The University of Melbourne

Doctoral Thesis

The Coupling Time for the Ising Heat-Bath Dynamics & Efficient Optimization for Statistical Inference

Author:
Timothy Hyndman

Supervisors:
Prof. Peter Taylor
Prof. Aurore Delaigle
Assoc. Prof. Tim Garoni

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Operations Research
School of Mathematics and Statistics

September 2018

THE UNIVERSITY OF MELBOURNE

Abstract

Faculty of Science
School of Mathematics and Statistics

Doctor of Philosophy

The Coupling Time for the Ising Heat-Bath Dynamics & Efficient Optimization for Statistical Inference

by Timothy Hyndman

The title page must be followed by an abstract of 300–500 words in English. The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too.

Declaration of Authorship

This is to certify that:

1. the thesis comprises only my original work towards the PhD except where indicated in the Preface,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed: ____

Date: ____

Preface

If applicable, a Preface page includes a statement of:

- Work carried out in collaboration indicating the nature and proportion of the contribution of others and in general terms the portions of the work which the candidate claims as original
- Work submitted for other qualifications
- Work carried out prior to PhD candidature enrolment
- any third party editorial assistance, either paid or voluntary (as limited to the Editing of Research Theses by Professional Editors guidelines) and/or
- Where a substantially unchanged multi-author paper is included in the thesis a statement prepared by the candidate explaining the contributions of all involved. A signed copy by all authors must be included with the submission form.

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor…

Contents

List of Figures

For/Dedicated to/To my…

Chapter 1

Introduction to this Thesis

Initially, this thesis was intended to be made up entirely of the contents of Part II, along with what we hoped would be several significant further contributions to the study. However, the practicalities of a deadline, along with the challenging nature of the research, meant that the decision was made to augment this thesis with an essentially separate section of study. This is what makes up Part I.

The reader should view these two parts as standalone topics, to be read independently. However, they are not without any commonality. Both are within the realm of stochastic mathematics, Part I being a study of a random variable constructed from a stochastic process, and Part II being a study of probability distributions that maximize certain statistical objective functions.

Part I

The Coupling Time for the Ising Heat-Bath Dynamics

Chapter 2

Introduction

2.1 The Ising Model

The Ising model is named after Ernst Ising who studied it in his 1924 thesis [1] under the supervision of Wilhelm Lenz who invented the model [2]. It was originally motivated by the phenomenon of ferromagnetism but it has since been re-interpreted to apply to numerous other situations in both physics and other fields [1].

The Ising model has enjoyed a prominent position in the statistical physics literature. This is largely due to the existence of a phase transition, a sharp transition in the large scale behaviour of the model as a parameter moves past a critical value. The transition was first shown to exist by Rudolph Peierls [4] in what was the first proof of the existence of a phase transition for any model in statistical mechanics. Additionally, the Ising model is both relatively simple, and also mathematically tractable in some non-trivial cases [5]. These qualities are rare among models with a phase transition and so the Ising model has become somewhat of a staple for both studying phase transitions and testing new statistical mechanics techniques.

The model is a probability distribution on spin configurations - assignments of $+1$ and -1 spins to each vertex in a finite graph $G = (V, E)$. The set of all possible configurations is

$$\Omega = \{-1, +1\}^V \tag{2.1}$$

and for a particular configuration, $\sigma \in \Omega$, we refer to the spin of a particular vertex $i \in V$ with $\sigma[i]$. Each configuration has an associated energy, given by

$$H_{G,\beta,h}(\sigma) = -\beta \sum_{ij \in E} \sigma[i]\sigma[j] - h \sum_{i \in V} \sigma[i] \tag{2.2}$$

[1] See [3, notes of Section 1.4.2] for a list of references concerning this.

where $\beta \in [0, \infty)$ is the inverse temperature, and $h \in \mathbb{R}$ is the magnetic field.

The Gibbs measure is the distribution on Ω that characterises the Ising model and it is defined by

$$\pi_{G,\beta,h}(\sigma) \propto \exp(-H_{G,\beta,h}(\sigma)). \tag{2.3}$$

In everything that follows, we will be concerned only with the zero-field ($h = 0$) Ising model. This gives us the slightly simpler form for the Gibbs measure,

$$\pi_{G,\beta}(\sigma) \propto \exp\left(\beta \sum_{ij \in E} \sigma[i]\sigma[j]\right), \qquad \sigma \in \{-1, 1\}^V. \tag{2.4}$$

2.1.1 The Phase Transition

An in depth study of the Ising phase transition and its associated critical temperature will not be needed for this work. However, we will still wish to refer to it occasionally and so here we give a workable description of the phase transition on lattices.

Consider the Gibbs measure with zero-field (2.4) in the limits $\beta \downarrow 0$ and $\beta \uparrow \infty$. It is easy to see that in the former limit, the measure is uniform across all configurations and in the latter limit, the measure assigns all weight to the constant configurations $\sigma^- = (-1, -1, \dots, -1)$ and $\sigma^+ = (+1, +1, \dots, +1)$. This leads to the following overly simplistic description of the phase transition. It is the change in distribution that occurs as we increase the temperature; from distributions concentrated on states whose spins mostly agree, to distributions producing states which have roughly equal numbers of plus and minus spins.

To be slightly more concrete we define quantities called the magnetization and magnetization density. The *magnetization* on a volume $\Lambda \subseteq V$ is defined as

$$M_\Lambda(\sigma) = \sum_{i \in \Lambda} \sigma[i]. \tag{2.5}$$

Normalizing this gives the *magnetization density*, $M_\Lambda(\sigma)/|\Lambda|$. On the d-dimensional torus with side length L, $G(L) = (\mathbb{Z}/L\mathbb{Z})^d$, the quantity

$$L(\beta) = \lim_{L \to \infty} \mathbb{E}_\beta \left| \frac{M_{G(L)}(\sigma)}{|G(L)|} \right| \tag{2.6}$$

depends on the inverse temperature β. When $d = 1$, $L(\beta) = 0$ for any β and there is no phase transition. However, when $d > 1$, there exists some critical $\beta_c(d)$ such that $L(\beta) = 0$ for $\beta < \beta_c(d)$ and $L(\beta) > 0$ for $\beta > \beta_c(d)$ [3]. This $\beta_c(d)$ is the critical inverse temperature at which we observe a phase transition.

2.2 Coupling from the Past

One of the primary concerns regarding the Ising model is how to efficiently sample from the Gibbs measure. Calculating the normalizing constant for (2.4), known as the partition function, is a #P-complete problem [6]. As such a direct approach to sampling is computationally intractable and so other methods must be employed instead. One such method is Markov Chain Monte Carlo (MCMC). This involves constructing a Markov chain whose states are elements of Ω and whose stationary distribution is given by (2.4). One can then obtain a sample by running this Markov chain for long enough that the output has distribution sufficiently close to (2.4). One difficulty in using MCMC is that, initially, one does not know how long to run the chain for. In principal, bounds on this time can be achieved, but in practise, proving these bounds can be very challenging.

An alternative to MCMC was introduced by Propp and Wilson called Coupling from the Past (CFTP) [7]. Unlike MCMC, CFTP not only has an automatically determined running time, but it has the additional advantage of outputting exact samples from the stationary distribution. This does not come without a cost - CFTP has a random running time. Therefore, a key question towards evaluating the effectiveness of CFTP is understanding the distribution of its running time, that is, the *coupling time*.

In Chapters 3 and 4, we will investigate the coupling time for the Ising heat-bath Glauber dynamics, both on the cycle, at any temperature, in Chapter 3, and on any vertex transitive graph, at sufficiently high temperatures, in Chapter 4. Our main result in each chapter will be proving that, when appropriately scaled, the coupling time essentially converges to a Gumbel distribution as the size of the graph increases.

2.2.1 Ising heat-bath Glauber dynamics

The continuous-time heat-bath Glauber dynamics for the Ising model is a Markov chain whose states are elements of Ω and whose stationary distribution is given by (2.4). For a given graph $G = (V, E)$, and a given inverse temperature, β, we can describe the dynamics as follows.

Initialize every vertex in V with a spin (for example, we could start in the all-plus configuration). To each vertex in V we give an i.i.d. rate-one Poisson clock. Define the probability

$$p_i(\sigma) = \frac{e^{\beta S_i(\sigma)}}{e^{\beta S_i(\sigma)} + e^{-\beta S_i(\sigma)}} \tag{2.7}$$

where

$$S_i(\sigma) = \sum_{j \sim i} \sigma[j] \tag{2.8}$$

is the sum of the spins of the neighbours of i, and $j \sim i$ denotes that j is connected to i with some edge $ij \in E$. Let σ_t denote the spin configuration at time t. When the clock of vertex i rings at some time t, we update $\sigma_t[i]$ to $+1$ with probability $p_i(\sigma_t)$, and to -1 otherwise.

2.2.2 The Coupling Time

We now describe the two coupled chains from which we define the coupling time of the Ising heat-bath Glauber dynamics. In order to do this, it will prove convenient to use a random mapping representation for the jump process. This will also help us outline an implementation of CFTP for our coupling.

Define $f : \Omega \times V \times [0,1] \mapsto \Omega$ via $f(\sigma, i, u) = \sigma'$ where $\sigma'[j] = \sigma[j]$ for $j \neq i$ and

$$\sigma'[i] = \begin{cases} 1, & u \leq p_i(\sigma), \\ -1, & u > p_i(\sigma). \end{cases} \tag{2.9}$$

Let $\mathscr{V}$ and U be independent, with $\mathscr{V}$ uniform on V and U uniform on $[0,1]$. Then, updating our chain at rate $n = |V|$, and performing updates from σ to σ' according to $\sigma' = f(\sigma, \mathscr{V}, U)$, we recover the dynamics described in Section 2.2.1.

We also note that f is monotonic, in the following sense. We define a partial ordering on Ω by writing that $\sigma \preceq \omega$ if $\sigma, \omega \in \Omega$ are such that $\sigma[i] \leq \omega[i]$ for all $i \in V$ (and similarly for $\sigma \succeq \omega$). Then for any fixed $i \in V$ and $u \in [0,1]$, if $\sigma \preceq \omega$ then $f(\sigma, i, u) \preceq f(\omega, i, u)$.

Let $(\mathscr{V}_k, U_k)_{k \geq 1}$ be an i.i.d. sequence of copies of $(\mathscr{V}, U)$. Define top and bottom chains, $(\mathscr{T}_t)_{t \geq 0}$ and $(\mathscr{B}_t)_{t \geq 0}$, with initial states

$$\mathscr{T}_0 = (1, 1, \ldots, 1) \tag{2.10}$$

$$\mathscr{B}_0 = (-1, -1, \ldots, -1) \tag{2.11}$$

that update together at rate n. On the kth update at time t_k, update $\mathscr{T}_{t_k}$ to $f(\mathscr{T}_{t_k}, \mathscr{V}_k, U_k)$ and update $\mathscr{B}_{t_k}$ to $f(\mathscr{B}_{t_k}, \mathscr{V}_k, U_k)$.

We call the coupled process, $(\mathscr{B}_t, \mathscr{F}_t)_{t \geq 0}$, *the Ising heat-bath coupling.* From the monotonicity of f, $\mathscr{T}_t \succeq \mathscr{B}_t$, for all $t \geq 0$.

A more descriptive explanation of the coupling is that the top and bottom chains share the same rate-one Poisson clocks at each vertex, and upon updating that vertex, we share the same uniform random variable U between the two chains to determine whether to update to a plus or minus according to (2.9).

The *coupling time* of the Ising heat-bath process is the random variable

$$T = \inf\left\{t : \mathscr{T}_t = \mathscr{B}_t\right\}. \tag{2.12}$$

This is the main object of interest for our analysis. Note that the coupling time is not just a property of the Ising heat-bath process, but also of the coupling we have chosen. In Section 3.1 we will make a change to the coupling we use to make the analysis easier. Some care will need to be taken to verify that the coupling time is not affected by this change.

2.2.3 Equivalence of Discrete and Continuous Coupling Time

So far we have stated that the running time of CFTP has the same distribution as the coupling time. In fact, we have glossed over one important detail. Namely, CFTP is exclusively run in discrete time, and our coupling time is defined by the continuous time dynamics. Therefore, for our motivation to be reasonable, we would like to show some sort of equivalence between the distributions of the discrete and continuous coupling times. We do this via Claim 2.2 which first requires the following Lemma.

Lemma 2.1. *Let $T(k)$ be the sum of k i.i.d. rate λ exponentials. For all $0 < \epsilon < 1$,*

$$\mathbb{P}\left(\left|\frac{T(k)\lambda}{k} - 1\right| \geq \epsilon\right) \leq 2\exp\left(-k\epsilon^2/4\right) \tag{2.13}$$

Proof.

$$\mathbb{P}\left(\left|\frac{T(k)\lambda}{k} - 1\right| \geq \epsilon\right) = \mathbb{P}\left(\frac{T(k)\lambda}{k} \leq 1-\epsilon\right) + \mathbb{P}\left(\frac{T(k)\lambda}{k} \geq 1+\epsilon\right) \tag{2.14}$$

Since $T(k)$ is the sum of k i.i.d. rate λ exponentials, its moment generating function is

$$M_k(t) = \left(\frac{\lambda}{\lambda - t}\right)^k, \qquad\qquad t < \lambda. \tag{2.15}$$

Using a Chernoff bound, for all $0 < t < \lambda$,

$$\begin{aligned}
\mathbb{P}\left(\frac{T(k)\lambda}{k} \geq 1+\epsilon\right) &= \mathbb{P}\left(T(k) \geq \frac{k}{\lambda}(1+\epsilon)\right) && (2.16)\\
&\leq \left(\frac{\lambda}{\lambda - t}\right)^k \exp\left(-\frac{tk}{\lambda}(1+\epsilon)\right) && (2.17)\\
&= \exp\left(k\left(\ln(\lambda/(\lambda - t)) - t(1+\epsilon)/\lambda\right)\right). && (2.18)
\end{aligned}$$

Taking $t = \lambda - \lambda/(1+\epsilon)$,

$$\mathbb{P}\left(\frac{T(k)\lambda}{k} \geq 1+\epsilon\right) \leq \exp\left(k(\log(1+\epsilon)-\epsilon)\right). \tag{2.19}$$

Similarly, for all $t < 0$,

$$\mathbb{P}\left(\frac{T(k)\lambda}{k} \leq 1-\epsilon\right) = \mathbb{P}\left(T(k) \leq \frac{k}{\lambda}(1-\epsilon)\right) \tag{2.20}$$

$$\leq \left(\frac{\lambda}{\lambda - t}\right)^k \exp\left(-\frac{tk}{\lambda}(1-\epsilon)\right) \tag{2.21}$$

$$= \exp\left(k\left(\ln(\lambda/(\lambda-t)) - t(1-\epsilon)/\lambda\right)\right). \tag{2.22}$$

Taking $t = \lambda - \lambda/(1-\epsilon)$,

$$\mathbb{P}\left(\frac{T(k)\lambda}{k} \leq 1-\epsilon\right) \leq \exp\left(k(\log(1-\epsilon)+\epsilon)\right). \tag{2.23}$$

Overall,

$$\mathbb{P}\left(\left|\frac{T(k)\lambda}{k} - 1\right| \geq \epsilon\right) \leq \exp\left(k(\log(1-\epsilon)+\epsilon)\right) + \exp\left(k(\log(1+\epsilon)-\epsilon)\right) \tag{2.24}$$

$$\leq 2\exp\left(k(\log(1+\epsilon)-\epsilon)\right) \tag{2.25}$$

$$\leq 2\exp\left(-k\epsilon^2/4\right) \tag{2.26}$$

for $0 < \epsilon < 1$. □

Claim 2.2. *Let $(N_n)_{n\in\mathbb{N}}$ be a sequence of positive random integers, and $(m_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence of integers such that $N_n \geq m_n$ for all n and $\lim_{n\to\infty} m_n = \infty$. Define $T(n)$ be the random time it takes for a rate λ Poisson clock to go off n times (this is the time it takes for n updates to occur in a continuous-time Markov Chain).*

Let a_n and b_n be positive deterministic sequences such that $b_n/a_n \to \infty$ and for any $\epsilon > 0$

$$\epsilon\frac{m_n a_n^2}{b_n^2} - \log\frac{b_n^2}{a_n^2} \to \infty. \tag{2.27}$$

Define

$$Y_n = \frac{T(N_n) - b_n}{a_n} \tag{2.28}$$

and

$$Z_n = \frac{N_n - \lambda b_n}{\lambda a_n}. \tag{2.29}$$

Let X be a random variable which doesn't place any mass at infinity. Then $Y_n \xrightarrow{d} X$ if and only if $Z_n \xrightarrow{d} X$.

Proof. In order to prove either direction, it is sufficient to show that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Z_n| > \epsilon) = 0. \tag{2.30}$$

First note that

$$|Y_n - Z_n| = \left|\frac{T(N_n) - b_n}{a_n} - \frac{N_n - \lambda b_n}{\lambda a_n}\right| \tag{2.31}$$
$$= \left|\frac{T(N_n)}{a_n} - \frac{N_n}{\lambda a_n}\right| \tag{2.32}$$
$$= \left|\frac{T(N_n)\lambda}{N_n} - 1\right| \frac{N_n}{\lambda a_n}. \tag{2.33}$$

So for any $\epsilon > 0$

$$\mathbb{P}(|Y_n - Z_n| > \epsilon) \le \mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \epsilon\frac{a_n}{4b_n}\right) + \mathbb{P}\left(\frac{N_n}{\lambda a_n} > \frac{4b_n}{a_n}\right). \tag{2.34}$$

We will show that both of the terms on the right hand side vanish as $n \to \infty$. We start with the first of these.

Since $N_n \ge m_n$,

$$\mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \epsilon\frac{a_n}{4b_n}\right) \le \mathbb{P}\left(\sup_{k\ge m_n}\left|\frac{T(k)\lambda}{k} - 1\right| > \epsilon\frac{a_n}{4b_n}\right) \tag{2.35}$$
$$= \mathbb{P}\left(\bigcup_{k\ge m_n}\left\{\left|\frac{T(k)\lambda}{k} - 1\right| > \epsilon\frac{a_n}{Cb_n}\right\}\right) \tag{2.36}$$
$$\le \sum_{k=m_n}^{\infty} \mathbb{P}\left(\left|\frac{T(k)\lambda}{k} - 1\right| > \epsilon\frac{a_n}{4b_n}\right). \tag{2.37}$$

To apply Lemma 2.1, we need that $\epsilon a_n/(4b_n) < 1$. However, since $a_n/b_n \to 0$, we can ensure this holds by taking n large enough. Continuing,

$$\sum_{k=m_n}^{\infty} \mathbb{P}\left(\left|\frac{T(k)\lambda}{k} - 1\right| > \epsilon\frac{a_n}{4b_n}\right) \le 2\sum_{k=m_n}^{\infty} \exp\left(-k\epsilon^2\frac{a_n^2}{64b_n^2}\right) \tag{2.38}$$
$$= 2\frac{\exp\left(-\epsilon^2 a_n^2(m_n-1)/\left(64b_n^2\right)\right)}{\exp\left(\epsilon^2 a_n^2/\left(64b_n^2\right)\right) - 1}. \tag{2.39}$$

Since $x \le \exp(x) - 1$ for $0 \le x \le 1$, for sufficiently large n,

$$2\frac{\exp\left(-\epsilon^2 a_n^2(m_n-1)/\left(64b_n^2\right)\right)}{\exp\left(\epsilon^2 a_n^2/\left(64b_n^2\right)\right) - 1} \le \frac{128b_n^2}{a_n^2\epsilon^2}\exp\left(-\epsilon^2 a_n^2(m_n-1)/\left(64b_n^2\right)\right) \tag{2.40}$$
$$\le 256\frac{b_n^2}{a_n^2\epsilon^2}\exp\left(-m_n\epsilon^2 a_n^2/\left(64b_n^2\right)\right) \tag{2.41}$$

By (2.27), this goes to zero as $n \to \infty$.

To bound the second term in (2.34), we will treat the two directions of the proof separately. Firstly, assume that $Z_n \xrightarrow{d} X$. Then note that

$$\mathbb{P}\left(\frac{N_n}{\lambda a_n} > \frac{4b_n}{a_n}\right) = \mathbb{P}\left(Z_n > 3\frac{b_n}{a_n}\right) \tag{2.42}$$

and since $b_n/a_n \to \infty$, and Z_n converges to a distribution which places no mass at ∞,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{N_n}{\lambda a_n} > \frac{4b_n}{a_n}\right) = 0. \tag{2.43}$$

Now assume that $Y_n \xrightarrow{d} X$. Note that if $T(N_n)/a_n < \epsilon/2$ and $|T(N_n)\lambda/N_n - 1| < 1/2$, then $N_n/(\lambda a_n) < \epsilon$. So

$$\mathbb{P}\left(\frac{N_n}{\lambda a_n} > \frac{4b_n}{a_n}\right) \leq \mathbb{P}\left(\frac{T(N_n)}{a_n} < \frac{2b_n}{a_n}\right) + \mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \frac{1}{2}\right) \tag{2.44}$$

$$= \mathbb{P}\left(Y_n < \frac{b_n}{a_n}\right) + \mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \frac{1}{2}\right). \tag{2.45}$$

As above, since $a_n/b_n \to \infty$, and Y_n converges to a distribution which places no mass at ∞, the first term vanishes in the limit. The second disappears since

$$\mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \frac{1}{2}\right) \leq \mathbb{P}\left(\left|\frac{T(N_n)\lambda}{N_n} - 1\right| > \epsilon\frac{a_n}{4b_n}\right) \tag{2.46}$$

for large enough n.

□

Remark 2.3. We apply Claim 2.2 to the coupling time of the Glauber heat-bath dynamics in the following way. Take N_n to be the discrete coupling time on a graph of size n. The continuous time coupling time is given by $T(N_n)$. Note that $N_n \geq m_n = n$ since each vertex must be updated at least once for coupling to occur. Finally Theorems 3.1 and 4.1 establish the limiting distribution of the continuous-time coupling time using scaling and shifting sequences a_n and b_n whos ratio is

$$\frac{b_n}{a_n} = \log n \tag{2.47}$$

and thus (2.27) is satisfied. This means that, appropriately scaled, the discrete-time coupling time has the same limiting distribution as the continuous-time coupling time.

2.2.4 Summary of CFTP

We are now in a position to give a brief summary of the CFTP method, as it applies to the Ising heat-bath coupling. It should be noted that we include this summary of CFTP for completeness. None of the details regarding the implementation of CFTP are required outside of this section. It serves only as motivation for the study of the coupling time.

Let $f : \Omega \times V \times [0,1] \mapsto \Omega$ and $(\mathcal{V}, U)$ be as defined in Section 2.2.2. Let $(\mathcal{V}_k, U_k)$ be an i.i.d. sequence of copies of $(\mathcal{V}, U)$ and define

$$f_{-k} = f(\cdot, \mathcal{V}_k, U_k). \tag{2.48}$$

We construct the composition

$$F_{-k} = f_0 \circ f_1 \circ \cdots \circ f_{k-1} \tag{2.49}$$

and define the *backwards coupling time* to be

$$T_{\text{BACK}} = \min\{k \in \mathbb{N} : F_{-k}(\mathscr{B}_0) = F_{-k}(\mathscr{T}_0)\}. \tag{2.50}$$

The state $F_{-T_{\text{BACK}}}(\mathscr{B}_0) = F_{-T_{\text{BACK}}}(\mathscr{T}_0)$ is the output of the CFTP algorithm, and was shown by Propp and Wilson [7] to be an exact sample from the chain's stationary distribution. To see why this is so, observe that by the monotonicity of f, if $F_{-k}(\mathscr{B}_0) = F_{-k}(\mathscr{T}_0)$, then $F_{-k}(\sigma) = F_{-k}(\mathscr{B}_0)$ for any $\sigma \in \Omega$. If we let σ_π be a random sample from the stationary distribution π, then $F_{-k}(\mathscr{B}_0) = F_{-k}(\mathscr{T}_0) = F_{-k}(\sigma_\pi)$ must also have distribution π, which in our case is given by (2.4).

If we reverse the composition to construct

$$F_k = f_k \circ f_{k-1} \circ \cdots \circ f_1 \tag{2.51}$$

we can define the usual discrete time coupling time as

$$T_{\text{DIS}} = \min\{k \in \mathbb{N} : F_k(\mathscr{B}_0) = F_k(\mathscr{T}_0)\}. \tag{2.52}$$

The forwards coupling time, T_{DIS}, has the same distribution as the backwards coupling time, T_{BACK}, although in general, $F_{T_{\text{DIS}}}(\mathscr{B}_0) = F_{T_{\text{DIS}}}(\mathscr{T}_0)$ does not have distribution (2.4).

In practise, one runs the CFTP algorithm by starting both the top and bottom chains from some point in the past to time zero. This is repeated for increasingly more distant times in the past until both chains agree at time 0. The sequence of times at which one restarts this process need not be $-1, -2, -3, \ldots$, rather, any monotonic natural sequence $a_1, a_2, \ldots$ can be used. See [8], [9], and [10] for further discussion.

2.3 Information percolation

A cornerstone to the proofs contained in Chapters 3 and 4 is the framework of information percolation, first introduced by Lubetzky and Sly in 2016 [11]. In this paper, Lubetzky and Sly managed to achieve much sharper results, in much more generality, regarding the mixing time for the Glauber dynamics for the Ising model than had been achieved before. In this section we provide a brief summary of their paper before laying out the basic framework, in the context of the Ising heat-bath dynamics, that will be required for Chapters 3 and 4.

2.3.1 Information percolation and cutoff for the stochastic Ising model

In order to define cutoff, the central phenomenon of study in Lubetzky and Sly's 2016 paper titled, 'Information percolation and cutoff for the stochastic Ising model', we first have to define the total-variation mixing time. Given a parameter ϵ, a Markov Chain Y_t has mixing time

$$t_{\mathrm{MIX}}(\epsilon) = \inf\left\{ t : \max_{x_0 \in \Omega} ||\mathbb{P}(X_t \in \cdot | X_0 = x_0) - \pi||_{\mathrm{TV}} \leq \epsilon \right\} \tag{2.53}$$

where the total variation distance $||\nu_1 - \nu_2||_{\mathrm{TV}}$ is defined as

$$\max_{A \in \Omega} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{\sigma \in \Omega} |\nu_1(\sigma) - \nu_2(\sigma)|. \tag{2.54}$$

A family of Markov chains (Y_t) indexed by n is said to exhibit cutoff if

$$t_{\mathrm{MIX}}(\epsilon) = (1 + o(1)) t_{\mathrm{MIX}}(\epsilon'), \tag{2.55}$$

for any fixed $0 < \epsilon, \epsilon' < 1$. A *cutoff window* is a sequence w_n where

$$t_{\mathrm{MIX}}(\epsilon) = t_{\mathrm{MIX}}(1 - \epsilon) + \mathcal{O}(w_n) \tag{2.56}$$

for any $0 < \epsilon < 1$.

Historically, proving cutoff has proven to be highly challenging. In a survey on the topic, Diaconis [12] wrote 'proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors'. It is therefore worth noting the significant gap between the strength of the results regarding cutoff achieved using information percolation, and those that existed previously.

Previous to [11], the best result known for general graphs was that cutoff occurs with a $\mathcal{O}(1)$ window in the simple case when $\beta = 0$ [13]. However, no results were known for $\beta > 0$, despite a conjecture by Levin et al. in 2009 [8, Section 23.2] that cutoff occurs on any sequence of transitive graphs when the mixing time is of order $\log n$ (as one would expect when $\beta < c_0$ for some $c_0 > 0$ that depends on the sequence of graphs). On lattices, the first results to appear were due to Lubetzky and Sly in 2013 who established cutoff up to the critical temperature for dimensions $d \leq 2$ with only a $\mathcal{O}(\log \log n)$ window [14].

Using information percolation, Lubetzky and Sly proved the existence of cutoff for the continous time Glauber dynamics for the Ising model with an $\mathcal{O}(1)$ window on $\mathbb{Z}^d$ for all temperatures up to the critical temperature. In a companion paper [15], they extended this result to include any graph for sufficiently high temperature. Since these papers, information percolation has also been used to establish cutoff for the Swendsen-Wang dynamics on the lattice [16], suggesting that the technique is effective on a broader class of problems than simply Glauber dynamics for Ising.

2.3.2 The framework

At its core, information percolation is a way of tracking how the dependencies of the final spins of the Glauber heat-bath dynamics percolate through the graph over time. These dependencies are traced backwards through time from some designated time t^* on the space-time slab $V \times [0, t^*]$ (see Figure 2.1 for example) to create the update history. These histories are made in such a way so that, if for any $j \in V$ no path exists connecting (i, t^*) to $(j, 0)$, then the spin of i does not depend on the initial state (and thus at time t^* vertex i takes $+1$ and -1 spins with equal probability by symmetry). The main constructs used to create this history are the update sequence, and the update support function which we will now define.

2.3.2.1 The update sequence

Recalling our random mapping representation from Section 2.2.2, we can encode an update of our coupled process with the tuple $(\mathcal{V}, U, t)$, where t is the time of the update, $\mathcal{V}$ is the vertex that is updated, and U is the value of the uniform random variable that tells us whether $\mathcal{V}$ is a plus or minus according to (2.9). The *update sequence* along an

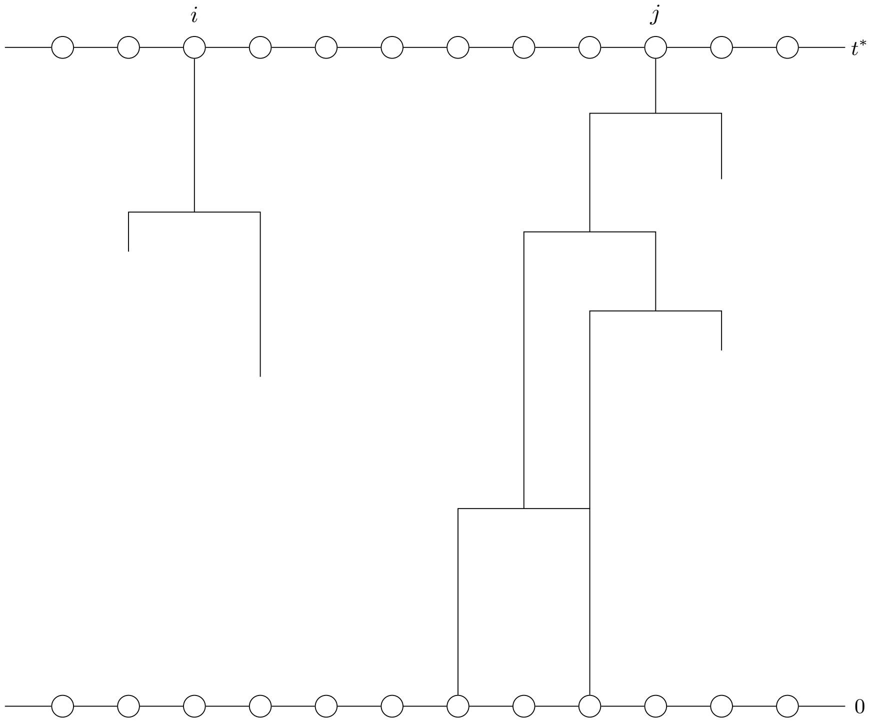

Figure 2.1 – *A section of the space-time slab $V \times [0, t^*]$ along with a typical appearance of the update histories for two vertices on the cycle. Time runs vertically from bottom to top, and the vertices are represented by circles, laid out horizontally. If there is a path in the update history of v between points (u, t) and (v, t^*), then the spin of v at time t^* depends on the spin of u at time t. In this example, since there is no path from vertex (i, t^*) to time 0, the final spin at i does not depend on the initial configuration whereas the final spin at j does.*

interval $(t_0, t_1]$ is the set of these tuples with $t_0 < t \leq t_1$. Given the state of our Markov Chain at time t_0, Y_{t_0}, the update sequence along $(t_0, t_1]$ contains all the information we need to contruct Y_{t_1}. In particular, given the update sequence along the interval $(0, t_1]$, Y_{t_1} is a deterministic function of Y_0.

2.3.2.2 The update support function

Given the update sequence along the interval $(t_1, t_2]$, the *update support function*, $\mathscr{F}(A, t_1, t_2)$, is the minimal set of vertices whose spins at time t_1 determine the spins of the vertices in A at time t_2. That is, $i \in \mathscr{F}(A, t_1, t_2)$ if and only if there exist states $Y_{t_1}, Y'_{t_1} \in \{-1, +1\}^V$ that differ only at i and such that when we construct Y_{t_2} and Y'_{t_2} using the update sequence, $Y_{t_2} \neq Y'_{t_2}$.

In particular, if $\mathscr{F}(i, 0, t) = \emptyset$ then the spin at vertex i at time t does not depend on the initial state and so for any two coupled chains Y and Y', $Y_t[i] = Y'_t[i]$. As a consequence of the monotonicity of our coupling, we can make the stronger statement that $\mathscr{T}_t[i] = \mathscr{B}_t[i]$ if and only if $\mathscr{F}(i, 0, t) = \emptyset$ which of course means that

$$\mathbb{P}[\mathscr{T}_t[i] \neq \mathscr{B}_t[i]] = \mathbb{P}[\mathscr{F}(i, 0, t) \neq \emptyset]. \tag{2.57}$$

For ease of notation, we will often use the shorthand

$$\mathcal{H}_i(t) := \mathscr{F}(i, t, t^*). \tag{2.58}$$

where t^* is some target time that should be clear from context. We call this the *update history of vertex i at time t*. Tracing $\mathcal{H}_i(t)$ backwards in time from t^* produces a subgraph of $\Omega \times [0, t^*]$ which we write as $\mathcal{H}_i$ and which we simply call the *update history of vertex i*. To be slightly more precise, to produce $\mathcal{H}_i$ we connect (j, t) to (j, t') if $j \in \mathcal{H}_i(t)$ and there are no updates of j along $(t', t]$ and we connect (j, t) to (j', t) if there was an update at (j, t), $j \in \mathcal{H}_i(t)$, $j' \notin \mathcal{H}_i(t)$, and $j' \in \mathcal{H}_i(t + \epsilon)$ for any sufficiently small $\epsilon > 0$.

To give some intuition to the definitions above, consider how we might construct the update history of a vertex i from some target time t^*. We have at our disposal the update sequence along $(0, t^*]$ which we place in order of decreasing time. If vertex i does not appear in the update sequence then we create a temporal edge between (i, t^*) and $(i, 0)$ and our update history is complete - vertex i was never updated and so it simply takes its initial value. Otherwise, we create temporal edge between (i, t^*) and (i, t_i) where t_i is the last time vertex i was updated. At this point we note from (2.9) that the spin that vertex i takes due to this update depends on the spins of its neighbours. So we add spatial edges from (i, t_i) to (j, t_i) for each $j \sim i$. Finally, we can iterate this process for each neighbour until every history has reached time 0.

In Figure 2.2 we have followed this procedure to show how we might create the update history from a single vertex on the cycle. This construction certainly contains every vertex that might influence the final spin at i, that is, it contains $\mathcal{H}_i$ as a subgraph. However, it is possible for updates to occur that do not depend on neighbouring spins. These updates cause temporal edges leading up to them to terminate without branching out to the neighbouring vertices. These type of updates are called *oblivious updates*.

2.3.2.3 Oblivious updates

Generally speaking, an update to a vertex is oblivious if we do not need to know the configuration of its neighbours to determine the spin of that vertex. More precisely, an

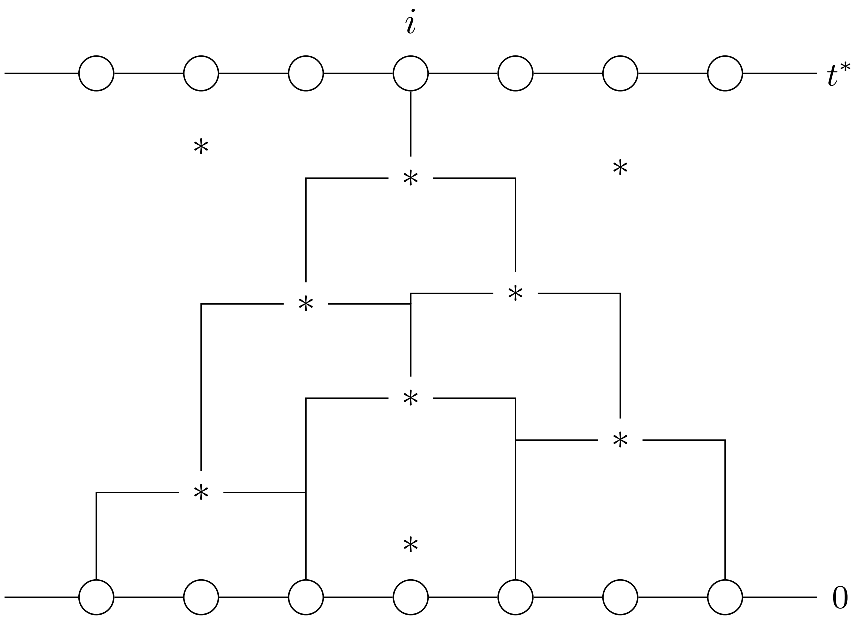

Figure 2.2 – *A naive construction of the update history of i. Each update $(\mathcal{V}, U, t)$ in the update sequence is represented by a $*$ at $(\mathcal{V}, t)$.*

update, $(\mathcal{V}, U, t)$, is oblivious if

$$f(\sigma, \mathcal{V}, U)[\mathcal{V}] = f(\sigma', \mathcal{V}, U)[\mathcal{V}] \tag{2.59}$$

for all $\sigma, \sigma' \in \Omega$, where f is as defined in (2.9).

Consider how these updates occur under our random mapping representation. Let Δ_i denote the degree of a vertex i. Recalling (2.7),

$$\frac{e^{-\beta\Delta_i}}{e^{\beta\Delta_i} + e^{-\beta\Delta_i}} \leq p_i(\sigma) \leq \frac{e^{\beta\Delta_i}}{e^{\beta\Delta_i} + e^{-\beta\Delta_i}}, \tag{2.60}$$

with equality holding for the lower and upper limits when the neighbours have spins all minus and all plus respectively. So for a particular update $(\mathcal{V}, U, t)$, if $U \leq \frac{e^{-\beta\Delta_\mathcal{V}}}{e^{\beta\Delta_\mathcal{V}} + e^{-\beta\Delta_\mathcal{V}}}$ then $\mathcal{V}$ is updated to a plus regardless of the configuration of its neighbours. Hence $(\mathcal{V}, U, t)$ is an oblivious update. Similarly, if $U > \frac{e^{\beta\Delta_\mathcal{V}}}{e^{\beta\Delta_\mathcal{V}} + e^{-\beta\Delta_\mathcal{V}}}$ then $\mathcal{V}$ is updated to a minus regardless of the configuration of its neighbours and hence $(\mathcal{V}, U, t)$ is an oblivious update. It is easy to see that these are the only types of oblivious updates.

The rate of these updates at vertex i is

$$\theta_i = 1 - \left(\frac{e^{\beta\Delta_i}}{e^{\beta\Delta_i} + e^{-\beta\Delta_i}} - \frac{e^{-\beta\Delta_i}}{e^{\beta\Delta_i} + e^{-\beta\Delta_i}} \right) \tag{2.61}$$

$$= 1 - \tanh(\beta\Delta_i). \tag{2.62}$$

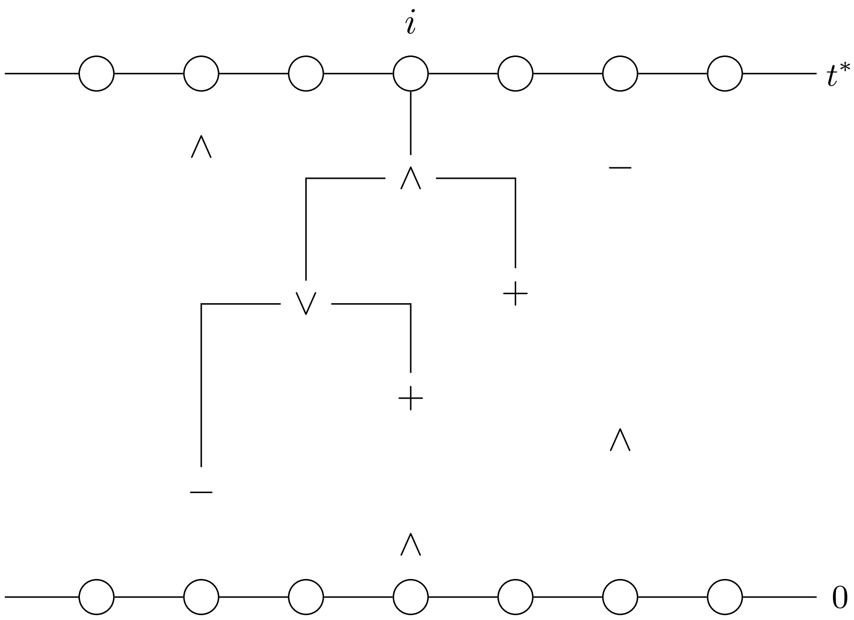

Figure 2.3 – *The update sequence for a section of the cycle and the corresponding update history from vertex i. For this particular update sequence, i takes a final spin of $+1$ regardless of the initial configuration.*

If G is a Δ-regular graph (as will be the case in the following chapters) then we can drop the subscript and write $\theta = 1 - \tanh(\beta\Delta)$ for the rate of oblivious updates at each vertex.

As noted earlier, oblivious updates cause temporal edges leading up to them in the update history to terminate. If $j \in \mathcal{H}_i(t)$, then an oblivious update (j, u, t) removes j from $\mathcal{H}_i(t)$ without adding any of its neighbours. In Figure 2.3 we construct an update history from a single vertex i as in Figure 2.2, but this time terminate branches at oblivious updates. To help us represent the updates in our update sequence more precisely, note that on the cycle, the function defined in (2.9) can be rewritten as

$$\sigma'[i] = \begin{cases} 1 & U \leq \theta/2, \\ \sigma[i-1] \vee \sigma[i+1] & \theta/2 < U \leq 1/2, \\ \sigma[i-1] \wedge \sigma[i+1] & 1/2 < U \leq 1-\theta/2, \\ -1 & U > \theta/2. \end{cases} \tag{2.63}$$

We can therefore represent each update $(\mathcal{V}, U, t)$ in the update sequence by placing at $(\mathcal{V}, t)$ one of the symbols $+$, $\vee$, $\wedge$, or $-$ choosen according to U. We then trace back from time t^*, branching to either side when we encounter a $\vee$ or $\wedge$, and terminating whenever we encounter a $+$ or $-$.

It is worth remarking that oblivious updates are not necessarily the only updates that

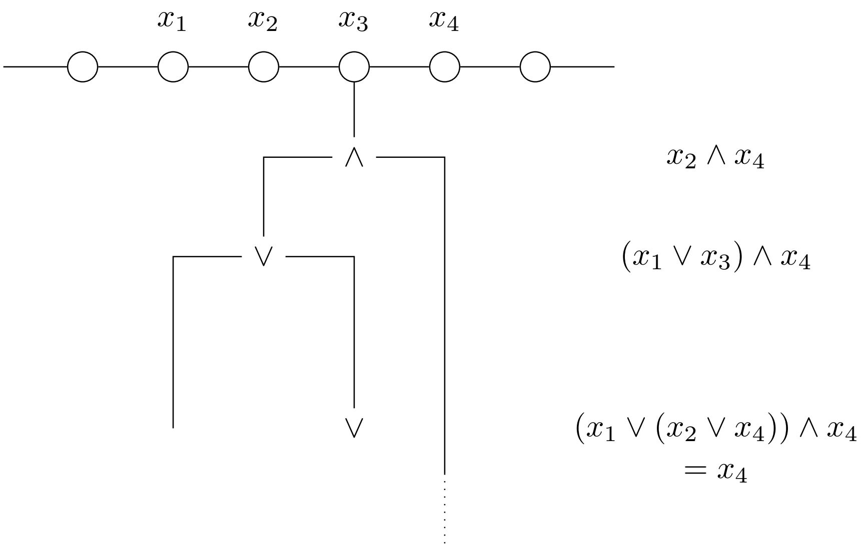

Figure 2.4 – *A non-oblivious update that shrinks the size of the update history. On the right is written the final spin of x_3 as a function of the configuration at that time. The update $x_3 \mapsto x_2 \vee x_4$ causes the entire function to collapse to x_4, and so removes x_1 and x_3 from the update history.*

can shrink the size of the update history of i (see for example Figure 2.4). However, for our analysis they will be the only such updates we will be concerned with. Indeed, in Chapter 3 we will use a different coupling so that these are the only updates that shrink the size of the update history, and in Chapter 4 we will use an alternative construction that bounds the true update history, in which updates are either oblivious or branch out to all Δ neighbours.

Chapter 3

The Coupling Time on the Cycle

In this chapter we consider the Ising heat-bath Glauber dynamics (as described in Section 2.2.1) on the cycle $G_n = (\mathbb{Z}/n\mathbb{Z})$. The object of interest is the coupling time, T_n, which was defined in Section 2.2.2 but whose definition will be modified slightly in Section 3.1 to allow for simpler analysis. The main result is Theorem 3.1 which establishes that T_n converges in distribution to a Gumbel distribution at all temperatures. This confirms, for $d = 1$, a conjecture by Collevecchio et al. that the coupling time of the Ising heat-bath process on the lattice $G_L = (\mathbb{Z}/L\mathbb{Z})^d$ converges to a Gumbel distribution as $L \to \infty$ for all $\beta < \beta_C$ [17, Conjecture 7.1] (We treat higher dimensions, and more generally any vertex transitive graphs, in Chapter 4). Of course, in one dimension, all temperatures are part of the high temperature regime [CITE], and likewise our result holds for any inverse-temperature β.

There is some intuition behind why we might expect that the coupling time converges to a Gumbel distribution. It is based on the belief that when the temperature is in the high-temperature regime, the dynamics behave qualitatively as if $\beta = 0$. In the $\beta = 0$ case, the spins update independently of their neighbours, and thus the top and bottom chains can be coupled so that they agree on each vertex that has been updated. The coupling time is then precisely the time it takes for each vertex to be updated. This corresponds to the coupon collector's problem, which is known to have a Gumbel limit [18].

As mentioned in Section 2.2.4, the coupling time is of practical interest since its distribution is the same as that of the running time of the coupling from the past (CFTP) algorithm. Our result shows that when running the Glauber heat-bath dynamics for the Ising model on a large enough cycle, the running time of CFTP can be approximated by a Gumbel distribution. We note that even though one is typically more interested in the Ising model on lattices of dimension at least two (so that there exists a phase transition), the one dimensional case proves to be a useful test case for the proof techniques. Furthermore, the applicability of Theorem 3.1 to the full high temperature regime justifies a separate treatment to the higher dimensional case in Chapter 4.

Theorem 3.1. *Let T_n be the coupling time for the continuous-time Ising heat-bath Glauber dynamics for the zero-field ferromagnetic Ising model on the cycle $(\mathbb{Z}/n\mathbb{Z})$. Then for any inverse-temperature β,*

$$\lim_{n\to\infty} \mathbb{P}\left[T_n < \frac{z+\ln n}{\theta}\right] = e^{-C_\theta e^{-z}} \tag{3.1}$$

where $\theta = 1-\tanh(2\beta)$ and C_θ is a positive constant bounded by

$$\frac{1}{2\sqrt{\frac{4}{\theta}-1}-1} \leq \lambda \leq 1. \tag{3.2}$$

The proof of Theorem 3.1 will be given in Section 3.3 after the essential preliminaries are presented. In Section 3.1 we describe some modified dynamics and show that the coupling time we construct from these has the same distribution as the coupling time defined in Section 2.2.2. Then in Section 3.2 we outline the overall approach to the proof and define some essential quantities. Finally, Section 3.4 contains additional lemmas that are used in Section 3.3.

3.1 A new coupling on the cycle

On the cycle, we will use a different coupling of $\mathscr{T}_t$ and $\mathscr{B}_t$ via a new set of update rules that will replace those from (2.9). The new update rules simplify our update histories greatly by ensuring that each of the update histories never contain more than one vertex at any one time. However, we must be cautious. The coupling time is not just a property of the heat-bath dynamics, but also of the specific coupling we chose. Hence, we will have to verify that switching to our new rules does not change the distribution of T_n.

The new update rules are defined by using almost the same construction as in Section 2.2.2. The one difference is that we replace (2.9) as follows. When vertex i updates, instead of comparing U to the probability $p_i(\sigma)$ to determine the new spin, we instead

$\mathscr{T}_t = \cdot$ / $\mathscr{B}_t = \cdot$ \ $\mathbb{P}[(\mathscr{T}_t[i]', \mathscr{B}_t[i]') = \cdot\,]$	(1,1)	(1,-1)	(-1,-1)
$(\dots,1,\sigma_i,1,\dots)$ $(\dots,1,\sigma_i,1,\dots)$	$1-\theta$	0	$\frac{\theta}{2}$
$(\dots,1,\sigma_i,1,\dots)$ $(\dots,1,\sigma_i,-1,\dots)$	$\frac{1}{2}$	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$
$(\dots,1,\sigma_i,1,\dots)$ $(\dots,-1,\sigma_i,1,\dots)$	$\frac{1}{2}$	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$
$(\dots,1,\sigma_i,1,\dots)$ $(\dots,-1,\sigma_i,-1,\dots)$	$\frac{\theta}{2}$	$1-\theta$	$\frac{\theta}{2}$
$(\dots,1,\sigma_i,-1,\dots)$ $(\dots,1,\sigma_i,-1,\dots)$	$\frac{1}{2}$	0	$\frac{1}{2}$
$(\dots,1,\sigma_i,-1,\dots)$ $(\dots,-1,\sigma_i,-1,\dots)$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{1}{2}$
$(\dots,-1,\sigma_i,1,\dots)$ $(\dots,-1,\sigma_i,1,\dots)$	$\frac{1}{2}$	0	$\frac{1}{2}$
$(\dots,-1,\sigma_i,1,\dots)$ $(\dots,-1,\sigma_i,-1,\dots)$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{1}{2}$
$(\dots,-1,\sigma_i,-1,\dots)$ $(\dots,-1,\sigma_i,-1,\dots)$	$\frac{\theta}{2}$	0	$1-\theta$

Table 3.1 – *Probabilities of updating from* $(\mathscr{T}_t, \mathscr{B}_t)$ *to* $(\mathscr{T}_t', \mathscr{B}_t')$ *given vertex* i *updates at time* t.

chose a new spin σ_i' via

$$\sigma_i' = \begin{cases} +1 & U < \theta/2, \\ \sigma_{i-1} & \theta/2 \leq U < 1/2, \\ \sigma_{i+1} & 1/2 \leq U < 1-\theta/2, \\ -1 & U \geq 1-\theta/2. \end{cases} \tag{3.3}$$

where $U \in [0,1]$ is an independent uniform random variable as before. It is easy to see that these update rules give rise to the same transition rates as those in (2.9). To show that the coupling time is unchanged, it is sufficient to verify that the joint jump probabilities of $(\mathscr{T}_t[i], \mathscr{B}_t[i])$ are unchanged for each possible configuration of spins of vertices $i-1$ and $i+1$. There are only nine possible configurations for the two neighbours of i in the top and bottom chain since $\mathscr{B}_t[i] \leq \mathscr{T}_t[i], \forall t$. Likewise, there are only three possible configurations for the updated spins $(\mathscr{T}_t[i]', \mathscr{B}_t[i]')$. Hence, given vertex i updates at time t, we can easily calculate all the required jump probabilities as shown in Table 3.1. These are unchanged whether using (2.9) or (3.3) and so the new rules do not change the coupled dynamics.

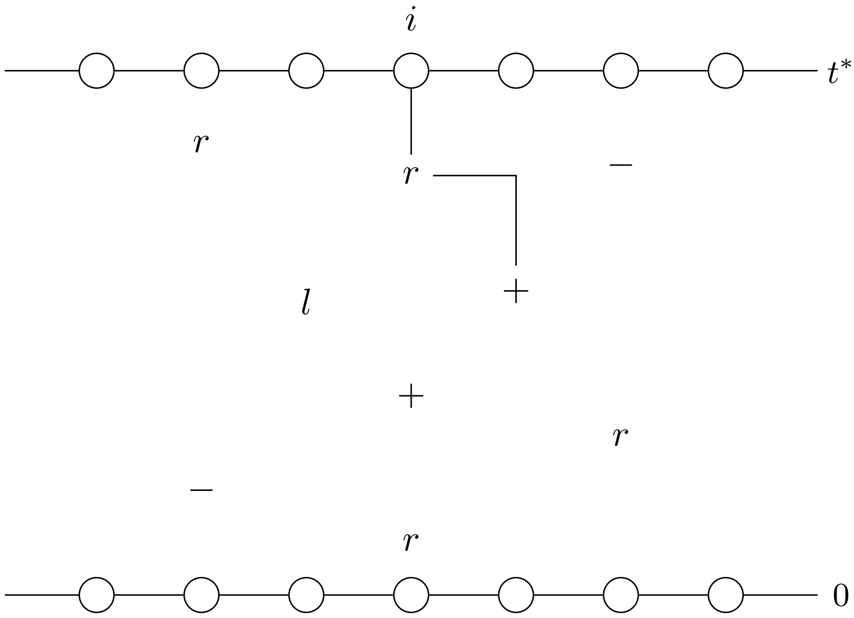

Figure 3.1 – *The update sequence for a section of the cycle and the corresponding update history from vertex i using the new update rules. Vertex i takes the same spin as the spin it terminates at, in this case $+1$.*

3.1.1 Update histories on the cycle

Under the update rules in (3.3), each time a vertex is updated, it is either an oblivious update with probability θ, or it takes the spin of a uniformly chosen neighbour. Unlike the histories considered earlier (for example Figure 2.3), this time a non-oblivious update does not cause the history to branch out to both its neighbours. Rather, given a non-oblivious update to some vertex v, we only need to know the spins of one of its neighbours to update it (the left spin if $U < 1/2$ and the right if $U \geq 1/2$). So the history simply moves either right or left without branching. As before, encountering an oblivious update causes $\mathcal{H}_i$ to terminate. An example history using these new rules is shown in Figure 3.1. In a similar vein to Figures 2.3 and 2.4 we represent each update $(\mathcal{V}, U, t)$ in the update sequence by placing at $(\mathcal{V}, t)$ one of the symbols $+$, r, l, or $-$ choosen according to U. We then trace back from time t^*, moving left or right when we encounter a l or r respectively, and terminating whenever we encounter a $+$ or $-$.

We now see that as t decreases from t^*, $\mathcal{H}_i(t)$ is a continuous-time random walk that dies at rate θ, moves left at rate $(1-\theta)/2$, and moves right at rate $(1-\theta)/2$. The probability that $\mathcal{H}_i(0) \neq \emptyset$ is simply the probability that the continuous-time random walk survives until time $t = 0$. This immediately gives us the following probability which we will use repeatedly in what follows. Recalling (2.57),

$$\mathbb{P}\left[\mathscr{B}_{t^*}[i] \neq \mathscr{T}_{t^*}[i]\right] = \mathbb{P}\left[\mathcal{H}_i(0) \neq \emptyset\right] = e^{-\theta t^*}. \tag{3.4}$$

since our histories die at rate θ.

3.2 Problem Setup

In order to prove Theorem 3.1, we will actually prove a stronger statement using Theorem 3.2. The general idea is that at some fixed time t^* we will count the number of vertices at which the bottom and top chains differ. This number is a random variable, which we will call W, and we can bound the total variation distance of its distribution with that of an appropriate compound poisson distribution. As a special case, we can use this bound as a bound on the probability that W is zero. Of course, if W is zero then the top and bottom chains must have coupled and so we can use this to establish Theorem 3.1.

Bounding the total variation distance between W and the compound Poisson will be done using compound Poisson approximation as described in [19]. This paper reviews a number of different methods by which approximations may be made. The specific method that we will employ is based on Stein's method for the compound Poisson distribution, introduced in [20].

We now make precise the ideas stated above. Fix z and a time of interest, $t_* = (z + \ln n)/\theta$. For each vertex $i \in G$, define indicators

$$X_i = \begin{cases} 1 & \mathscr{B}_{t^*}[i] \neq \mathscr{T}_{t^*}[i], \\ 0 & \mathscr{B}_{t^*}[i] = \mathscr{T}_{t^*}[i] \end{cases} \tag{3.5}$$

and set $W = \sum_{i \in V} X_i$. Note that from (3.4) we get

$$\mathbb{P}[X_i = 1] = e^{-\theta t_*} = \frac{e^{-z}}{n}. \tag{3.6}$$

For each $i \in V$, decompose W into $W = X_i + U_i + Z_i + W_i$ where

$$U_i = \sum_{j \in B_i} X_j, \qquad Z_i = \sum_{j \in C_i} X_j, \qquad W_i = \sum_{j \in D_i} X_j. \tag{3.7}$$

and B_i, C_i, and D_i are the vertex sets

$$B_i = \{j \neq i : |j - i| \leq b_n\}, \tag{3.8}$$

$$C_i = \{j \notin B_i \cup \{i\} : |j - i| \leq c_n\}, \tag{3.9}$$

$$D_i = V \setminus (B_i \cup C_i \cup \{i\}). \tag{3.10}$$

We have some freedom in choosing b_n and c_n; they are chosen to control the assymptotics of various quantities to be defined later. For this chapter, we will choose $b_n = \ln(n)$ and $c_n = \ln(n)^2$.

We now define the quantities

$$\lambda = \sum_{i \in V} \mathbb{E}\left[\frac{X_i}{X_i + U_i} I[X_i + U_i \geq 1]\right], \tag{3.11}$$

$$\mu_l = \frac{1}{l\lambda} \sum_{i \in V} \mathbb{E}\left[X_i I[X_i + U_i = l]\right], \qquad l \geq 1, \tag{3.12}$$

which will be the parameters of the approximating compound Poisson distribution to W. We also define

$$\delta_1 = \sum_{i \in V} \sum_{k \geq 0} \mathbb{P}[X_i = 1, U_i = k] \mathbb{E}\left|\frac{\mathbb{P}[X_i = 1, U_i = k | W_i]}{\mathbb{P}[X_i = 1, U_i = k]} - 1\right|, \tag{3.13}$$

$$\delta_4 = \sum_{i \in V} \left(\mathbb{E}[X_i Z_i] + \mathbb{E}[X_i]\mathbb{E}[X_i + U_i + Z_i]\right), \tag{3.14}$$

which we desire to be small for the compound Poisson approximation to be good.

The following theorem (reworked from [19]) bounds the distance between the distributions of W and the approximating compound Poisson.

Theorem 3.2 ([19])**.** *Let W, λ, $\boldsymbol{\mu}$, δ_1 and δ_4 be as defined above. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{CP}(\lambda, \boldsymbol{\mu})) \leq (\delta_1 + \delta_4)e^{\lambda}. \tag{3.15}$$

Note that W is zero precisely when $T < t^*$ and so the events $\{W = 0\}$ and $\{T \leq t^*\}$ are the same. Furthermore, if W' is compound Poisson with parameters λ and $\boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is supported only on the positive integers (as in (3.12)), then $\mathbb{P}[W' = 0] = e^{-\lambda}$. These observations lead to the following corollary of Theorem 3.2.

Corollary 3.3. *Let T_n be the coupling time of the continuous-time heat-bath Glauber dynamics for the zero-field Ising model at inverse-temperature β on the cycle $(\mathbb{Z}/n\mathbb{Z})$ and let δ_1, δ_4 and λ be as defined above. Then*

$$\left|\mathbb{P}\left[T_n \leq \frac{z + \ln(n)}{\theta}\right] - e^{-\lambda}\right| \leq (\delta_1 + \delta_4)e^{\lambda}, \tag{3.16}$$

where $\theta = 1 - \tanh(2\beta)$.

3.3 Proof of Theorem 3.1

In this section we use Corollary 3.3 to prove Theorem 3.1 by bounding λ and showing that δ_1 and δ_4 go to zero as $n \to \infty$. This is done in Lemmas 3.4, 3.5, and 3.6. The proofs of these require some additional lemmas concerning properties of the update histories which have been deferred to Section 3.4.

We begin by bounding λ.

Lemma 3.4. *Using the above setup*

$$\lim_{n\to\infty} \lambda = C_\theta e^{-z} \tag{3.17}$$

where

$$\frac{1}{2\sqrt{\frac{4}{\theta}-1}-1} \leq C_\theta \leq 1. \tag{3.18}$$

Proof. Beginning with the definition of λ, we have

$$\lambda = \sum_{i\in V} \mathbb{E}\left[\frac{X_i}{X_i+U_i} I[X_i+U_i \geq 1]\right] \tag{3.19}$$

$$= \sum_{i=1}^{n} \mathbb{P}(X_i = 1)\mathbb{E}\left[\frac{1}{1+U_i}|X_i = 1\right] \tag{3.20}$$

$$= \sum_{i=1}^{n} \frac{e^{-z}}{n}\mathbb{E}\left[\frac{1}{1+U_i}|X_i = 1\right] \tag{3.21}$$

$$= e^{-z}\mathbb{E}\left[\frac{1}{1+U_i}|X_i = 1\right] \tag{3.22}$$

where we have used that X_i is either zero or one, (3.6), and the transitivity of the graph in each step respectively. Clearly

$$\mathbb{E}\left[\frac{1}{1+U_i}|X_i = 1\right] \leq 1. \tag{3.23}$$

By Jensen's inequality

$$\mathbb{E}\left[\frac{1}{1+U_i}|X_i = 1\right] \geq \frac{1}{\mathbb{E}[1+U_i|X_i=1]} \tag{3.24}$$

$$= \frac{1}{1+\mathbb{E}[U_i|X_i=1]}. \tag{3.25}$$

Now

$$\mathbb{E}[U_i|X_i = 1] = \sum_{j\in B_i} \mathbb{P}[X_j = 1|X_i = 1] \tag{3.26}$$

$$= \sum_{k=1}^{\lfloor b_n \rfloor} \sum_{|j-i|=k} \mathbb{P}[X_j = 1|X_i = 1] \tag{3.27}$$

$$= 2\sum_{k=1}^{\lfloor b_n \rfloor} \mathbb{P}[X_{i+k} = 1|X_i = 1] \tag{3.28}$$

where we have used the symmetry of X_{i+k} and X_{i-k} in the last step. From Lemma 3.10,

$$\mathbb{E}[U_i|X_i = 1] \leq 2\sum_{k=1}^{\lfloor b_n \rfloor} \left(\frac{e^{-z}}{n} + 2\left(\frac{2-\sqrt{\theta(4-\theta)}}{2-\theta} \right)^k \right) \tag{3.29}$$

$$< 2\sum_{k=1}^{\lfloor b_n \rfloor} \frac{e^{-z}}{n} + 4\sum_{k=1}^{\infty} \left(\frac{2-\sqrt{\theta(4-\theta)}}{2-\theta} \right)^k \tag{3.30}$$

$$= \frac{2\lfloor b_n \rfloor}{n} e^{-z} + 2\left(\sqrt{\frac{4}{\theta} - 1} - 1 \right). \tag{3.31}$$

Finally, as $n \to \infty$ the first term vanishes and

$$\liminf_{n\to\infty} \mathbb{E}\left[\frac{1}{1+U_i} |X_i = 1 \right] \geq \frac{1}{2\sqrt{\frac{4}{\theta} - 1} - 1}. \tag{3.32}$$

To complete the proof, we need to show that the limit of

$$\mathbb{E}\left[\frac{1}{1+U_i} |X_i = 1 \right] \tag{3.33}$$

as $n \to \infty$ exists. Since we already have bounds, it is sufficient to show that it is monotonic.

[NEED TO SHOW THAT THING ACTUALLY CONVERGES]

□

The next two lemmas prove that δ_1 and δ_4 go to zero as $n \to \infty$. Since from Lemma 3.4 we know that λ is bounded above by a constant, this is enough for (3.16) to show that

$$\lim_{n\to\infty} \mathbb{P}\left[T_n < \frac{z + \ln n}{\theta} \right] = e^{-\lambda}. \tag{3.34}$$

Lemma 3.5. *Let δ_1 be as defined above in (3.13). Then*

$$\lim_{n\to\infty} \delta_1 = 0. \tag{3.35}$$

Proof. Starting with the definition of δ_1, we have

$$\delta_1 = \sum_{i=1}^{n}\sum_{k=0}^{2\lfloor b_n\rfloor} \mathbb{P}[X_i = 1, U_i = k]\mathbb{E}\left|\frac{\mathbb{P}[X_i = 1, U_i = k|W_i]}{\mathbb{P}[X_i = 1, U_i = k]} - 1\right|, \tag{3.36}$$

$$= n\sum_{k=0}^{2\lfloor b_n\rfloor} \mathbb{E}\left|\mathbb{P}[X_i = 1, U_i = k|W_i] - \mathbb{P}[X_i = 1, U_i = k]\right| \tag{3.37}$$

by the transitivity of the cycle. Let

$$C_i^c = \{j : |j - i| \le (c_n + b_n)/2\} \tag{3.38}$$

be the set of vertices within distance $(b_n + c_n)/2$ of i and define the events

$$A_1 = \{\exists j \in B_i \cup \{i\}, \exists t \in [0, t^*] : \mathcal{H}_j(t) \nsubseteq C_i^c\} \tag{3.39}$$

and

$$A_2 = \{\exists j \in D_i, \exists t \in [0, t^*] : \mathcal{H}_j(t) \cap C_i^c \neq \emptyset\} \tag{3.40}$$

as well as their intersection

$$A = A_1 \cap A_2. \tag{3.41}$$

From Lemma 3.7,

$$\mathbb{P}[X_i = 1, U_i = j|A^{\complement}, W_i] = \mathbb{P}[X_i = 1, U_i = j|A^{\complement}]. \tag{3.42}$$

Continuing on from (3.37), we split the probabilities into

$$\begin{aligned}
\delta_1 &= n\sum_{k=0}^{2\lfloor b_n\rfloor} \mathbb{E}\Big|\mathbb{P}[X_i = 1, U_i = k|W_i, A]\mathbb{P}[A|W_i] - \mathbb{P}[X_i = 1, U_i = k|A]\mathbb{P}[A] + \qquad (3.43)\\
&\qquad \mathbb{P}(X_i = 1, U_i = k|A^{\complement})(\mathbb{P}[A^{\complement}|W_i] - \mathbb{P}[A^{\complement}])\Big| \\
&\le n(2\lfloor b_n\rfloor + 1)\mathbb{E}\left[\mathbb{P}[A|W_i] + \mathbb{P}[A] + \left|\mathbb{P}[A^{\complement}|W_i] - \mathbb{P}[A^{\complement}]\right|\right] \qquad (3.44)\\
&= n(2\lfloor b_n\rfloor + 1)\mathbb{E}\left[\mathbb{P}[A|W_i] + \mathbb{P}[A] + |1 - \mathbb{P}[A|W_i] - (1 - \mathbb{P}[A])|\right] \qquad (3.45)\\
&\le n(2\lfloor b_n\rfloor + 1)\mathbb{E}\left[\mathbb{P}[A|W_i] + \mathbb{P}[A] + \mathbb{P}[A|W_i] + \mathbb{P}[A])\right] \qquad (3.46)\\
&= 2n(2\lfloor b_n\rfloor + 1)\left(\mathbb{E}[\mathbb{P}[A|W_i]] + \mathbb{P}[A]\right) \qquad (3.47)\\
&= 4n(2\lfloor b_n\rfloor + 1)\mathbb{P}[A]. \qquad (3.48)
\end{aligned}$$

For either A_1 or A_2 to hold, there must exists a history that spreads at least distance $(c_n - b_n)/2$ away from its starting vertex. By a union bound

$$\mathbb{P}[A] \leq \sum_{j=1}^{n} \mathbb{P}[\mathcal{H}_i \not\subseteq B(i,(c_n - b_n)/2) \times [0,t^*]] \tag{3.49}$$

$$= n\mathbb{P}\left[\bigcup_{u\in[0,t^*]} \mathcal{H}_i(t^* - u) \not\subseteq B(i,(c_n - b_n)/2)\right] \tag{3.50}$$

Combining this with Lemma 3.8, and recalling our choices of $b_n = \ln(n)$ and $c_n = \ln(n)^2$ we get that

$$\delta_1 \leq 4n^2(2\lfloor b_n\rfloor + 1)\exp\left(3(z + \ln n)/\theta - \ln 2(c_n - b_n)/2\right) \tag{3.51}$$

$$\leq 8\exp(3z/\theta)n^{3+3/\theta+\ln 2/2-\ln n/2} \tag{3.52}$$

which goes to 0 as $n \to \infty$. □

Lemma 3.6. *Let δ_4 be as defined above in (3.14). Then*

$$\lim_{n\to\infty} \delta_4 = 0. \tag{3.53}$$

Proof. Starting with the definition of δ_4, we have

$$\delta_4 = \sum_{i=1}^{n} \left(\mathbb{E}[X_i Z_i] + \mathbb{E}[X_i]\mathbb{E}[X_i + U_i + Z_i]\right), \tag{3.54}$$

$$= n\mathbb{P}[X_i = 1]\mathbb{E}[Z_i|X_i = 1] + e^{-z}\sum_{j\in\{i\}\cup B_i\cup C_i} \mathbb{E}[X_j], \tag{3.55}$$

$$= e^{-z}\mathbb{E}[Z_i|X_i = 1] + \frac{(2\lfloor c_n\rfloor + 1)e^{-2z}}{n}. \tag{3.56}$$

Now

$$\mathbb{E}[Z_i|X_i = 1] = \sum_{j\in C_i} \mathbb{P}[X_j = 1|X_i = 1], \tag{3.57}$$

$$= 2\sum_{k=\lfloor b_n\rfloor+1}^{\lfloor c_n\rfloor} \mathbb{P}[X_{i+k} = 1|X_i = 1]. \tag{3.58}$$

From Lemma 3.10,

$$\mathbb{E}[Z_i|X_i = 1] \leq 2 \sum_{k=\lfloor b_n \rfloor+1}^{\lfloor c_n \rfloor} \left(\frac{e^{-z}}{n} + 2 \left(\frac{2 - \sqrt{\theta(4-\theta)}}{2-\theta} \right)^k \right), \tag{3.59}$$

$$\leq \frac{2(c_n - b_n + 1)e^{-z}}{n} + 4(c_n - b_n + 1) \left(\frac{2 - \sqrt{\theta(4-\theta)}}{2-\theta} \right)^{b_n+1}. \tag{3.60}$$

Altogether,

$$\delta_4 \leq \frac{2(c_n - b_n + 1)e^{-z}}{n} + 4(c_n - b_n + 1) \left(\frac{2 - \sqrt{\theta(4-\theta)}}{2-\theta} \right)^{b_n+1} + \frac{(2c_n + 1)e^{-2z}}{n} \tag{3.61}$$

which, recalling that $b_n = \ln(n)$ and $c_n = \ln(n)^2$, goes to 0 as $n \to \infty$. □

3.4 Additional Lemmas

This section contains the proofs for a number of lemmas concerning properties of the update histories on the cycle.

Lemma 3.7. *Let i be a vertex in a graph and let*

$$C_i^c = \{j : |j - i| \leq (b_n - c_n)/2\} \tag{3.62}$$

be the set of vertices within distance $(b_n + c_n)/2$ of i. Define the events

$$A_1 = \{\exists j \in B_i \cup \{i\}, \exists t \in [0, t^*] : \mathcal{H}_j(t) \nsubseteq C_i^c\} \tag{3.63}$$

and

$$A_2 = \{\exists j \in D_i, \exists t \in [0, t^*] : \mathcal{H}_j(t) \cap C_i^c \neq \emptyset\} \tag{3.64}$$

as well as their union

$$A = A_1 \cup A_2. \tag{3.65}$$

Then

$$\mathbb{P}[X_i = 1, U_i = j|A^{\complement}, W_i] = \mathbb{P}[X_i = 1, U_i = j|A^{\complement}]. \tag{3.66}$$

Proof. If $A_1^{\complement}$ holds, then the events $\{X_1 = 1\}$ and $\{U_i = j\}$ depend only on the values of the update sequence inside C_i^c. If $A_2^{\complement}$ holds then the events $\{W_i = k\}$, $k \geq 0$, depend only on the values of the update sequence outside of C_i^c. Since the update sequences of each vertex are independent of each other vertex, if $A_2^{\complement}$ holds, conditioning on W_i does not affect the update sequences inside C_i^c and so

$$\mathbb{P}[X_i = 1, U_i = j | A^{\complement}, W_i] = \mathbb{P}[X_i = 1, U_i = j | A^{\complement}]. \tag{3.67}$$

[WOULD BE NICE TO SAY THIS MORE MATHEMATICALLY.] □

The following Lemma bounds how fast updates can percolate through the cycle. The proof is a slight modification of a similar one in [11].

Lemma 3.8. *Let $B(i,l)$ indicate the set of vertices at distance l or smaller from vertex i. The probability that the history of vertex i escapes $B(i,l)$ in time s is bounded by*

$$\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^* - u) \not\subseteq B(i,l)\right] \leq \exp\left(3s - l \ln 2\right). \tag{3.68}$$

Proof. Let $\mathcal{W} = \{\boldsymbol{w} = (w_1, w_2, \ldots, w_l) : w_1 = i, ||w_{k-1} - w_k|| = 1\}$ be the set of length l sequences of adjacent vertices starting at vertex i. If $\mathcal{H}_i$ contains any vertex outside $B(i,l)$ at a time $u \in [t^* - s, t^*]$ then there must be some sequence $w \in \mathcal{W}$ such that each w_i was updated at some time $t^* > t_i > t^* - s$ and $t_{k-1} > t_k$. Call this event M_w. For any particular sequence w,

$$\mathbb{P}[M_w] = \mathbb{P}[\mathrm{Po}(s) \geq l] \tag{3.69}$$

where $\mathrm{Po}(s)$ is Poisson with rate s. By a union bound over $\mathcal{W}$,

$$\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^* - u) \not\subseteq B(i,l)\right] \leq 2^{l-1}\mathbb{P}[\mathrm{Po}(s) \geq l]. \tag{3.70}$$

The moment generating function of a poisson random variable with rate s is

$$M(t) = \exp\left(s\left(e^t - 1\right)\right). \tag{3.71}$$

Using a Chernoff bound we have for every $t > 0$,

$$\mathbb{P}[\mathrm{Po}(s) \geq l] \leq \exp\left(s\left(e^t - 1\right) - tl\right). \tag{3.72}$$

Overall we have

$$\begin{aligned}\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^* - u) \not\subseteq B(i,l)\right] &\leq 2^{l-1}\exp\left(s\left(e^t - 1\right) - tl\right) &(3.73)\\ &\leq \exp\left(s\left(e^t - 1\right) + l(\ln 2 - t)\right). &(3.74)\end{aligned}$$

Choosing $t = 2\ln 2$,

$$\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^*-u) \not\subseteq B(i,l)\right] \leq \exp\left(3s - l\ln 2\right). \tag{3.75}$$

□

Lemma 3.9. *Let i and j be the indices of two vertices that are separated by distance k. Then*

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_j = 1] \leq 2\left(\frac{2-\sqrt{\theta(4-\theta)}}{2-\theta}\right)^k. \tag{3.76}$$

where $k = |i-j|$.

Proof. We first must deal with the effect that conditioning on $X_j = 1$ has on the probability that the two update histories merge. For the history of vertex j to survive, all updates along the history must be non-oblivious updates. We also note that conditioning on the history of vertex j surviving should not result in an increase in the overall rate of updates (since each update is a chance that the history will die). By this reasoning,

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_j = 1] \leq \mathbb{P}[\mathcal{H}_i \cap \bar{\mathcal{H}}_j] \tag{3.77}$$

where $\bar{\mathcal{H}}_j$ is an undying random walk that starts at j and going backwards in time from t^* moves right at rate $1/2$ and left at rate $1/2$.

We note that, while $\mathcal{H}_i$ survives, the distance between $\mathcal{H}_i$ and $\bar{\mathcal{H}}_j$ is a birth and death process that starts at $k = |i-j|$ and has birth and death rates, $\lambda = \mu = (2-\theta)/2$. Let $P(t)$ be such a process and define $s_0 = \inf\{t : P(t) = 0\}$ to be the first time the process reaches zero (this corresponds to the update histories merging). Let s_d be exponentially distributed with rate θ (this corresponds to the update history of vertex i dying). Then

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_j = 1] \leq \mathbb{P}[\mathcal{H}_i \cap \bar{\mathcal{H}}_j] \leq 2\mathbb{P}_k(s_0 < s_d) \tag{3.78}$$

where $\mathbb{P}_k$ indicates that $P(0) = k$. The factor of two comes from the fact that the update histories may meet by going the other direction around the cycle. We also note that we allow $P(t)$ to continue beyond $t = t^*$, unlike our update histories which stop at time 0. This does not present a problem as the effect of allowing this is to increase the size of our upper bound.

At any time before s_d there are three possibilities for what can happen to P next. Either the next event is a birth with probability $(2-\theta)/4$, the next event is a death with the same probability or we reach time s_d with probability $\theta/2$. Writing $\zeta_k = \mathbb{P}_k(s_0 < s_d)$

this gives us the recurrence relation

$$\zeta_k = \frac{2-\theta}{4}\zeta_{k-1} + \frac{2-\theta}{4}\zeta_{k+1} \tag{3.79}$$

which is subject to the conditions

$$\zeta_0 = 1 \tag{3.80}$$

$$\zeta_k \leq 1,\ \forall k \in \mathbb{N}. \tag{3.81}$$

This recurrence has characteristic equation

$$x^2 - \frac{4}{2-\theta}x + 1 = 0 \tag{3.82}$$

which has roots

$$r_1 = \frac{2+\sqrt{\theta(4-\theta)}}{2-\theta} \tag{3.83}$$

$$r_2 = \frac{2-\sqrt{\theta(4-\theta)}}{2-\theta} \tag{3.84}$$

and so

$$\zeta_k = ar_1^k + br_2^k \tag{3.85}$$

where a and b are constants to be determined from (3.80) and (3.81). We note that $r_1 \geq 1, \forall \theta \in [0,1]$ and so from (3.81) we have that $a = 0$. Finally from (3.80), $b = 1$ and so

$$\zeta_k = \left(\frac{2-\sqrt{\theta(4-\theta)}}{2-\theta}\right)^k. \tag{3.86}$$

□

Lemma 3.10.

$$\mathbb{P}[X_{i+k} = 1 | X_i = 1] \leq \frac{e^{-z}}{n} + 2\left(\frac{2-\sqrt{\theta(4-\theta)}}{2-\theta}\right)^k. \tag{3.87}$$

Proof. There are two ways in which the update history of vertex $i+k$ can survive until time 0. The update history can survive without intersecting with the update history of vertex i or the update history of vertex $i+k$ can merge with the update history of vertex i (whose survival we are conditioning on). Breaking up the probability this way we have

$$\mathbb{P}[X_{i+k} = 1 | X_i = 1] = \mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} = \emptyset | X_i = 1] + \mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} \neq \emptyset | X_i = 1]. \tag{3.88}$$

If the histories of i and j do not intersect, then conditioning on $\mathcal{H}_i$ surviving does not affect the chance that $\mathcal{H}_j$ survives. However, conditioning on $\mathcal{H}_i$ surviving does make it more likely that the histories intersect. So

$$\mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} = \emptyset | X_i = 1] \leq \mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} = \emptyset] \tag{3.89}$$

[DOES THIS NEED FIXING?] and so

$$\mathbb{P}[X_{i+k} = 1 | X_i = 1] \leq \mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} = \emptyset] + \mathbb{P}[X_{i+k} = 1, \mathcal{H}_i \cap \mathcal{H}_{i+k} \neq \emptyset | X_i = 1], \tag{3.90}$$

$$\leq \mathbb{P}[X_{i+k} = 1] + \mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_{i+k} \neq \emptyset | X_i = 1]. \tag{3.91}$$

The result follows from (3.6) and Lemma 3.9. □

Chapter 4

The Coupling Time on Vertex Transitive Graphs

In this chapter we prove the following theorem.

Conjecture 4.1. *Let T_L be the coupling time for the continuous-time Ising heat-bath dynamics for the zero-field ferromagnetic Ising model on the torus $(\mathbb{Z}/L\mathbb{Z})^d$. Then for any small enough inverse-temperature β,*

$$\lim_{L\to\infty} \mathbb{P}[T_L < a_L z + b_L] = e^{-e^{-z}} \tag{4.1}$$

where a_L and b_L have yet to be determined.

Proof. □

Heuristically, in the high-temperature regime, we expect the dynamics to be similar to those when $\beta = 0$. This is like coupon collector blahblah etc...

4.1 Information Percolation in higher dimensions

In the previous chapter, we showed that on the cycle, there was a coupling that made the update history of a single vertex to be a continuous-time random walk that died at rate θ. On lattices of dimension $d > 2$, we can no longer use this coupling and so the updates histories are significantly more complex.

Recall from Section 2.3.2.2 that given a target time t^*, the update history of a vertex set A at time t, $\mathcal{H}_A(t)$, is the set of vertices whose spins at time t determine the spins of A

at time t^*. Developing this history backwards in time from $t = t^*$ produces a subgraph of $\Omega \times [0, t^*]$ which we write as $\mathcal{H}_A$ and call the update history of vertex set A. This history can be constructed using the update sequence along $(t, t^*]$.

In practise, we may choose to construct this history as follows: For each $i \in A$, create a temporal edge between (i, t^*) and (i, t_i) where t_i is the time of the latest update to i (or 0 if i is never updated). Then for each update (i, u, t_i), we either terminate the edge if u is such that the update is oblivious, or we add spatial branches to each of the neighbours of i. We repeat this process recursively for the neighbours of i until every branch has been terminated due to an oblivious update or has reached time 0.

However, it is possible for vertices to be removed from $\mathcal{H}_A(t)$ from updates that are not oblivious. [PUT EXAMPLE IN]. Since our method above for constructing the history does not take this into account, the history it produces will possibly be larger than $\mathcal{H}_A$. To ensure a distinction between the two, the history that results from the above construction we will denote $\hat{\mathcal{H}}_A$, and likewise $\hat{\mathcal{H}}_A(t)$ for the history at time t that results from the above construction. We have that

$$\mathcal{H}_A(t) \subseteq \hat{\mathcal{H}}_A(t) \tag{4.2}$$

and also that $\mathcal{H}_A$ is a subgraph of $\hat{\mathcal{H}}_A$.

4.1.1 Magnetization

One quantity which we used multiple times in Chapter 3 was $\mathbb{P}[X_i = 1]$. Although it was not required earlier, we would now like to make clear that this is in fact the magnetization at time t^*.

The magnetization at vertex $i \in V$ at time $t > 0$ is defined to be

$$m_t(i) = \mathbb{E}[\mathcal{T}_t[i]] \tag{4.3}$$

where $(\mathcal{T}_t)_{t \geq 0}$ is the dynamics starting from the all-plus configuration. Given a monotonically coupled chain $(\mathcal{B}_t)_{t \geq 0}$, starting in the all minus configuration and such that $\mathcal{T}_t[i] \geq \mathcal{B}_t[i]$ for all $t \geq 0$ and $i \in V$, we can split up this expectation by conditioning on the event $A_t = \{\mathcal{T}_t[i] \neq \mathcal{B}_t[i]\}$. We obtain that

$$\begin{aligned}
m_t(i) &= \mathbb{E}[Y_t^+[i]] && (4.4)\\
&= \mathbb{P}\left[A_t\right]\left(\mathbb{P}\left[Y_t^+[i] = 1|A_t\right] - \mathbb{P}\left[Y_t^+[i] = -1|A_t\right]\right) && (4.5)\\
&\quad + \mathbb{P}\left[A_t^{\complement}\right]\left(\mathbb{P}\left[Y_t^+[i] = 1|A_t^{\complement}\right] - \mathbb{P}\left[Y_t^+[i] = -1|A_t^{\complement}\right]\right).
\end{aligned}$$

Now if event $A_t^{\complement}$ holds, $\mathcal{T}_t[i] = \mathcal{B}_t[i]$, and so by symmetry vertex i must take values -1 and $+1$ uniformly. Furthermore, by the monontonicity of our coupling, if A_t holds, we must have that $\mathcal{T}_t[i] = +1$ and $\mathcal{B}_t[i] = -1$. So

$$m_t(i) = \mathbb{P}[A_t]. \tag{4.6}$$

Finally, given a target time t^*, X_i is defined such that $\{X_i = 1\} = A_{t^*}$. So

$$\mathbb{P}[X_i = 1] = m_{t^*}(i). \tag{4.7}$$

This motivates the following restatement of part of Lemma 2.1 from [11].

Lemma 4.2 ([11], Lemma 2.1). *There exist some constant $c_{\beta,d} > 0$ such that for any $t > 0$,*

$$m_t \leq 2e^{-c_{\beta,d}t} \tag{4.8}$$

Corollary 4.3.

$$\mathbb{P}[X_i = 1] \leq 2e^{-c_{\beta,d}t^*} \tag{4.9}$$

4.2 Setup

Define the time

$$t_c(n) = \inf\left\{t > 0 : m_t = \frac{1}{n}\right\}. \tag{4.10}$$

Fix z and a time of interest $t^* = t_c(n) + z$.

Lemma 4.4 ([15], Claim 3.3). *On any graph with maximum degree Δ, for any $t, s > 0$ we have*

$$e^{-2s} \leq \frac{\sum_i m_{t+s}[i]^2}{\sum_i m_t[i]^2} \leq e^{-2(1-\beta\Delta)s}. \tag{4.11}$$

The following corollary is then straightfoward.

Corollary 4.5. *On any vertex transitive graph with degree Δ, m_{t^*} can be bounded as follows:*

For $z \geq 0$,

$$\frac{e^{-z}}{n} \leq m_{t^*} \leq \frac{e^{-(1-\beta\Delta)z}}{n}. \tag{4.12}$$

For $z \leq 0$,

$$\frac{e^{-(1-\beta\Delta)z}}{n} \leq m_{t^*} \leq \frac{e^{-z}}{n}. \tag{4.13}$$

Corollary 4.6. *On any vertex transitive graph with degree Δ, for $\beta < 1/\Delta$*

$$\ln(n) \leq t_c(n) \leq \frac{\ln(n)}{1 - \beta\Delta} \tag{4.14}$$

4.3 Proof of Theorem 4.1

Lemma 4.7. *Using the above setup*

$$\leq \lim_{n\to\infty} \lambda \leq nm_{t^*} \tag{4.15}$$

Proof.

$$\begin{aligned}
\lambda &= \sum_{i\in V} \mathbb{E}\left[\frac{X_i}{X_i + U_i} I[X_i + U_i \geq 1]\right] &(4.16)\\
&= \sum_{i=1}^{n} \mathbb{P}(X_i = 1)\mathbb{E}\left[\frac{1}{1+U_i} | X_i = 1\right] &(4.17)\\
&= nm_{t^*}\mathbb{E}\left[\frac{1}{1+U_i} | X_i = 1\right] &(4.18)
\end{aligned}$$

where we have used that X_i is zero-one, (4.7), and the transitivity of the graph. Clearly

$$\mathbb{E}\left[\frac{1}{1+U_i} | X_i = 1\right] \leq 1 \tag{4.19}$$

and so $\lambda \leq nm_{t^*}$.

By Jensen's inequality

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{1+U_i} | X_i = 1\right] &\geq \frac{1}{\mathbb{E}[1 + U_i | X_i = 1]} &(4.20)\\
&= \frac{1}{1 + \mathbb{E}[U_i | X_i = 1]}. &(4.21)
\end{aligned}$$

so in order to find a lower bound for λ we will find an upper bound to $\mathbb{E}[U_i|X_i = 1]$. Now

$$\begin{aligned}
\mathbb{E}[U_i|X_i = 1] &= \sum_{j\in B_i} \mathbb{P}[X_j = 1|X_i = 1] &(4.22)\\
&= \sum_{k=1}^{\lfloor b_n \rfloor} \sum_{|j-i|=k} \mathbb{P}[X_j = 1|X_i = 1] &(4.23)\\
&\leq \sum_{k=1}^{\lfloor b_n \rfloor} \sum_{|j-i|=k} \left(C_{z,\epsilon} e^{2\alpha} n^{-\epsilon} + e^{-k+2+2\alpha}\right) &(4.24)
\end{aligned}$$

by [BLAH]. From [BLAH]

$$\mathbb{E}[U_i|X_i=1] \leq \sum_{k=1}^{\lfloor b_n \rfloor} P(k)\left(C_{z,\epsilon}e^{2\alpha}n^{-\epsilon}+e^{-k+2+2\alpha}\right) \tag{4.25}$$

$$\leq C_{z,\epsilon}e^{2\alpha}b_nP(b_n)n^{-\epsilon}+e^{2+2\alpha}\sum_{k=1}^{\infty}P(k)e^{-k}. \tag{4.26}$$

As $n \to \infty$, the first term vanishes and we are left with [COMPLETE] $\square$

Lemma 4.8. *Let δ_1 be as defined above in [REF]. Then*

$$\lim_{n\to\infty} \delta_1 = 0. \tag{4.27}$$

Proof. Starting with the definition of δ_1, we have

$$\delta_1 = \sum_{i=1}^{n}\sum_{k=0}^{|B_i|}\mathbb{P}[X_i=1,U_i=k]\mathbb{E}\left|\frac{\mathbb{P}[X_i=1,U_i=k|W_i]}{\mathbb{P}[X_i=1,U_i=k]}-1\right| \tag{4.28}$$

$$= n\sum_{k=0}^{|B_i|}\mathbb{E}\left|\mathbb{P}[X_i=1,U_i=k|W_i]-\mathbb{P}[X_i=1,U_i=k]\right| \tag{4.29}$$

by the transitivity of the graph. Let

$$C_i^c = \{j : |j-i| \leq (c_n+b_n)/2\} \tag{4.30}$$

be the set of vertices within distance $(b_n+c_n)/2$ of i and define the events

$$A_1 = \{\exists j \in B_i \cup \{i\}, \exists t \in [0,t^*] : \mathcal{H}_j(t) \nsubseteq C_i^c\} \tag{4.31}$$

and

$$A_2 = \{\exists j \in D_i, \exists t \in [0,t^*] : \mathcal{H}_j(t) \cap C_i^c \neq \emptyset\} \tag{4.32}$$

as well as their intersection

$$A = A_1 \cap A_2. \tag{4.33}$$

From Lemma 3.7,

$$\mathbb{P}[X_i=1,U_i=j|A^{\complement},W_i] = \mathbb{P}[X_i=1,U_i=j|A^{\complement}]. \tag{4.34}$$

Continuing on from (4.29), we split the probabilities into

$$\begin{aligned}
\delta_1 &= n \sum_{k=0}^{|B_i|} \mathbb{E} \Big| \mathbb{P}[X_i = 1, U_i = k | W_i, A] \mathbb{P}[A | W_i] - \mathbb{P}[X_i = 1, U_i = k | A] \mathbb{P}[A] + \quad (4.35) \\
&\qquad \mathbb{P}(X_i = 1, U_i = k | A^{\complement})(\mathbb{P}[A^{\complement} | W_i] - \mathbb{P}[A^{\complement}]) \Big| \\
&\leq n(|B_i| + 1) \mathbb{E} \left[\mathbb{P}[A | W_i] + \mathbb{P}[A] + \left| \mathbb{P}[A^{\complement} | W_i] - \mathbb{P}[A^{\complement}] \right| \right] \quad (4.36) \\
&= n(|B_i| + 1) \mathbb{E} \left[\mathbb{P}[A | W_i] + \mathbb{P}[A] + |1 - \mathbb{P}[A | W_i] - (1 - \mathbb{P}[A])| \right] \quad (4.37) \\
&\leq n(|B_i| + 1) \mathbb{E} \left[\mathbb{P}[A | W_i] + \mathbb{P}[A] + \mathbb{P}[A | W_i] + \mathbb{P}[A]) \right] \quad (4.38) \\
&= 2n(|B_i| + 1) \left(\mathbb{E}[\mathbb{P}[A | W_i]] + \mathbb{P}[A] \right) \quad (4.39) \\
&= 4n(|B_i| + 1) \mathbb{P}[A] \quad (4.40)
\end{aligned}$$

For either A_1 or A_2 to hold, there must exists a history that spreads at least distance $(c_n - b_n)/2$ away from its starting vertex. By a union bound

$$\begin{aligned}
\mathbb{P}[A] &\leq \sum_{j=1}^{n} \mathbb{P}[\mathcal{H}_i \nsubseteq B(i, (c_n - b_n)/2) \times [0, t^*]] \quad (4.41) \\
&= n \mathbb{P} \left[\bigcup_{u \in [0, t^*]} \mathcal{H}_i(t^* - u) \nsubseteq B(i, (c_n - b_n)/2) \right] \quad (4.42)
\end{aligned}$$

Combining this with Lemma 4.14, and recalling our choices of $b_n = \ln(L)$ and $c_n = \ln(L)^2$ we get that

$$\begin{aligned}
\delta_1 &\leq 4n^2(|B_i| + 1) \exp(t^* \Delta - \ln \Delta (c_n - b_n)/2) \quad (4.43) \\
&\leq 4n^{2 + \Delta/(1 - \beta\Delta)}(|B_i| + 1) \exp(\Delta z) \exp(-\ln \Delta (c_n - b_n)/2) \quad (4.44)
\end{aligned}$$

which goes to 0 as $n \to \infty$.

□

Lemma 4.9. *Let δ_4 be as defined above in [REF]. Then*

$$\lim_{n \to \infty} \delta_4 = 0. \quad (4.45)$$

Proof. Starting with the definition of δ_4, we have

$$\delta_4 = \sum_{i=1}^{n} \left(\mathbb{E}[X_i Z_i] + \mathbb{E}[X_i]\mathbb{E}[X_i + U_i + Z_i]\right) \tag{4.46}$$
$$= n\mathbb{E}[X_i Z_i] + nm_{t^*}^2 \left(1 + |B_i| + |C_i|\right) \tag{4.47}$$
$$= nm_{t^*}\mathbb{E}[Z_i|X_i = 1] + nm_{t^*}^2 \left(1 + |B_i| + |C_i|\right) \tag{4.48}$$
$$\leq C_z\mathbb{E}[Z_i|X_i = 1] + \frac{C_z^2}{n} \left(1 + |B_i| + |C_i|\right) \tag{4.49}$$

where

$$C_z = \max(e^{-z}, e^{-(1-\beta\Delta)z}). \tag{4.50}$$

Now from Lemmas 4.12 and 4.13, the second term above vanishes as $n \to \infty$. So we turn our attention to the first term. We have

$$\mathbb{E}[Z_i|X_i = 1] = \sum_{j \in C_i} \mathbb{P}[X_j = 1|X_i = 1] \tag{4.51}$$
$$= \sum_{k=\lfloor b_n \rfloor + 1}^{\lfloor c_n \rfloor} \sum_{j:|i-j|=k} \mathbb{P}[X_j = 1|X_i = 1]. \tag{4.52}$$

By Lemma 4.15, for any $0 < \epsilon < 1$, there exists an α such that for sufficiently small β,

$$\mathbb{E}[Z_i|X_i = 1] \leq \sum_{k=\lfloor b_n \rfloor + 1}^{\lfloor c_n \rfloor} \sum_{j:|i-j|=k} \left(C_{z,\epsilon} e^{2\alpha} n^{-\epsilon} + e^{-k+2+2\alpha}\right) \tag{4.53}$$
$$\leq \sum_{k=\lfloor b_n \rfloor + 1}^{\lfloor c_n \rfloor} P(k) \left(C_{z,\epsilon} e^{2\alpha} n^{-\epsilon} + e^{-k+2+2\alpha}\right) \tag{4.54}$$
$$\leq (c_n - b_n + 1)P(b_n) \left(C_{z,\epsilon} e^{2\alpha} n^{-\epsilon} + e^{-b_n+2+2\alpha}\right) \tag{4.55}$$

where

$$C_{z,\epsilon} = \exp(-\epsilon z)\max(1, e^{\beta\Delta z}). \tag{4.56}$$

□

4.4 Additional Lemmas

The first of our additional Lemmas comes from [21, Lemma 3.1]. We have slightly modified the statement and proof.

Lemma 4.10 ([21]). *Let*

$$\chi(\mathcal{H}_i) = \#\{((u,t),(v,t)) \in \mathcal{H}_i\} \tag{4.57}$$

count the total number of horizontal edges in $\mathcal{H}_i$ *and let*

$$\mathcal{L}(\mathcal{H}_i) = \sum_{i\in V} \int_0^{t^*} I_{(i,t)\in\mathcal{H}_i} \,\mathrm{d}t \tag{4.58}$$

be the sum of the lengths of all the horizontal edges in $\mathcal{H}_i$*. Then for any* $0 \leq \eta < 1$*,* $\lambda \in \mathbb{R}$*,* $\alpha > -\ln(1-\eta)$*, there exists* $\beta_0 = \beta_0(d,\eta,\lambda,\alpha)$ *such that if* $\beta < \beta_0$ *then for any* $A \subseteq V$*,*

$$\mathbb{E}[\lambda \exp(\chi(\mathcal{H}_A) + \eta\mathcal{L}(\mathcal{H}_A))] \leq \exp(\alpha|A|) \tag{4.59}$$

where $\mathcal{H}_A$ *is defined as*

$$\mathcal{H}_A = \bigcup_{i\in A} \mathcal{H}_i. \tag{4.60}$$

Proof. We first relax our histories to our alternative construction by observing that

$$\chi(\mathcal{H}_i) \leq \chi(\hat{\mathcal{H}}_i), \qquad Z(\mathcal{H}_i) \leq Z(\hat{\mathcal{H}}_i). \tag{4.61}$$

Let $W_s = |\hat{\mathcal{H}}_i(t^*-s)|$, let $Y_s = \chi(\hat{\mathcal{H}}_i \cap V \times [t^*-s, t^*])$ count the total number of spatial edges observed in the history by time t^*-s and let $Z_s = \mathcal{L}(\hat{\mathcal{H}}_i \cap V \times [t^*-s, t^*])$.

Initially, $W_0 = 1$, $Y_0 = 0$, and $Z_0 = 0$. Recall that an oblivious update of a vertex causes it to be removed from the history and that a non-oblivious update causes the history to branch out to its Δ neighbours. Oblivious updates occur at rate θW_s and cause W_s to decrease by 1. Non-oblivious updates occur at rate $(1-\theta)W_s$ and cause both W_s and Y_s to increase by no more than Δ. The length, Z_s, grows as $\mathrm{d}Z_s = W_s\,\mathrm{d}s$. Therefore we can create a coupled process $(\bar{W}_s, \bar{Y}_s, \bar{Z}_s)$ such that $\bar{W}_s \geq W_s$, $\bar{Y}_s \geq Y_s$, and $\bar{Z}_s \geq Z_s$ in the following way. We start with $(\bar{W}_s, \bar{Y}_s, \bar{Z}_s) = (|A|, 0, 0)$ and at rate $\theta\bar{W}_s$, $\bar{W}_s$ decreases by 1; at rate $(1-\theta)\bar{W}_s$, both $\bar{W}_s$ and $\bar{Y}_s$ increase by Δ; and $\bar{Z}_s$ grows as $\mathrm{d}\bar{Z}_s = \bar{W}_s\,\mathrm{d}s$.

Let $Q_s = \exp\left(\alpha\bar{W}_s + \lambda\bar{Y}_s + \eta\bar{Z}_s\right)$ where α, λ, and η are some fixed constants yet to be determined, and $\alpha > -\ln(1-\eta)$. We will show that Q_s is a supermartingale. Let h be some small time-step. Then [FIX THIS]

$$\begin{aligned}\mathbb{E}\left[Q_{s_0+h} - Q_{s_0} | Q_{s_0}\right] &= h\theta\bar{W}_{s_0}\left(\exp(\alpha(\bar{W}_{s_0}-1)+\lambda\bar{Y}_{s_0}) - \exp(\alpha\bar{W}_{s_0}+\lambda\bar{Y}_{s_0})\right) + \quad (4.62)\\ &\quad h(1-\theta)\bar{W}_{s_0}\left(\exp(\alpha(\bar{W}_{s_0}+\Delta)+\lambda(\bar{Y}_{s_0}+\Delta)) - \exp(\alpha\bar{W}_{s_0}+\lambda\bar{Y}_{s_0})\right) +\\ &\quad \mathcal{O}(h^2)\\ &= \left(\eta + \theta(e^{-\alpha}-1) + (1-\theta)(e^{(\alpha+\lambda)\Delta}-1)\right) h\bar{W}_{s_0}Q_{s_0} + \mathcal{O}(h^2). \quad (4.63)\end{aligned}$$

Dividing through by h and taking h to 0, we have

$$\frac{d}{ds}\mathbb{E}\left[Q_s | Q_{s_0}\right]\bigg|_{s=s_0} = \left(\eta + \theta(e^{-\alpha}-1) + (1-\theta)(e^{(\alpha+\lambda)\Delta}-1)\right)\bar{W}_{s_0}Q_{s_0} \quad (4.64)$$

which is negative for θ sufficiently close to 1. Hence Q_s is a supermartingale for sufficiently large θ.

Define the stopping time

$$\tau = \inf\{s : \bar{W}_s = 0\}. \quad (4.65)$$

Finally, from optional stopping [FIGURE THIS OUT PROPERLY],

$$\begin{aligned}\mathbb{E}[\exp(\lambda\bar{Y}_\tau + \eta\bar{Z}_\tau)] &\leq \mathbb{E}[Q_0] \quad (4.66)\\ &= \exp(\alpha|A|). \quad (4.67)\end{aligned}$$

□

Lemma 4.11. *The number of vertices at distance k on a degree Δ transitive graph is*

$$P(k) = ??? \quad (4.68)$$

Lemma 4.12. *The size of B_i is bounded by*

$$|B_i| \leq \quad (4.69)$$

Lemma 4.13. *The size of C_i is bounded by*

$$|C_i| \leq \quad (4.70)$$

In [11], Lubetzky and Sly proved something similar...

Lemma 4.14. *Let $B(i,l)$ indicate the set of vertices at distance l or smaller from vertex i. The probability that the history of vertex i escapes $B(i,l)$ in time s is bounded by*

$$\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^*-u) \not\subseteq B(i,l)\right] \le \exp\left(s\Delta^2 - l\ln\Delta\right). \tag{4.71}$$

Proof. Let $\mathcal{W} = \{\boldsymbol{w} = (w_1, w_2, \ldots, w_l) : w_1 = i, ||w_{k-1} - w_k|| = 1\}$ be the set of length l sequences of adjacent vertices starting at vertex i. If $\mathcal{H}_i$ contains any vertex outside $B(i,l)$ at a time $u \in [t^* - s, t^*]$ then there must be some sequence $w \in \mathcal{W}$ such that each w_i was updated at some time $t^* > t_i > t^* - s$ and $t_{k-1} > t_k$. Call this event M_w. For any particular sequence w,

$$\mathbb{P}[M_w] = \mathbb{P}[\mathrm{Po}(s) \ge l] \tag{4.72}$$

where $\mathrm{Po}(s)$ is Poisson with rate s. By a union bound over $\mathcal{W}$,

$$\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^*-u) \not\subseteq B(i,l)\right] \le \Delta^{l-1}\mathbb{P}[\mathrm{Po}(s) \ge l]. \tag{4.73}$$

The moment generating function of a poisson random variable with rate s is

$$M(t) = \exp\left(s\left(e^t - 1\right)\right). \tag{4.74}$$

Using a Chernoff bound we have for every $t > 0$,

$$\mathbb{P}[\mathrm{Po}(s) \ge l] \le \exp\left(s\left(e^t - 1\right) - tl\right). \tag{4.75}$$

Overall we have

$$\begin{aligned}\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^*-u) \not\subseteq B(i,l)\right] &\le \Delta^{l-1}\exp\left(s\left(e^t - 1\right) - tl\right) &(4.76)\\ &\le \exp\left(s\left(e^t - 1\right) + l(\ln\Delta - t)\right). &(4.77)\end{aligned}$$

Choosing $t = 2\ln\Delta$,

$$\begin{aligned}\mathbb{P}\left[\bigcup_{u\in[0,s]} \mathcal{H}_i(t^*-u) \not\subseteq B(i,l)\right] &\le \exp\left(s\left(\Delta^2 - 1\right) - l\ln\Delta\right) &(4.78)\\ &\le \exp\left(s\Delta^2 - l\ln\Delta\right). &(4.79)\end{aligned}$$

□

Lemma 4.15. *For any $0 < \epsilon < 1$, there exists an α such that for sufficiently small β,*

$$\mathbb{P}[X_j = 1 | X_i = 1] \leq C_{z,\epsilon} e^{2\alpha} n^{-\epsilon} + e^{-k+2+2\alpha}. \tag{4.80}$$

where $k = |i - j|$ is the distance between vertices i and j and

$$C_{z,\epsilon} = \exp(-\epsilon z) \max(1, e^{\beta \Delta z}). \tag{4.81}$$

Proof. There are two ways in which the update history of vertex j can survive until time 0. The update history can survive without intersecting with the update history of vertex i or the update history of vertex j can merge with the update history of vertex i (whose survival we are conditioning on). Breaking up the probability this way we have

$$\begin{aligned}\mathbb{P}[X_j = 1 | X_i = 1] &= \mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] \\ &\quad + \mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1] &(4.82) \\ &\leq \mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] + \mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1]. &(4.83)\end{aligned}$$

The result follows from Lemmas 4.17 and 4.16. □

This proof is closely based on that in [21].

Lemma 4.16. *Let i and j be the indices of two vertices separated by distance k. Then*

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1] \leq e^{-k+2+2\alpha}. \tag{4.84}$$

Proof. Firstly,

$$\begin{aligned}\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1] &= \frac{\mathbb{P}[\{\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset\} \cap \{X_i = 1\}]}{\mathbb{P}[X_i = 1]} &(4.85) \\ &\leq m_{t^*}^{-1} \mathbb{P}[\{\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset\} \cap \{X_i = 1\}] &(4.86) \\ &\leq m_{t^*}^{-1} \mathbb{P}[\{\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset\} \cap (\{X_i = 1\} \cup \{X_j = 1\})] &(4.87) \\ &= m_{t^*}^{-1} \mathbb{P}[\{\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset\} \cap \{\mathcal{H}_i(0) \cup \mathcal{H}_j(0) \neq \emptyset\}] &(4.88)\end{aligned}$$

since

$$\{X_i = 1\} = \{\mathcal{H}_i(0) \neq \emptyset\}. \tag{4.89}$$

Proceeding backwards from t^*, define S to be the random time at which $\mathcal{H}_i(t) \cup \mathcal{H}_j(t)$ first reduced to less than two vertices, or define $S = 0$ if the combined histories contain at least two vertices all the way to time 0. Since $\mathcal{H}_i(S) \cup \mathcal{H}_j(S)$ contains at most one vertex,

$$\mathbb{P}\left[\mathcal{H}_i(0) \cup \mathcal{H}_j(0) \neq \emptyset | S = t_s\right] \leq m_{t_s}. \tag{4.90}$$

FIX THIS STEP... ASSUMING WE CAN DO THIS...

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1] \leq m_{t^*}^{-1} \mathbb{E}[I_{\hat{\mathcal{H}}_i \cap \hat{\mathcal{H}}_j \neq \emptyset} m_S] \tag{4.91}$$
$$\leq m_{t^*}^{-1} \mathbb{E}[I_{\chi(\mathcal{H}_{ij}) \geq k-1} m_S] \tag{4.92}$$
$$\leq m_{t^*}^{-1} m_{t^*} \mathbb{E}[I_{\chi(\mathcal{H}_{ij}) \geq k-1} e^{t^*-S}] \tag{4.93}$$
$$= \mathbb{E}[I_{\chi(\mathcal{H}_{ij}) \geq k-1} e^{t^*-S}] \tag{4.94}$$
$$\leq \mathbb{E}[e^{\chi(\mathcal{H}_{ij})-(k-1)} e^{t^*-S}] \tag{4.95}$$
$$\leq e^{-(k-1)} \mathbb{E}[e^{\chi(\mathcal{H}_{ij})+(1/2)L(\mathcal{H}_{ij})+1}] \tag{4.96}$$
$$= e^{-k+2} \mathbb{E}[e^{\chi(\mathcal{H}_{ij})+(1/2)L(\mathcal{H}_{ij})}] \tag{4.97}$$

From Lemma 4.10, there exists an $\alpha > \ln(2)$ such that for small enough β,

$$\mathbb{P}[\mathcal{H}_i \cap \mathcal{H}_j \neq \emptyset | X_i = 1] \leq e^{-k+2+2\alpha}. \tag{4.98}$$

□

Lemma 4.17. *There exists an α such that for β sufficiently small*

$$\mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] \leq C_{z,\epsilon} \exp(2\alpha) n^{-\epsilon}. \tag{4.99}$$

Proof.

$$\mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] = m_{t^*}^{-1} \mathbb{P}[X_j = 1, X_i = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset] \tag{4.100}$$
$$\leq m_{t^*}^{-1} \mathbb{P}\left[\frac{1}{2}\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j) \geq t^*\right] \tag{4.101}$$
$$= m_{t^*}^{-1} \mathbb{P}\left[\frac{1+\epsilon}{2}\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j) \geq (1+\epsilon)t^*\right] \tag{4.102}$$

for any $0 < \epsilon < 1$. Continuing

$$\mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] \leq m_{t^*}^{-1} \mathbb{E}\left[I_{(1+\epsilon)\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j)/2 \geq (1+\epsilon)t^*}\right] \tag{4.103}$$
$$= m_{t^*}^{-1} \mathbb{E}\left[\exp((1+\epsilon)\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j)/2 - (1+\epsilon)t^*)\right] \tag{4.104}$$
$$= m_{t^*}^{-1} \exp(-(1+\epsilon)t^*) \mathbb{E}\left[\exp\left(\frac{1+\epsilon}{2}\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j)\right)\right]. \tag{4.105}$$

Since $t^* \geq \ln(n) + z$, and $m_{t^*} \geq \min(e^{-z}, e^{-(1-\beta\Delta)z})/n$,

$$\mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] \leq C_{z,\epsilon} n^{-\epsilon} \mathbb{E}\left[\exp\left(\frac{1+\epsilon}{2}\mathcal{L}(\mathcal{H}_i \cup \mathcal{H}_j\right)\right] \tag{4.106}$$

where

$$C_{z,\epsilon} = \exp(-\epsilon z)\max(1, e^{\beta\Delta z}). \tag{4.107}$$

From Lemma 4.10, there exists an α such that for β sufficiently small

$$\mathbb{P}[X_j = 1, \mathcal{H}_i \cap \mathcal{H}_j = \emptyset | X_i = 1] \leq C_{z,\epsilon}\exp(2\alpha)n^{-\epsilon}. \tag{4.108}$$

□

Chapter 5

Conclusion

Part II

Efficient Optimization for Statistical Inference

Bibliography

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

[2] Wilhelm Lenz. Beiträge zum verständnis der magnetischen eigenschaften in festen körpern. *Phys. Z.*, 21:613–615, 1920.

[3] Sacha Friedli and Yvan Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, November 2017.

[4] R Peierls. On ising's model of ferromagnetism. *Math. Proc. Cambridge Philos. Soc.*, 32(3):477–481, October 1936.

[5] Lars Onsager. Crystal statistics. i. a Two-Dimensional model with an Order-Disorder transition. *Phys. Rev.*, 65(3-4):117–149, February 1944.

[6] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

[7] James Gary Propp and David Bruce Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. 1996.

[8] David A Levin, Yuval Peres, and Elizabeth L Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.

[9] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, May 2002.

[10] Mark Jerrum. Mathematical foundations of the markov chain monte carlo method. In Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 116–165. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[11] Eyal Lubetzky and Allan Sly. Information percolation and cutoff for the stochastic ising model. *J. Amer. Math. Soc.*, 2016.

[12] P Diaconis. The cutoff phenomenon in finite markov chains. *Proc. Natl. Acad. Sci. U. S. A.*, 93(4):1659–1664, February 1996.

[13] David Aldous. Random walks on finite groups and rapidly mixing markov chains. In *Séminaire de Probabilités XVII 1981/82*, pages 243–297. Springer Berlin Heidelberg, 1983.

[14] Eyal Lubetzky and Allan Sly. Cutoff for the ising model on the lattice. *Invent. Math.*, 191(3):719–755, March 2013.

[15] Eyal Lubetzky and Allan Sly. Universality of cutoff for the ising model. *Ann. Probab.*, 45(6A):3664–3696, November 2017.

[16] Danny Nam and Allan Sly. Cutoff for the Swendsen-Wang dynamics on the lattice. May 2018.

[17] Andrea Collevecchio, Eren Metin Elci, Timothy M Garoni, and Martin Weigel. On the coupling time of the heat-bath process for the Fortuin-Kasteleyn random-cluster model. May 2017.

[18] P Erdős and A Renyi. On a classical problem of probability theory. *Publ. Math. Inst. Hung. Acad. Sci.*, Ser. A 6:215–219, 1961.

[19] A D Barbour and O Chryssaphinou. Compound poisson approximation: a user's guide. *Ann. Appl. Probab.*, 11(3):964–1002, August 2001.

[20] A D Barbour, Louis H. Y. Chen, and Wei-Liem Loh. Compound poisson approximation for nonnegative random variables via stein's method. *Ann. Probab.*, 20(4): 1843–1866, 1992.

[21] Eyal Lubetzky and Allan Sly. An exposition to information percolation for the ising model. December 2014.

[22] Bruce G Lindsay. The geometry of mixture likelihoods: A general theory. *Ann. Stat.*, 11(1):86–94, March 1983.

[23] Bruce G Lindsay. The geometry of mixture likelihoods, part II: The exponential family. *Ann Stat*, 11(3):783–792, September 1983.

[24] Aurore Delaigle and Peter Hall. Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):231–252, January 2016.