(\log m)$ in the supercritical case. Moreover, introducing a small modification, states with different boundary conditions can be coupled to align locally, i.e. with reasonable distance from the boundary.

7.5.1 Coupling time measurements

Consider the rectangular subgraph $G = (V_m, E_m)$ of $\mathbb{Z}^2$ with some boundary conditions and parameters $p \in (0,1)$, $q \in \mathbb{N}$ with $q \geq 2$. Let H be the Swendsen-Wang dynamics grand coupling introduced above, and consider the induced coupled Markov chain (X^t, Y^t) for some initial states $X^0, Y^0 \subseteq E_m$. We investigate the coupling time

$$\tau_{coup} = \tau_{coup}(H, 1/4) = \inf\left\{t \colon \mathbb{P}[X^t \neq Y^t \text{ for any } X^0, Y^0 \subseteq E_m] \leq 1/4\right\}$$

of H in a numerical way. For a monotone coupling, it would suffice to consider the maximal initial states $X^0 = \emptyset$ and $Y^0 = E_m$, however this is not the case for H. For increasing m, the number of possible initial states grows exponentially fast, therefore we have to make an approximation here. Consider

$$\tau_{coup}\left(\{X_1^0, \dots X_n^0\}\right) = \inf\left\{t \colon \mathbb{P}\left[X_1^t = \dots = X_n^t\right] \geq 3/4\right\}$$

as a discrete approximation. We obviously have $\tau_{coup}(\{X_1^0, \dots X_n^0\}) \leq \tau_{coup}$, but hope that both quantities are reasonably close for sufficient conditions on the set $\{X_1^0, \dots X_n^0\}$.

As a quick overview, we begin with the parameter set $m = 32$, $q = 2$, $p \in \{0.05, 0.1, \dots, 0.9, 0.95\}$, $\{\emptyset, E_m\}$ as initial conditions and free boundary condition. We run the simulation $N = 1000$ times and record the coupling times to estimate the quantity $P(t) = \mathbb{P}[X_1^t = \dots = X_n^t]$, with the results seen in Figure 7.1. For low and high p, the coupling probability increases very fast in t. For $p \in \{0.5, 0.55\}$, we see that $P(t)$ is not increasing at all up to $t = 100$. Note that the critical value for our parameter settings is $p_c(2) \approx 0.5858$. This simulation already gives a lot of insight into the properties of the given coupling. In the subcritical phase, the coupling time seems to behave well for $p \leq 0.45$, which is already a great improvement to the known result by Huber. In the supercritical phase, this also seems to be the case for $p > 0.6$.
We continue with a finer increment scheme given by $p \in \{0.4, 0.41, \dots, 0.7\}$. Running the simulation again for $N = 1000$ times, we get the coupling time approximations in Figure 7.2 (the experiment was aborted after $t = 400$ timesteps).
It is evident that in the range $p \in (0.5, 0.6)$ the coupling time does not behave well. We can repeat the same experiment with wired boundary conditions and essentially

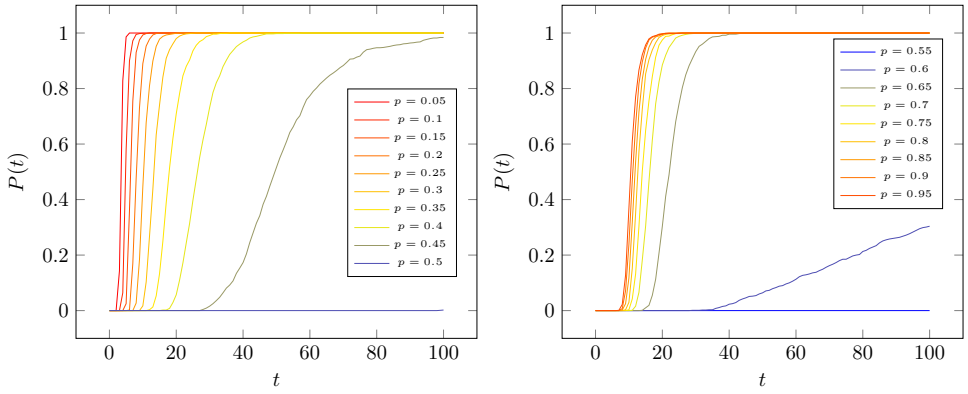

Figure 7.1: Coupling probability $P(t)$ for different values of p. For $p \in \{0.5, 0.55, 0.6\}$ the Swendsen-Wang coupling does not couple fast. This is in the vicinity of the critical value $p_c(2) \approx 0.5858$.

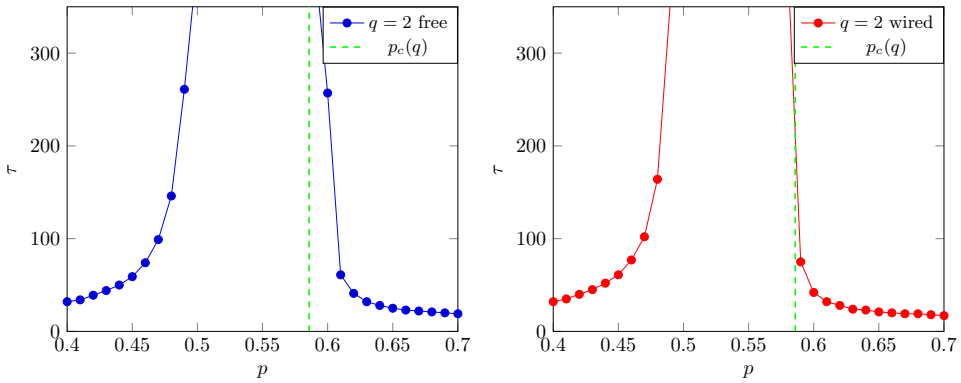

Figure 7.2: Coupling time approximations for $m = 32$, $q = 2$, $p \in \{0.40, 0.41, \ldots, 0.7\}$ and free (left) or wired (right) boundary conditions.

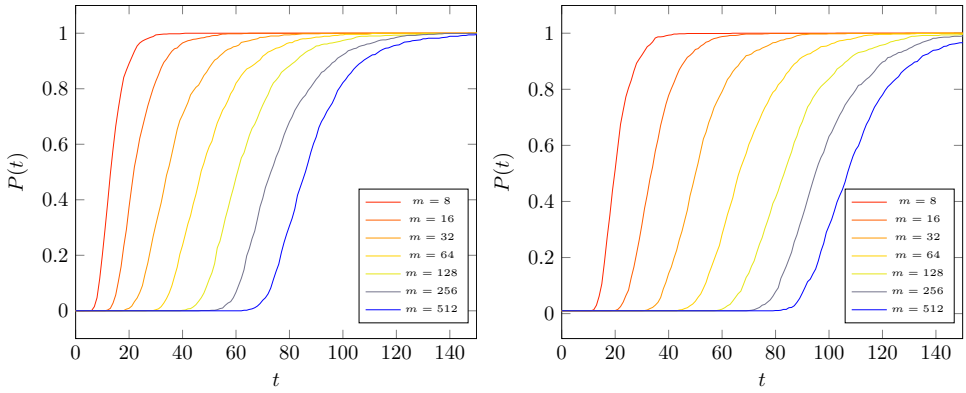

Figure 7.3: On the left: Coupling probability $P(t)$ for increasing m. This simulation had parameters $q = 2$, $p = 0.6$, wired boundary conditions and initial conditions $\{\emptyset, E_m\}$. Averages were obtained from $N = 1000$ samples. On the right: Increased initial conditions X.

get the same picture, with a slightly different critical range for p, see Figure 7.2. Due to $\tau_{coup(\emptyset,E_m)}$ being a lower bound of τ_{coup}, these experiments show the limitations of the coupling in the range $p \in (0.5, 0.6)$ for this parameter set.
The optimal mixing time of the Swendsen-Wang algorithm is already known in the subcritical phase, therefore we will focus on the supercritical phase from now on. We continue to investigate the coupling time approximation, this time for the parameter set $q = 2$, $p = 0.6$, $\{\emptyset, E_m\}$ as initial conditions and wired boundary conditions, with varying system size $m \in \{8, 16, 32, 64, 128, 256, 512\}$.
It is easy to see from Figure 7.3 (left) that the coupling time approximation roughly satisfies

$$\tau_{coup}(\{\emptyset, E_m\}) \in \mathcal{O}(\log_2 m)\,,$$

in this specific case, though it is unclear whether the initial configuration set is sufficient to allow for conclusions regarding the coupling time. We therefore repeat the experiment with an increased initial conditions set

$$X = \{\emptyset, E_m\} \cup \bigcup_{i=0}^{9} \bigcup_{j=0}^{9} X_i^j\,,$$

where X_i^j is distributed according to the percolation law $\mu_{G,p_i,1}$ with $p_i = 0.05 + 0.1 \cdot i$, and independent from X_i^k for $j \neq k$. The results in Figure 7.3 (right) indicate that the qualitative behaviour does not change, which increases the numerical evidence of $\tau_{coup}(\emptyset, E_m)$ being a good approximation to τ_{coup}. Repeating the experiment for $p \in \{0.7, 0.8, 0.9\}$ produces the same behavior, see Figure 7.4.

We continue with a variation of the parameter q. We expand the first experiment with $m = 32$, free and wired boundary conditions respectively, initial conditions $\{\emptyset, E_m\}$ and record the coupling time approximation for $q \in \{2, 3, 4, 5, 6, 7, 8, 9\}$ and $p \in \{0.01, 0.02, \ldots 0.99\}$, repeating the experiment $N = 1000$ times, see Figure 7.6. For increasing q, the range of p-values that do not have a good coupling time slightly

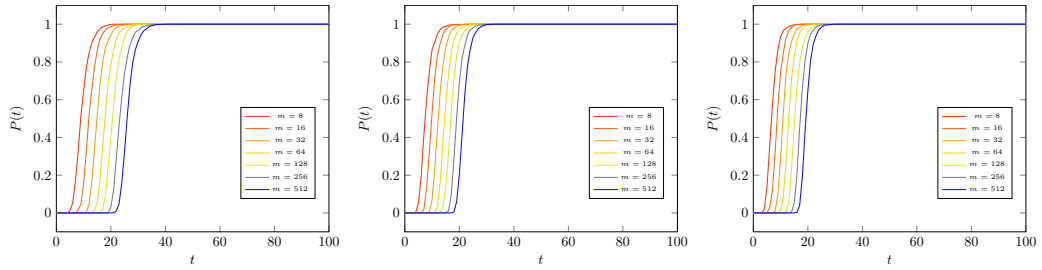

Figure 7.4: Coupling probability $P(t)$ for increasing m, for $p = 0.7$ (left), $p = 0.8$ (middle) and $p = 0.9$ (right). This simulation had parameters $q = 2$, wired boundary conditions and initial conditions $\{\emptyset, E_m\}$. Averages were obtained from $N = 1000$ samples.

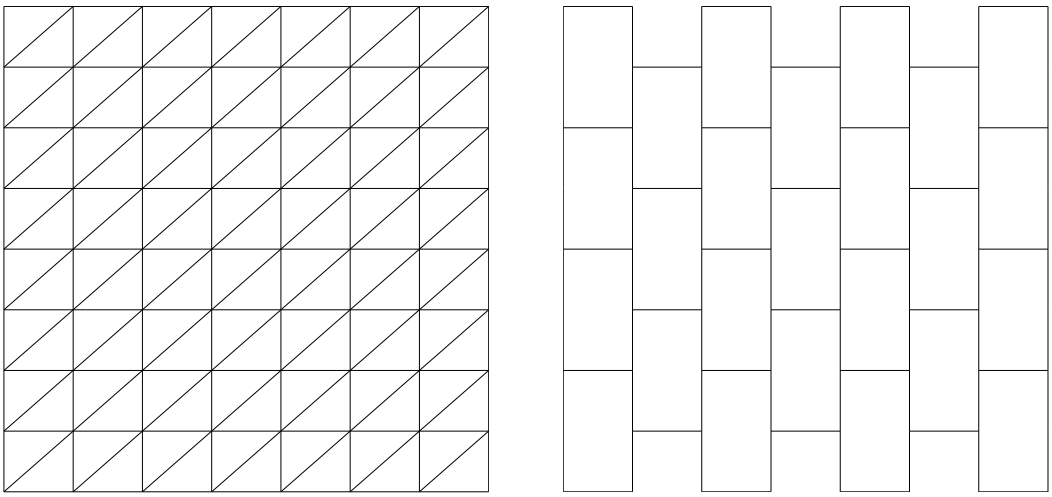

Figure 7.5: Triangular- and hexagonal lattice structures in the plane.

shifts to the right on the p-axis. It is evident that this non-amenable range is tied to the critical value $p_c(q)$. This gives another evidence to the claim that the coupling is sensitive to the phase transition of the random cluster model, and that the coupling works in the supercritical phase of the model.

For $q = 4$ with free boundary conditions, we repeat the experiment at increasing system size $m \in \{8, 16, \ldots, 512\}$ to see how close to the threshold $p_c(q)$ the coupling is still providing good results, see Figure 7.7. It can be seen that the graphs of the coupling time approximation are intersecting each other around $p = 0.7$, which means that the coupling time approximation is not increasing monotonously in m close to the critical value $p_c(q) = 2/3$. This indicates that the coupling increases its efficiency for increasing system size, at a fixed value of $p > p_c(q)$ (in the long run for $m \to \infty$ we expect the $\mathcal{O}(\log m)$ regime to hold regardless).

As a final step, we test the coupling on the triangular and hexagonal lattice, see Figure 7.5. From the theoretical point of view, the random cluster model on these graph classes is similar to the rectangular one, in the sense that the model has a subcritical and a supercritical phase, separated by a sharp phase transition at the critical point $p_c(q)$ which can be computed exactly, see [BDC12].

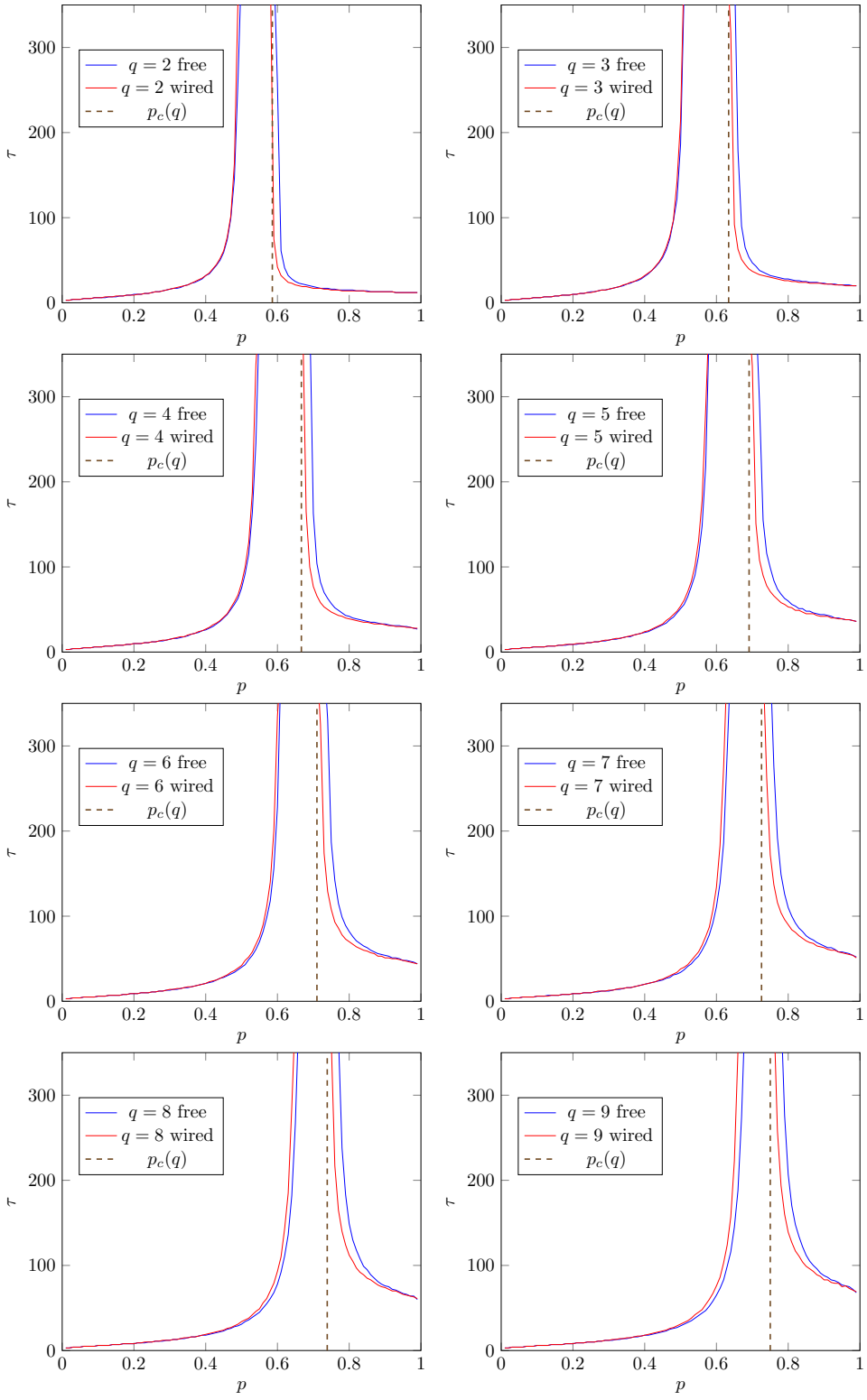

Figure 7.6: Coupling time approximations $\tau = \tau_{coup}(\{\emptyset, E_m\})$ for $q \in \{2, 3, \ldots, 9\}$ and $p \in \{0.01, 0.02, \ldots, 0.99\}$, at fixed system size $m = 32$. Free and wired boundary conditions are taken into account. Results were obtained from $N = 1000$ samples.

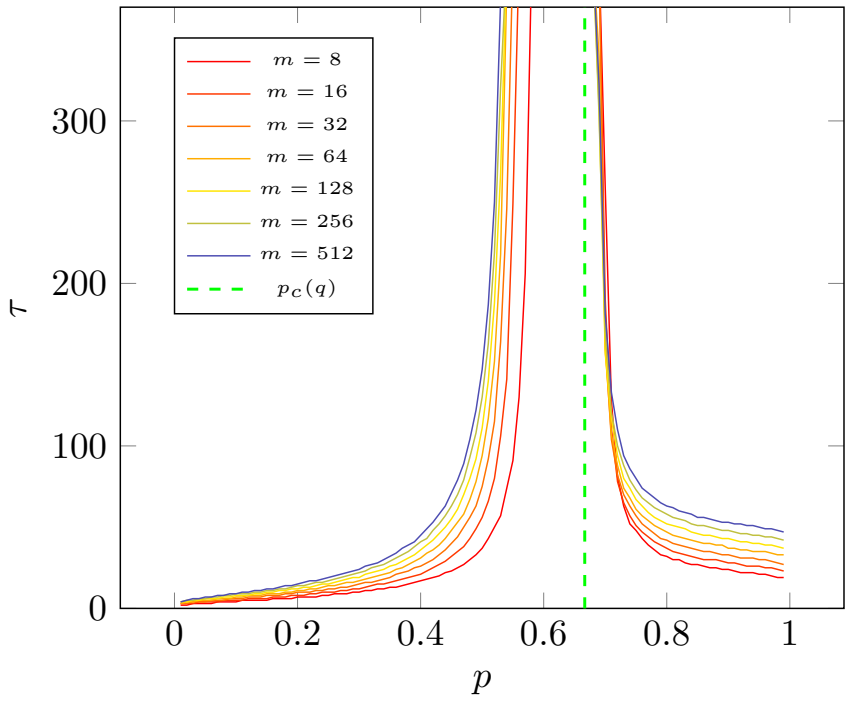

Figure 7.7: Coupling time approximations $\tau = \tau_{coup}(\{\emptyset, E_m\})$ for $q = 4$ and $p \in \{0.01, 0.02, \ldots, 0.99\}$, at varying system size $m \in \{8, 16, 32, 64, 128, 256, 512\}$. Free boundary conditions are taken into account. Results were obtained from $N = 1000$ samples.

We proceed with a rigorous definition of the graphs used. Let $G_\Delta^{m,n} = (V^{m,n}, E_\Delta^{m,n})$ be the graph with vertex set

$$V^{m,n} = \{(v_1, v_2) \in \mathbb{Z}^2 \colon 0 \le v_1 \le m,\, 0 \le v_2 \le n\}$$

and edge set

$$E_\Delta^{m,n} = \{(v, w) \in V \times V \colon v \neq w,\, 0 \le w_1 - v_1 \le 1,\, 0 \le w_2 - v_2 \le 1\}\,.$$

With this definition, the left graph from Figure 7.5 becomes $G_\Delta^{7,8}$. We similarly define $G_\bigcirc^{m,n} = (V^{m,n}, E_\bigcirc^{m,n})$ to have the same vertex set $V^{m,n}$, together with the edge set

$$\begin{aligned} E_\bigcirc^{m,n} = \quad & \{(v, w) \in V \times V \colon v_1 = v_2 \text{ and } w_2 - v_2 = 1\} \\ & \cup \{(v, w) \in V \times V \colon w_1 - v_1 = 1,\, v_2 = w_2 \text{ and } v_1 + w_1 \text{ is an even integer}\}\,. \end{aligned}$$

The right graph in Figure 7.5 is the denoted by $G_\bigcirc^{7,8}$. As is, the graphs are stated with free boundary conditions. For wired boundary conditions, the outermost vertices of $V^{m,n}$ are treated as being connected regardless of the underlying state, and we use the same notation in this case.
We repeat the simulation that produced Figure 7.6 for the triangular graph $G_\Delta^{m,m}$ with $m = 64$ (12416 edges), and for the hexagonal graph $G_\bigcirc^{m-1,m}$ with $m = 100$ (15000 edges). The results can be observed in Figures 7.8 and 7.9. In both cases, the overall behavior aligns with the rectangular lattice case. The Swendsen-Wang coupling H is sensitive to the phase transition and seems to be amenable throughout the supercritical phase.

7.5.2 Limit distribution and locality of the coupling

Consider the usual rectangular subgraph $G = (V_m, E_m)$ of $\mathbb{Z}^2$, together with the random cluster measures μ_0^m, μ_1^m which correspond to free and wired boundary conditions, respectively. Using the coupling procedure of Lemma 5.1, it is possible to define a coupling (X, Y) such that

$$X \sim \mu_0^m, \quad Y \sim \mu_1^m, \quad X \le Y\,,$$

using the Glauber dynamics Markov chain. Moreover, this coupling can be modified such that for any edge e with endpoints v, w, it holds that if v or w is not connected to the boundary via the Y configuration, $X(e) = Y(e)$ follows (here we write $X(e) = 1$ iff $e \in X$ and $X(e) = 0$ iff $e \notin X$). In other words, the coupling aligns on all edges which do not belong to the Y-cluster that is connected to the boundary. In the subcritical phase, this results in an exponential decay of $\mathbb{P}[X(e) \neq Y(e)]$, with respect to the distance of e to the boundary. This was an essential property for the coupling time proofs of Chapter 5. We will call it *locality*, because locally, i.e. away from the boundary, the coupling aligns. Sadly, for the given coupling it is unclear whether this property is present in the supercritical phase, because the range of the Y-cluster that connects to the boundary will encompass the whole system, independent of m.

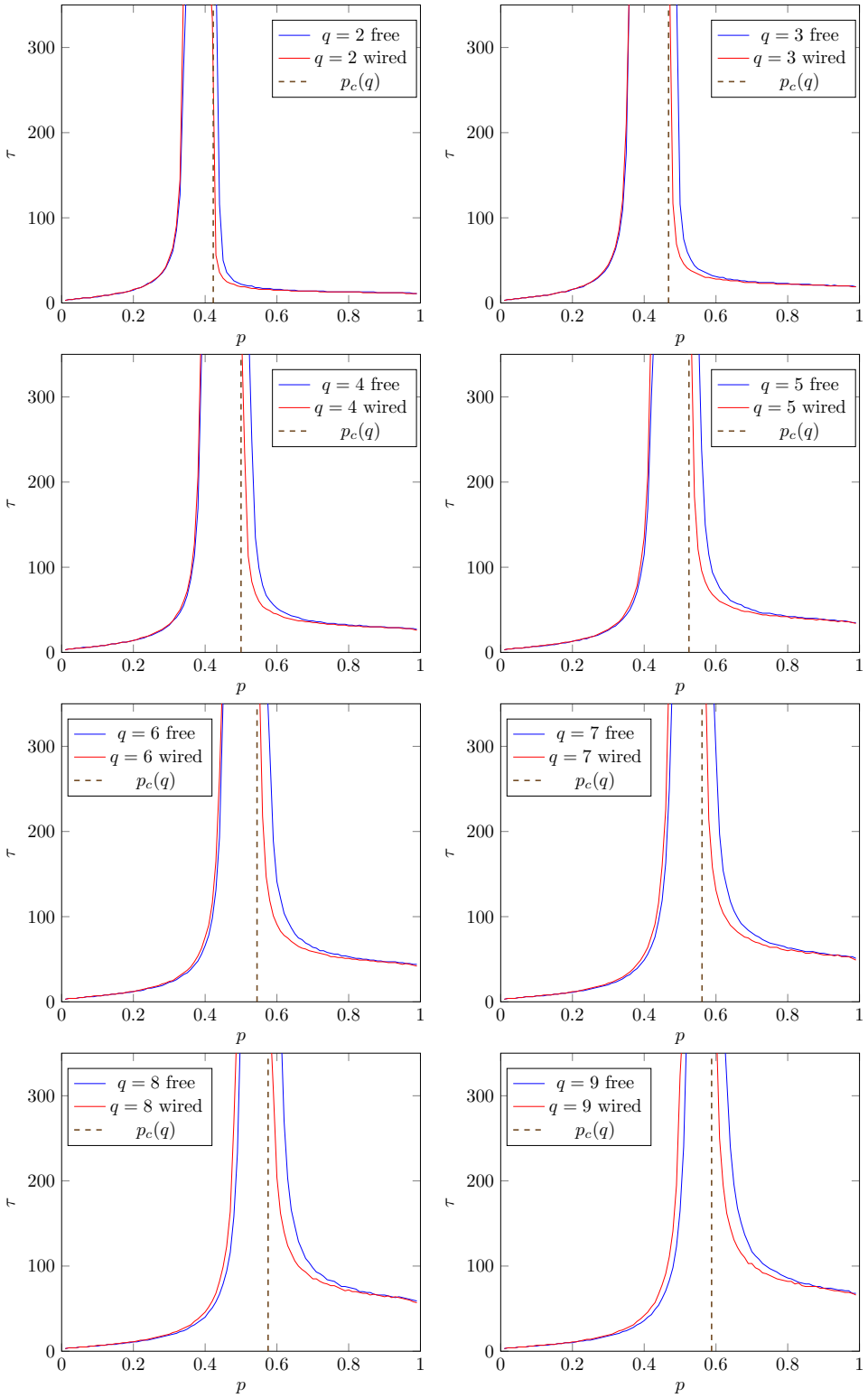

Figure 7.8: Coupling time approximations for the triangular lattice graph $G_{\Delta}^{m,m}$ with $\tau = \tau_{coup}(\{\emptyset, E_{\Delta}^{m,m}\})$ for $q \in \{2, 3, \ldots, 9\}$ and $p \in \{0.01, 0.02, \ldots, 0.99\}$, at fixed system size $m = 64$. Free and wired boundary conditions are taken into account. Results were obtained from $N = 1000$ samples.

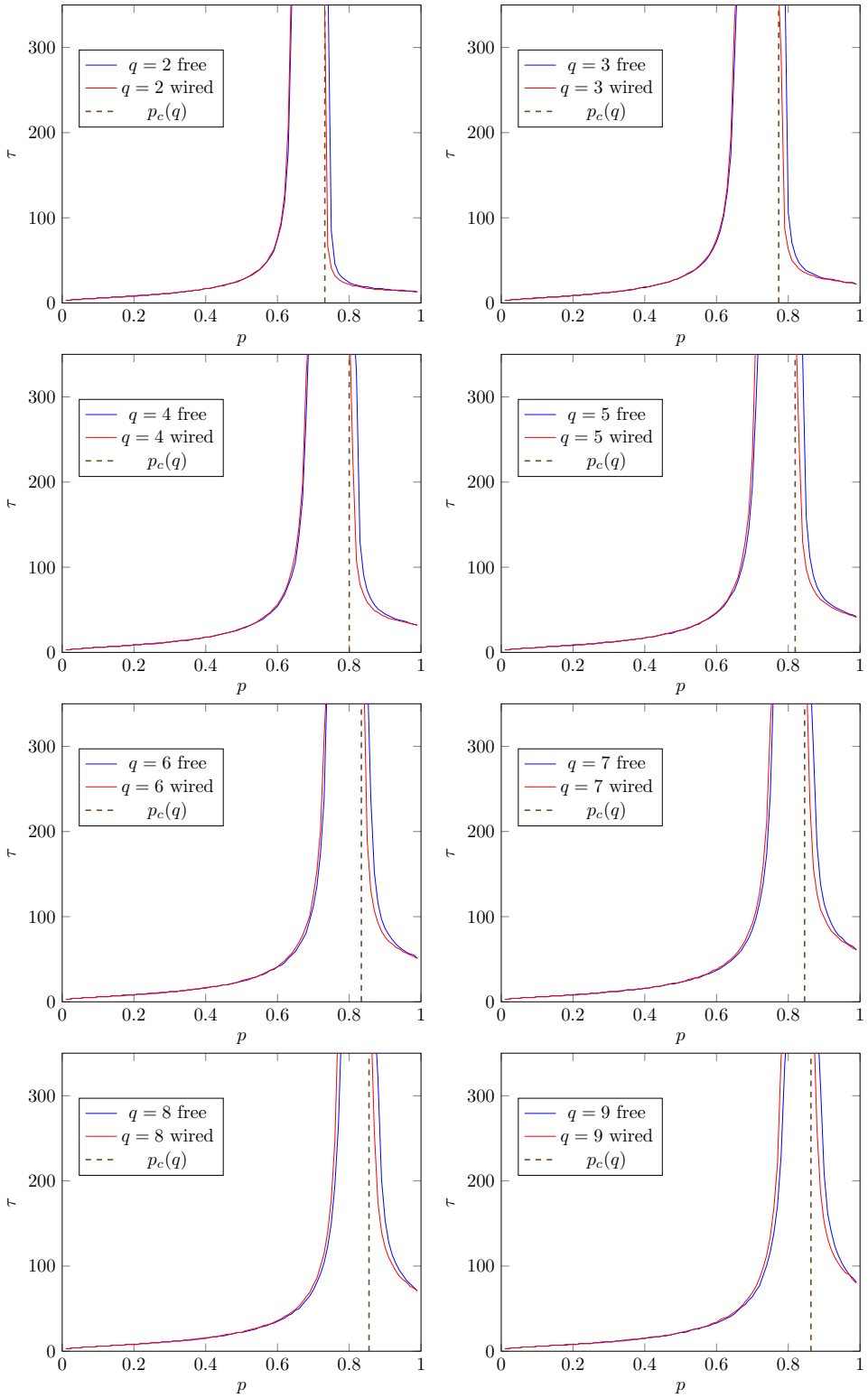

Figure 7.9: Coupling time approximations for the hexagonal lattice graph $G_{\hexagon}^{m-1,m}$ with $\tau = \tau_{coup}(\{\emptyset, E_{\hexagon}^{m-1,m}\})$ for $q \in \{2, 3, \ldots, 9\}$ and $p \in \{0.01, 0.02, \ldots, 0.99\}$, at fixed system size $m = 100$. Free and wired boundary conditions are taken into account. Results were obtained from $N = 1000$ samples.

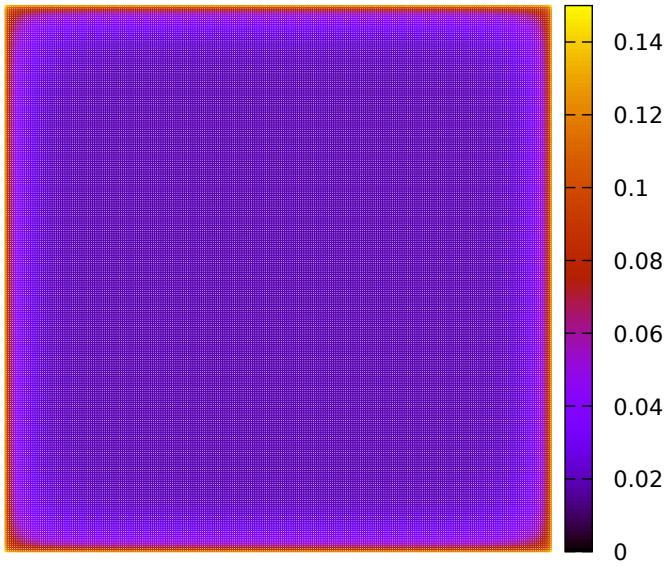

Figure 7.10: Approximation of $D(e)$ for the limit distribution of the H-coupling applied with free and wired boundary conditions. Parameters are given by $q = 2$, $p = 0.6$, $m = 128$. It is evident that the coupling does not have the locality property, because $D(e)$ is not decreasing to 0 with increasing distance to the boundary.

Nevertheless, we can try to find couplings that have this nice property. For instance, we can consider the coupling H and apply it to states X and Y, but *with respect to different boundary conditions*. In this case, clearly the coupling will almost never align completely, and even if it does, this does not imply that subsequent states are aligned. Instead, we can simulate the coupling up to equilibrium and investigate the limit distribution of the coupling and its locality features. We run the following experiment: Let $q = 2$, $p = 0.6 > p_c(q)$, $m = 128$ and $G_m = (V_m, E_m)$ the usual rectangular subgraph of $\mathbb{Z}^2$. Let $X^0 = \emptyset$ and $Y^0 = E_m$ be the initial states. We apply the H coupling to X^0 with respect to free boundary conditions and to Y^0 with respect to wired boundary conditions to obtain the coupled Markov chains (X^t, Y^t). After a burn-in time of $t_0 = 10000$, we track the observables $\mathbb{1}[X^t(e) \neq Y^t(e)]$ for all edges $e \in E_m$ up to $t_1 = 10000000 + t_0$ to approximate the function

$$D(e) = \mathbb{P}[X(e) \neq Y(e)]\,,$$

where (X, Y) is distributed to the limit distribution of the given coupling. The result is given in Figure 7.10.
Close to the boundary $D(e)$ is high. It then decays with increasing distance to the boundary, down to a value $D(e) \approx 0.018$, clearly *not* displaying the desired locality property. Going into the details of the given coupling, we can find a reason why this is the case. Assume that (X, Y) is distributed to the limit distribution, then both X and Y will have a unique biggest (global) cluster (in terms of vertex count) C_X and C_Y. Even if C_X and C_Y align with increasing distance to the boundary, if the random enumeration generated by the coupling implies different root vertices for C_X and C_Y, the disagreement from the boundary can propagate to the whole configuration in

one Step, due to both global clusters getting different spin assignments.
In an attempt to mitigate this problem, we introduce the following modification to the H coupling:

1. Generate a random spin configuration $\sigma\colon V_m \to \{1,\dots,q\}$, and a random enumeration $i = (v_1,\dots,v_M)$ of the vertices.

2. For a given configuration $A \subseteq E$, assign to each open cluster C of A a spin $s = f(C,\sigma,i) \in \{1,\dots,q\}$, where

$$f(C,\sigma,i) = \sigma(v_j), \text{ with } j = \min\{k : v_k \in C\}\,.$$

*. Choose $s^* \in \{1,\dots,q\}$ uniformly at random. Let C_A be the open cluster of A that has the most vertices (if there are multiple such open clusters, choose one of them uniformly at random). Assign to it the spin s^*.

3. For $e \in F$, set $e \in A'$ if and only if $e \in D$. For $e \notin F$, draw $r_e \in [0,1]$ uniformly at random and perform the assignment

$$\begin{cases} e \notin A' & \text{both endpoints have differing spin or } r_e > p\,, \\ e \in A' & \text{else}\,. \end{cases}$$

We call the procedure with the added $*$-Step the H^* coupling. In the supercritical phase of the model, the probability that many biggest clusters exist is practically zero. The modification attempts to align the global clusters of the given cluster configurations. The definition of the global cluster can also be implemented in the following way to avoid the uniqueness issue: A cluster is called global if it connects to all four boundary sides. If such a cluster exists, it is unique for rectangular graphs. Or, in the case of wired boundary conditions, one can define the global cluster to be the unique cluster connecting to any boundary vertex. In the supercritical phase, all of these methods are practically the same. We will however stick to the above definition in this section because this is the version we used in the simulations.
The H-coupling satisfies the following property: If A and B align on two clusters C_1, C_2 that are next to each other, then after applying H, the resulting A' and B' align on all edges belonging to C_1, C_2 *and* all edges connecting C_1 and C_2. For the H^*-coupling, this only holds for non-global clusters C_1, C_2, however an additional property also holds: If A and B align on a cluster C and its boundary to the respective global clusters of A and B, then the resulting A' and B' align on C and on this boundary. This means that in the event that A and B disagree only close to the system boundary, this disagreement will not travel across the whole configuration in one step.
Before we make this statement more rigorous, we repeat the simulation with the H^*-coupling, see Figure 7.11. In this case, the function $D(e)$ has better decay properties. To further investigate this decay, we show an axis-aligned slice through the lattice that passes the origin in Figure 7.12. It is evident that $D(e)$ for the H^*-coupling has very good decay properties. It decays exponentially fast right up to the precision of this MC-simulation, which is $t_1^{-1} = 10^{-7}$ in this case.

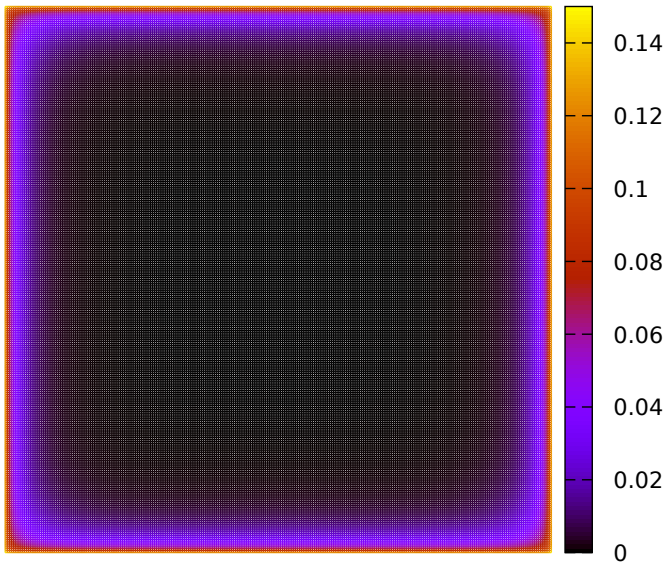

Figure 7.11: Approximation of $D(e)$ for the limit distribution of the H^*-coupling applied with free and wired boundary conditions. Parameters are given by $q = 2$, $p = 0.6$, $m = 128$. The coupling seems to have the locality property, because $D(e)$ is decreasing to 0 with increasing distance to the boundary.

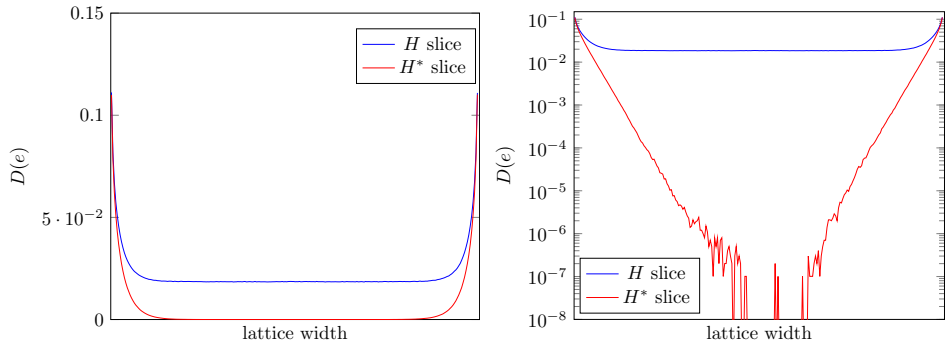

Figure 7.12: Comparison of $D(e)$ for the limit distribution of the H-coupling and H^*-coupling applied with free and wired boundary conditions. Parameters are given by $q = 2$, $p = 0.6$, $m = 128$. The data is taken from an axis-aligned slice through the middle of the lattice. On the left, we have the absolute values of $D(e)$, on the right the value-axis has been scaled logarithmically. Exponential decay for the H^*-coupling seems to be present.

7.6 The H^*-coupling in the supercritical phase

The H^*-coupling seems to have the locality property in the supercritical phase, though it is hard to prove this due to the lack of monotonicity of the Swendsen-Wang dynamics. Nevertheless, we can try to specify the reach of the disagreement percolation for this specific coupling. For this purpose, we will rely on the exponential decay of non-global clusters property from Lemma 5.22. Here, we as well define the global cluster in the *.-Step to be the unique cluster connecting to the boundary.

Theorem 7.7. *Let $G = (V_m, E_m)$ be a rectangular subgraph of $\mathbb{Z}^2$ with wired boundary conditions. Let μ_m be the associated random cluster measure, with $p > p_c(q)$. Let $A^0 \sim \mu_m$ and let B^0 differ from A^0 at a specific edge $e \in E_m$. Let (A^t, B^t) be the time evolution of the H^*-coupling and let*

$$E(e,t) = \bigcup_{s=0}^{t} \{e \in E_m \colon A^s(e) \neq B^s(e)\}$$

be the union of disagreeing edges up to time t. then one has

$$\mathbb{E}[|E(e,t)|] \in \mathcal{O}(t^b)$$

for all $b > 2$.

Proof. Let $\Gamma(e,t)$ be the set of vertices adjacent to $E(e,t)$, i.e. all vertices that are an endpoint of an edge in $E(e,t)$. Let u be an endpoint of e and define $R(t) = \max_{v \in \Gamma(e,t)} \|u - v\|_1$ to be the radius of $\Gamma(e,t)$. We want to bound

$$\mathbb{E}[R(t) \mid R(t-1)]$$

using a tower argument, and therefore need to consider the probability

$$\mathbb{P}[v \in \Gamma(e,t) \mid R(t-1)]\,.$$

This can only be the case if an adjacent edge f is contained in $E(e,t)$. So let f be an edge with endpoints v, w and consider the cases for which $f \in E(e,t)$ is possible. We claim the following: $f \in E(e,t)$ is only possible if v or w is connected to a vertex in $\Gamma(e, t-1)$ via a non-global cluster (in A^{t-1} or B^{t-1}). Assume that this is not the case. Then the A^{t-1}-cluster associated to v is either a non-global cluster not connecting to $\Gamma(e, t-1)$, which implies that this cluster aligns with the B^{t-1}-cluster associated to v. Or it is the global cluster, which implies that the B^{t-1}-cluster associated to v is also the global cluster. In both cases, the spin assignment of Steps 2 and (*) of the H^*-coupling assign the same spin to v in both Markov chains. The same holds for w, therefore the status of edge f will be the same in both Markov chains, which implies $f \notin E(e,t)$.
We have shown that $f \in E(e,t)$ implies that v or w are connected to a vertex in $\Gamma(e, t-1)$ via a non-global cluster. Therefore we get

$$\mathbb{P}[v \in \Gamma(e,t) \mid \Gamma(e,t-1)] \leq \sum_{w \in \Gamma(e,t-1)} C \exp(-c\|v - w\|_2) \leq C' \exp(-c'd)$$

with $d = d(v, \Gamma(e, t-1)) = \min_{w \in \Gamma(e,t-1)} \|v - w\|_1$ and adjusted constants $C', c' > 0$. Now, we proceed as in Lemma 5.16:

$$\begin{aligned}
&\mathbb{E}[R(t) - R(t-1) \mid R(t-1)] \\
&\leq \mathbb{E}\left[\sum_{r \geq 1} \mathbb{1}[\exists v \in \Gamma(e,t) \text{ with } \|u - v\|_1 = R(t-1) + r] \mid R(t-1)\right] \\
&= \sum_{r \geq 1} \mathbb{P}[\exists v \in \Gamma(e,t) \text{ with } \|u - v\|_1 = R(t-1) + r \mid R(t-1)] \\
&\leq \sum_{r \geq 1} \min\left\{1, 4(R(t-1) + r)C' \exp(-c'r)\right\} ,
\end{aligned}$$

where the last inequality comes from a summation over vertices with distance $R(t-1) + r$ to u and the exponential bound from above. As in Lemma 5.16, we get that this sum can be bound by an $\mathcal{O}(\log(R(t-1)))$ term, which implies

$$\mathbb{E}[R(t) \mid R(t-1)] \leq R(t-1) + \mathcal{O}(\log(R(t-1)))$$

and therefore $\mathbb{E}[R(t)] \in \mathcal{O}(t^b)$ for all $b > 1$. The result follows from $|E(e,t)| \in \mathcal{O}(R(t)^2)$. □

8 Conclusion

In this Chapter, we give a short summary of this thesis and provide an outlook for possible extensions of this research.

8.1 Summary

Throughout the course of this thesis, we considered Markov chains for the Fortuin-Kasteleyn random cluster model. Being a classical example of a statistical mechanics model, the only reliable method for the approximation of quantities of interest is given by Markov chain Monte-Carlo methods. In the planar $\mathbb{Z}^2$ case, the model undergoes an order/disorder phase transition, and the main goal was to understand the effect of the phases on the mixing time behavior of Markov chains for this model. To this end, we collected necessary tools for the analysis of Markov chains. We gave an overview of different Monte-Carlo integration schemes for statistical models, and derived basic error estimates for these methods. These are dependent on the mixing behavior of the given Markov chain, and in order to analyze this we introduced the coupling method as a powerful tool. The coupling method can be used to prove mixing time estimates in theory, but it can also be used in practice to obtain insights for the Markov chain at hand.
Afterwards, we gave a thorough introduction to the Fortuin-Kasteleyn random cluster model. We discussed monotonicity features of the model on general graphs, its connection to the Ising and q-state Potts models, as well as planar duality and infinite volume limits. In the planar case, the phase picture of this model is quite well-understood due to recent results concerning the different phases of the model. We stated the essential results for the model on the $\mathbb{Z}^2$-graph, as well as their implication for rectangular subgraphs of $\mathbb{Z}^2$.
Next, we introduced popular Markov chains for the random cluster model and related models. These range from standard Glauber dynamics over block dynamics to systematic scans, and of course the famous Swendsen-Wang algorithm. One main tool for the analysis of these chains is given by monotone couplings, and we discussed in which cases such a coupling is possible.
In the next chapter, we gave an extensive discourse on mixing time results for the introduced monotone Markov chains and block dynamics. Many of the results have been published in parallel to the time this thesis was written, and we outlined

the central proof techniques concerning coupling methods. We contributed with an alternate approach to optimal mixing time results using a speed of disagreement percolation argument, which can be used to bootstrap polynomial mixing time bounds to optimal order. Moreover, we were able to transfer ideas to the supercritical phase of the model in a certain sense, showing new results here.
As an application, we introduced the new down-up model and showed that it has a rich monotone structure that can be exploited in a similar way. This model can be viewed as a generalization of the Ising model for non-integer $q \geq 2$. We showed that the monotone Swendsen-Wang Markov chain has a generalization for this model that is monotone, and used phase properties coming from the random cluster model to prove mixing time results.
In the final chapter, we concerned ourselves with the Swendsen-Wang algorithm, a Markov chain that does not have a monotone coupling and therefore cannot be treated with the methods of the previous chapters. After a discussion on results for this Markov chain (which are mainly concerning the subcritical phase), we focused on the supercritical case of the model. We investigated a grand coupling using numerical methods, attempting to measure quantities related to the mixing time. The results showed that the given grand coupling performs exceedingly well throughout the supercritical phase, and we solidified this observation with an extensive variation of parameters, as well as a variation of underlying graph structures. We discussed a modification of the coupling that numerically shows exponential decay of disagreement with respect to the boundary, a major indication pointing towards optimal mixing time results. We also proved a polynomial speed of disagreement percolation result for this grand coupling similar to those in Chapter 5.

8.2 Outlook

Many of the stated results in this thesis concern monotone Markov chains, requiring monotonicity features of the model. These are present in the random cluster model, Ising model and down-up model. Monotonicity of the model and the Markov chains is a great feature and makes it relatively easy to show local/global connections and properties of the limit distribution for a coupling. A natural extension would be to generalize the results to the Potts model with $q \geq 3$. Monotonicity is missing in this case, however local to global coupling results and speed of disagreement percolation results are still present.
The down-up model is interesting on its own, and its critical behavior should be similar to the random cluster model. This means that this model opens new possibilities for the study of discrete conformal invariance. Smirnov has shown that single $-/+$ interfaces in the rectangular Ising model converge to the Schramm-Loewner-Evolution SLE(3) at criticality, and similar statements might be possible for single ∇/Δ interfaces in the down-up model. Moreover, the model allows the study of observables that relate in a monotonous way in the down-up model, but not in the random cluster model, adding a new layer of depth.
For the Swendsen-Wang algorithm in the supercritical phase, using the given grand coupling should allow us to give optimal mixing time results. One big step to this end would be to show the locality of the coupling, which manifests itself through

exponential decay of disagreement with respect to the boundary. Here as well, monotonicity is not present, and new methods need to be discovered to rigorously validate the observed mixing behaviour.

Acknowledgement

I thank everyone who helped me to complete this thesis. I thank my advisor Prof. Dr. Michael Griebel for introducing me to the topic of high-dimensional integration, as well as for giving me the freedom to explore the related topic of statistical mechanics. Furthermore, I thank Prof. Dr. Michael Griebel and Prof. Dr. Carsten Urbach for many fruitful discussions regarding this thesis. Many thanks to the IT-Staff of the Institute for Numerical Simulation, who never ceased to provide technical support, especially with regard to home office issues during the pandemic.
I thank my family for their constant support over the last years. Most of all, I thank Christina for carefully reading my thesis and being there for me in every situation.

Bibliography

[Ale04] Kenneth S. Alexander. Mixing properties and exponential decay for lattice systems in finite volumes. *Ann. Probab.*, 32(1A):441–487, 01 2004.

[BCP+20] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen-Wang dynamics, 2020.

[BCSV19] Antonio Blanca, Pietro Caputo, Alistair Sinclair, and Eric Vigoda. Spatial mixing and nonlocal Markov chains. *Random Structures & Algorithms*, 55(3):584–614, 2019.

[BCV18] Antonio Blanca, Zongchen Chen, and Eric Vigoda. Swendsen-Wang dynamics for general graphs in the tree uniqueness region, 06 2018.

[BD97] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231, 1997.

[BDC12] Vincent Beffara and Hugo Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probability Theory and Related Fields*, 153(3-4):511 – 542, August 2012. 27 pages, 10 figures.

[Bol96] L. Boltzmann. Vorlesungen über gastheorie, bd. i. 1896.

[BS16] A. Blanca and A. Sinclair. Random-cluster dynamics in $\mathbb{Z}^2$. In *SODA (Symposium On Discrete Algorithms)*, 2016.

[DC17] Hugo Duminil-Copin. Lectures on the Ising and Potts models on the hypercubic lattice, 2017.

[DCGH+16] Hugo Duminil-Copin, Maxime Gagnebin, Matan Harel, Ioan Manolescu, and Vincent Tassion. Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. 11 2016.

[DCST16] Hugo Duminil-Copin, Vladas Sidoravicius, and Vincent Tassion. Continuity of the phase transition for planar random-cluster and Potts models with $1 \leq q \leq 4$. *Communications in Mathematical Physics*, 349(1):47–107, October 2016.

[DSVW02] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. In José D. P. Rolim and Salil Vadhan, editors, *Randomization and Approximation Techniques in Computer Science*, pages 149–163, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[EmcW13] Eren Metin Elçi and Martin Weigel. Efficient simulation of the random-cluster model. *Phys. Rev. E*, 88:033303, Sep 2013.

[FK72] C.M. Fortuin and P.W. Kasteleyn. On the random-cluster model: I. introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[Gib02] J. Williard Gibbs. Elementary principles in statistical mechanics. 1902.

[GOPS11] Timothy M. Garoni, Giovanni Ossola, Marco Polin, and Alan D. Sokal. Dynamic critical behavior of the Chayes–Machta algorithm for the random-cluster model, i. two dimensions. *Journal of Statistical Physics*, 144(3):459–518, Jul 2011.

[Gri02] Geoffrey Grimmett. *The Random-Cluster Model*, volume 110. 06 2002.

[Hub03] Mark Huber. A bounding chain for Swendsen-Wang. *Random Struct. Algorithms*, 22:43–59, 01 2003.

[Isi25] E Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, 31:253–258, 1925.

[JOA19] Pierre E. Jacob, John O'Leary, and Yves F. Atchadé. Unbiased Markov chain Monte Carlo with couplings, 2019.

[KW41] H. A. Kramers and G. H. Wannier. Statistics of the Two-Dimensional Ferromagnet. Part I. *Physical Review*, 60(3):252–262, August 1941.

[LP17] David Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, October 2017.

[Max60] J. C. Maxwell. V. Illustrations of the dynamical theory of gases.—part i. On the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 19(124):19–32, 1860.

[NS19] Danny Nam and Allan Sly. Cutoff for the Swendsen–Wang dynamics on the lattice. *Ann. Probab.*, 47(6):3705–3761, 11 2019.

[Ons44] Lars Onsager. Crystal statistics. i. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, Feb 1944.

[Pei36] R. Peierls. On Ising's model of ferromagnetism. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):477–481, 1936.

[PW96] James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1-2):223–252, 1996.

[SW87] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, Jan 1987.

[Swe83] Mark Sweeny. Monte carlo study of weighted percolation clusters relevant to the Potts models. *Phys. Rev. B*, 27:4445–4455, Apr 1983.

[Sze52] G. Szegö. On certain hermitian forms associated with the Fourier series of a positive function. 1952.

[Ull12a] Mario Ullrich. Rapid mixing of Swendsen-Wang dynamics in two dimensions. *Dissertationes Mathematicae*, 12 2012.

[Ull12b] Mario Ullrich. Swendsen-Wang is faster than single-bond dynamics. *arXiv e-prints*, page arXiv:1201.5793, January 2012.

[Ull13] Mario Ullrich. Comparison of Swendsen-Wang and heat-bath dynamics. *Random Structures & Algorithms*, 42(4):520–535, 2013.

[Yan52] C. N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.*, 85:808–816, Mar 1952.