Algorithmic and geometric aspects of the random-cluster model

Elci, Eren Metin

Award date:
2015

Awarding institution:
Coventry University

Algorithmic and geometric aspects of the random-cluster model

Elçi, E.

Algorithmic and geometric aspects of the random-cluster model

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

(Physics)

by

Eren Metin Elçi

B.Sc., University of Mainz

Applied Mathematics Research Centre
Department of Mathematics and Physics
Faculty of Engineering and Computing
Coventry University
United Kingdom

August 2015

Contents

Abstract

In this thesis we investigate the geometric and algorithmic aspects of the random-cluster model, a correlated bond percolation model of great importance in the field of mathematics and statistical mechanics. We focus on the computational and statistical efficiency of the single-bond or heat-bath Markov chain for the random-cluster model and develop algorithmic techniques that allow for an improvement from a previously known polynomial to a poly-logarithmic runtime scaling of updates for general graphs. The interplay between the (critical) cluster structure of the random-cluster model and algorithmic, as well as statistical, efficiencies is considered, leading to new exact identities. A complementary analysis of certain fragility properties of the Fortuin-Kasteleyn clusters provides new insights into fragmentation phenomena, culminating in a revised scaling relation for a related fragmentation power law exponent, previously only shown for the marginal bond percolation case. By utilising the established structural results, a dynamic fragmentation process is studied that allows for an extraction of characteristics of the equilibrium cluster structure by a careful analysis of the limiting fragments, as well as the entire evolution of the fragmentation process.

Besides focussing on structural and computational aspects, in this dissertation we also analyse the efficiency of the coupling from the past perfect sampling algorithm for the random-cluster model via large-scale numerical simulations. Two key results are the particular, close to optimal, efficiency in the off-critical setting and the intriguing observation of its superiority compared to the alternative Chayes-Machta-Swendsen-Wang approach in three dimensions. Governed by a random runtime, the efficiency of the coupling from the past algorithm depends crucially on the fluctuations of the runtime. In this connection a compelling appearance of universal Gumbel fluctuations in the distribution of the runtime of the coupling from the past algorithm is established, both at and off criticality. Fluctuations at a tricritical point and at a discontinuous phase transition are shown to deviate from this Gumbel law. The above findings in two and three dimensions are supported by a rigorous analysis of certain aspects of the algorithm in one dimension, including a proof of the limiting Gumbel law.

Acknowledgement

This dissertation would have not been possible without the help of many people. First of all, I would like to thank my advisory team consisting of Martin Weigel, Nikolaos Fytas and Christian von Ferber. I am grateful for their guidance, valuable and fruitful discussions and their willingness to pass on their expertise. My deepest gratitude goes to Nikolaos Fytas in whom I have not only found a great scientific teacher and collaborator but also a true friend. Friends and family are the most important thing and I do not want to miss to thank my loved ones for their warmth and support: My mother, Franzi, my brother Levent, Annette & André Mildner, Annelore & Peter Mildner, Bernd & Marion Heinze, Brigitte and Maria Faeth for the wonderful afternoons we spent in Wörth and lastly my uncle Adnan Elçi for supporting me financially at the beginning of my studies.
I am also particularly honoured that I had the chance to learn from many experts of various fields over the last three years: Huseyin Acan, Andrea Collevecchio, Youjin Deng, Timothy Garoni, Nikolay Izmailian, Malte Henkel, Wolfhard Janke, Anastasios Malakis, Ralph Kenna, Alan Sokal and Taras Yavors'kii.
In particular I want to thank Timothy Garoni for inviting me to visit the School of Mathematical Sciences at Monash University in Melbourne where I learned a lot about Markov chain mixing and rigour. I also thank Andrea Colleveccio for inviting me to visit him at the Università Ca' Foscari Venezia in Venice, where the rigorous part of our coupling from the past project is rooted.
Special thanks goes also to my fellow PhD students and colleagues for countless hours of interesting conversations and fun: Kélig Aujogue, Lee Braiden, Pádraig Mac Carron, Hamid Khoshbakht, Kacper Kornet, Ravinder Kumar, Chris Lawton, Richard Looney, Robin de Regt, Emilio José Flores Sola, Joseph Yose, Zongzheng Zhou and Johannes Zierenberg. Many thanks to the many proofreaders of this thesis, you helped me to achieve this. I thank the members of the Institut für Theoretische Physik at the University of Leipzig for their hospitality during my visit at the beginning of this year. Lastly, I wish to thank my "Australian family" Betty Wayne & Bill Knight for their wonderful hospitality during my stay in Melbourne.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Eren Metin Elçi
August 14, 2015

Für meine Mutter,
die immer an mich glaubt.

Chapter 1

Introduction

The simple but powerful idea of utilising a random walk on large combinatorial or sample spaces as a specific type of computational simulation has changed physics and related fields over the last few decades. It is safe to say that computational simulation is now established as the third cornerstone of science, complementing theory and experiment. Theoretically and experimentally intractable problems can often be efficiently approximated using computational methods, new theoretical ideas can be tested in a computer simulation and certain features of experiments can be modelled and reproduced with a computer model. A prominent computational method, the Markov chain Monte Carlo (MCMC) simulation, has emerged as one of the most versatile and widely-used simulation method in condensed matter physics. It has in particular become an indispensable tool for the theory of critical phenomena, the study of continuous phase transitions, and contributed significantly to the advancement of the field [1].

The use of the MCMC method naturally introduces a dynamical dimension into the problem, because one implicitly relies on the Ergodic Theorem for Markov chains [2], that is the equivalence of time and spatial averages. In other words averaging a measured quantity along the trajectory of the Markov chain approximates the correctly weighted spatial average in the state space under study. It has been realised early that the structural complexity of critical phenomena influences the dynamical scales of the utilised Markov chain in the MCMC method [3], in form of slow relaxation and the expansion of temporal correlations. This is a potential bottleneck for the applicability of the MCMC method and is commonly known as critical slowing-down [1]. In particular, it is expected that at a continuous phase transition the inherent (Markov chain) time scales show a power law scaling in the volume or size parameter of the system, and therefore impede the study of large problem instances, that are needed to approximate properties of the thermodynamic limit. For instance, a particularly severe critical slowing-down occurs for localised spin flip dynamics for the Ising or Potts model [4] in statistical physics where the dynamical critical exponent is $z \approx 2$ and even larger in the presence of a conserved quantity such as the magnetisation. The intuition behind this slowing-down is the diffusive nature of the

localised changes made by the Markov chain which need to propagate on all length scales of the system.

A drastic improvement of the critical slowing-down effect for the Potts model was achieved by the invention of the Swendsen-Wang dynamics [5], and later generalised to a larger class of graphical models by Chayes and Machta [6, 7]. The crucial idea in these approaches is to loosen the restriction to localised updates and revert to large scale alterations of the spin configuration, which are particularly efficient at criticality due to the existence of large scale structures found there.
The Swendsen-Wang and Chayes-Machta dynamics rely essentially on a graphical representation of the Potts model known as the random-cluster model. The random-cluster model was introduced in 1969 by Fortuin and Kasteleyn who studied this model in a series of papers [8, 9, 10]. It is a correlated bond percolation model or in other words a family of probability mass functions defined on the space of spanning sub-graphs $\Omega = \{A | A \subseteq E\}$ of a graph $G = (V, E)$, where V and E are the sets of vertices and edges, respectively. We denote by n and m the number of vertices and edges in G, respectively. The model has two parameters $0 \leq p \leq 1$ and $q > 0$, referred to as the bond density and cluster weight, respectively. The probability assigned to a specific spanning sub-graph (V, A) equals

$$\pi_{p,q}(A) = \frac{q^{K(A)} p^{|A|} (1-p)^{m-|A|}}{Z_{p,q}} = \frac{q^{K(A)} v^{|A|}}{\tilde{Z}_{v,q}}, \tag{1.0.1}$$

where $v = p/(1-p)$ and $K(A)$ is the number of connected components in the spanning sub-graph (V, A) (counting isolated vertices), and $|\cdot|$ denotes the set cardinality, that is $|A|$ is the number of edges in A. The normalisation constants $Z_{p,q}$ and $\tilde{Z}_{p,q} = Z_{p,q}/(1-p)^m$ are also called partition functions. As shown by Swendsen and Wang in [5] and formalised and extended by Edwards and Sokal in [11], one obtains a random-cluster configuration from a Potts spin configuration by first adding edges between interacting vertices with the same spin, and then sparsifying this graph by randomly and independently removing edges with probability $e^{-\beta}$, where β is the inverse temperature of the Potts model[1]. Conversely, one obtains a Potts spin configuration from a random-cluster configuration by assigning to each connected component in the random-cluster configuration a randomly and independently chosen spin. Yet, being defined for general positive real cluster weights, the random-cluster model goes beyond the Potts model as well as the bond percolation model and is rather a continuous family of graphical models. It has played a major role in the study of the geometry of the critical phase, see for example [12, 13, 14, 15], and has recently emerged in the rigorous study of two dimensional critical phenomena via its connection to Schramm-Loewner evolution (SLE) [16, 17, 18, 19].

In 1982, 5 years before Swendsen and Wang published their seminal paper [5] in 1987, Mark Sweeny published his results [20] on a study of the Metropolis dynamics for the random-cluster model. In contrast to the Swendsen-Wang and Chayes-Machta approaches,

[1] Here we assume that all spins have the same interaction strength which we set to 1.

this Markov chain again operates with local updates, however uses the random-cluster rather then the spin representation. Remarkably, Sweeny finds that his simulation results in two dimensions are not affected by any critical slowing-down. 25 years later Deng et al. systematically studied in [21] the dynamical critical behaviour of the Sweeny dynamics in two and three dimensions and discovered the surprising result that the involved dynamical critical exponents for the Sweeny dynamics are even slightly smaller than for the Chayes-Machta dynamics. Furthermore, they found the remarkable phenomenon of critical speeding-up. This corresponds to the observation that certain non-local observables can exhibit a fast decorrelation below the scale of sweeps, in the sense that the involved dynamical critical exponent is negative.

The price one has to pay for this improved statistical efficiency is a possible algorithmic slowing-down. As pointed out first by Sweeny, a connectivity check is needed in order to calculate the required transition probability, checking whether the proposed update changes the cluster structure. This introduces a non-local aspect into the dynamics, albeit any two subsequent configurations in the Markov chain sequence differ at most by one edge. Sweeny considered the two dimensional square lattice in [20] and utilised planarity and the medial loop representation [22] of a random-cluster configuration. He represented the loop configuration in a adapted skip-list data structure, which allowed him to perform a configuration update with worst case computational effort logarithmic in the number of vertices. It is clear that this loop construction has limited applicability in higher dimensions and it is therefore desirable to find more versatile algorithmic solutions to the inherent connectivity problem in the Sweeny dynamics. This algorithmic perspective on the single-bond Markov chain approach to the random-cluster model is one of the main themes of this thesis.

We present a study of algorithmic solutions to the connectivity problem in the single-bond chain, or Sweeny dynamics, and investigate the strong interplay between the cluster structure and algorithmic and statistical efficiencies. In particular in Chapter 4 we embark on a study of various algorithmic solutions to the connectivity problem. We start with a simple traversal based breadth-first-search method and discuss its potential bottlenecks. We make the intriguing observation that the expected running time per update step exhibits a power law scaling in the system size, where the exponent is clearly relatable to standard equilibrium critical exponents for the random-cluster model [15]. The appearance of such a power law scaling in the running time of update operations introduces the algorithmic analogue to the statistical critical slowing-down, which we therefore simply refer to as computational critical slowing-down. In order to reduce this computational critical slowing-down we consider improved algorithmic techniques in the subsequent sections of Chapter 4. These approaches are based on an interleaved variant of the breadth-first search in combination with the well known union-find data structure, successfully utilised for bond percolation in [23]. Albeit the running time exponents can indeed significantly be reduced, we still find a polynomial running time scaling. In other words none of the approaches is capable of suppressing the computational slowing-down

completely. To overcome the computational critical slowing-down ultimately we therefore utilise a recent fully dynamic poly-logarithmic dynamic connectivity algorithm [24] that, as the name suggests, supports all required operations with a computational effort poly-logarithmic in the system size. We carefully discuss its application to the Sweeny dynamics and moreover find that the rather complex structure of the algorithm can be drastically simplified and adjusted to the particular setting of critical random-cluster models. This allows us to achieve a significant reduction in memory consumption, while practically even improving the running time of the original connectivity algorithm. Equipped with the poly-logarithmic computational complexity we then compare the Chayes-Machta and Sweeny approaches using a combined computational and statistical efficiency measure.

An analysis of the algorithmic efficiency of a connectivity algorithm for the Sweeny dynamics leads to the study of certain structural fragmentation properties of the random-cluster model. We therefore devote Chapter 5 to the study of fragmentation of Fortuin-Kasteleyn clusters. In particular, we consider the probability of splitting a cluster into two parts upon removal of a randomly chosen edge. This naturally leads to the study of bridges and non-bridges in the stationary random-cluster model, which we analyse by examining their respective densities. We derive an exact identity that relates these to the overall density of edges which is valid for any graph and any choice of p and q. We derive our bridge-edge identity by utilising a random-cluster model variant of the Russo-Margulis formula, originally developed for independent bond percolation [25, 26, 27], which is of independent interest. Now, this exact identity then allows us to derive the asymptotic densities for the square lattice at the critical point $p_{\mathrm{sd}}(q) = \sqrt{q}/(1+\sqrt{q})$ [28] by exploiting duality. In particular we find the simple and concise formulas for the asymptotic density of bridges and non-bridges:

$$\frac{1}{2(1+\sqrt{q})} \quad \text{and} \quad \frac{\sqrt{q}}{2(1+\sqrt{q})}.$$

Beyond that, we also use the bridge-edge identity for a non-asymptotic analysis and derive finite-size corrections to the asymptotic densities, and relate various previously seemingly unrelated quantities in the literature. Moreover we study the scale of fluctuations in the number of bridges by applying the Russo-Margulis method to the variance in the number of bridges. This allows us to reveal an interesting and unexpected finite size scaling for the critical $q < 1$ random-cluster model in two dimensions.
In addition to the analysis of bridges and non-bridges, we also consider the size distributions of the two fragment clusters obtained by a removal of a randomly chosen bridge. This has a clear relevance for traversal based implementations of Sweeny's algorithm, as it determines the running time of, say, a breadth-first search, following the removal of a bridge. The key quantity for our study is the break-up kernel $b_{s'|s}$ which is the probability that, given a fragmentation of a cluster consisting of s vertices, one of the daughter clusters consists of s' vertices. We show that the following scaling ansatz, originally introduced for critical bond percolation [11], holds to high precision also for the fragmentation of critical

Fortuin-Kasteleyn clusters in the thermodynamic limit

$$b_{s'|s} \sim s^{-\phi} \mathcal{F}(s'/s).$$

Additionally, we derive and confirm the following novel scaling relation between the fragmentation exponent ϕ [11] and standard fractal dimensions:

$$\phi = 2 - \frac{d_R}{d_F},$$

where d_R and d_F are the red bond and cluster fractal dimension, respectively. Remarkably, it turns out that a previously assumed scaling relation due to Edwards et al. [11, 29, 30] for bond percolation, does not naturally extend to the random-cluster model, but is rather a marginal case of our derived scaling relation.
Besides edge-induced fragmentation we also consider the interesting problem of vertex fragmentation. In contrast to the binary nature of edge fragmentation, vertex fragmentation allows for higher order fragmentation events. We analyse related asymptotic and finite-size quantities such as the density of the different types of cut vertices, and consider a partial extension of the bridge-edge identity to cut vertices.

While chapter 5 is entirely concerned with fragmentation properties in equilibrium, we consider in Chapter 7 a non-equilibrium process of repeated or iterated fragmentation of critical Fortuin-Kasteleyn clusters. To this end, we introduce a cut-off that suppresses the fragmentation of clusters below a minimal size. In particular, we study the cut-off dependence of the distribution of "stable" fragments which reveals footprints of the equilibrium cluster structure, an observation of obvious practical relevance. Moreover we find numerical evidence in favour of a recursive applicability of the equilibrium break-up kernel $b_{s'|s}$, beyond the one-step equilibrium fragmentation. This is one of the standard assumptions in rate equation approaches to fragmentation, which are mean-field models that describe the time evolution of the fragment size distribution in a fragmentation process. Our numerical findings suggest that certain features of the iterated fragmentation of fractal random structures are describable by rate equation approaches.
Insight into the fragmentation process can also be obtained by analysing the splitting tree associated to a fragmentation process, which encodes the genealogy of fragments. By carefully analysing particular morphological properties of the splitting tree of the iterated fragmentation process of fractal random structures we can in fact extract certain stationary properties of the fragmented object. We further establish interesting similarities between the iterated fragmentation process and the so called beta splitting model [31], which is a perfectly recursive mean-field fragmentation model.

The random-cluster model is also amenable to a particular type of exact or perfect sampling algorithms, both for $0 < q < 1$ [32], and for $q \geq 1$, the latter known as the "coupling from the past algorithm" [33] due to Propp & Wilson. Unfortunately this sampling method

for the random-cluster model seems to have been unnoticed in the statistical physics literature, probably due to algorithmic complications described above. Having an efficient poly-logarithmic connectivity algorithm at our disposal we report in Chapter 6 on a study of the efficiency of the coupling from the past algorithm in two and three dimensions. This perfect sampling algorithm is crucially based on the heat-bath version of the Sweeny dynamics and we describe this connection in detail. We consider the two dimensional square lattice as well as the simple cubic lattice and study the random running time τ it takes for the algorithm to produce a (independent) perfect sample. We examine both off-critical and on-critical settings. The analysis of certain statistical measures of τ allows us to infer connected statistical efficiency statements for the underlying Sweeny Markov chain. Likewise, we find a strong interplay between τ and the cluster structure. The two salient features that emerge from our analysis are the close to optimal efficiency for off critical settings and the appearance of Gumbel fluctuations in the coupling time τ, both off and at criticality. The observation regarding the distribution of the coupling time τ appears to be universal and holds, in an appropriate rescaling, asymptotically in two and three dimensions, irrespective of the cluster weight and bond density, as long as the model does not undergo a discontinuous phase transition. In the latter setting we have strong indications for the appearance of Frèchet fluctuations. We complete our analysis by a rigorous treatment of the coupling process in one dimension on the cycle graph. In this setting we are in particular able to prove two salient features found in higher dimensions.

Besides that we also prove a number of intuitive facts related to the Sweeny dynamics. We establish a duality result for relaxation times [34] and we show that the Metropolis and heat-bath variants are in the same dynamical universality class. Furthermore we prove the useful fact that for monotone Markov chains such as the heat-bath chain for the random-cluster model or the Glauber dynamics for the Ising model one can explicitly construct observables that probe the slowest mode. In the case of the random-cluster model this can be achieved by using the observable that counts the number of open edges, whereas for the Ising Glauber dynamics it is naturally the magnetisation. We believe that the rigorous establishment of this result can be of interest for a future (rigorous) analysis of the relaxation or mixing of the heat-bath chain.

The thesis is organised as follows. In Chapter 2 we review the Markov chain Monte Carlo method with particular emphasis on the random-cluster model and prove some important related aspects. Chapter 3 develops the algorithmic background needed for the subsequent analysis of the computational and statistical efficiency of the Sweeny dynamics in Chapter 4. Chapter 5 describes our results relating to the fragmentation of Fortuin-Kasteleyn clusters, Chapter 6 describes our findings on the coupling from the past algorithm and Chapter 7 discusses the non-equilibrium fragmentation process. Finally in Chapter 8 we conclude and discuss open questions and possible future projects.

Chapter 2

Markov Chain Monte Carlo for the random-cluster model

In this chapter we discuss the Markov chain Monte Carlo method in general and then focus on two particular algorithms for the random-cluster model. We briefly describe the computational complexity of calculating the corresponding partition function and proceed with the introduction of the probably two most important aspects of Markov chains, that is the relaxation towards the unique stationary distribution and correlations in Markov chains. We then focus on the heat-bath Markov chain for the random-cluster model, where we establish an interesting duality result between heat-bath chains for dual bond densities and derive the Li-Sokal bound. We then briefly describe the alternative Chayes-Machta-Swendsen-Wang MCMC algorithm. Finally we elaborate on a recently established connection between the latter and heat-bath algorithm and show why recent comparison results between the two do not provide a conclusive statement about which method is more efficient in practise.

2.1 Computational complexity of the random-cluster model and Markov chain Monte Carlo

In order to understand why numerical approaches, such as the *Markov Chain Monte Carlo* (MCMC) method, are needed to investigate the random-cluster model on non-trivial graphs, we will briefly (and informally) discuss the computational complexity of computing the partition function $Z_{p,q}$ of the random-cluster model with parameters $p, q > 0$. Recall that, $Z_{p,q}$ is formally defined as a sum of 2^m terms corresponding to all spanning sub-graphs of the graph $G = (V, E)$ with m edges, each of the form $q^{k(A)}p^{|A|}(1-p)^{m-|A|}$. Any such term can clearly be calculated in time polynomial in $\max\{m, n\}$. However the vast amount of possible configurations casts an exact calculation of $Z_{p,q}$ in most cases intractable in polynomial time. Strong evidence in favour of this expectation comes from rigorous results on the computational complexity of computing the Tutte polynomial [35].

To see this, note that the partition function of the random-cluster model is actually a specific evaluation of the Tutte polynomial $T(x,y)$ for the graph G, defined as

$$T(x,y) = \sum_{A\subseteq E}\Big(x-1\Big)^{r(E)-r(A)}(y-1)^{|A|-r(A)}, \tag{2.1.1}$$

where $r(A) = n - k(A)$ is the rank of the spanning sub-graph (V,A). It is not hard to see that indeed for a connected graph G, i.e. $k(E) = 1$, one has

$$Z_{p,q} = \frac{1}{q}\ \frac{p}{1-p}\Big)^{n-1}\ T\ \ G;1+\frac{q(1-p)}{p},\frac{1}{1-p}\Big)\Big(\tag{2.1.2}$$

The authors of Ref. [35] show that the Tutte polynomial for a finite graph G belongs to the class of #P-*hard* computational problems for all points in the complex (x,y)-plane, except for a specific family of points, for which the exact calculation of the Tutte polynomial can be done in polynomial time. Applied to the random-cluster model partition function with $p,q > 0$ one finds that for the following special choices of p and q the exact computation of $Z_{p,q}$ can be done with polynomial computational effort for any finite graph G:

- $q = 1$ and $0 \le p \le 1$: Independent bond percolation on G. The partition function is trivial.

- $p,q \to 0$ such that $q/p \to 0$: Number of spanning trees of G (Kirchhoff's matrix tree theorem [36])

Additionally, for the class of bipartite planar graphs it was also shown that $q = 2$, that is the Ising model, has polynomial complexity [37, 35]. Note that for other points one can, in a strict sense, not rule out the existence of polynomial time algorithms, however, the existence of such a polynomial time algorithm would yield polynomial time algorithms for any other algorithm in the class #P, which in particular includes such notoriously intractable problems as counting truth assignments for a Boolean formula in conjunctive normal form, also known as #SAT, or counting the number of Hamiltonian paths in a graph. It is therefore very plausible that the exact computation of $Z_{p,q}$ is intractable except for very special choices of G or p and q. Further, the intractability of computing $Z_{p,q}$ extends directly to the probability distribution $\pi_{p,q}$ on $\Omega = \{A|A \subseteq G\}$. In order to estimate expectations of observables or sample from the desired probability distribution one can thus not expect to rely on exact computations involving $Z_{p,q}$. Remarkably, in some cases one can sample exactly from a desired probability distribution without having to perform exact calculations involving the partition function. We will describe a numerical and rigorous analysis of such an approach in the case of the random-cluster model in chapter 6. Yet, even in the situations where such methods are not available one can utilise approximation schemes of which the Markov chain Monte Carlo method is probably the most versatile approach.

2.1.1 The Markov chain Monte Carlo method

In this thesis we almost exclusively focus on the versatile approximation method known as *Markov Chain Monte Carlo* (MCMC). The basic idea is to construct a sequence of random variables with a certain memoryless property, that has a limiting stationary distribution which equals the desired probability distribution. We start with general considerations regarding Markov chains.
Let Ω be a finite space and $\mathbf{X} = (X_0, X_1, \dots)$ a sequence of random variables in Ω. The sequence $\mathbf{X}$ is a *Markov chain* with state space Ω and transition matrix P if for all $x, y \in \Omega$, all $t \geq 1$, and all events $H_{t-1} = \bigcap_{s=0}^{t-1}\{X_s = x_s\}$ satisfying $\mathbb{P}[H_{t-1} \cap \{X_t = x\}] > 0$, we have

$$\mathbb{P}[X_{t+1} = y | H_{t-1} \cap \{X_t = x\}] = \mathbb{P}[X_{t+1} = y | X_t = x] = P(x, y). \tag{2.1.3}$$

In words, the distribution of X_{t+1} only depends on the value of the sequence attained at time t, and in particular any dependence on the sequence at times before t is "lost". We note that P is necessarily stochastic, that is we have for any $x \in \Omega$, $\sum_{y\in\Omega} P(x, y) = 1$.
The transition matrix P of a finite Markov chain is called *irreducible*, if for any $x, y \in \Omega$ there exists an integer $t_{x,y}$ such that $\mathbb{P}\left[X_{t_{x,y}} = y | X_0 = x\right] = P^{t_{x,y}}(x, y) > 0$. The *period* of a state x is defined to be the greatest common divisor of the integers in $\{t \geq 1 : P^t(x, x) > 0\}$. Notice that it is possible to show, see e.g. Lemma 1.6 in [2], that if P is irreducible then all states in Ω must have the same period. Markov chains for which all states have period 1 are referred to as being *aperiodic*. We can now state the three fundamental theorems underlying the Markov chain Monte Carlo method:

Theorem 2.1.1. *(See [2]). A finite Markov chain with irreducible transition matrix P has a unique stationary distribution π on Ω such that $\pi = \pi P$ and $\pi(x) > 0$ for all $x \in \Omega$.*

While the previous theorem assures the existence of a unique stationary distribution, the next theorem, commonly known as *Convergence Theorem*, states that if the chain is irreducible and aperiodic, then $P^t(x, \cdot)$ converges to its unique stationary distribution π for any $x \in \Omega$:

Theorem 2.1.2. *(See [2]). For a finite Markov chain with irreducible and aperiodic transition matrix P with stationary distribution π one has for any $x, y \in \Omega$*

$$\lim_{t\to\infty} P^t(x, y) = \pi(y). \tag{2.1.4}$$

In what follows we write $\mathbb{E}_\mu[f] = \sum_{x\in\Omega} \mu(x) f(x)$ for any real-valued function (observable) f and probability distribution μ on Ω. Moreover when we are concerned with a Markov chain $(X_t)_{t\geq 0}$ we write $\mathbb{P}_\mu[\cdot]$ for the probability of the event $\cdot$, when X_0 has law μ. Now, the *Ergodic Theorem* intuitively states that "time averages equal space averages" for irreducible Markov chains:

Theorem 2.1.3. *(See [2]). Let f be an observable defined on Ω. If $(X_t)_{t\geq 0}$ is an irreducible Markov chain, then for any starting distribution μ on Ω,*

$$\mathbb{P}_\mu\left[\lim_{T\to\infty}\frac{1}{T}\sum_{s=0}^{T-1} f(X_s) = \mathbb{E}_\pi(f)\right] = 1. \tag{2.1.5}$$

The MCMC method now exploits the above statements and usually works as follows. We are given a distribution π we want to sample from, but the sample space is too large to do so or π itself cannot be computed efficiently. In order to sample from π construct an irreducible and aperiodic Markov chain, let it run "long enough" to come "close to stationarity" and start to use "enough" samples, e.g. for estimating $\mathbb{E}_\pi[f]$. It is clear from the above Theorems that if we could wait infinitely long, then this method would be equivalent to sampling from π (Theorem 2.1.2) or calculating $\mathbb{E}_\pi[f]$ exactly (Theorem 2.1.3). Finding answers to what precisely "long enough" and "close to stationarity" means is the study of the speed of convergence for Markov chains or commonly known as Markov chain mixing [2]. However, before we discuss related concepts, let us consider the question what "enough" samples means under the assumption that the chain is stationary, i.e. $\mathbb{P}[X_0 = x] = \pi(x)$. Suppose further we want to estimate the expectation of an observable f. A natural estimator is

$$Y_T \equiv \frac{1}{T}\sum_{s=0}^{T-1} f(X_s), \quad T > 0, \tag{2.1.6}$$

and it is clear that $\mathbb{E}_\pi[Y_T] = \mathbb{E}_\pi[f]$, in other words Y_T is unbiased. In order to study the fluctuations of Y_T around $\mathbb{E}_\pi[f]$ we can apply Chebyshev's inequality [38], that is we have for any $\eta > 0$:

$$\mathbb{P}\Big[\,|Y_T - \mathbb{E}_\pi[f]| > \eta\Big] \leq \frac{\mathrm{Var}[Y_T]}{\eta^2} \tag{2.1.7}$$

Thus in order to study the likelihood of such deviations one needs $\mathrm{Var}[Y_T]$, which can be shown, [39], to be equal to:

$$\mathrm{Var}[Y_T] = \frac{1}{T^2}\sum_{k=-(T-1)}^{T-1}\big(T - |k|\big)C_f(k). \tag{2.1.8}$$

Here for each integer k, $C_f(k)$ is the covariance of $f(X_t)$ and $f(X_{t+k})$ for the stationary Markov chain under study. Now under the assumption that $\sum_{k=-\infty}^{\infty} C_f(k)$ converges absolutely, which is often the case as many Markov chains have an approximate asymptotic exponential decay for $C_f(k)$ (see section 4.3.1 for a particular example), one can show by means of the dominated convergence theorem [39] that $\lim_{T\to\infty} T\,\mathrm{Var}[Y_T] = 2\tau_{\mathrm{int},f}\,\mathrm{Var}_\pi[f]$ and hence

$$\mathrm{Var}[Y_T] \sim \frac{2}{T}\tau_{\mathrm{int},f}\,\mathrm{Var}_\pi[f]. \tag{2.1.9}$$

The quantity $\tau_{\mathrm{int},f}$ above is the *integrated autocorrelation time* of f and equals

$$\tau_{\mathrm{int},f} \equiv \frac{1}{2} + \sum_{k=1}^{\infty} \rho_f(k), \tag{2.1.10}$$

and $\rho_f(k) = C_f(k)/\mathrm{Var}_\pi[f]$ is the *normalised autocorrelation time* at time-lag k. We note that when the sequence $(X_t)_{t\geq 0}$ is independent, one has $\tau_{\mathrm{int},f} = \frac{1}{2}$ for any f, for which (2.1.9) and (2.1.7) imply the well-known weak law of larger numbers [2]. We remark that for the particular case of (reversible) heat-bath Markov chains, considered in this thesis, we can do an explicit calculation for $\mathrm{Var}[Y_t]$ which is based on the spectral representation for reversible Markov chains (see next section). In particular, it can be verified that one has the spectral expansion $C_f(k) = \sum_{i=2}^{|\Omega|} c_i^2 \lambda_i^k$, where $0 \leq \lambda_i \leq 1$ is the i'th eigenvalue (the lower bound does not hold in general for reversible chains, however it holds for the class of heat-bath chains as described below) and $c_i \in \mathbb{R}$ (depending also on f). This clearly implies $C_f(k) \geq 0$, and we therefore obtain the following bound

$$\mathrm{Var}[Y_T] = \frac{1}{T^2} \sum_{k=-(T-1)}^{T-1} (T-|k|) C_f(k) \leq \sum_{k=-\infty}^{\infty} C_f(k) = C_f(0) + 2\sum_{k\geq 1} C_f(k) \equiv 2\tau_{\mathrm{int},f}\,\mathrm{Var}_\pi[f].$$

Therefore, for heat-bath chains, which applies to the Sweeny dynamics considered in this thesis, one has under the above assumptions

$$\mathbb{P}\Big[|Y_T - \mathbb{E}_\pi[f]| > \eta\Big] \Big(\leq \frac{2\,\mathrm{Var}_\pi[f]\tau_{\mathrm{int},f}}{T\eta^2}, \tag{2.1.11}$$

where for reversible Markov chains only asymptotic equivalence to the upper bound holds. Now in order to assure that the likelihood of such a deviation does not exceed $\epsilon > 0$ one would need a sample of length at least $2\,\mathrm{Var}_\pi[f]\tau_{\mathrm{int},f}/(\epsilon\eta^2)$, which makes it essential to obtain a precise understanding of $\tau_{\mathrm{int},f}$ and how it depends on the size of the state space. Note that also the size dependence of $\mathrm{Var}_\pi[f]$ plays an important role. However, because it is relevant in both the uncorrelated and correlated setting, it is not a particular issue of MCMC methods. Further support for the importance of $\tau_{\mathrm{int},f}$ comes from a central limit theorem for reversible stationary Markov chains, generalising the theorem for sequences of i.i.d. random variables. More precisely one has the following asymptotic law due to Kipnis and Varadhan [40] (see also [41]):

Theorem 2.1.4. *(Kipnis & Varadhan, 1986) For a stationary, irreducible and reversible Markov chain for which $v^2 \equiv \sum_{k=-\infty}^{\infty} C_f(k) < \infty$ one has for $T \to \infty$*

$$\sqrt{T}\,(Y_T - \mathbb{E}_\pi[f]) \xrightarrow{D} \mathcal{N}(0, v^2),$$

where $\xrightarrow{D}$ denotes convergence in distribution and $\mathcal{N}(0,v^2)$ is a normal random variable with mean 0 and variance v^2.

Recalling that $v^2 \sim 2\tau_{\text{int},f}\,\text{Var}_\pi[f]$, we conclude with the important result that $\tau_{\text{int},f}$ is required to obtain error bars for the estimator Y_T of $\mathbb{E}_\pi[f]$ used in MCMC.
It is clear that for the above to be valid, we need stationarity. However, as indicated before, in practise one can not run a Markov chain forever and stationarity is therefore merely an approximation. It is therefore necessary to study the bias induced by sampling from the chain after a finite period. This leads us directly to the question of the meaning of closeness to stationarity. In order to define what is close we need a notion of distance between two probability distributions on the same state space. A widely used metric is the *total variation distance* [2]. Let μ and ν be two probability distributions on Ω, then the total variation distance between μ and ν is defined to be the worst case absolute difference the two distributions μ and ν assign to any event $F \subset \Omega$, that is

$$\left\|\mu-\nu\right\|_{TV} \equiv \max_{F\subset\Omega} |\mu(F)-\nu(F)|. \tag{2.1.12}$$

We note that if $\left\|\mu-\nu\right\|_{TV} < \epsilon$, then the two distributions cannot differ more than ϵ at any $x \in \Omega$. The total variation distance has alternative definitions [2] of which the following is of particular use [2] (because of its explicit nature):

$$\left\|\mu-\nu\right\|_{TV} = \frac{1}{2}\sum_{x\in\Omega}|\mu(x)-\nu(x)|. \tag{2.1.13}$$

This is a direct consequence of the fact that the event F' that maximises $|\mu(F)-\nu(F)|$ is given by $F' = \{x \in \Omega : \mu(x) \geq \nu(x)\}$. Furthermore note that $\mu(F') - \nu(F') = \nu(F'^c) - \mu(F'^c)$, where $F'^c = \Omega \setminus F'$ (the complement of F'), hence we have by (2.1.12)

$$\begin{aligned}
\left\|\mu-\nu\right\|_{TV} &= \mu(F')-\nu(F'),\\
&= \nu(F'^c)-\mu(F'^c),\\
&= \frac{1}{2}\left[\mu(F')-\nu(F')+\nu(F'^c)-\mu(F'^c)\right],\\
&= \frac{1}{2}\sum_{x\in F'}[\mu(x)-\nu(x)] + \frac{1}{2}\sum_{x\in F'^c}[\nu(x)-\mu(x)],\\
&= \frac{1}{2}\sum_{x\in\Omega}\left|\mu(x)-\nu(x)\right|.
\end{aligned}$$

To give an example on how sensitive the total variation distance is to the difference between the two probability distributions, let us consider the following example. Suppose we want to generate samples from the independent bond percolation model on a graph (V,E) with m edges and parameter p. It is not hard to see that the corresponding distribution μ on $\Omega = \{A : A \subseteq E\}$ is $\mu(A) = p^{|A|}(1-p)^{m-|A|}$. Suppose we generate a sample A with distribution μ, by independently removing edges from E with probability $1-p$. However our computer code has a flaw or bug and never deletes edge $f \in E$. The corresponding distribution is $\nu(A) = \mathbb{1}_{\{f\in A\}}p^{|A|-1}(1-p)^{m-|A|}$. The total variation distance $\left\|\mu-\nu\right\|_{TV}$ now certainly fulfils

the following:

$$\begin{aligned}\left\|\mu-\nu\right\|_{TV} &= \frac{1}{2}\sum_{A\subseteq E}|\mu(A)-\nu(A)|,\\ &\geq \frac{1}{2}\sum_{\substack{A\subseteq E:\\ f\notin A}}\mu(A),\\ &= \frac{1}{2}\pi_{p,q=1}[f\notin A],\\ &= \frac{1-p}{2}.\end{aligned}$$

Which in particular holds for $m\to\infty$, thus the total variation distance remains finite even if the edge f intuitively contributes a vanishing fraction when $m\to\infty$. Equipped with this notion of distance one can now formalise the notation of "a Markov chain is close to its stationary distribution", by using the total variation distance between $P^t(x,\cdot)$ and π which we write as $d_x(t)=\left\|P^t(x,\cdot)-\pi\right\|_{TV}$. Given a starting state $x\in\Omega$ one defines the *mixing time* of state x by

$$t_{\text{mix}}^{(x)}(\epsilon)=\min\{t: d_x(t)\leq\epsilon\}\quad \epsilon>0,$$

i.e. $t_{\text{mix}}^{(x)}(\epsilon)$ is the first time $P^t(x,\cdot)$ and π come within ϵ in total variation distance. We note that it is possible to show that $d_x(t+1)\leq d_x(t)$, that is increasing t beyond $t_{\text{mix}}^{(x)}(\epsilon)$ can only bring the chain closer to π, c.f. e.g. [2]. Now, *the* mixing time of a Markov chain is the maximal mixing time over all starting states in Ω, i.e. $t_{\text{mix}}(\epsilon)\equiv\max_{x\in\Omega}t_{\text{mix}}^{(x)}$. We follow a standard convention [2] and write $t_{\text{mix}}\equiv t_{\text{mix}}(1/4)$. Note, that once t_{mix} is known one can push the total variation further down to any $\epsilon<1/4$ because one has generically that $t_{\text{mix}}(\epsilon)\leq\lfloor\log_2\epsilon^{-1}\rfloor t_{\text{mix}}$, c.f. e.g. [2]. In typical settings in physics, e.g., in the study of critical phenomena in statistical physics, one is interested in the size dependence of expectations of certain observables. In situations where this is approximated by utilising the MCMC method, it is thus crucial to understand (at least heuristically or numerically) how t_{mix} depends on the size of system under study, as we consider a sequence of Markov chains with increasing state spaces. For instance, for the random-cluster model on a graph G we know that $|\Omega|=2^m$, hence in order for the MCMC to be an efficient method to use we demand that the mixing time is polynomial in m, or alternatively poly-logarithmic in the size of the state space. This coincides with the widely used notion of *rapidly mixing* Markov chains. The mathematical study of mixing times is a beautiful and young theory which has seen significant progress in recent years [2, 42].

2.1.2 Reversible Markov chains

A particular important class of Markov Chains, mainly due to their rich mathematical structure outlined below, are *reversible* Markov chains, defined by the following *detailed*

balance equations:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \text{ for all } x,y \in \Omega. \tag{2.1.14}$$

More precisely a Markov chain with transition matrix P that fulfils the detailed balance equations (2.1.14) with respect to π is called reversible and it follows immediately that π is *a* stationary distribution of P

$$\sum_{x\in\Omega} \pi(x)P(x,y) = \pi(y) \sum_{x\in\Omega} P(y,x) = \pi(y).$$

We emphasise that (2.1.14) are only sufficient conditions for π to be the stationary distribution of P and we do not claim that a probability distribution π, that fulfils the detailed balance equations with respect to a transition matrix P, is unique. The reason for the prominence of reversible Markov chains, in particular in the MCMC setting, is two-fold. First of all the detailed balance equations (2.1.14) provide an easy way to assure that a desired probability distribution π is a stationary distribution of P and we only need to be able to determine ratios $\pi(x)/\pi(y)$, which in particular avoids the need of calculating the intractable partition function (normalisation constant). Often one actually starts with π and constructs P such that (2.1.14) are fulfilled with additional considerations regarding irreducibility. Second of all, reversible Markov chains have a beautiful, well established, mathematical structure which is based on the observation that one can find a spectral representation of P. We briefly review the important facts and refer the reader for more details to e.g. [2]. To start with, reversibility of P implies that the matrix A defined via $A(x,y) \equiv \sqrt{\pi(x)}P(x,y)/\sqrt{\pi(y)}$ is symmetric. This in turn allows us to apply the spectral theorem for symmetric matrices [43]. Hence we can conclude that the inner product space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle)$ has an orthonormal basis of eigenvectors[1] of A, $\{\phi_j\}_{j=1}^{|\Omega|}$ with real eigenvalues λ_j. Here $\langle\cdot,\cdot\rangle$ is the standard inner product given by $\langle f,g\rangle = \sum_{x\in\Omega} f(x)g(x)$. By definition, orthonormality of ϕ_i, ϕ_j is equivalent to $\langle\phi_i,\phi_j\rangle = \delta_{i,j}$, where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise. Further, it is not hard to see that $\sqrt{\pi}$ is an eigenfunction of A with corresponding eigenvalue 1, thus we set $\phi_1 = \sqrt{\pi}$ which clearly has $\lambda_1 = 1$. Now the spectral decomposition allows us to write any function $f : \Omega \to \mathbb{R}$ as $f(x) = \sum_{j=1}^{|\Omega|} \langle f,\phi_j\rangle\phi_j(x)$. With D_π being the diagonal matrix that has $\sqrt{\pi}$ along its diagonal, we have $A = D_\pi P D_\pi^{-1}$. It is easy to verify that $f_j = D_\pi^{-1}\phi_j$ is an eigenvector of P with eigenvalue λ_j for $1 \le j \le |\Omega|$:

$$Pf_j = PD_\pi^{-1}\phi_j = D_\pi^{-1}\left(D_\pi P D_\pi^{-1}\right)\phi_j = D_\pi^{-1}A\phi_j = \lambda_j D_\pi^{-1}\phi_j = \lambda_j f_j.$$

The eigenvectors f_i, f_j can also be shown to be orthonormal with respect to a modified inner product $\langle f,g\rangle_\pi = \sum_{x\in\Omega} f(x)\pi(x)g(x)$. Now orthonormality of $\{f_i\}_{j=1}^{|\Omega|}$ follows from:

$$\delta_{ij} = \langle\phi_i,\phi_j\rangle = \langle D_\pi f_i, D_\pi f_j\rangle = \langle f_i, f_j\rangle_\pi.$$

[1] We use the term eigenvectors and eigenfunctions synonymously. This is because any vector in the inner product space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle)$ can be thought of as a mapping/function from Ω to $\mathbb{R}$.

We conclude that the eigenvectors of a reversible transition matrix P yield an orthonormal basis for the inner product space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle_\pi)$. It is then in particular possible to decompose $P^t(x,y)$ as [2]

$$P^t(x,y) = \pi(y) + \sum_{j=2}^{|\Omega|} f_j(x) f_j(y) \pi(y) \lambda_j^t \quad \forall x, y \in \Omega,$$

where we have used that $f_1 = D_\pi^{-1}\phi_1 = \mathbf{1}$ (the constant vector with elements 1). Now, in order for $P^t(x,y) \to \pi(y)$, it suffices that $|\lambda_j| < 1$ for $j \geq 2$. Indeed it can be shown that for irreducible and aperiodic transition matrices P, all eigenvalues are in magnitude not larger than 1, -1 is not an eigenvalue and the vector space of eigenfunctions corresponding to eigenvalue 1 is the one-dimensional space generated by $\mathbf{1}$ [2]. Whence we can label the eigenvalues of an irreducible, aperiodic and reversible transition matrix P in decreasing order:

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} > -1.$$

With $\lambda_* \equiv \max\{|\lambda_2|, |\lambda_{|\Omega|}|\}$ we define the *absolute spectral gap* by $\gamma_* \equiv 1 - \lambda_*$. Similarly the *spectral gap* γ is defined by $\gamma \equiv 1 - \lambda_2$. The *relaxation time* t_{rel} of a reversible Markov chain with absolute spectral gap γ_* is defined to be $t_{\text{rel}} \equiv \gamma_*^{-1}$. It turns out that the relaxation time t_{rel} and the mixing time $t_{\text{mix}}(\epsilon)$ provide the correct notion of what it means to "wait sufficient long" and collect "enough samples", in the sense of the following non-asymptotic generalisation of the Ergodic theorem 2.1.3

Theorem 2.1.5. *(Theorem 12.19 in [2].) Let (X_t) be a reversible Markov chain. If $r \geq t_{\text{mix}}(\epsilon/2)$ and $t \geq [4\,\text{Var}_\pi[f]/(\eta^2\epsilon)]t_{\text{rel}}$, then for any state $x \in \Omega$ and any function $f : \Omega \to \mathbb{R}$,*

$$\mathbb{P}_x\left[\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) - \mathbb{E}_\pi[f]\right| \geq \eta\right] \leq \epsilon.$$

We underline that Theorem 2.1.5 does not need irreducibility or aperiodicity. Further, compare this result to our previous inequality (2.1.11), which assumed stationarity. The current result relaxes this assumption and merely demands an "equilibration" phase of length at least $t_{\text{mix}}(\epsilon/2)$. Another particular point concerns the relation of $\tau_{\text{int,f}}$ and t_{rel}, the latter being manifestly independent of f, and we expect t_{rel} to be an upper bound for any $\tau_{\text{int},f}$. In order to render this more precisely we need to introduce another important quantity. The *exponential* autocorrelation time of a real-valued function f on the state space Ω is defined by, [39]:

$$\tau_{\text{exp},f} \equiv \limsup_{k\to\infty} \frac{k}{-\log\left(|C_f(k)|\right)}.$$

The intuition behind this definition is that one anticipates that $C_f(k)$ often roughly decays like $\exp\left(-k/\tau_{\text{exp},f}\right)$ for large k, which we will explicitly verify in the setting of the heat-bath chain for independent bond percolation in section 4.3.1. *The* exponential autocorrelation time of a Markov chain is defined to be the supremum over all real-valued functions on

Ω, i.e.

$$\tau_{\text{exp}} \equiv \sup_f \tau_{\text{exp},f},$$

hence it is an upper bound for $\tau_{\text{exp},f}$ of any f and characterises the slowest time scale. Now we can relate τ_{exp} to t_{rel} in particular for reversible, irreducible and aperiodic chains. In order to see this, note that it can be shown that (c.f. e.g. Section 9.2.3. in [39]) $\exp(-1/\tau_{\text{exp}})$ equals the largest eigenvalue, in magnitude, of $Q \equiv P - \Pi$, where Π is the matrix in which all rows equal to π. It is not hard to see that any constant function is in the kernel of Q. Now we know that the set of eigenfunctions of P forms an orthonormal basis of the inner product space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle_\pi)$. So any function (vector) f in that space can be expanded in terms of the eigenfunctions f_j, $f = \sum_{j=1}^{|\Omega|} a_j f_j$. Because $f_1 = (1, \cdots, 1)^T$ we have $\langle f_1, f\rangle_\pi = \mathbb{E}_\pi[f]$ thus orthogonality of f_i, f_j for any $i \neq j$ is equivalent to $\mathbb{E}_\pi[f_i] = \Pi f_i = (0, \cdots, 0)^T$ for $i > 1$. Hence we have that Q equals P for any function with zero projection on f_1 and any function parallel to f_1 (thus corresponding to eigenvalue 1) is in the kernel of Q. It follows that the spectrum of Q is given by the spectrum of P without the largest eigenvalue $\lambda_1 = 1$. For that reason it follows that for any reversible and irreducible finite Markov chain we have:

$$\lambda_* = e^{-\frac{1}{\tau_{\text{exp}}}}.$$

In order to relate τ_{exp} we make use of the following standard bounds for the logarithm, which follow from the Taylor expansion of $\log(1+x)$, and are valid whenever $|x| \leq 1$, c.f. [38]:

$$\frac{x}{1+x} \leq \log(1+x) \leq x.$$

Setting $x = -\gamma_*$ yields

$$t_{\text{rel}} - 1 \leq \tau_{\text{exp}} \leq t_{\text{rel}}.$$

The last step is necessary to relate $\tau_{\text{int}f}$ and t_{rel} is an inequality between τ_{exp} and $\tau_{\text{int},f}$. In order to obtain such an inequality we note that it is possible to relate a modified exponential autocorrelation time $\tilde{\tau}_{\text{exp}}$, which is defined via λ_2 instead of λ_*, (c.f. e.g. Section 9.2.3. in [39]), to $\tau_{\text{int,f}}$ via:

$$\tau_{\text{int},f} \leq \tilde{\tau}_{\text{exp}}\left(1 + O(1/\tilde{\tau}_{\text{exp}})\right) \tag{2.1.15}$$

Where in general one has $\tilde{\tau}_{\text{exp}} \neq \tau_{\text{exp}}$. However, for the particular class of heat-bath chains it was recently[2] established that the corresponding transition matrix has no negative eigenvalues [45]. Hence there is no difference between the two quantities and we can conclude

$$\tau_{\text{int},f} \leq t_{\text{rel}}\,(1 + O(1/t_{\text{rel}})) \tag{2.1.16}$$

Hence we established that the relaxation time t_{rel} provides a safe upper bound for any integrated autocorrelation time. The question of how far the integrated autocorrelation

[2] Apparently A. D. Sokal was aware of the fact in 1996, as can be seen in [44] footnote 38.

time is apart from the actual upper bound is non-trivial and depends on the properties of f. In fact in this thesis, in particular in section 4.4, we show that certain, "slow", observables actually attain the upper bound. On the other hand, we also observe the phenomenon of critical-speeding up, as discussed in 4.3, which effectively "projects" quantities away from the slowest "mode" in the system. It is also important to note that the relaxation and exponential autocorrelation time can be related to the mixing time. In particular, as shown for instance in [2], one can relate both quantities for reversible, irreducible and aperiodic chains:

$$(t_{\text{rel}} - 1)\log\left(\frac{1}{2\epsilon}\right) \leq t_{\text{mix}}(\epsilon) \leq \log\left(\frac{1}{\pi_{\min}\epsilon}\right) t_{\text{rel}}, \tag{2.1.17}$$

where $\pi_{\min} \equiv \min_{x\in\Omega} \pi(x)$. A key point to note is the presence of the factor $\log 1/\pi_{\min}$ which can be large, thus worsening the tightness of the bounds. We note that t_{mix} and t_{rel} are a priori two different quantities and there is no generally applicable argument showing their equivalence, hence it is not surprising that such comparison inequalities leave some scope. Let us come back to the Ergodic Theorem, which in particular required an equilibration phase of length at least $t_{\text{mix}}(\epsilon/2)$ to assure that the likelihood of a deviation of more than η does not exceed ϵ. We can now use (2.1.17) and express this in terms of τ_{exp} and conclude that an equilibration length of $\log(2/(\epsilon\pi_{\min}))(\tau_{\text{exp}} + 1)$ steps is enough to ensure the desired quality.

Before we proceed we remark that often one has no theoretical (rigorous) arguments for either $\tau_{\text{exp}}, t_{\text{rel}}$ and t_{mix} and thus the relevant notions of relaxation and correlations have to be heuristically/numerically explored. In the case for the heat-bath chain for the random-cluster model there is only very restricted theoretical knowledge and we will devote part of this thesis to the numerical and analytical investigation of related relaxation and correlation questions and in particular investigate whether τ_{exp} or equivalently t_{rel} is far off t_{mix} and how close $\tau_{\text{int},f}$ is to τ_{exp} for a class of functions f of general interest in the study of critical phenomena. We emphasise that our primary interest lies in critical phenomena and properties of the heat-bath chain close to a second order, continuous, phase transition.

Let us now introduce the heat-bath chain for the random-cluster model in some detail, which is the only chain on which we focus analytically[3] in this thesis. This is mainly due to its more amenable mathematical structure as compared to the Metropolis chain (monotonicity for instance), however we provide in chapter 4 relevant comparison statements that allow us to translate relevant properties of the heat-bath chain to the Metropolis chain.

[3] On a first view a seemingly conflicting fact is that numerically we consider the Metropolis chain. This was simply intended to complement the study Deng et al. in [21]. Our analytical study succeeded the computational part.

2.2 Heat-bath Markov chain for the random-cluster model (or Sweeny's algorithm)

The heat-bath Markov chain (or Glauber dynamics or Gibbs sampler [27, 2]) is a versatile MCMC algorithm. It provides a general recipe to construct a reversible Markov chain for state spaces Ω of a certain structure: Suppose S and V are two finite sets and $\Omega \subseteq S^V$. One can think of a configuration as an assignment of a value in S to any element of V. Examples of models that posses such a structure are the Ising and Potts model as well as proper colourings or anti-ferromagnetic models. To see that the configuration space of the random-cluster model has this structure recall that $\pi_{p,q}$ assigns positive probability to elements in $\{A : A \subseteq E\}$, where as usual $G = (V, E)$ is a finite graph. Any element $A \in \Omega$ can thus be uniquely mapped to a $\{0,1\}$ valued vector, indexed by the edges in E, such that for $e \in E$ the corresponding element is 1 if and only if $e \in A$. Equivalently one can view the state space as the m-dimensional hypercube $\{0,1\}^m$, where m is the number of edges in G. In what follows we refer to the value e attained in $\{0,1\}$ as the state of e in configuration A and denote it by $A(e)$.

Now the heat-bath Markov chain chooses an edge $e \in E$ uniformly at random and then updates the state of e according to the stationary distribution of e, conditioned on the state of all other edges. More precisely for $A \in \Omega$ define

$$\Omega_{A,e} \equiv \{A' \in \Omega : A'(f) = A(e) \ \forall f \in E \setminus \{e\}\ ,$$

which is the set of configurations that differ from A at most in the state of e. With the standard notation $\pi_{p,q}(S) \equiv \sum_{x \in S} \pi_{p,q}(x)$ for any $S \subseteq \Omega$, we define for $e \in E$

$$P_e(A, A') = \frac{\pi_{p,q}(A')}{\pi_{p,q}(\Omega_{A,e})} \mathbb{1}_{\{A' \in \Omega_{A,e}\}}, \tag{2.2.1}$$

where $\mathbb{1}_{\{A' \in \Omega_{A,e}\}}$ equals 1 if $A' \in \Omega_{A,e}$ and 0 otherwise. An important point to highlight is that the ratio $\pi_{p,q}(A')/\pi_{p,q}(\Omega_{A,e})$ does not depend on $Z_{p,q}$ and hence avoids the computational challenge related to the calculation of $Z_{p,q}$ outlined at the beginning of this chapter. Another crucial observation is that P_e is reversible with respect to $\pi_{p,q}$ for any e. This follows from the fact that if $A' \in \Omega_{A,e}$ then $\Omega_{A,e} = \Omega_{A',e}$. However it is not hard to see that for fixed e the Markov chain described by P_e is not irreducible (the chain can only alternate between the two configurations in $\Omega_{A,e}$ for a fixed A and e). To cure this and establish irreducibility one can, among other possibilities, use a sequence of randomly selected edges performing an update prescribed by P_e for each of them. In this case we obtain the following transition matrix P:

$$P(A, A') = \frac{1}{m} \sum_{e \in E} P_e(A, A') \tag{2.2.2}$$

Note that reversibility of P_e naturally extends to the reversibility of P. To show that P is irreducible, we fix two configurations $A, A' \in E$ and write $S = A \triangle A'$ for the symmetric difference between A and A', that is all edges either in A or in A', but not in both. Choose an arbitrary order of edges in S and denote the i'th edge by e_i for $1 \leq i \leq |S|$. Then consider the following sequence of configurations:

$$A \to A \triangle \{e_1\} \to \Big(A \triangle \{e_1\}\Big) \triangle \{e_2\} \to \cdots \to A \triangle S \equiv A'$$

The last step follows from the fact that the symmetric difference is commutative and associative. Define recursively $A_i = A_{i-1} \triangle \{e_i\}$ for $i \geq 1$ and $A_0 = A$. We then have $A_i \in \Omega_{A_{i-1}, e_i}$ and hence, by (2.2.1), $P(A_{i-1}, A_i) > 0$ for $i \geq 1$. Now the sequence $\{A_0, A_1, \cdots, A_{|S|}\}$ has probability (conditioned on the chain starting in A_0)

$$\prod_{i=1}^{|S|} P_{e_i}(A_{i-1}, A_i) > 0,$$

which shows that for $t = |S| \leq m$ one has $P^t(A, A') > 0$. It follows that P is irreducible. Last but not least, the fact that $P(A, A) > 0$ for any $A \in \Omega$ demands that $1 \in \{t \geq 1 : P^t(A, A)\}$ which in turn shows that the period of A is 1 and hence P is aperiodic.

Lemma 2.2.1. *The heat-bath Markov chain for the random-cluster model with parameters $p, q > 0$ on a finite graph is reversible, irreducible and aperiodic.* ■

In order to derive an explicit expression for (2.2.1) we observe that $\Omega_{A,e} = \{A_e, A^e\}$, where we defined $A_e \equiv A \setminus \{e\}$ and $A^e \equiv A \cup \{e\}$, of which one necessarily equals A. Now for two configurations $A \in \Omega$ and $B \in \Omega_{A,e}$ we have

$$\begin{aligned}
\frac{\pi_{p,q}(B)}{\pi_{p,q}(\Omega_{A,e})} &= \frac{q^{k(B)} v^{|B|}}{q^{k(A^e)} v^{|A^e|} + q^{k(A_e)} v^{|A_e|}}, \\
&= \frac{q^{k(B)-k(A_e)} v^{|B|-|A_e|}}{q^{k(A^e)-k(A_e)} v + 1}, \\
&= \begin{cases} \frac{q^{k(B)-k(B_e)} v}{q^{k(B)-k(B_e)} v + 1} & \text{if } e \in B, \\ 1 - \frac{q^{k(B^e)-k(B)} v}{q^{k(B^e)-k(B)} v + 1} & \text{if } e \notin B. \end{cases}
\end{aligned}$$

Hence not so surprisingly, $P_e(A, B)$ depends on whether $e \in B$ or $e \notin B$. However, as opposed to the Metropolis chain, there is no dependence on whether $e \in A$ or not. The most important point to observe is that and how the transition matrix depends on the importance of e for the connected component structure in B. More precisely, note that one clearly has $0 \leq k(A_e) - k(A^e) \leq 1$. We call an edge $e \in E$ for which $k(A_e) - k(A^e) = 1$ *pivotal* to (or for) A and *non-pivotal* otherwise. Note that this definition does not depend on whether $e \in A$ or not; It depends on A_e or (equivalently A^e) in a non-local way: In contrast to other heat-bath chains, such as for the Ising model or proper colourings, the heat-bath chain transition matrix P_e does *not* depend on a local and fixed set of other variables on

edges or vertices, but depends explicitly on the connectivity structure of A. This property makes the algorithm special and so efficient in many cases, and on the other hand also notoriously hard to analyse. Moreover, if $e \in A$ we call the edge a *bridge* and *non-bridge* otherwise. We illustrate the definitions in Figure 2.1.

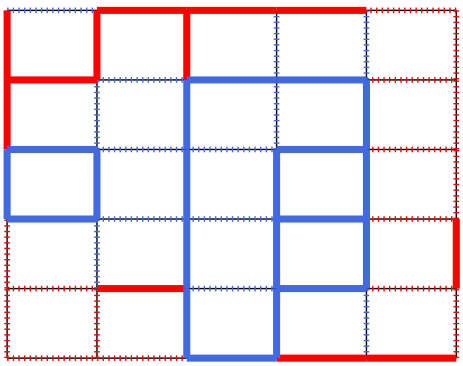


Figure 2.1: *A spanning sub-graph A of the 6×6 grid with pivotal edges in red and non-pivotal edges in blue. Moreover open edges ($e \in A$) are drawn with solid and closed edges ($e \notin A$) with dashed lines. Hence solid red lines correspond to bridges and solid blue lines to non-bridges.*

With these definitions we can henceforth write the transition matrix in (2.2.1) as

$$P_e(A,B) = \begin{cases} a(A;e) & \text{if } e \in B, \\ 1 - a(A;e) & \text{if } e \notin B, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2.3}$$

where we have defined

$$a(A;e) = \begin{cases} \tilde{p}(p,q) \equiv \frac{p}{p+(1-p)q} & \text{if } e \text{ is pivotal to } A, \\ p & \text{otherwise.} \end{cases} \tag{2.2.4}$$

The quantity $\tilde{p}(p,q)$ is less than p for $q > 1$, equals p for $q = 1$ and is larger than p for $q < 1$. This means that the heat-bath dynamics for $q > 1$ has a tendency towards non-pivotal edges, i.e. the probability of inserting an edge at a given step is larger for non-pivotal edges than for pivotal edges. This reflects the intuition that the random-cluster model favours configurations with a higher number of connected components for $q > 1$ than for $q \leq 1$. Note that for $q < 1$ the opposite behaviour applies.

Before we proceed with the discussion of general probabilistic and analytic aspects of the heat-bath chain for the random-cluster model, we pause to discuss the computational

obstacle entailed by the unequal weighting of pivotal and non-pivotal edges. Suppose the current configuration is A. In order to decide whether a given edge e should be inserted into the next configuration B, we need to determine whether e is pivotal to A or equivalently B. Pivotality, by definition, states that the number of clusters changes upon deletion/insertion of e from/into A. This property is however not local, in the sense that it depends on the connectivity structure of the graph. In order to decide whether e is pivotal to A, it is sufficient to determine if the vertices incident to e, say x and y, are connected in A_e. In this case it follows that the number of components does not change upon inserting or deleting e into/from A; there is at least one alternative path in A_e that connects x and y and does not use e. Alternatively one can consider A^e and check if it contains a cycle to which e belongs. Both methods, however, demand a non-local analysis of a configuration. This becomes a severe complication for configurations with macroscopic components (of order n, the number of vertices). Such a large component, often called *giant* component, occurs in particular for simulations where the equilibrium random cluster model possesses a second order phase transition [46]. For instance for the random-cluster model on hypercubic lattices $\mathbb{Z}_L^d$ with periodic boundary conditions it is expected [15] that the average cluster size diverges as $\approx L^{\gamma/\nu}$ and the size of the largest component as $\approx L^{d-\beta/\nu}$. We will devote chapter 4 to a careful analysis of this phenomenon and algorithmic methods, presented in the next chapter 3, to improve this *computational slowing down* to an effective logarithmic slowing down. These algorithmic observations underline the importance of a sensible definition of a time scale taking into account the computational complexity of a single Markov step. We will have more to say about this in chapter 4, where we present results on a numerical analysis of such a generalised efficiency measure.
Coming back to the analytic study of the heat-bath chain, we now discuss a duality result in two dimensions, which, roughly, links the statistical efficiency of the heat-bath chain in the low-temperature to the high-temperature regime.

2.3 Duality

In this section we show how the concept of duality in graph theory can be used to derive a relationship between relaxation times of heat-bath chains on dual graphs with corresponding dual bond densities. We then apply this idea in particular to the graph $\mathbb{Z}_L^2$ embedded into the torus, which is the most relevant graph to this thesis. This in turn allows us to establish a strong relationship between sub-critical, that is $p < p_{\mathrm{sd}}(q)$, and super-critical, i.e. $p > p_{\mathrm{sd}}(q)$, mixing times[4].

Let G be a finite graph embedded into a surface. Define the dual graph $G^\star = (V^\star, E^\star)$ as follows: Place dual vertices on the faces in the embedding of G, and for each edge $e \in E$ add a dual edge to $E^\star$ for all pairs of dual vertices corresponding to two faces bordered by e. Let Ω be the power set of E and $\Omega^\star$ the corresponding power set of $E^\star$ (here we associate

[4]The author found out about the equivalent result in [34], after the result for $\mathbb{Z}_L^2$ was established.

a spanning sub-graph with its edge set). We construct the usual bijection $\mathcal{D}$ between Ω and $\Omega^\star$ by the following rule: For any $A \subseteq E$ the associated dual configuration $A^\star$ consists of dual edges in $E^\star$ whose corresponding edge in E is not in A. In other words an edge e is open in A if and only if its dual edge $e^\star$ is closed in $A^\star$. We write $A^\star = \mathcal{D}(A)$. Furthermore, it will turn out to be convenient to explicitly highlight the p and q dependence of the heat-bath transition matrix in (2.2.2) and denote it by $P_{p,q}$.

The starting point is the observation that once an embedding for the graph G is fixed, it is possible to construct a Markov chain for the dual graph $G^\star$ on $\Omega^\star$, based on the heat-bath chain for G on Ω. This is achieved by mapping the configurations of the heat-bath chain on Ω via $\mathcal{D}$ to configurations in $\Omega^\star$, that is to their respective dual configurations. In other words we consider the sequence of dual configurations of configurations in the original Markov chain sequence. Clearly, the constructed stochastic process is Markovian and its corresponding transition matrix $P^\star_{p,q}$ is related to transition matrix $P_{p,q}$ of the primal heat-bath chain. Now, if X_0, X_1 denote two consecutive states of the heat-bath chain on Ω and Y_0, Y_1 the corresponding states of the induced chain on $\Omega^\star$, then we clearly have for any $A, B \in \Omega^\star$:

$$\begin{aligned} P^\star_{p,q}(A,B) &\equiv \mathbb{P}[Y_1 = B | Y_0 = A], \\ &= \mathbb{P}\Big[X_1 = B^\star | X_0 = A^\star\Big], \\ &= P_{p,q}(A^\star, B^\star). \end{aligned}$$

This defines the transition matrix $P^\star_{p,q}$ of the heat-bath-induced Markov chain on $\Omega^\star$. Moreover based on the above result we can show the intuitive fact that $P^\star_{p,q}$ and $P_{p,q}$ have the same spectrum. Suppose $f : \Omega \to \mathbb{R}^{|\Omega|}$ is an eigenfunction of $P_{p,q}$ with eigenvalue λ, then $f^\star \equiv f \circ \mathcal{D}^{-1} : \Omega^\star \to \mathbb{R}^{|\Omega|}$ is an eigenfunction of $P^\star_{p,q}$ with the same eigenvalue. To see this note that we have for any $A \in \Omega^\star$:

$$\begin{aligned} \sum_{B \in \Omega^\star} P^\star_{p,q}(A,B) f^\star(B) &= \sum_{B \in \Omega^\star} P_{p,q}(A^\star, B^\star) f(B^\star), \\ &= \sum_{B \in \Omega} P_{p,q}(A^\star, B) f(B), \\ &= \lambda f(A^\star), \\ &= \lambda f^\star(A). \end{aligned}$$

In the second step we used the fact that $\Omega^\star$ is in one-one correspondence with Ω. Now, we know by Lemma 2.2.1 that $P_{p,q}$ is reversible, irreducible and aperiodic. These properties transfer also to $P^\star_{p,q}$, which hence together with the above spectral observation implies that $\pi^\star_{p,q}$ is the unique stationary distribution of $P^\star_{p,q}$ on $\Omega^\star$. So far this holds for any pair of dual graphs constructed from an embedding into a surface. Random-cluster models on planar graphs have a special property in that one has, exploiting Euler's formula for planar graphs [47] (see also section 5.2.3), that $P^\star_{p,q}$ is itself a random-cluster model heat-bath

chain on $\Omega^\star$ with parameters $p^\star(p,q)$ and q, and moreover we have $\pi^\star_{p,q}(A) = \pi_{p^\star(p,q),q}(A)$ where

$$p^\star(p,q) \equiv \frac{(1-p)q}{(1-p)q+p}. \tag{2.3.1}$$

We note that here $\pi_{p^\star(p,q),q}$ has to be interpreted as the random-cluster model probability distribution of the dual graph $G^\star$. See Figure 2.2 for plots of $p^\star$ for various values of q.

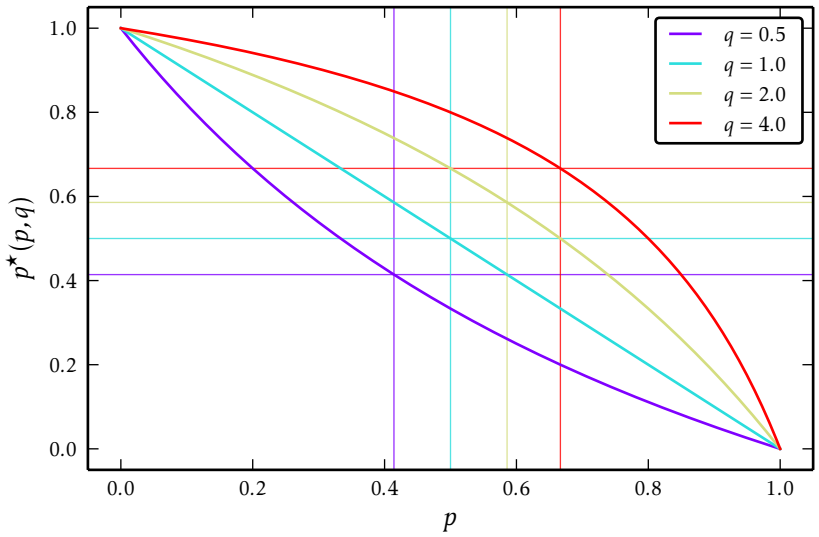


Figure 2.2: *Dual bond density $p^\star(p,q)$ for 4 representative values of q in $[0,4]$. The straight line for $q=1$ corresponds to $p^\star(p,1) = 1-p$. For values of $q<1$ one has $p^\star(p,q) < p$ and for $q>1$ $p^\star(p,q) > p$. The self-dual point $p_{\mathrm{sd}}(q) = \sqrt{q}/(1+\sqrt{q})$ is the fix-point of $p^\star$ for fixed q. This is indicated by the horizontal and vertical lines, which correspond to the values of $p_{\mathrm{sd}}(q)$.*

In words, the dual model of a random-cluster model with parameters p,q on a planar graph G, is a random-cluster model with parameters $p^\star(p,q),q$ on $G^\star$. It is well known that the graph $\mathbb{Z}^2_L$ is not planar. However, having genus 1, it can be embedded into the torus. This embedding has the nice property that $(\mathbb{Z}^2_L)^\star = \mathbb{Z}^2_L$, i.e. $\mathbb{Z}^2_L$ is self-dual when embedded into the torus. In this case one has the modified Euler formula [28] for any $A \in \Omega$:

$$K(A) = |V| - |A| + F(A) - 1 + \delta(A),$$

where $\delta(A) \in \{0,1,2\}$ depends on the homotopy classes of cycles in the configuration A, as defined in [28], where the case $\delta(A)$ evaluates to 0 if and only if the induced spanning sub-graph (V,A) is planar. However, for the following discussion we do not need the precise definition of δ and we only need the fact that it is bounded (for more on cycles on the torus consider section 5.3). It is possible to show, see for example [28], that for $A \in \Omega^\star$

one has:

$$\begin{aligned}\pi^{\star}_{p,q}(A) &\equiv \pi_{p,q}(A^{\star}),\\ &= \frac{1}{Z_{p,q}} q^{K(A^{\star})} p^{|A^{\star}|}(1-p)^{m-|A^{\star}|},\\ &= \frac{1}{Z_{p,q}} q^{F(A)} p^{m-|A|}(1-p)^{|A|},\\ &= q^{1-\delta(A)-n} \frac{p^{\star}(p,q)^{|A|}(1-p^{\star}(p,q))^{m-|A|} q^{k(A)}}{Z_{p,q}}.\end{aligned}$$

Here we used that the number of components in $(V^{\star}, A^{\star})$ equals the number of faces of (V, A), as well as $p^{\star}(p,q)p/\Big[(1-p^{\star}(p,q))(1-p)\Big] = q$. Further it can be verified that, assuming $q \geq 1$, that for any $A \in \Omega^{\star}$:

$$q^{-2}\pi_{p^{\star}(p,q),q}(A) \leq \pi^{\star}_{p,q}(A) \leq q^{2}\pi_{p^{\star}(p,q),q}(A). \tag{2.3.2}$$

Note, by reversing the direction of the inequalities in (2.3.2), the above holds for $q < 1$. In what follows we focus on $q \geq 1$, as the adaption to $q < 1$ is straightforward. In other words, the inequalities (2.3.2) show that $\pi^{\star}_{p,q}$ has the same order of magnitude as $\pi_{p^{\star}(p,q),q}$. This also allows us to show that also the spectral gaps of $P^{\star}_{p,q}$ and $P_{p^{\star}(p,q),q}$ must have the same order of magnitude, using a standard comparison result for reversible Markov chains, see for instance Lemma 13.22 in [2]. Now, in order to apply this comparison technique we need to introduce the *Dirichlet form*. More precisely the *Dirichlet form* of a reversible transition matrix P with stationary distribution π is defined for functions f and g on Ω by

$$\mathcal{E}(f,g) \equiv \langle (\mathbb{1} - P)f, g\rangle_{\pi}. \tag{2.3.3}$$

We need in particular to consider $\mathcal{E}(f)$ defined by

$$\mathcal{E}(f) \equiv \frac{1}{2}\sum_{A,A' \in \Omega} [f(A) - f(A')]^2 \pi(A) P(A,A'). \tag{2.3.4}$$

It is possible to show, c.f. e.g. Lemma 13.11 in [2], that $\mathcal{E}(f) = \mathcal{E}(f,f)$. Now, Lemma 13.22 in [2] applied to our setting states that, if the Dirichlet forms $\mathcal{E}^{\star}(f)$ of the pair $(P^{\star}_{p,q}, \pi^{\star}_{p,q})$ and $\mathcal{E}(f)$ of $(P_{p^{\star}(p,q),q}, \pi_{p^{\star}(p,q),q})$ fulfil the following inequality for an $\alpha > 0$ and any f:

$$\mathcal{E}^{\star}(f) \leq \alpha \mathcal{E}(f)$$

then

$$\gamma\Big(P^{\star}_{p,q}\Big) \leq \left[\max_{A \in \Omega^{\star}} \frac{\pi_{p^{\star}(p,q),q}(A)}{\pi^{\star}_{p,q}(A)}\right] \alpha\, \gamma\Big(P_{p^{\star}(p,q),q}\Big)$$

where $\gamma\Big(P^{\star}_{p,q}\Big), \gamma\Big(P_{p^{\star}(p,q),q}\Big)$ are the spectral gaps of the transition matrices $P^{\star}_{p,q}$ and $P_{p^{\star}(p,q),q}$, respectively. This immediately yields an analogous result for the corresponding relaxation

times $t_{\mathrm{rel}}\left(P^{\star}_{p,q}\right)$ and $t_{\mathrm{rel}}\left(P_{p^{\star}(p,q),q}\right)$

$$t_{\mathrm{rel}}\left(P^{\star}_{p,q}\right) \geq \frac{t_{\mathrm{rel}}\left(P_{p^{\star}(p,q),q}\right)}{\alpha}\left[\min_{A\in\Omega}\frac{\pi^{\star}_{p,q}(A)}{\pi_{p^{\star}(p,q),q}(A)}\right]$$

It remains to be proven that the above conditions of Lemma 13.22 have been fulfilled. We start by writing both Dirichlet forms for arbitrary function f

$$\begin{aligned} 2\mathcal{E}^{\star}(f) &= \sum_{\substack{A,B\in\Omega^{\star}\\ |A\triangle B|=1}} [f(A)-f(B)]^2\,\pi^{\star}_{p,q}(A)P^{\star}_{p,q}(A,B), \\ 2\mathcal{E}(f) &= \sum_{\substack{A,B\in\Omega^{\star}\\ |A\triangle B|=1}} [f(A)-f(B)]^2\,\pi_{p^{\star}(p,q),q}(A)P_{p^{\star}(p,q),q}(A,B). \end{aligned}$$

Note that $\triangle$ denotes the symmetric difference, which restricts the sum only to pairs of configurations A, B, which differ in only one edge. It is easy to see by employing (2.2.1) that these are the only pairs which contribute to the sum. In the following we indeed show that $\pi^{\star}_{p,q}(A) \leq c(p,q)\pi_{p^{\star}(p,q),q}(A)$ for any A as well as $P^{\star}_{p,q}(A,B) \leq c'(p,q)P_{p^{\star}(p,q),q}(A,B)$ for any pair $A, B \in \Omega$ which differ in only one edge, i.e. $|A\triangle B| = 1$. The first inequality follows with $c(p,q) = q^2$ from (2.3.2). For the second inequality note that it follows from (2.2.1) that for any $A, A' \in \Omega$ with $|A\triangle A'| = 1$:

$$\begin{aligned} P^{\star}_{p,q}(A,B) &= P_{p,q}(A^{\star},B^{\star}) \\ &= \frac{1}{m}\frac{\pi_{p,q}(B^{\star})}{\pi_{p,q}(A^{\star})+\pi_{p,q}(B^{\star})}, \\ &\leq \frac{1}{m}\frac{q^2\pi_{p^{\star}(p,q),q}(B)}{q^{-2}\pi_{p^{\star}(p,q),q}(A)+q^{-2}\pi_{p^{\star}(p,q),q}(B)} \\ &= q^4 P_{p^{\star}(p,q),q}(A,B). \end{aligned}$$

In the second step we used that there is only one term contributing to (2.2.2) for $|A\triangle B| = 1$, namely the one corresponding to the edge in which A and B differ. This yields also $c'(p,q) = q^4$ and therefore establishes the above conditions for $\alpha = q^6$, which in turn implies

$$t_{\mathrm{rel}}\left(P^{\star}_{p,q}\right) \geq \frac{t_{\mathrm{rel}}\left(P_{p^{\star}(p,q),q}\right)}{q^6}\left[\min_{A\in\Omega^{\star}}\frac{\pi^{\star}_{p,q}(A)}{\pi_{p^{\star}(p,q),q}(A)}\right] \tag{2.3.5}$$

Now finally inequality (2.3.2) allows us further to bound $\pi^{\star}_{p,q}(A)/\pi_{p^{\star}(p,q),q}(A)$ from below by q^{-2} so that we conclude

$$t_{\mathrm{rel}}\left(P^{\star}_{p,q}\right) \geq \frac{t_{\mathrm{rel}}\left(P_{p^{\star}(p,q),q}\right)}{q^8} \tag{2.3.6}$$

Note that $t_{\mathrm{rel}}\left(P^{\star}_{p,q}\right) = t_{\mathrm{rel}}\left(P_{p,q}\right)$ thus inequality (2.3.6) therefore bounds the relaxation time of a heat-bath chain with parameters p, q from below by the relaxation time of a heat-bath chain with parameters $p^{\star}(p,q)$. The last point to consider is that the duality identity $p^{\star}\left(p^{\star}(p,q),q\right) = p$ yields inequality (2.3.6) with p replaced by $p^{\star}$ and vice versa. Thus we

obtain the key result:

$$q^8 t_{\mathrm{rel}}\Big(P_{p^\star(p,q),q}\Big) \geq t_{\mathrm{rel}}\Big(P_{p,q}\Big) \geq \frac{t_{\mathrm{rel}}\Big(P_{p^\star(p,q),q}\Big)}{q^8} \tag{2.3.7}$$

This result shows that for fixed q the relaxation of heat-bath chains for $\mathbb{Z}_L^2$ with bond densities p and $p^\star(p,q)$ have the same order of magnitude, or equivalently we have $\tau_{\exp}\Big(P_{p^\star(p,q),q}\Big) \asymp \tau_{\exp}\Big(P_{p,q}\Big)$. This result suggests that the non-planarity of $\mathbb{Z}_L^2$ does not cause a too severe effect deviation from the "perfect" relaxation duality for planar graphs; Here by "perfect" we mean that $t_{\mathrm{rel}}\Big(P_{p,q}\Big) = t_{\mathrm{rel}}\Big(P_{p^\star(p,q),q}\Big)$ however this has to be interpreted as an equality of relaxation times for the heat-bath chains on the primal and dual graph (which not necessarily equal) with respective bond densities p and $p^\star(p,q)$. Moreover, the self-dual point $p_{\mathrm{sd}}(q)$ is known to be the critical point on the square lattice for $q \geq 1$ [28]. To appreciate the result 2.3.7 on $\mathbb{Z}_L^2$ consider the case $q = 2$, which corresponds to the Ising model. For the (Spin-)Glauber dynamics for the Ising model, which directly works on the space $\{-1,1\}^{|V|}$, it is rigorously established [48], that for a finite $L \times L$ sub-box in $\mathbb{Z}^2$ with arbitrary boundary conditions the mixing time is constant for $\beta < \beta_c$, polynomial in L for $\beta = \beta_c$ and exponential in L for $\beta > \beta_c$, where $\beta_c = \log(1+\sqrt{2})/2$. This clearly does not show any mixing time symmetry between sub- and super-critical values of β. We emphasise, that similar arguments as above can be used to relate spectral gaps of relaxation times for heat-bath chains on general pairs of dual graphs, which however only in the special case of self-duality turn out to be equal.

2.4 Li-Sokal bound for Sweeny's algorithm

The previously introduced Dirichlet form is of further use in this section, where we establish the so-called *Li-Sokal bound*. This is a lower bound for the relaxation time (or exponential autocorrelation time) of the heat-bath chain for the random-cluster model as well as for autocorrelation times of "energylike" observables. It was originally established by Li and Sokal for the Swendsen-Wang dynamics in [49]. However the authors mention that the same lower bound applies to the heat-bath chain for the random-cluster model (without proof). Here we provide the derivation for completeness. The starting point is the Min-Max-Theorem which provides a variational characterisation of eigenvalues of Hermitian matrices. Applied to the spectral gap γ of a reversible transition matrix P with stationary distribution π it reads (c.f. Lemma 13.12 in [2])

$$\gamma = \min_{\substack{f \in \mathbb{R}^{|\Omega|} \\ \mathrm{Var}_\pi(f) \neq 0}} \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)} \tag{2.4.1}$$

where $\mathcal{E}(f)$ is the Dirichlet form of the pair (P,π). The above expression allows us to bound γ from above

$$\gamma \leq \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)} \tag{2.4.2}$$

for any $f \in \mathbb{R}^{|\Omega|}$ with $\mathrm{Var}_\pi(f) \neq 0$. Recall, that $\mathcal{E}(f)$ is equal to

$$\mathcal{E}(f) \equiv \frac{1}{2} \sum_{A,B\in\Omega} \Big([f(A) - f(B)]^2 \, \pi(A) P(A,B). \tag{2.4.3}$$

Now, for the heat-bath chain for the random-cluster model with transition matrix $P_{p,q}$ we have that that $P_{p,q}(A,B) = 0$ for any $A, B \in \Omega$ which differ in more than one edge ($|A \triangle B| > 1$). In particular, when we choose the function f to be the number of edges in A, i.e. $f(A) = N(A) = \#\{\text{edges in } A\} = |A|$, we clearly have that for any pair A, B such that $P_{p,q}(A,B) \neq 0$:

$$[N(A) - N(B)]^2 \leq 1. \tag{2.4.4}$$

Hence we obtain for the Dirichlet form $\mathcal{E}_{p,q}(N)$ of N for the pair $(P_{p,q}, \pi_{p,q})$:

$$\begin{aligned}
\mathcal{E}_{p,q}(N) &\equiv \frac{1}{2} \sum_{A,B\in\Omega} [N(A) - N(B)]^2 \, \pi_{p,q}(A) P_{p,q}(A,B) \\
&= \frac{1}{2} \sum_{\substack{A,B\in\Omega \\ |A\triangle B|\leq 1}} \Big([N(A) - N(B)]^2 \, \pi_{p,q}(A) P_{p,q}(A,B) \\
&\leq \frac{1}{2} \sum_{\substack{A,B\in\Omega \\ |x\triangle y|\leq 1}} \pi_{p,q}(A) P_{p,q}(A,B) \\
&= \frac{1}{2}
\end{aligned}$$

We conclude therefore with

$$t_{\mathrm{rel}}\Big(P_{p,q}\Big) \Big(\geq 2\,\mathrm{Var}_{p,q}(N). \tag{2.4.5}$$

The intuition behind the Li-Sokal bound and similar Min-Max-Theorem based arguments for Markov chains goes as follows. Because the chain is restricted to localised modifications of the configuration, that is single edge updates, it can only make unit steps in N-space. We can roughly visualise the corresponding process in N-space as a simple random walk with unit steps. In order for the chain to relax, it must at least "diffuse" through the region of N-space covered by the majority of probability mass of the distribution of N, which in turn is roughly $\sqrt{\mathrm{Var}_{p,q}(N)}$ wide. Using the well known relationship between the squared displacement and the number of steps (elapsed time) for a simple random walk, we expect that this takes roughly time $\sqrt{\mathrm{Var}_{p,q}(N)}^2 = \mathrm{Var}_{p,q}(N)$. Moreover it is generally expected (conjectured) that at a second order phase transition, for a large hypercubic lattice with linear dimension L in d dimensions, one has $\mathrm{Var}_{p_c(q),q}(N)/L^d \approx L^{\alpha/\nu}$ [50] which implies, assuming the typical scaling form $\tau_{\mathrm{exp}} \sim L^{z_{\mathrm{exp}}}$ (here τ_{exp} is measured in sweeps that is dL^d steps):

$$z_{\mathrm{exp}} \geq \frac{\alpha}{\nu}. \tag{2.4.6}$$

In two dimensions, in particular for the graph $\mathbb{Z}_L^2$, we can utilise Coulomb gas arguments [51] which show that the width of the distribution of N is of order L^d for $q < 2$ that is $\mathrm{Var}_{p_{\mathrm{sd}}(q),q}(N)/L^d \to \text{const}$ in the sense that α/ν above is negative (for $q = 1$ we have

$\alpha/\nu = -1/2$ in two dimensions). The N-space diffusion argument does therefore not enforce a critical slowing down, which together with the intuition (which in some sense will be made rigorous in section 4.4) that the number of edges has the slowest dynamical scale suggests the absence of slowing down at all. Note that as a reference we can consider the case of independent bond percolation $q = 1$ for which $\mathrm{Var}_{p,1}[N] \asymp L^d$. As we will see in the subsequent parts of this thesis the regime $1 \leq q < 2$ in two dimensions is particularly intriguing from a dynamical point of view. We devote chapter 4 to an in-depth analysis of the dynamical critical behaviour of the heat-bath chain and confirm the Li-Sokal bound, which remarkably in some cases turns out to be surprisingly sharp.

Lastly, let us show how the Li-Sokal bound also applies to the integrated autocorrelation time of the observable N [44]. First note that we have the following spectral expansion of $\rho_g(k)$ for arbitrary function f and $k \geq 1$:

$$\rho_g(k) = \frac{C_g(k)}{\mathrm{Var}_{p,q}[g]} = \frac{\sum_{j=2}^{|\Omega|} \hat{g}_j^2 \lambda_j^k}{\sum_{j=2}^{|\Omega|} \hat{g}_j^2}.$$

Where $\hat{g}_j \equiv \mathbb{E}_{p,q}[g f_j]$, the projection onto the j'th eigenfunction of $P_{p,q}$. Now recall that the heat-bath chain has no negative eigenvalues, hence we can apply Jensen's inequality and find

$$\rho_g(k) \geq \rho_g(1)^k.$$

This implies the following inequality for the integrated autocorrelation time of g:

$$\tau_{\mathrm{int},g} = \frac{1}{2} + \left(\sum_{k \geq 1} \rho_g(k) \right) \geq \frac{1}{2} + \left(\sum_{k \geq 1} \rho_g(1)^k \right) = \frac{1}{1-\rho_g(1)} - \frac{1}{2} = \frac{1}{2} \frac{1+\rho_g(1)}{1-\rho_g(1)}.$$

Additionally one can easily verify that

$$1 - \rho_g(1) = \frac{\mathcal{E}_{p,q}(g)}{\mathrm{Var}_{p,q}[g]}.$$

We obtain therefore

$$\tau_{\mathrm{int},g} \geq \frac{\mathrm{Var}_{p,q}[g]}{\mathcal{E}_{p,q}(g)} - \frac{1}{2}.$$

Applying this to $g = N$ allows us to use the same arguments as above and obtain the equivalent Li-Sokal bound for $\tau_{\mathrm{int},N}$:

$$\tau_{\mathrm{int},N} \geq 2\,\mathrm{Var}_{p,q}[N] - \frac{1}{2}.$$

In the next section we discuss a different MCMC approach for the random-cluster model when $q \geq 1$, due to Chayes, Machta, Swendsen and Wang.

2.5 Chayes-Machta-Swendsen-Wang chain and its relation to the Sweeny dynamics

The idea of Swendsen and Wang [5], which was later generalised by Chayes and Machta [6, 7], is to introduce auxilary spin (colour) variables and consider a Markov chain on the larger joint space of colour and bond variables. The chain is iterated by alternately updating the colour and bond degrees of freedom in a specific way, precisely incorporating the cluster weighting. The Chayes-Machta chain, in its simplest form [50] iterates from one random-cluster configuration A to the next by the following steps:

1. *Activate* each component in (V, A) with probability $1/q$.
2. Remain all inactive components unchanged.
3. Do independent bond percolation with parameter p on the active vertex induced sub-graph of G.

It can be verified that this dynamics is in fact reversible with respect to $\pi_{p,q}$ [7] whenever $q \geq 1$ and is also irreducible. It follows that for any initial configuration, the Chayes-Machta chain converges to its unique stationary distribution $\pi_{p,q}$. The closely related Swendsen-Wang dynamics, originally considered as a Markov chain for the Potts model, and hence defined only for integer q operates as follows:

1. Assign each component in (V, A) independently one spin chosen uniformly at random among the q different spins.
2. For each spin induced sub-graph of G do independent bond percolation with probability p.

The Swendsen-Wang and Chayes-Machta dynamics can also be shown to be governed by the Li-Sokal bound [49] and hence are not a priori free of any critical slowing-down. We will compare the efficiencies of the Sweeny and Chayes-Machta-Swendsen-Wang approach in chapter 4. The Chayes-Machta and heat-bath chains seem to be two a priori unrelated Markov chains for the random-cluster model. However, it is possible to find a common mathematical structure and relate their relaxation times, as first shown in [34] for the Swendsen-Wang chain and recently generalised to the Chayes-Machta chain in [52]. To start with, let $t_{\text{rel}}^{(\text{CM})}$ and $t_{\text{rel}}^{(\text{HB})}$ be the relaxation times of the Chayes-Machta and heat-bath chain, respectively for an arbitrary value of p and $q \geq 1$, and fix any finite graph. Then they key result, as stated in Claim 5.4 in [52], reads:

$$\frac{q^2}{p + (q-1)p} t_{\text{rel}}^{(\text{HB})} \geq t_{\text{rel}}^{(\text{CM})} \geq \frac{t_{\text{rel}}^{(\text{HB})}}{8m \log(m)}. \tag{2.5.1}$$

Firstly, note that the first bound is in a certain sense trivial, as it compares one step of the heat-bath algorithm, which consists of one edge update, to a complete colour-bond cycle

in the Chayes-Machta chain. The second bound is more interesting, because it restricts how far the Chayes-Machta chain can be more efficient than the heat-bath chain.
Here we adopt the natural convention of one "sweep", that is m edge updates in the heat-bath algorithm, and therefore compare $t_{\text{rel}}^{(\text{HB})}/m$ to $t_{\text{rel}}^{(\text{CM})}$. There are two scenarios consistent with the right bound in (2.5.1).

1. The Chayes-Machta algorithm is (statistically) more efficient than the heat-bath sweep approach, that is $t_{\text{rel}}^{(\text{CM})} \leq t_{\text{rel}}^{(\text{HB})}/m$. However, in this case the bound in (2.5.1) states that the former can at most by a factor proportional to $\log(m)$ be faster than the latter. On the level of dynamical critical exponents, we would need to conclude that $z_{\text{exp}}^{(\text{CM})} = z_{\text{exp}}^{(\text{HB})}$.

2. Heat-bath sweeps are more efficient than the Chayes-Machta approach, that is $t_{\text{rel}}^{(\text{CM})} \geq t_{\text{rel}}^{(\text{HB})}/m$.

The marginal case happens, for instance, in the simplest case of bond percolation $q = 1$, where one can verify that the Chayes-Machta chain coincides with a sequential bond percolation algorithm, which has $t_{\text{rel}}^{(\text{CM})} = 1$, whereas on the other hand $t_{\text{rel}}^{(\text{HB})} = m$, as shown in section 4.3.1. Our numerical study presented in chapter 4, also appeared in [53], as well as the earlier (and independent) study in [21], find numerically for two and three dimensions that, at criticality, $z_{\text{exp}}^{(\text{HB})} < z_{\text{exp}}^{(\text{CM})}$, thus supporting the second scenario. However, there is also another notable aspect which makes the heat-bath chain even more effective in particular situations. This is the phenomenon of critical speeding-up, described in detail in section 4.3 and first observed in the random-cluster model setting in [21]. Briefly summarised, it describes the observation that certain global observables, say f, have integrated autocorrelation time that is dominated by m, in other words $\tau_{\text{int},f}/m \to 0$ for $m \to \infty$. Therefore, in a setting where one is merely interested in sampling from the distribution of f, one can effectively sample on a smaller scale than sweeps. In fact, we often find a stronger notion of critical speeding up, that is $\tau_{\text{int},f} \approx m^{1-c}$, where $c > 0$. It is clear that Chayes-Machta algorithm is insensitive to any (critical) speeding-up effect, as the natural intrinsic time scale is one sequential sweep.

Lastly, as we already pointed out earlier, any efficiency statement or comparison of Markov chains has to consider, both the relaxation time and the computational complexity of a single step. The algorithmic aspects of the Chayes-Machta algorithm are straightforward: One colour-bond cycle can be done with $O(m)$ worst-case computational effort utilising, e.g., an union-find data structure or a depth- or breadth-first search based connected component identification. On the other hand we devote the next chapter to a detailed study of algorithmic solutions to the required pivotality determination in the heat-bath chain for the random-cluster model. Once the necessary algorithmic tools are set up, we embark in chapter 4 on a numerical study of both computational and statistical efficiency of Sweeny's algorithm, which concludes with the analysis of a joint efficiency measure.

Chapter 3

Graph algorithms for the random-cluster model

The efficiency of the Sweeny dynamics (both heat-bath and Metropolis) does not solely depend on statistical properties of the underlying Markov chain, but depends also crucially on the computational complexity of the underlying connectivity algorithm. We therefore discuss and collect in this chapter the theoretical background for the various approaches used in our efficiency analysis of the Sweeny dynamics in Chapter 4. This includes (partially) dynamic connectivity algorithms based on depth- and breadth-first searches, union-find data structures as well as poly-logarithmic fully-dynamic connectivity algorithms. Thus in this sense this chapter complements the previous Chapter 2 by the consideration of the computational complexity of a single heat-bath step. Further, for our analysis of static fragmentation properties in Chapter 5 we need somewhat specialised, at least in the physics-literature, algorithms, such as an algorithm for the identification of bridges and cut vertices [54], which we briefly describe. One reason for the need of an alternative method is that previous related studies [55, 56] were based on a "home-made" bridge identification algorithm tailored to $\mathbb{Z}_L^2$ and planar graphs. This clearly has limited applicability in higher dimensions, let alone for its potential worse-than-linear expected running time of the identification of all bridges.

3.1 Why do we need connectivity algorithms?

It comes not as a surprise that in a computational study of the random-cluster model graph algorithms play an important role. Our main application of graph algorithms lies in the implementation of Markov chain Monte Carlo algorithms. To illustrate this consider the heat-bath chain for the random-cluster model on a finite graph $G = (V, E)$, introduced in the previous chapter. Recall that this Markov chain iterates from a configuration A to B by the following steps:

- Choose $e \in E$ uniformly at random.
- Set $B = A \cup \{e\} \equiv A^e$ with probability $a(A;e)$ and $B = A \setminus \{e\} \equiv A_e$ otherwise.

The crucial observation is that a relates to the graph structure of A in a non-local way:

$$a(A;e) \equiv \begin{cases} \frac{p}{p+(1-p)q} & \text{if } e \text{ is pivotal to } A, \\ p & \text{otherwise.} \end{cases} \tag{3.1.1}$$

As explained in the previous chapter, to determine whether a given edge is pivotal to a configuration is a non-local operation. Here by non-local we mean that, unlike other heat-bath chains for which the calculation of a only depends on a (typically) small and fixed[1] set of variables, a depends explicitly on structural non-local properties of A. This poses a particular problem at a point of a second order phase transition, where one observes that the average size of A increases with a power in the number of vertices or linear dimension of the lattice [46, 15].

Now suppose we have a graph algorithm that answers connectivity queries, i.e., it determines if two vertices x and y, not necessarily neighbours, are connected. We can abstract the algorithm into the function $\mathcal{C}(A;x,y)$ which equals 1 if $x \overset{A}{\leftrightarrow} y$ and 0 otherwise. Here $x \overset{A}{\leftrightarrow} y$ denotes the event that x is connected to y by a path of open edges in (V,A). Equipped with $\mathcal{C}$ we can re-formulate the definition of a in (3.1.1) for $e = (x,y)$:

$$a(A;e) = \begin{cases} \frac{p}{p+(1-p)q} & \text{if } \mathcal{C}(A_e;x,y) \neq \mathcal{C}(A^e;x,y), \\ p & \text{otherwise.} \end{cases}$$

Note that necessarily one of the two configurations A_e and A^e equals A. So it follows that the determination of the insertion probability a can be achieved by one call to a connectivity algorithm. It is therefore desirable, especially having second order phase transitions in mind, to find efficient connectivity algorithms [53, 57, 15, 21, 20].

Note that graph algorithms, not particularly connectivity algorithms, can be broadly classified into dynamic and static algorithms. As the names suggest, static algorithms work with a time independent graph. Here one has an initial construction phase of the graph followed by a sequence of non-manipulative operations such as connectivity or shortest path queries. The ultimate goal of such algorithms is to minimize the computational cost for the non-manipulative operations and not so much on the construction phase. Between static and fully dynamic algorithms there are also partially dynamic algorithms. A prominent example is the *disjoint-set data structure* or *union-find data structure*, described below. It turns out that such a union-find data structure belongs actually to the class of *incrementally* partially dynamic graph algorithms. Translated to the graph-theoretic setting

[1] Consider the heat-bath chain for the Ising model on a graph with maximal degree Δ. Any vertex chosen to be updated has at most Δ neighbours and hence it is a-priori clear how many other vertices have to be considered to calculate the transition probability.

this means it allows not only for non-manipulative connectivity- or cluster-size-queries but also for the insertion of edges. A recent application of this data structure to the study of site- and bond-percolation can be found in [23]. Moreover we will show in Chapter 4 how this data structure can be combined with graph-traversal algorithms to yield a fully dynamic connectivity algorithm, which allows for both dynamic insertions and deletions (however with polynomial running time). Note that opposed to the incremental partial dynamic graph algorithms there exist also decremental algorithms, see for instance the recent work [58] for an application of the union-find data structure to a decremental connectivity algorithm for planar graphs. The two probably most famous connectivity algorithms are the *depth-first search* (DFS) and *breadth-first search* (BFS). Both algorithms explicitly traverse the graph to answer queries and will be described in the section. Moreover they are not only of practical relevance but also theoretically very interesting. In particular, we will describe a recent algorithm [54] in section 3.6 that exploits the DFS to identify (all) bridges and cut vertices in a graph in linear time.

3.2 Graph traversals

The depth-first search or short DFS supports the exhaustive traversal of a connected graph, both directed and undirected, and belongs to the class of backtracking algorithms. Backtracking here is related to the inherent property of following a branch of the graph as long as possible, that is until a dead-end or a previously visited vertex is encountered. Before we go into the details of this algorithm we would like to mention a few general aspects related to its applicability. First of all the restriction to connected graphs can be loosened to graphs consisting of multiple connected components. One simply applies the DFS algorithm to all such components. Secondly in the case where G is not connected, and hence the answer to the question if two vertices x and y are connected is not (by definition) immediate clear, one can use the DFS to obtain an answer to a connectivity query. Here one could start an exhaustive DFS in the component of vertex x, denoted C_x, and keep track of all visited vertices. Once C_x is determined one can simply use

$$y \in C_x \Leftrightarrow x \overset{G}{\longleftrightarrow} y. \tag{3.2.1}$$

to answer connectivity queries. This is actually the definition of the connected component of x, in other words $C_x \equiv \{y | y \in V, x \overset{G}{\longleftrightarrow} y\}$. In practical terms, however, one would use a slight modification of the DFS algorithm which directly terminates with a positive answer as soon as the target vertex y has been visited. This saves the overhead of maintaining a list of explored vertices and more importantly usually terminates before the whole component C_x is exhausted. Besides being used for connectivity related problems the DFS has also important theoretical applications such as for planarity testing, 2-vertex and 2-edge-connectivity or the identification of bridges and cut vertices. The latter application will be of importance to the fragmentation studies of the random-cluster model in Chapter

5. Furthermore, DFS has also recently been used to provide a simple proof of the phase transition in the random graph model of Erdős & Rényi [59].

The DFS algorithm is best described by actual computer code, so below we show pseudo code for the recursive variant of the DFS algorithm [47]: The recursive formulation nicely

Algorithm 1 Recursive depth first search

```
function DFS(v)
    if v unvisited then
        mark v as visited
        for all (v, w) ∈ E do
            DFS(w)
        end for
    end if
end function
```

reflects the "follow a branch as long as possible and backtrack" idea underlying the DFS. A call to DFS starting at vertex v will, assuming v has not been visited before, mark v as visited and continue with a call to the DFS on the first neighbour w encountered. This recursion continues until a vertex is encountered that has already been visited. Note that this does not necessarily correspond to the existence of a cycle in the graph. To see this suppose the graph is a tree, hence has no cycles, and the order edges are iterated is from left to right in a planar embedding. In this case the DFS will initially follow the left-most branch until it reaches a leaf ℓ. The next call to DFS(ℓ) then will mark ℓ as visited and iterate through all of its edges. Suppose p is the direct ancestor of ℓ in the tree and hence has been visited before ℓ in the depth-first search. In other words ℓ has been reached through the edge (p, ℓ). Now because G is a tree and ℓ a leaf it follows that the edge iteration in DFS(ℓ) only consists of (ℓ, p) and hence is followed by a call DFS(p) which itself does not execute any loop because p has already been visited. At this step in the program flow the algorithm continues with the edge iteration in the call to DFS(p) and will explore possibly other sub-branches.

It is standard to show that the DFS algorithm is a valid (exhaustive) graph traversal algorithm for connected G and moreover works in linear time. Instead of proving this we introduce a generalised graph traversal algorithm which includes the DFS as a special case. We then show that this general graph traversal algorithm works in linear time and is exhaustive. Besides being interesting in its own right, the analysis of the more general algorithm also implies that the breadth-first search algorithm, another well known graph traversal algorithm, is a valid graph traversal algorithm. Our treatment closely resembles [60].

The first step in generalising the DFS is an iterative formulation of the algorithm which rests upon the concept of a stack. A stack is a so-called *Last-In-First-Out* (LIFO) data structure and typically provides two functions Push and Pop. The former will put a new

element into the stack and the latter returns (and removes) the last element inserted via Push.

Algorithm 2 Iterative depth first search

```
function DFS(v)
    Push(v)
    while stack is non-empty do
        v ← Pop
        if v unvisited then
            mark v as visited
            for all (v, w) ∈ E do
                Push(w)
            end for
        end if
    end while
end function
```

It follows from the LIFO property of the stack that both, the iterative and recursive, variants describe exactly the same algorithm. The key observation that leads to a generalisation of Algorithm 2 is that the order of elements returned by the data structure, accessed and manipulated via Push and Pop, is not relevant for the algorithm to be correct and linear time. The important property is that each vertex is exactly visited once. This, as will be shown below, ensures that the algorithm is both exhaustive and works in linear time. The LIFO nature of the stack used in the DFS merely facilitates the depth-first heuristic. Instead one could use a *First-In-First-Out* (FIFO) data structure, that is push still inserts an element but Pop returns and removes the first element inserted. Such data structures are naturally called *queues* [61], where Pop and Push are usually termed Dequeue and Enqueue, respectively. Using a queue instead of a stack transforms the DFS into a breadth-first search or BFS. As the name suggests the BFS explores first all neighbours of a given vertex and then continues with the next-nearest neighbours and so forth. This follows naturally from the FIFO property, which will return vertices added to the data structure chronologically. Hence there is no possibility to follow one branch without extending all other branches too. In Figure 3.1 we show an example which compares the depth- and breadth-first search tree for the grid graph, a subset of the infinite square lattice $\mathbb{Z}^2$.

More generally now assume instead of a stack or queue we allow for a generic data structure, which we call *bag* [60]. The bag provides two functions Get and Give, the pendants to the stack methods Pop and Push, respectively[2]. Beyond that we make no further assumption, that is the order in how elements from the bag are returned can be arbitrary and even random. Now, equipped with even such a universal bag we show below that Algorithm 3 is visits all vertices in time linear in the number of edges and vertices.

[2] In a strict sense one would need another function IsEmpty to check whether the bag is empty. However we assume that the function GET is implemented such that it returns a unique element, say NULL, whenever the bag is empty.

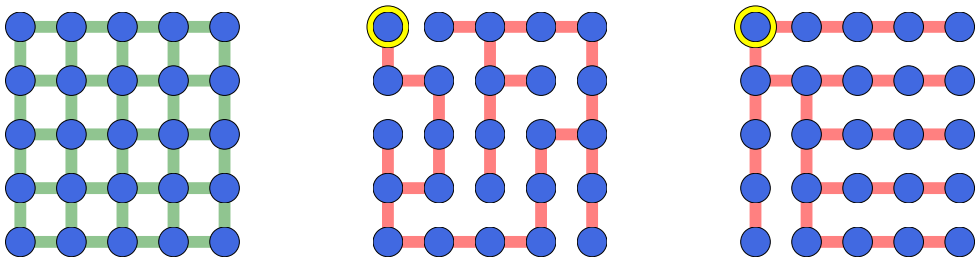


Figure 3.1: *The 5×5 grid graph in the left panel and one of its depth-first-search trees rooted at the highlighted vertex is shown in the central panel. The rightmost panel shows one breadth-first-search tree rooted at the same highlighted vertex. The qualitative difference between both trees is also evident, i.e. the DFS-tree is usually deep and has a small branching structure whereas the BFS tree is often characterised by a wide branching structure. Moreover note that for a connected graph both the depth- and breadth-first-tree are spanning-trees.*

Algorithm 3 Generic traversal algorithm

```
function TRAVERSAL(v)
    Give(v)
    while bag is non-empty do
        v ← Get
        if v unvisited then
            mark v as visited
            for all (v, w) ∈ E do
                Give(w)
            end for
        end if
    end while
end function
```

Lemma 3.2.1. *Let G be a finite connected graph. The generic traversal algorithm defined in Algorithm 3 visits every vertex of G exactly once.*

Proof. It is clear that TRAVERSAL visits every vertex at most once. To show that every vertex is visited at least once suppose vertex $x \in V$ has not been visited after TRAVERSAL terminates. Furthermore let $v \in V$ be the vertex passed to TRAVERSAL. Note that because v is the very first vertex put into the bag, and a priori not visited, it will always be returned by the bag directly afterwards and hence marked as visited. Now, from the definition of TRAVERSAL it follows that an unvisited vertex must be surrounded by unvisited neighbours. This is because any visited neighbour w must have iterated over the edge (w, x) connecting it to x. In this case x would have been inserted into the bag. However the algorithm can not terminate before all elements have been removed from the bag. This in particular also

applies to x which then would be marked visited. This contradicts the initial assumption. Thus any neighbour of x must be unvisited too and we can inductively iterate this to vertex v, with which TRAVERSAL started (this is always possible because G is connected and hence there exists at least one path between any two vertices in G). This contradicts the assumption. Therefore any vertex in G has been visited at least once. ■

With the knowledge that every vertex must be visited exactly once we can conclude that the inner loop in Algorithm 3 is executed exactly $|V|$ times. The number of iterations in the inner loop for vertex v is exactly $\deg(v)$, the degree of vertex v. If we use an adjacency list to represent the graph we can assume that every inner-loop iteration incurs a constant computational effort. Note that for any undirected graph one has $2|E| = \sum_{v \in V} \deg(v)$. If moreover both GIVE and GET functions of the bag data structure also work in constant time it follows that the running time of TRAVERSAL is $O(|E|)$, because for any connected graph one has $|E| \geq |V| - 1$.

Corollary 3.2.2. *The generic traversal algorithm 3 as well as the depth- and breadth-first search return an answer to a connectivity query in $O(|E|)$ worst case running time.*

3.3 Amortised computational complexity

In the Sweeny MCMC setting based on the Sweeny dynamics for the random-cluster model one performs a long sequence of operations such as insertions and deletions of edges interleaved with connectivity queries. Having a worst-case bound for each of the elementary operations is of course always desirable but in practise a bound on the running time of the overall sequence is often as useful. This idea is formalised in the theory of *amortised computational complexity* [62] and has been very powerful in the analysis of several important algorithms [63, 64]. In contrast to the worst-case analysis, the amortised analysis considers the running time of entire sequences of operations. In particular, when the worst-case scenario is sufficiently rare, the amortised view yields naturally improved bounds. Additionally, focusing on bounds for entire sequences of operations often facilitates the design of efficient algorithms. This is because one has the freedom to use computationally expensive parts in the sequence to reduce the cost of future operations and spread (amortise) the cost of expensive operations in a sequence onto a priori cheaper ones. The previously introduced depth- or breadth-first search based connectivity algorithm does not exploit this idea; any connectivity query implies a traversal of part of the graph but beyond the search for an answer to the current connectivity query no optimisation for future operations is done. To illustrate this consider the following example.
Suppose the sequence of operations consists of a construction phase where two components, each of size $O(n)$ are constructed. Assume furthermore the cost of this phase is $O(n)$. Then in a sequence of length $O(n^2)$ alternately an edge e connecting both components is inserted and deleted. Any time the edge e is deleted a traversal of part of the graph is necessary, which imposes a cost of $O(n)$ for any such deletion. Thus the sequence of all

operations takes $O(n^3)$ computational effort.
In the *aggregate analysis* [62], a particular form of the amortised analysis, one distributes the overall cost $T(\ell)$ of a sequence of length ℓ equally on all operations, which translates into an amortised bound $T(\ell)/\ell$ for any operation in the sequence. Applied to previous example this yields an amortised cost of $O(n)$ for any operation in the sequence. This in particular coincides with the worst-case bound of the depth first search, established above. Both, the union-find data structure and the poly-logarithmic fully dynamic connectivity algorithm described below exploit in one or the other form the idea of amortisation and the corresponding running time bounds are amortised.

3.4 The union-find data structure

Any incremental dynamic connectivity algorithm can be cast into a problem of maintaining disjoint sets under set unions. For this problem the union-find data structure supports all necessary operations. There exists various flavours of the union-find data structure with different worst-case or amortised worst-case bounds. It turns out that one particular version, described below, supports all operations in "almost" constant time. This is a very strong result, especially having in mind that in our typical applications the size of components in a graph can scale as a power of the total number of vertices.
The main observation required to view the incremental connectivity problem as a disjoint set problem is that connectivity in a graph $G = (V, E)$ naturally induces an equivalence relation on the set of vertices V. The induced equivalence classes, corresponding to connected components in G, partitions V. Moreover the insertion of edges and vertices can also be formulated in a set-theoretic way. Firstly, inserting a new vertex v into the graph introduces a new singleton equivalence class. Secondly the insertion of an edge $e = (x, y)$ with $x, y \in V$ can merge two equivalence classes, which happens precisely when x and y are not connected in the graph without e, in other words x and y are members of two disjoint equivalence classes $\mathcal{C}_x$ and $\mathcal{C}_y$. In this case e becomes a bridge in the modified graph $(V, E + e)$. All previously mentioned set operations are supported by the union-find data structure[3]. In what follows we use the set-theoretic notation to describe the data structure and call the elements of any set *objects* and denote by n the number of objects. Furthermore for an object x we denote by A_x the set containing it.
The union-find data structure chooses for any set A an unique representative object $\rho(A) \in A$, which is used to identify the set A. Now the creation of a singleton set is supported by `Makeset`(x), which hence creates the new set $\{x\}$ together with its obvious representative object x, i.e. $\rho(\{x\}) = x$. For any $x \in A$ the function `Find`(x) returns $\rho(A_x)$, the unique representative object of A_x. Finally `Union`(x, y) replaces the two disjoint sets A_x, A_y by their union $A_x \cup A_y$ and chooses a new representative object $\rho(A_x \cup A_y)$.
A natural way to represent a set A is to use a rooted tree $\mathcal{T}_A$, where each vertex in $\mathcal{T}_A$

[3] We speak about *the* union-find data structure as an abstract algorithm independent of the actual implementation flavour.

corresponds to a unique object in A and in particular $\rho(A)$ is represented by the root of $\mathcal{T}_A$. Hence the collection of disjoint sets corresponds to a rooted forest. Based on this representation it is straightforward to implement the above functions. `Makeset`(x) simply creates a tree consisting of a single vertex, `Find`(x) traverses the tree up the root and returns it. We remark that on a technical level the tree can be stored as a pointer-based data structure, where we store for any object x a pointer to the direct ancestor of x in the $\mathcal{T}_{A_x}$. `Union`(x,y) redirects the parent pointer of the root of $\mathcal{T}_{A_x}$ to the root of $\mathcal{T}_{A_y}$. This choice of merging the rooted trees of $\mathcal{T}_{A_x}$ and $\mathcal{T}_{A_y}$ is by no means the only one. It turns out that this particular choice has both amortised and worst-case cost of magnitude n for `Find`. This is because the running time of `Find` clearly depends on the maximal depth of the representing tree and it is not hard to construct a sequence yielding a chain like tree representation. A straightforward improvement considers the depth of both trees and makes the shallower one the child of the deeper one. To do so we need to augment the data structure by a depth variable associated to any vertex. Here the depth of a vertex x is the maximal graph theoretic distance to any leaf in the sub-tree rooted at x. Now given this depth information `Union` always redirects the shallower tree to the deeper one by comparing the depth of the roots of both trees. Intuitively this keeps the trees balanced and in the following Lemma 3.4.1 we show that this *union by depth* heuristic reduces the worst-case time of `Union` and `Find` to $\Theta(\log n)$:

Lemma 3.4.1. *[60]. The union-find data structure in a pointer based tree representation with union by depth heuristic runs `Find` and `Union` in $\Theta(\log n)$ worst-case. `Makeset` runs in $\Theta(1)$ worst-case.*

Proof. The running time of `Union` is dominated by `Find`, because `Union`(x,y) needs to determine the representative objects of the sets corresponding to x and y. Once these are known `Union` uses $O(1)$ to redirect one pointer and possibly updates the depth of the root in the former larger tree. We will show that for a representative object $\rho(A)$ whose vertex has depth d in $\mathcal{T}_A$ the number of objects in A is at least 2^d. On the other hand a given set can at most contain n objects, as this is the maximal number of vertices. This directly implies that the maximal depth is $\Theta(\log(n))$. It remains to show the relationship between the depth of $\rho(A)$'s vertex and the size of A. We prove this by induction on the depth d. The base case $d = 0$ is obviously true, because only trees corresponding to singleton sets have a root with depth 0. Now the depth of the vertex corresponding to $\rho(A)$, the root of $\mathcal{T}_A$, can only increase, by one, when it merges with another tree of same depth. Suppose the depth of both trees is $d-1$. By the inductive hypothesis we have that both trees contain each at least 2^{d-1} objects. Thus the new tree contains at least $2^{d-1} + 2^{d-1} = 2^d$ vertices. Successive unions with trees of smaller depth do only increase the size, which does not violate the statement. ∎

This logarithmic time bound is not the best possible. A significant improvement is based on another observation. To determine the representative element of the set A_x, containing

object x, `Find`(x) traverses, starting from the vertex of x, the tree $\mathcal{T}_{A_x}$ up until it reaches its root. It is obvious that at the time `Find`(x) reaches the root of $\mathcal{T}_{A_x}$, i.e. determines $\rho(A_x)$, one actually determines simultaneously the answer to `Find`(y) for any object y whose corresponding vertex is an ancestor of x's vertex in $\mathcal{T}_{A_x} = \mathcal{T}_{A_y}$. It causes a constant additional overhead per step to redirect the parent pointer of any such y directly to $\rho(A_x)$. This heuristic is called *path-compression*, due to the contraction or compression of possibly long tree paths into paths of length 1. Note when combined with the union by depth heuristic the depth variable looses its meaning. However the relationship shown in Lemma 3.4.1 is still true when the same rule to update the depth variable is used, i.e. only when two trees with same depth are linked does the depth of one of the two increase by one. To avoid confusion the name depth is usually replaced by *rank* [65] and the union by depth heuristics becomes the union by rank heuristic. The path-compression heuristic, completely in the spirit of amortisation, uses the computational effort in `Find` operations to speed up future `Find` operations. Intuitively, one expects that this heuristic improves the amortised cost of `Find` and hence also `Union` operations. Indeed, it can be shown that the combination of the path-compression and union by rank heuristics yields the optimal amortised union-find data structure:

Theorem 3.4.2. *[65]. For any $k \in \mathbb{N}$ the running time of m `Union` and `Find` operations in a universe of n objects with path-compression and union by rank heuristic is $O(km + 2nJ_k(\lceil \log_2 n \rceil))$.*

The integer functions J_k are defined as $J_0(r) = \lceil (r-1)/2 \rceil$ and recursively for $k > 0$

$$J_k(r) = \begin{cases} J_{k-1}(r) & \text{if } J_{k-1}(r) \leq 1, \\ 1 + J_k(\lceil \log_2(J_{k-1}(r)) \rceil) & \text{if } J_{k-1}(r) > 1. \end{cases} \tag{3.4.1}$$

We emphasise that Theorem 3.4.2 is valid for any $k \in \mathbb{N}$ hence there are infinitely many choices to bound the running time. The functions $J_k(r)$ for fixed $k \geq 1$ are extremely slowly increasing in r (even for small k). For example it is easy to verify that for $r < 100$ both $J_1(r)$ and $J_2(r)$ do not exceed 2. Now this allows to cover values of $n < 2^{100} \approx 10^{30}$ in Theorem 3.4.2. To eliminate the dependence on k an optimal balance between the term km, obviously increasing with k, and $2nJ_k(\lceil \log_2 n \rceil))$, which can be easily shown to decrease with k, is found by the choice $k = \alpha_s(m, n)$, where

$$\alpha_S(m, n) \equiv \min\{k \in \mathbb{N} | J_k(\lceil \log_2 n \rceil)) \leq 1 + m/n\}. \tag{3.4.2}$$

In this particular choice we obtain a $O(n + m\alpha_s(m, n))$ bound in Theorem 3.4.2. Note that the linear term in n can be attributed to n `Makeset` invocations, necessary to create the data structure. It follows that in the aggregate analysis `Makeset`, `Find` and `Union` have amortised $O(\alpha_s(m, n))$ running time. We stated above that the union-find data structure with path-compression and union by rank heuristics is optimal. It was indeed shown by

Tarjan, for a certain pointer-based model of computation, in [64] that the upper bound above is tight.[4].

The union-find data structure has applications beyond the incremental connectivity algorithm. In fact it is possibly to use the union-find data structure efficiently in a decremental connectivity algorithm for planar graphs as shown recently in [58]. In the next subsection we mention another, novel, application of the union-find data structure to the analysis of the loop structure in the medial graph of a planar graph. We will utilise this algorithm for the study of fragmentation of Fortuin-Kasteleyn clusters in Chapter 5, where finite-size effects can be understood by studying a classification of edges based on the medial graph construction.

3.4.1 An efficient loop configuration algorithm

In this section we define the loop configuration associated to a graph and then provide an efficient data structure that allows us to check if two loop segments, the fundamental building blocks in the loop configuration, belong to the same loop. For most applications in the theory of critical phenomena the typical length of loops diverges with the linear dimension of the system [15]. Hence the analysis based on a naive traversal becomes inefficient. The data structure we provide benefits from the quasi constant running-time of the union-find data structure, described before, and solves the computational problem efficiently. We remark that this algorithm only works efficiently in an incremental or static setting. We will use this algorithm for our analysis of fragility properties of Fortuin-Kasteleyn cluster in section 5.

To start with, fix a planar[5] graph $G = (V, E)$ together with a planar embedding. Now the medial graph $\mathcal{M}(G) = (\mathcal{V}, \mathcal{E})$ of G is constructed as follows. For any edge $e \in E$ there exists a vertex $\hat{e} \in \mathcal{V}$. For any face in the embedding of G we traverse the edges surrounding it in an arbitrary consecutive order, and any consecutive pair of edges $e_1, e_2 \in E$ corresponds to an edge $(\hat{e}_1, \hat{e}_2) \in \mathcal{E}$. Note that it is crucial to include the outer, infinite face, too. Moreover, edges in E that touch only one face correspond to loops in $\mathcal{M}(G)$. Furthermore $\mathcal{M}(G)$ can contain multiple edges albeit G is a simple graph. Last but not least it is not hard to see that any vertex in $\mathcal{V}$ has degree 4 (counting loops twice). In the left panel of Figure 3.2 we provide an example of the construction. Note that the medial graph of $\mathbb{Z}^2$ is itself $\mathbb{Z}^2$ tilted by $\pi/4$. Now instead of providing a formal description of the construction of the loop-configuration $\mathcal{L}(G)$ of G, we revert to Figure 3.2, which shows how $\mathcal{L}(G)$ can be obtained from $\mathcal{M}(G)$. In what follows we describe the construction of the data structure in the case of the square lattice (or grid graph) with the canonical planar embedding. The

[4]Tarjan uses the inverse Ackermann function $\alpha_T(m, n)$ instead of $\alpha_S(m, n)$. However, as mentioned in [65], their difference is minor and one can show that they asymptotically equivalent.

[5]A geometric dual as well as the associated medial graph can also be constructed in the more general setting of a cellularly embedded graph [66] such as $\mathbb{Z}^2_L$, the square lattice with periodic boundary conditions embedded on the torus. However, for the sake of clarity we present the concept in the setting of planar graphs.

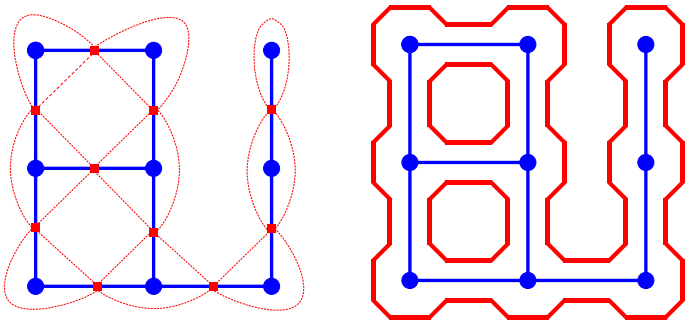


Figure 3.2: *The left panel shows a spanning subgraph of the 3×3 grid graph (in blue) together with its medial graph. The right panel shows the corresponding loop configuration (in red).*

data structure assigns four auxiliary vertices, called *strokes*, to any vertex in G. This is illustrated in Figure 3.3. The strokes for each vertex are denoted by $1, 2, 3, 4$. Now consider a fixed stroke such as for example stroke 2 of vertex x. Depending on the status of edge (x, y), stroke 2 of x either connects to stroke 1 of vertex y or two stroke 3 of vertex x. These two options are mutually exclusive; consider the left and right panel of Figure 3.3. Similar constructions apply for stroke 2 of vertex x when considering the edge leading upwards from x. An analogous construction can obviously be done for any stroke in the graph. The important observation is that this operation is completely local. Changing the viewpoint from strokes to edges, we note that the occupation of a given edge in G influences not only one but actually four strokes. Thus alternatively one can iterate over all edges of G and wire the strokes involved correspondingly. An example for an horizontal edge, denoted by (x, y), is shown in Figure 3.3. Now it is not hard to see that this wiring of strokes precisely

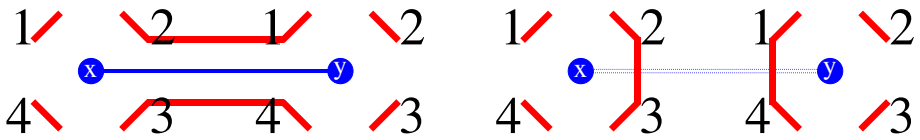


Figure 3.3: *Illustration of the wiring of strokes in the construction of the loop configuration data structure.*

recovers the loop configuration $\mathcal{L}(G)$ and in particular allows us to check if two loop arcs, corresponding to a given edge in G, belong to the same loop in $\mathcal{L}(G)$. For instance, to check whether the two loop segments associated to edge (x, y) belong to the same loop in $\mathcal{L}(G)$ it suffices to check if one of either the strokes 1 of y or 2 of x is connected to at least one stroke of either 3 of x or 4 of y. All the operations described above, the wiring of strokes, and the determination of the loop a given stroke belongs to fit into the framework of incremental dynamic connectivity algorithms. Indeed, the involved operations are precisely the ones supported by an union-find data structure. This in particular allows us to perform all required operations with quasi constant (in the spirit of Theorem 3.4.2) running time and in particular avoids the traversal of possibly long loops in $\mathcal{L}(G)$.

An immediate application of our loop configuration algorithm is the identification of bridges and non-bridges in planar graphs. Recall, any non-bridge must reside in at least one cycle and in any planar embedding cycles enclose faces, which in turns implies that the two loop-segments of a non-bridge must be in two different loops [55]. Hence to determine if edge $e = (x, y) \in E$ is a non-bridge it suffices to check whether the two associated loop segments are in two different loops in $\mathcal{L}(G)$, which can be performed efficiently with the above union-find based data structure. A complication occurs in the case G is not planar, such as for instance for $\mathbb{Z}_L^2$, where this equivalence does not hold [55]. We have more to say about this observation and its impact on finite-size corrections for the density of bridges, at criticality, in Chapter 5. In order to avoid a case-by-case consideration for the identification of bridges in graphs, we review in section 3.6 a recent algorithm [54] that provides a universal method to detect all bridges in a graph in linear time.

Before we proceed with the next section we remark that the data structure used by Mark Sweeny in his paper [20] is also based on the medial graph construction. He represents the loop configuration in a tailored skip list data structure which allows him to perform all operations with worst case computational effort $\log(n)$. This therefore provides a fully dynamic version of our medial graph algorithm, however with a worse running time bound. For our applications, where we generated a large number of Fortuin-Kasteleyn configurations in advance which where then later analysed in a "static setting", we clearly benefit from the union-find approach.

3.5 Poly-logarithmic fully dynamic connectivity algorithm

The *poly-logarithmic deterministic fully dynamic connectivity algorithm* of [24, 57] follows the idea of utilising expensive operations to reduce the cost of some future operations and achieves the following amortised complexity:

Theorem 3.5.1. *(Theorem 3 in [24]). Given a graph G with m edges and n vertices, there exists a deterministic fully dynamic algorithm that answers connectivity queries in $O(\log(n)/\log(\log(n)))$ time worst-case, and uses $O(\log(n)^2)$ amortised time per insert or delete.*

This algorithm (and similar variants of it [67]) consist of the following ingredients, described in more detail below:

- Spanning forest: For each connected component in the graph maintain a tree. Insertions and deletions of edges are translated into amalgamations and splits of trees.
- Euler-tour representation: An efficient data structure that allows to perform the necessary tree operations with $O(\log(n))$ computational cost.

- Edge level hierarchy: Implements the idea of using expensive operations to cheapen similar future operations.

We now explain the details of the algorithm and prove Theorem 3.5.1. For a given graph $G = (V, E)$ a spanning forest is a subgraph $\mathcal{F} = (V, \mathcal{E})$ with $\mathcal{E} \subseteq E$ and the property that for any pair $x, y \in V$ which is connected in G, there exists exactly one path that connects x and y in $\mathcal{F}$. Hence $\mathcal{F}$ must be a forest, i.e. a component of G corresponds to a tree in $\mathcal{F}$. A spanning forest is in general clearly not unique, yet once fixed it introduces a natural partition of the edges in E into *tree* (edges in $\mathcal{E}$) and *non-tree* edges (not in $\mathcal{E}$), see also Figure 3.4.

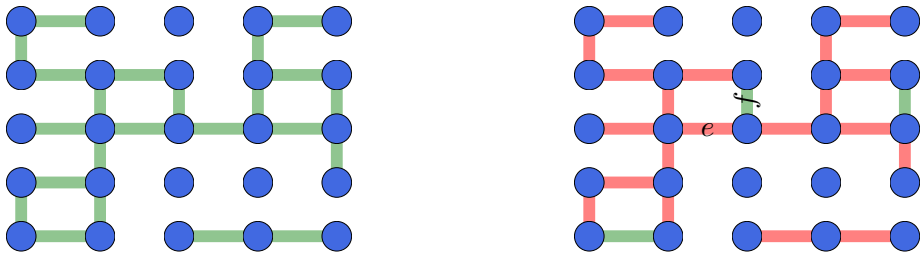


Figure 3.4: *A graph (left panel) together with one of its spanning forest (only red edges) shown in the right panel. Non-tree edges correspond to the the additional green edges in the right panel. The labeled non-tree edge f in the right panel is a replacement edge for the tree edge e.*

It follows directly from the definition of a spanning forest that connectivity queries in G and $\mathcal{F}$ are equivalent. Moreover the insertion of edges into G can also be translated straightforwardly into modification of $\mathcal{F}$ as follows. Suppose we want to extend G by one edge $e = (x, y)$. If x and y are not connected in G (equivalently in $\mathcal{F}$) we know that two components in G will merge. This must also be the case for the two trees in $\mathcal{F}$ corresponding to x and y. Thus we simply link the two trees with e. This preserves the spanning forest property, i.e. $(V, \mathcal{E} + e)$ is still a spanning forest of $(V, E + e)$. In case x and y are already connected before e is added to G we can not add e to $\mathcal{F}$ because this will violate the cycle-free property. However we still have that $\mathcal{F}$ is a spanning forest of $(V, E + e)$.

To support the deletions of edges with the desired computational efficiency, the data structure needs further modification. Here only the deletion of non-tree edges can be performed without further complication. Deleting a non-tree edge, that is $e \in E - \mathcal{E}$, does not demand any modification of $\mathcal{F}$ but still preserves the spanning forest property. The complication happens when a tree edge is deleted. Temporarily the tree $\mathcal{T}$ containing x and y in $\mathcal{F}$ splits into two trees $\mathcal{T}_x$ and $\mathcal{T}_y$. However it is not a priori clear if the component

containing x and y in G also splits, viz. there can be a non-tree edge $f = (u, v)$ with $u \in \mathcal{T}_x$ and $v \in \mathcal{T}_y$ that holds the component in G together. Thus the forest $\mathcal{F}$ must not split but rather needs to be replaced by $(V, \mathcal{E} \triangle \{e, f\})$. For an example see the right panel of Figure 3.4. Note that obviously any replacement edge in a undirected graph G must have its two incident vertices in T_x and T_y. This in particular allows us to only consider the smaller of the two trees. Thus when the trees are stored in data structures that support augmentation it suffices to scan the set of non-tree edges incident to the smaller tree.
Before we proceed with the discussion of the edge level hierarchy we briefly discuss the Euler-tour representation of a tree and how it can be used to devise a data structure that provides all required operations for the manipulations of the spanning forest in the poly-logarithmic connectivity algorithm.

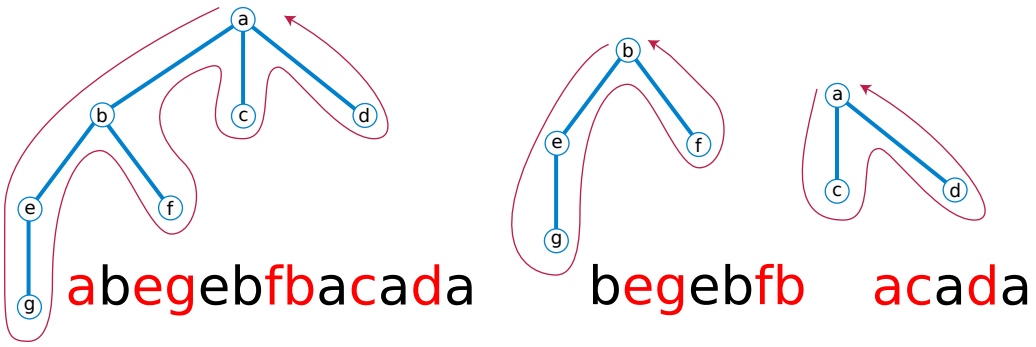


Figure 3.5: *The left panel shows a tree together with a possible Euler tour. The sequence of vertices corresponds to the visits of vertices in order dictated by the arrow. The red vertices are the representative vertices, one for each vertex. The center and right panel show the two trees, obtained from the left tree by the removal of edge (a, b), together with each one of their Euler tours.*

As the name suggests the Euler-tour data structure rests upon the concept of an Eulerian cycle. An Eulerian cycle is a trail in a graph that visits every edge exactly once and which starts and ends at the same vertex. A trail is a walk in which all edges are distinct and a walk in turn is a sequence of vertices and edges, where each edge's endpoints are the preceding and following vertices in the sequence. The following is a standard result [47] going partially back to Euler, which we state without proof:

Theorem 3.5.2. *(Theorem 6.1 in [47]). An undirected (multi-)graph G has an Eulerian cycle if and only if it is connected and all vertices have even degree.*

In order to represent each tree of a spanning forest by an Eulerian cycle we need, by Theorem 3.5.2, that all vertices in a given tree have even degree. Clearly this can not be true in general. However we modify the tree such that we replace any edge by two copies of it. This casts the tree to a multi-graph and ensures the existence of an Eulerian cycle. The left panel in Figure 3.5 illustrates this and shows one way a tree can be stored as a

ordered sequence of vertices. The crucial point is that the amalgamation of two trees in the forest or the split of one tree into two can be done by at most 2 splits and 2 concatenations of the original Euler tours [24], as illustrated in Figure 3.5. In order to support the concatenation and split of Euler tours in worst-case $O(\log(n))$ the sequence of vertices can be represented by a balanced search tree [61]. In our studies we used a self-adjusting binary search tree [63] which performs all the desired operations in amortised $O(\log(n))$. Another important observation is that in order to store the incident non-tree edges for any tree in the spanning forest we designate one occurrence of any vertex as a so called *representative* vertex which stores an adjacency list of incident non-tree edges. Moreover we adjust the balanced search tree representation such that iterating from one non-tree edge to the next incurs $O(\log(n))$ computational effort [24]. Note that this improves over a traversal based search (DFS or BFS) for a replacement edge, which have $O(n)$ worst case computational effort to iterate from one non-tree edge to the next. We omit the technical details and refer the reader to [24].
To summarise, an augmented Euler-Tour representation of a spanning forest supports the following operations with $O(\log(n))$ amortised computational effort:

- Insertion of edges
- Deletion of non-tree edges.
- Connectivity queries.
- Augmentation by "additional" edges not composing the forest (non-tree edges).
- Iteration from one non-tree edge to the next.

Yet there is one obstacle with the data structure presented so far in case of the deletion of tree edges. As elaborated above, a tree edge deletion can lead to the loss of the spanning property of the spanning forest used to represent the graph, in other words it is possible that there is a (former) non-tree edge that acts as a replacement edge in G. Hence it is necessary to start a search for such a replacement edge. Before we show how to do this in a way that guarantees the above complexity, note that for applications where the underlying graph is a tree, the Euler-Tour representation clearly yields a fully dynamic (worst-case) $O(\log(n))$ connectivity algorithm. This in particular can be interesting for simulations of the spanning forest model, which was shown to undergo a geometric ("ferromagnetic") phase-transition in dimensions three and above [68], as oposed to the two-dimensional case which only shows a "zero-temperature" transition in the spanning-tree limit.

What remains to show is how the above representation can be extended to yield the desired amortised poly-logarithmic fully dynamic connectivity algorithm. It is clear that a potential bottleneck is the search for replacement edges. It is easy to construct a sequence of operations (see example above) that considers a non-tree edge too often to be able to allocate only $O(\log(n)^2)$ to the insertion and deletion of that non-tree edge. Recall that Theorem 3.5.1 states that the cost of insertions and deletions is amortised

$O(\log(n)^2)$ and ultimately any sequence of operations only consists of insertions, deletions and connectivity queries. So any cost incurred by considering a non-tree edge as a possible replacement edge for a previously deleted tree edge has to be carried by the above operations.

The key observation that leads to the desired result is that any non-tree edge that we consider as not suitable during the search for a replacement edge must have both incident vertices in the tree we currently scan for replacement edges. Now, because we can easily augment the Euler-tour data structure to store the number of vertices in it, we can always choose the smaller tree as the one to operate on. This implies that any time we consider a non-tree edge as not suitable, we can (at least) half the number of tree edges in the current spanning forest that could consider the current non-tree edge in a future deletion call as a possible replacement edge. Therefore, we would ideally need a method to hide the non-tree edge from all other tree edges that are not part of the smaller tree. If this hiding mechanism can be iterated further it is not hard to see that a given non-tree edge can be at most considered $\lceil \log(n) \rceil$ times as a replacement edge.

To accommodate a systematic replacement edge search based on this observation an edge-level hierarchy is introduced. More precisely let $\ell_{\max} = \lfloor \log(n) \rfloor$ and associate to any edge $e \in E$ a level $\ell(e) \in \{0, 1, \ldots, \ell_{\max}\}$. For each $0 \leq i \leq \ell_{\max}$ let $\mathcal{F}_i$ denote the sub-forest of $\mathcal{F}$ induced by edges of level at least i, hence

$$\mathcal{F}_{\ell_{\max}} \subseteq \cdots \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_0 = \mathcal{F}. \tag{3.5.1}$$

Now we assume the graph starts with no edges, that is the construction phase of the graph is included in the amortisation argument. The insertion of an edge $e = (x, y)$ always happens at level 0. If x and y are already connected in level 0 then e becomes a non-tree edge, otherwise a tree edge. As described above this can be done with $O(\log(n))$ computational effort as well as the deletion of a non-tree edge.

When deleting a tree edge $e = (x, y)$ in level i we first cut all trees in levels less/equal i. Then we start the search for a replacement edge as follows. Suppose the tree in level i containing x and y is split by the deletion of e into $\mathcal{T}_x^{(i)}$ and $\mathcal{T}_y^{(i)}$ in level i. Let $T^{(i)}$ be the smaller of the two. Because of the hierarchy such a cut has to happen also in all levels below i. Now starting in level i we insert one copy of each tree edge in $T^{(i)}$ into level $i+1$ (if it exists). Once this is done we iterate over all non-tree edges incident to $T^{(i)}$. Let f be such a non-tree edge. If f is not suitable, i.e. does not reconnect $\mathcal{T}_x^{(i)}$ and $\mathcal{T}_y^{(i)}$, we remove it from level i and insert it into level $i+1$. In case f is a replacement edge we insert it as a tree edge in levels at most i, thus merging all $\mathcal{T}_x^{(j)}$, $\mathcal{T}_y^{(j)}$ with $j \leq i$, and terminate. However, if the search at level i does not yield a replacement edge we continue the same procedure, including the lifting of tree edges, at the next lower level $i-1$. Once level 0 is reached without finding a suitable replacement edge, the search terminates.

To summarise, it is precisely the idea of hiding non-trees from tree edges or vertices from which we know that the non-tree edges are not suitable for, what is achieved by the level-hierarchy and the upward drift of unsuitable non-tree edges. By first moving

the smaller tree one level up and then successively lifting non-suitable non-tree edges we reduce the circle of vertices/tree edges that will consider those non-tree edges as replacement candidates in later stages of the sequence of operations under consideration. This is precisely the idea of using computationally expensive parts to cheapen future operations, mentioned in the section on the amortised computational complexity.

Let us now conduct the amortised analysis of the above connectivity algorithm. Firstly, recall that the overall computational cost of a sequence of operations, consisting of insertions, deletions and connectivity queries, can be either attributed to the manipulation of Euler-Tour data structures or the search for replacement edges. Consider first the insertion of an edge, which either requests to link or to augment two Euler-Tour data structures, corresponding to the case of inserting a tree or non-tree edge, respectively. In both cases the direct cost is at most $O(\log(n))$, purely arising from the representation of the Euler-Tour data structure as a balanced binary search tree. Furthermore, any connectivity query has the same computational complexity dominated by the maximal depth of the balanced binary search tree representing a given Euler-Tour. It therefore remains to show that the computational cost of deletion operations can be amortised accordingly.
In the case of deletions it helps to distinguish the two cases of tree and non-tree edges. Firstly, suppose the edge to delete is a non-tree edge. In this case the immediate required computational cost is $O(\log(n))$, as two Euler-Tour data structures need to be modified. Here it is important to recall that a given non-tree edge resides only in two Euler-Tour data structures in its corresponding level, as opposed to tree edges, for which there are precisely $\ell + 1$ replicas, if the level of it is ℓ. To summarise the computational complexity so far, each of the previous operations "carried" a computational cost logarithmic in n, the number of vertices.
The most expensive case is clearly the deletion of tree edges, because of the involved search for replacement edges, the cuts of up to $\log(n)$ Euler-Tours and the lifting of tree and non-tree edges in the level hierarchy. However, we have the freedom to attribute part of the cost to insertions and connectivity queries. Firstly, the cut of Euler-Tours in at most $\ell_{\max} + 1$ levels incurs a computational cost of order $\log(n)^2$. We attribute this cost directly to the deletion of that particular edge. Secondly, in the sub-sequent methodical search for replacement edges in the level hierarchy, we first increase the level of all tree edges of the smaller of the two trees formely connected by the deleted tree edge. This means that we need to insert all such tree edges to the forest in the next higher level. On a first view this seems to be a bottleneck. However, one can easily verify that the level of an edge can never decrease, hence the involved cost of lifting tree edges in the level hierarchy is $O(\log(n)^2)$ per tree edge. We attribute this cost to insertions, which then casts the amortised cost of a single insertion to $O(\log(n)^2)$. Considering non-tree edges, we have, as already mentioned before, that a given non-tree edge can only be touched $O(\log(n))$ times, after which it resides in level $\ell_{\max}$. Each time it is touched, meaning it is considered as a possible replacement edge, we have a cost of $O(\log(n))$ for removing it from the current tree and augmenting the corresponding tree in the next higher level.

Additionally, note that while iterating from one (non-)tree edge to the next (non-)tree edge we have to spent at most $O(\log(n))$ each time. We can easily add this to the cost for lifting a (non-)tree edge and still remain with the $O(\log(n)^2)$ cost for insertions as before. Lastly, note that is also possible for a non-tree edge to become a tree edge, precisely in case it was found to be a suitable replacement edge for a previously deleted tree edge. Once such a replacement edge is found, we link all the trees below and in the current level, this incurs a computational cost of $O(\log(n)^2)$. However, in the subsequent sequence of operations we clearly have that the previous arguments used for tree edges still describe an upper bound on the computational cost involved. We can therefore conclude with the claimed poly-logarithmic amortised running times for all operations. We refer the reader to [24] for more details. ■

The probability mass function of the random-cluster model $\pi_{p,q}$ can naturally be formulated in terms of the numbers of non-tree and tree edges. To this end, note that the number of edges in *any* spanning forest (equivalently the number of tree edges) of (V, A) is equal to $t(A) \equiv n - k(A)$, and hence the number of non-tree edges equals $c(A) = |A| - n + k(A)$, where $c(A)$ is the circuit rank of (V, A) [69]. We therefore have for $\pi_{p,q}$:

$$\pi_{p,q}(A) \propto v^{|A|} q^{k(A)} = v^{c(A)} \left(\frac{v}{q}\right)^{t(A)} \propto q^{c(A)} \left(\frac{v}{q}\right)^{|A|}.$$

For the particular case of the random-cluster model on the square lattice we have that $v = \sqrt{q}$ is the critical manifold [28] and we hence obtain

$$\pi_{p_{\mathrm{sd}}(q),q} \propto \sqrt{q}^{c(A)} \left(\frac{1}{\sqrt{q}}\right)^{t(A)}.$$

Thus in the two dimensional critical setting we see that the random-cluster model favours (disfavours) data structure layouts that augment the Euler-tour spanning forest by non-tree edges when $q \geq 1$ ($q < 1$).

3.6 Bridge and cut vertex identification

The depth-first search has found applications as part of various more complex algorithms, e.g. [70, 71]. One particular example is the identification of bridges and cut vertices in linear time as presented in [72]. A simplified algorithm to identify cut vertices and bridges, still based on the DFS, was recently presented in [54]. Before we present the algorithm [54], we need two more DFS related concepts. The first is the so-called depth-first search index (DFI), which encodes the order in which vertices were visited in the DFS, thus DFI(v) is intuitively the *time* when vertex v was visited in the DFS. Note that DFI can obviously be used to relabel the vertices in G. The second concept is the one termed *back-edge*. A back-edge is an edge in G but not in T, where T denotes DFS tree. It is the analogue of the non-tree edge in the poly-logarithmic dynamic connectivity algorithm

above. As already stated, it immediately follows that a connected graph G with n and m edges has precisely $m-n+1$ back-edges. For an illustration of the two concepts consider Figure 3.6.

Now, let G be a simple and connected graph with n vertices and m edges. The first step is to run an exhaustive depth first search and extract the corresponding DFS tree together with the DFI order. Let T denote the DFS tree and suppose that all tree edges are directed towards the root of T (denoted r). All remaining edges are now back-edges and are added to T as directed edges oriented away from r. See Figure 3.6 for an illustration, where $r = f_0$.

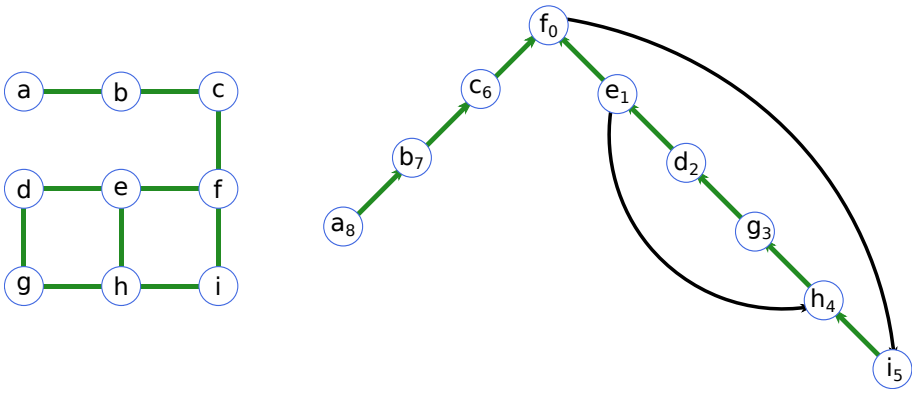


Figure 3.6: *A spanning sub-graph of the* 3×3 *grid (left panel) and one of its DFS-trees, together with all back-edges (right panel). The subscript in the DFS tree denotes the DFI.*

The next step is the construction of a *chain decomposition*[6] of the set of edges. A chain can be either a cycle or a path. The chain decomposition consists of the edges in all chains and is constructed as follows. To start with, all vertices are marked unvisited. Then vertices are visited in increasing DFI order. Let v be the current vertex. Mark v visited if not already. Any back-edge b incident to v is traversed and vertices in the directed cycle, b belongs to, are visited (and marked visited) until an already visited vertex is encountered. All edges traversed are now associated with the current chain. It is not hard to see that the complete chain decomposition can be constructed in linear time. Moreover if the first vertex (in DFI order) is incident to a back-edge, then the corresponding chain is always a cycle. This is because any vertex encountered during the traversal of the directed cycle is unvisited. For example the chain decomposition of the example depicted in Figure 3.6 consists of two chains:

$$\begin{aligned} C_1 &= \{(f,i),(i,h),(h,g),(g,d),(d,e),(e,f)\}, \\ C_2 &= \{(e,h)\}. \end{aligned}$$

[6]The term decomposition is misleading here. Strictly speaking, the collection of chains does only decompose (partition) the set of edges of G when G is 2-edge-connected, as shown in [54]. However when we consider the chain decomposition as a composition of the vertex set of the graph then it is a proper decomposition. This was in fact done in [54].

Here, C_1 is a cycle and C_2 a path. Note that it is not necessary that all edges are covered by the chain-decomposition, and the above construction is completely sufficient for the purpose of identifying bridges and cut vertices. Indeed, a bridge is an edge that is not part of any chain. This clearly can be exploited to detect bridges, simply by checking for a membership in any chain decomposition:

Lemma 3.6.1. *(Lemma 5 in [54].) Let C be a chain decomposition of a simple connected graph G. An edge e in G is a bridge if and only if e is not in any chain in C.*

Proof. If e is a bridge it can not be in any cycle [54], thus there can not be any chain e belongs to. When e is an edge that is not in any chain, then e must be an edge in T. Denote by x the vertex incident to e that is farther away from r and let $T(x)$ be the sub-tree rooted at x. Any back-edge with one vertex incident to $T(x)$ must also have the second incident to $T(x)$, because otherwise e would be in a chain. In other words back-edges incident to $T(x)$ must not be originated at a ancestor of x if e is not in any chain. It follows that when e is removed, all vertices in $T(x)$ are disconnected from r, which, by definition, implies that e is a bridge. ■

As we also study some aspects of vertex fragmentation in Chapter 5, we need a method to identify whether a given vertex is a cut vertex. A *cut vertex*, similarly to a bridge, is a vertex that increases the number of connected components upon deletion. A subtle point with that definition is that vertices with degree 1 are not covered by this definition. However, later we will loosen this restriction and refer to such vertices as *fragmenting* vertices. In the left panel of Figure 3.6 vertex a is a fragmenting but not a cut vertex. It follows that any vertex incident to a bridge and with degree at least 2, is a cut vertex. However, not all cut vertices correspond to bridges, an essential observation for the vertex fragmentation discussed in 5. Intuitively speaking, a vertex v that is not incident to a bridge and has degree at least 2, is a cut vertex when at least two cycles overlap at v and nowhere else. Now, the analogue to Lemma 3.6.1 reads (without proof):

Lemma 3.6.2. *(Lemma 6 in [54].) Let C be a chain decomposition of a simple connected graph G with minimal degree 2. A vertex v in G is a cut vertex if and only if v is either incident to a bridge or v is the first vertex of a cycle in $C - C_1$, where C_1 is the first chain in the decomposition.*

Chapter 4

Computational and statistical analysis of Sweeny's algorithm

This chapter is devoted to a study of the heat-bath and Meropolis Markov chain Monte Carlo algorithm for the random-cluster model, commonly known as Sweeny's algorithm [20]. In the following, we refer to the heat-bath and Metropolis algorithm in the random-cluster model setting simply as variants of Sweeny's algorithm. In the first part of this chapter we address the question of an efficient implementation of Sweeny's algorithm, a non-trivial problem due to the global nature of the required connectivity information. We utilise the algorithmic and graph-theoretic concepts discussed in Chapter 3, and adjust the standard approaches to the particular structure of the random-cluster model near criticality. In fact our computational analysis reveals a strong and intriguing connection between structural properties of the random-cluster model and the computational efficiency of Sweeny's algorithm. In the second part of this chapter we embark on a study of the statistical properties of Sweeny's algorithm, in particular its dynamical critical behaviour. In fact one of the distinct features of the Markov chain is the phenomenon of critical speeding-up. This phenomenon corresponds to the decorrelation of specific "global" observables, that is the corresponding integrated autocorrelation times have negative dynamical critical exponents. We present a corresponding numerical study in two dimensions, establishing new rigorous results for monotone Markov chains, which are of independent interest. We then continue with an analysis of a joint efficiency measure, that considers both the statistical and computational results established in this section. This will provide a meaningful way to compare Sweeny's approach to the alternative Chayes-Machta-Swendsen-Wang chain. Finally, we present two optimisations of the standard approaches tailored to the structure of random-cluster models at criticality.

4.1 General considerations and the difference between heat-bath and Metropolis variants of Sweeny's algorithm

The efficiency of the local-bond approach for the random-cluster model, in heat-bath or Metropolis variant, to which we refer in what follows simply as Sweeny's algorithm or Sweeny's chain [20], relies on the availability of efficient algorithmic solutions to determine the pivotality of edges. As shown in the previous chapter, this can be solved by a connectivity algorithm. One such connectivity algorithm is based on the breadth-first search, that has worst-case and amortised computational complexity, which is linear in the number of vertices and edges.
Yet, what we have not taken into account, so far, is the fact that the underlying cluster structure is actually random! This suggests a probabilistic analysis of running times [73] of the involved connectivity algorithms. In this section we will in particular provide numerical evidence and scaling arguments that show that an adapted breadth- or depth-first search based implementation has expected running time that is sub-linear in the volume, i.e. in the number of edges or vertices. To be specific, we consider in this section the stationary Sweeny chain on the square lattice with periodic boundary conditions $\mathbb{Z}_L^2$. The main reason for considering $\mathbb{Z}_L^2$ is the rich interplay between rigorous mathematics, mathematical physics and exact statistical mechanics [28, 27, 74, 22, 51] for two dimensional critical phenomena, which in particular yields exact values for various critical exponents [51] for the random-cluster model on $\mathbb{Z}_L^2$. Further, the exact solution for the Ising case ($q = 2$), due to Onsager, and various exact results for finite lattices [75] provide a solid testing ground for devising new computational approaches in the random-cluster model. Additionally, there are many open questions related to the random-cluster model even under such mild circumstances as planarity, (self-)duality and regularity, as given for $\mathbb{Z}^2$. In the subsequent chapters we investigate a few of such open geometric questions. We remark that the arguments established in this section are expected to extend naturally to other graphs such as $\mathbb{Z}_L^d$ with $d \geq 2$ with appropriate replacements, e.g., critical exponents.

4.1.1 Avoiding cluster-traversals

Before we investigate the interplay between the (random) cluster structure and the running time of various implementations of Sweeny's algorithm, we elaborate on an optimisation that, in many cases allows to avoid expensive cluster traversal with an asymptotically non-vanishing probability.
To start, consider the heat-bath version of Sweeny's algorithm and recall, (2.2.4), that in order to iterate the chain from configuration A_t to A_{t+1}, an edge e is chosen uniformly at random, and inserted into the next configuration A_{t+1} with probability $a(A_t; e)$ equal to $\tilde{p} \equiv p/(p + (1-p)q)$ when e is pivotal to A_t, and p otherwise. The main computational obstacle stems from the determination of e's pivotality to A_t. Assume $a(A_t; e)$ is determined, then one usually uses a uniform random number U in $[0, 1]$ and sets $A_{t+1} = A_t \cup \{e\}$ if $U \leq a(A_t)$, and $A_{t+1} = A_t \setminus \{e\}$ otherwise. Note that the status of e is reseted and not

'flipped" or reversed as in the Metropolis variant. Clearly, one has $\mathbb{P}[U \leq a(A_t)] = a(A_t)$, which together with the fact that $a(A_t)$ is two-valued, means e is open in A_{t+1}, whenever $U \leq \min\{\tilde{p}, p\}$, regardless of the pivotality of e. Similarly, whenever $U > \max\{\tilde{p}, p\}$, it is clear that $e \notin A_{t+1}$ is independent of A_t. To be specific, let us consider the case $q \geq 1$, where one has $\tilde{p} \equiv p/(p + (1-p)q) \leq p$, and it follows that, given edge e, only in a fraction $\mathbb{P}[\tilde{p} < U \leq p] = p - \tilde{p}$ of all cases, one needs to determine the pivotality of e.
We show in Figure 4.1 how the difference $p - \tilde{p}$ varies with p for some values of q. Trivially,

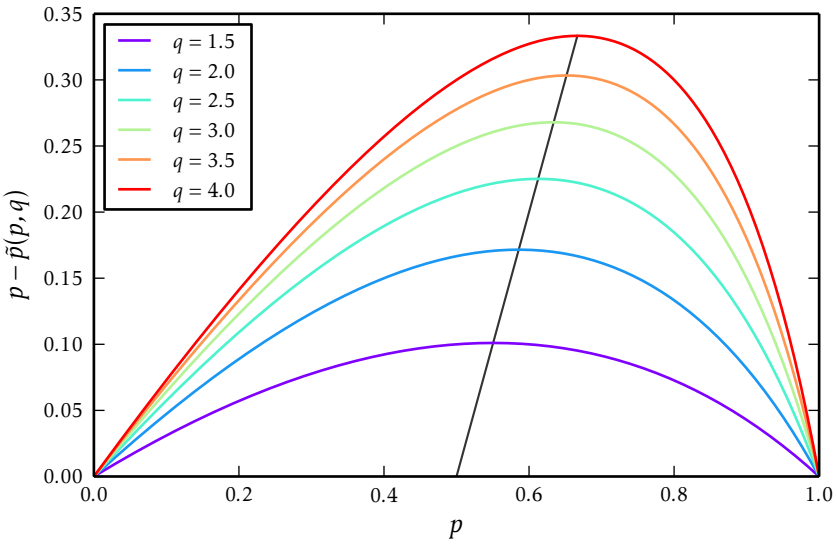


Figure 4.1: *The difference of the insertion probability of non-pivotal edges and pivotal edges, i.e. $p - \tilde{p}$. Or, equivalently, the probability of the necessity of analysing the cluster structure in the heat-bath dynamics. The solid black line corresponds to the line $(p_{\mathrm{sd}}(q), p_{\mathrm{sd}}(q) - \tilde{p}(p_{\mathrm{sd}}(q), q))$ and therefore shows the critical difference of p and $\tilde{p}$ as the cluster weight q is varied.*

for $q = 1$ one has $p = \tilde{p}$, hence, as expected, there is no dependence of $p(A_t)$ on the cluster structure, i.e. the heat-bath chain performs independent bond percolation. Figure 4.1 also shows that the maximum of $p - \tilde{p}$, for fixed q, is reached at the self-dual point $p_{\mathrm{sd}}(q)$. We stress that this result is independent of the particular graph. Further, it is known that $p_{\mathrm{sd}}(q)$ is precisely the critical point of the random-cluster model on $\mathbb{Z}_L^2$ [28] for $q \geq 1$. Thus in addition to the intricate cluster structure at the critical point, that as we show below, causes a computational critical slowing down, we have that the heat-bath chain is most inefficient in avoiding expensive cluster analysis whenever $p = p_{\mathrm{sd}}(q)$. We can precisely evaluate the difference along $p = p_{\mathrm{sd}}(q)$ and obtain

$$p_{\mathrm{sd}}(q) - \tilde{p}(p_{\mathrm{sd}}(q), q) = \frac{\sqrt{q} - 1}{1 + \sqrt{q}},$$

which is valid for $q \geq 1$, where the difference is further increasing with q, translating into an increasing probability of the need to determine whether e is pivotal. For the sake of completeness, one obtains $p_{\mathrm{sd}}(q) - \tilde{p}(p_{\mathrm{sd}}(q), q) = (1 - \sqrt{q})/(1 + \sqrt{q})$ for $q < 1$ which is

decreasing with q. We show the corresponding self-dual difference for $q \in [0,4]$ in Figure 4.2.

Alternatively, one can consider the Metropolis variant of Sweeny's algorithm, which in

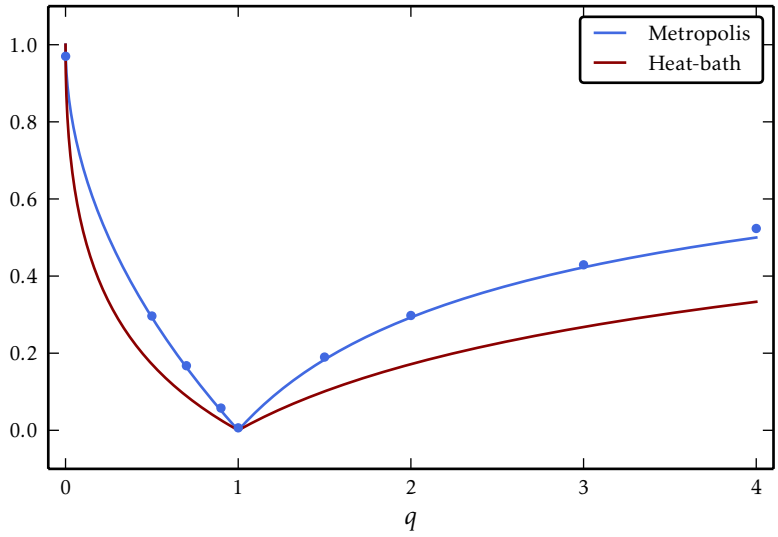


Figure 4.2: *Probability of a cluster structure dependent update step for the heat-bath and Metropolis variants of Sweeny's algorithm. The data (dots) corresponds to the ratio of running times of an interleaved breadth-first search Sweeny algorithm exploiting the unconditional acceptance, to the corresponding running time of the same variant of Sweeny's algorithm without this optimisation.*

the simplest form, chooses an edge $e \in E$ uniformly at random and proposes a "flip" of the status of e, i.e. $A_t \to A_{t+1} = A_t - e$ when $e \in A_t$, or $A_t \to A_{t+1} + e$ when $e \notin A_t$. The proposed flip is accepted with probability

$$\min\{v^{|A_{t+1}-A_t|} q^{k(A_{t+1})-k(A_t)}\},$$

where $v = p/(1-p)$. Considering $p = p_{\mathrm{sd}}(q)$, that is $v = \sqrt{q}$, we obtain that the Metropolis algorithm avoids the determination of Δk with probability $\min\{1/\sqrt{q}, \sqrt{q}\}$. Note that the Metropolis version of Sweeny's algorithm can only avoid determining Δk by an unconditional acceptance, opposed to the heat-bath version, which benefits from both unconditional rejections and acceptances. The corresponding q dependence is also shown in Figure 4.2. We conclude that at the self-dual point, the heat-bath chain is more efficient than the Metropolis chain, in avoiding cluster structure traversals. On the other hand, we remark that this is not the only efficiency measure. Another perspective arises from comparison inequalities between autocorrelation times of the heat-bath and Metropolis algorithm, which we evaluate here at the self-dual point.

4.1.2 Comparison of autocorrelation times between heat-bath and Metropolis

To start with, one can easily verify that the transition matrix $P_{\mathrm{HB}}(A,B)$ of the heat-bath chain and $P_{\mathrm{M}}(A,B)$ of the Metropolis algorithm obey the following inequalities for any $A \neq B$ whenever $p = p_{\mathrm{sd}}(q)$ and $q \geq 1$:

$$\frac{1}{1+\sqrt{q}} P_{\mathrm{M}}(A,B) \leq P_{\mathrm{HB}}(A,B) \leq P_{\mathrm{M}}(A,B).$$

This intuitively means that the Metropolis algorithm is more mobile than the heat-bath chain at the self-dual point, i.e. given a state A the Metropolis version of Sweeny's algorithm has higher probability to move to any $B \neq A$ than the heat-bath variant. To make this statement precise, note that in [76] (Theorems A.2-A.3) it was shown that for two Markov chains on the same state space Ω, described by transition matrices P, P', which are reversible with respect to the same stationary distribution, and for which $P(A,B) \geq \alpha P'(A,B)$ holds for $\alpha > 0$ and any $A, B \in \Omega$ such that $A \neq B$, one has, for any observable $f : \Omega \to \mathbb{R}$, $\tau_{\mathrm{int},f}(P) + 1/2 \leq \alpha^{-1}\left(\tau_{\mathrm{int},f}(P') + 1/2\right)$. Applied to the Metropolis and heat-bath chain for the random-cluster model we immediately obtain

$$\tau_{\mathrm{int},f}\left(P_{\mathrm{M}}\right) \leq \tau_{\mathrm{int},f}\left(P_{\mathrm{HB}}\right) \leq (1+\sqrt{q})\left(\tau_{\mathrm{int},f}\left(P_{\mathrm{M}}\right) + \frac{1}{2}\right). \tag{4.1.1}$$

We emphasise, that this is only valid for the choice $p = p_{\mathrm{sd}} = \sqrt{q}/(1+\sqrt{q})$ and $q \geq 1$, that is the critical point of the random-cluster model on $\mathbb{Z}^2$. Other values of p need separate consideration. However, because our main interest lies in dynamical critical behaviour, the above statements suffice for the subsequent treatment. The result (4.1.1) states that autocorrelation times for any observable f in the Metropolis or heat-bath variant of Sweeny's algorithm are of the same order of magnitude at criticality.

We can extend the comparison also to *the* exponential autocorrelation time, where we can for instance use a comparison technique [2] for reversible Markov chains, which we already utilised in section 2.3 to establish the duality result for the heat-bath chain on $\mathbb{Z}_L^2$. In general, under the same conditions on the two transition matrices P and P', one can indeed show that $\tau_{\mathrm{exp}}(P') \geq \alpha \tau_{\mathrm{exp}}(P) + 1$. This allows us to conclude that

$$\tau_{\mathrm{exp}}\left(P_{\mathrm{HB}}\right) + 1 \geq \tau_{\mathrm{exp}}\left(P_{\mathrm{M}}\right) \geq \frac{1}{1+\sqrt{q}} \tau_{\mathrm{exp}}\left(P_{\mathrm{HB}}\right) - 1.$$

Which is the analogue statement to (4.1.1). Thus we have shown that, both the heat-bath and the Metropolis version of Sweeny's algorithm, belong to the same dynamic universality class [3], i.e. both share the same dynamical critical exponent.

4.2 Computational efficiency of Sweeny's algorithm

Let us now return to the actual analysis of running times of the variety of computational approaches outlined in chapter 3. In what follows we only consider the critical point. The reason for this is two-fold. Firstly, our target is the design of an efficient MCMC method to investigate critical phenomena in the random-cluster model, and ultimately show that it is possible to "beat" the alternative Swendsen-Wang-Chayes-Machta approach under certain circumstances. Secondly, the fractal structure of the critical random-cluster model on $\mathbb{Z}_L^2$ casts the study of the self-dual point into the most complicated and interesting scenario, in particular with regard to the interplay between algorithmic properties and the geometric structure. On the other hand, for off-critical values of p, we expect no real computational complication. As for sub-critical values, one has exponential decay of correlations [27] (finite correlation length) and hence at most logarithmically large components. In the super-critical regime, macroscopically large components exists, consisting of many (overlapping) cycles, such that the breadth-first search algorithm typically extends only across a few shells until it finds an alternative path.

4.2.1 Sequential breadth-first search

We first consider a plain breadth-first search implementation of Sweeny's algorithm, in what follows also called *sequential breadth-first search* approach (SBFS). This variant is the conceptually simplest approach described in this thesis. This is because, in addition to the actual breadth-first search traversal related queue data structure, one only needs to maintain a data structure that allows a check whether a given edge is active. Due to the sparseness of $\mathbb{Z}_L^2$ (or in general $\mathbb{Z}_L^d$) we have chosen the adjacency-list representation, which achieves the basic manipulation operations of insertion and deletion of edges with constant computational effort. Now, in order to determine whether an (randomly chosen) edge $e = (x, y)$ is pivotal to the current configuration A_t, we remove e from A_t temporarily (if $e \in A_t$) and perform a BFS starting at x. As shown in the previous chapter, whenever y is encountered during the traversal, we can conclude that e is not pivotal to A_t. Now, without further assumptions, we need to conclude that this algorithm has worst-case running time per operation that scales linear with the volume L^d.

However, in a probabilistic setting, given here for the random-cluster model, it turns out that a worst-case running time bound is often too pessimistic and crude. For instance, it is well known that at criticality in a large but finite box with linear dimension L, the expected size of the component a randomly chosen edge belongs to scales with $L^{\gamma/\nu}$ [15, 46].

The following arguments show that the expected running time of the sequential breadth-first search approach to Sweeny's algorithm has the same asymptotic $L^{\gamma/\nu}$ scaling as the typical cluster size.

Consider first the insertion of an edge $e = (x, y)$, which additionally we assume to be closed in the current configuration A_t, in other words $e \notin A_t$. The imposed breadth-first search traversal, that determines whether e is pivotal to A_t, chooses one site, say x, to

start the BFS. Crucially, as we will show in section 5.2, one has for the critical model on $\mathbb{Z}_L^2$, asymptotically, a fraction $\sqrt{q}/[2(1+\sqrt{q})]$ of disconnected nearest-neighbour pairs (or candidate bridges). Clearly, for such instances the BFS traversal terminates without finding a path between x and y. This in turn, combined with a constant fraction of such candidate bridges, implies a constant probability of a BFS traversal of asymptotic cost proportional to $L^{\gamma/\nu}$. Hence the insertion of candidate-bridges contributes $L^{\gamma/\nu}$ to the overall expected running time of the SBFS implementation of Sweeny's algorithm.
The same asymptotic scaling holds for the removal of bridges. Firstly, note that one has an asymptotically constant fraction $1/[2(1+\sqrt{q})]$ of bridges. Secondly, we will show in Chapter 5, a bridge removal typically creates two very unbalanced clusters (fragments), of which the larger is typically of size $L^{\gamma/\nu}$ and the smaller of size $L^{d_F - x_2} \ll L^{\gamma/\nu}$, see also [15]. Because, in this version of the SBFS it is a priori not clear which of the two vertices belongs to the smaller cluster, we have to conclude that the deletion of bridges also contributes $L^{\gamma/\nu}$ to the expected running time. We confirmed the $L^{\gamma/\nu}$ scaling for operations (insertions and deletions) on pivotal edges by numerically estimating the expected number of vertices visited by the breadth-first search. Furthermore, we fitted a power-law (in L) with exponent $y_{\mathrm{SBFS}}^{(e)}$ to the data using the method of least squares fitting. We summarise the extracted values in table 4.1. The numerical results are statistically consistent with $y_{\mathrm{SBFS}}^{(e)} = \gamma/\nu$, except for $q = 3, 4$, where the deviation becomes significant, however we attribute this to unconsidered corrections and the non-asymptotic nature of our numerical analysis.

It remains to analyse operations on non-pivotal edges, i.e. the removal of non-bridges and insertion of candidate-non-bridges. We first note that, as for the bridges and candidate-bridges, the corresponding densities of non-bridges and candidate-non-bridges are $\sqrt{q}/[2(1+\sqrt{q})]$ and $1/[2(1+\sqrt{q})]$, respectively. In order to obtain the leading size dependence of the running times for operations on non-pivotal edges, we need to consider how the breadth-first search iterates. As discussed in Chapter 3, the breadth-first search explores (incident) vertices sequentially corresponding to their graph-distance. For instance, starting at a vertex x, at first all vertices that are incident to x, are considered. Once the current, so called BFS "shell", is exhausted, the search continues with neighbours of neighbours of x, until the cluster is exhausted. It is not hard to see, that the path between any two vertices in the corresponding BFS tree is always a shortest path. Applied to the (attempted) insertion of an edge $e = (x, y)$ into a configuration A_t, where x and y are already connected, this means that the BFS, starting either at x or y, extends over $s(x, y; A_t)$ BFS shells, where $s(x, y; A_t)$ is the shortest path distance between the two vertices x and y in A_t. As shown by Grassberger [77, 78], for bond percolation[1], the probability that two nearby points on the lattice are connected by a shortest path of length ℓ, has large ℓ

[1] We expect the results to naturally extend to the random-cluster model with appropriate replacements in the exponents.

asymptotics $p(\ell) \approx l^{-\psi_\ell}\mathcal{L}(\ell/L^{d_{\min}})$, where

$$\psi_\ell = 1 + \frac{2\beta}{\nu d_{\min}} + \frac{g_1}{d_{\min}}.$$

Here, $d_{\min}$ is the shortest-path fractal dimension [79] and g_1 is the scaling exponent related to the density of growth sites [78]. Ziff [80] demonstrated that $g_1 = x_2 - 2\beta/\nu$, where x_2 is the two-arm scaling exponent [81, 82]. Hence $\psi_\ell = 1 + x_2/d_{\min}$. As a result, the average length of the shortest path between nearby points exhibits system-size scaling according to

$$\langle \ell \rangle \approx L^{d_{\min}-x_2}. \tag{4.2.1}$$

Now, in order to obtain the leading scaling of the number of vertices visited by a breadth-first search on critical percolation clusters, we need a relationship between the number of BFS-shells and the number of vertices within it. The number of sites touched by a BFS from i to j separated by a shortest path of length ℓ is expected to be $\ell^{\hat{d}}$, where $\hat{d} = d_F/d_{\min}$ is known as the *spreading dimension* [79]. Here, $d_F = d - \beta/\nu$ denotes the fractal dimension of the percolating cluster. Hence, the average number of sites touched by the BFS for an internal edge is

$$\langle \ell^{\hat{d}} \rangle \approx L^{d_F-x_2}. \tag{4.2.2}$$

Note that, while d_F and x_2 are exactly known [51, 46], this is not the case for $d_{\min}$ [15, 83]. This scaling result suggests that the expected running time for the insertion of candidate-non-bridges scales as $L^{d_F-x_2}$. Further, in case $e = (x, y)$ is a non-bridge, one removes e from the current configuration and performs an analogous BFS, as for the candidate-non-bridge case. It is not difficult to see that the leading running time scaling is also $L^{d_F-x_2}$ which eventually lets us conclude that the insertion and deletion of pivotal edges scales, in expectation and to leading order, as $L^{\gamma/\nu}$, whereas non-pivotal edges allow for a typical running time of $L^{d_F-x_2} \ll L^{\gamma/\nu}$.

Table 4.1 confirms our theoretical arguments. Firstly, it shows our estimates for the exponent $y^{(i)}_{\mathrm{SBFS}}$ in a power law scaling $L^{y^{(i)}_{\mathrm{SBFS}}}$ of the expected number of vertices visited by the BFS for update-operations on non-pivotal edges. Secondly, note that because both pivotal and non-pivotal edges contribute in average a fraction 1/2 of all edges[2], we have that for sufficiently large L, the $L^{\gamma/\nu}$ scaling, stemming from pivotal operations, dominates the overall expected running time of update operations, denoted by $\bar{t}$. In other words we have for the sequential BFS implementation of Sweeny's algorithm the following scaling:

$$\bar{t} \approx L^{\gamma/\nu}.$$

We tested the scaling relation for $\bar{t}$ for some values of q, shown in table 4.1, by fitting a power-law to the system size dependent average running time at the corresponding self-dual point. We denote the involved exponent by κ, that is $\bar{t} \approx L^{\kappa}$. More precisely, the

[2]Indeed, one has $\sqrt{q}/[2(1+\sqrt{q})] + 1/[2(1+\sqrt{q})] = 1/2$, which is precisely the asymptotic density of pivotal (non-pivotal) edges for the self-dual random-cluster model on $\mathbb{Z}^2_L$ in the limit $L \to \infty$.

corresponding exponents are shown in table 4.1 in the column for κ_{SBFS}. The values seem to deviate from γ/ν, however we emphasise that the average running time $\bar{t}$ is a mixture of two power-law contributions with exponents $d_F - x_2$ and γ/ν. Moreover, we observe that the exponent is closer to γ/ν than to $d_F - x_2$ in the entire regime of cluster weights considered. We explain the deviations precisely due to this mixture, and still expect the exponent γ/ν to distinguish itself for larger system sizes than considered here. In Figure 4.3 we show exact values of γ/ν and $d_F - x_2$ derived from the Coulomb gas formulation of critical two-dimensional random-cluster models [51]. We therefore conclude, that the

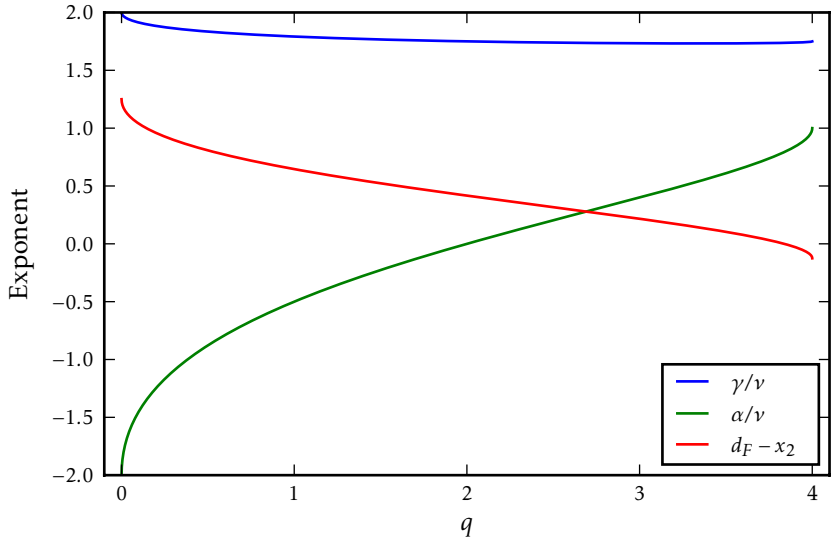


Figure 4.3: *Exact values for γ/ν, $d_f - x_2$ and α/ν for the two-dimensional critical random-cluster model as predicted by Coulomb gas arguments. Observe the significant variation of $d_F - x_2$ as compared to γ/ν.[51].*

sequential breadth-first search implementation of Sweeny's algorithm, albeit being simple and having a small additional data structure overhead, has expected running time almost linear in the volume. This introduces a drastic computational slowing down retarding any improvement in the dynamical critical behaviour of Sweeny's algorithm as compared to other Markov chains for the random-cluster model.

4.2.2 Interleaved breadth-first search

A simple and elegant trick to weaken or reduce the computational slowing down was first introduced in [15], and rests upon the observation that the cluster structure at criticality is fractal and one typically expects very unequal cluster sizes following the removal of a bridge, corresponding to $L^{\gamma/\nu} \ll L^{d_F - x_2}$ (consider Figure 4.3). For a more detailed discussion of this observation and related fragmentation phenomena we refer to Chapter 5. The main obstacle with the plain BFS approach is that a priori one cannot do better than choosing the larger of the two clusters attached to two neighbouring (disconnected)

Table 4.1: *Run-time scaling exponents in two dimensions for the SBFS, IBFS and the UF implementation. The scaling exponents γ/ν and $d_F - x_2$ are shown for reference and comparison. The y exponents correspond to scaling $\approx L^y$ of the number of vertices touched in a sequential (SBFS) and interleaved (IBFS) breadth-first cluster traversal for internal or non-pivotal (i) and external or pivotal (e) edges, respectively.*

q	κ_{SBFS}	κ_{IBFS}	κ_{UF}	$y^{(i)}_{\mathrm{IBFS}}$	$y^{(e)}_{\mathrm{IBFS}}$	$y^{(i)}_{\mathrm{SBFS}}$	$y^{(e)}_{\mathrm{SBFS}}$	$d_F - x_2$	γ/ν
0.0005	1.81(1)	1.18(3)	1.54(6)	1.25(1)	1.24(1)	1.25(1)	1.99(1)	1.23407	1.99296
0.005	1.80(2)	1.14(2)	1.70(6)	1.22(1)	1.21(1)	1.22(1)	1.98(1)	1.20021	1.97823
0.05	1.74(1)	1.02(3)	1.77(4)	1.10(1)	1.11(1)	1.11(1)	1.93(1)	1.09783	1.93580
0.1	1.72(1)	0.97(2)	1.78(3)	1.05(1)	1.05(1)	1.06(1)	1.91(1)	1.03881	1.91284
0.5	1.65(2)	0.71(3)	1.77(2)	0.82(1)	0.82(1)	0.82(1)	1.83(1)	0.80768	1.83449
0.7	1.63(2)	0.69(4)	1.77(4)	0.75(2)	0.76(1)	0.76(2)	1.81(1)	0.73541	1.81407
1	0	0	1.74(2)	0.66(1)	0.67(1)	0.67(1)	1.79(1)	0.64583	1.79167
1.5	1.56(2)	0.43(2)	1.71(3)	0.55(2)	0.56(1)	0.57(2)	1.75(1)	0.52298	1.76644
2	1.57(2)	0.35(3)	1.68(3)	0.46(2)	0.47(2)	0.48(3)	1.73(1)	0.41667	1.75000
3	1.52(4)	0.19(1)	1.67(2)	0.32(3)	0.30(4)	0.35(3)	1.69(2)	0.21667	1.73333
4	1.42(4)	0.13(2)	1.64(7)	0.22(11)	0.23(1)	0.26(1)	1.68(1)	-0.12500	1.75000

vertices, with probability 1/2. Therefore the running time is ruled by the typical cluster size $L^{\gamma/\nu}$.

Now, by starting two simultaneous breadth-first searches at the two vertices incident to a randomly chosen edge, one can reduce the expected running time to a scaling $L^{d_F - x_2}$. To start with, consider the removal of a bridge or the insertion of a candidate-bridge. Now, both cases revert to two simultaneous BFS to determine if the inspected edge is pivotal. In the former case one temporarily erases the bridge before the two BFS are started. Denote the edge by $e = (x, y)$. As soon as one of the two BFS traversals is exhausted, which is clearly the one operating on the smaller cluster, one can conclude that vertex x and y are not connected, hence e is pivotal (bridge in the former and candidate-bridge in the latter case) to the original configuration. As mentioned earlier, both cases contribute a constant fraction of operations, which in turn implies that the insertion and deletion of pivotal edges has $L^{d_F - x_2}$ contribution to the expected running time of the simultaneous BFS variant of Sweeny's algorithm.

We remark that one can either use true (trivial) parallelisation or interleave the two breadth-first searches sequentially. In both cases, the system size scaling is $L^{d_F - x_2}$, but the interleaved approach is (assuming perfect parallelisation) by a factor of two slower than the parallelised version. In what follows we only consider the interleaved variant, and abbreviate it by IBFS.

We confirmed the outlined improved run-time scaling for operations on pivotal edges numerically, and show in table 4.1 our numerical results, extracted from power-law fits to the estimated expected number of vertices visited by the two interleaved breadth-first searches in case of a deletion or insertion of a pivotal edge. The corresponding exponent is denoted by $y^{(e)}_{\mathrm{IBFS}}$. Further, Figure 4.4 shows a graphical confirmation of the scaling prediction $y^{(e)}_{\mathrm{IBFS}} = d_f - x_2$, for three values of q.

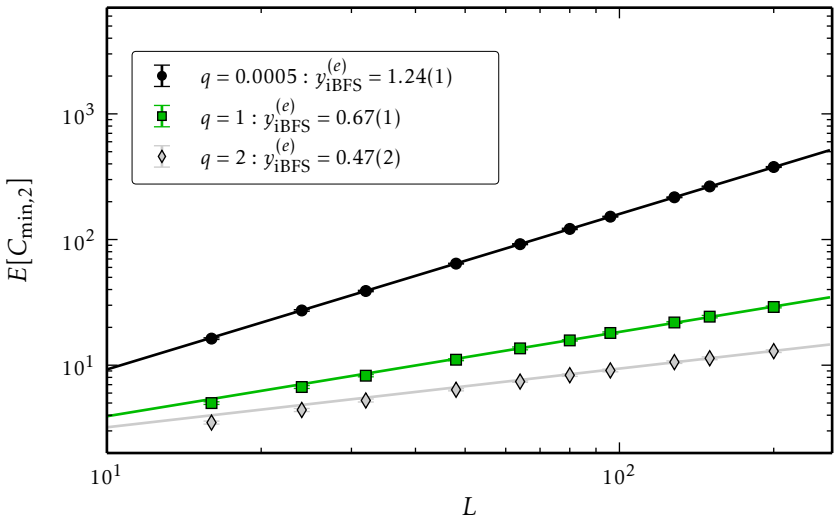


Figure 4.4: *Expected number of vertices $\mathbb{E}[C_{\text{min},2}]$ in the smaller of two adjacent clusters, as visited in a 2D simulation using IBFS for operations on pivotal edges. The lines show fits of the power-law form $C_{\text{min},2} = AL^{y^{(e)}_{IBFS}}$ to the data.*

Due to the fact that the IBFS approach has no additional data structure to maintain during edge updates, it is perfectly capable of exploiting the unconditional acceptance of update-steps, discussed in the beginning of this chapter. We numerically compared the running time of the IBFS algorithm using unconditional acceptances whenever possible, to the IBFS algorithm without such optimisation. In Figure 4.1 we show the ratio of the expected running times per operation of the optimised version to the plain IBFS version. The data clearly supports the theoretical arguments that showed that in a fraction of $\min\{1/\sqrt{q}, \sqrt{q}\}$ of all cases, one can unconditionally accept a proposed deletion/insertion of edge in the Metropolis chain.

The running time for the removal or insertion of a non-pivotal edge has the same asymptotic expected scaling $L^{d_F - x_2}$ as can be seen by the following scaling argument [15]: The running time is proportional to the number of sites visited until the two breadth first searches merge. If we denote the number of vertices visited by one of the BFS searches until the two BFS merge by B, we have $\mathbb{P}[B \geq s] \approx p_2(s^{1/d_F})$, where $p_2(R)$ is the probability that there are at least two distinct clusters emanating in an annulus from the inner boundary of radius $O(1)$ to the outer boundary of radius R, which has large R asymptotics $p_2(R) \approx R^{-x_2}$ [15]. We therefore expect $\mathbb{P}_q[B \geq s] \approx s^{-x_2/d_F}$ and hence $\mathbb{E}_q[B] \approx L^{d_F - x_2}$. The argument simply states that when it takes each of the two BFS roughly s vertices until they merge, then we can construct an annulus centred around the two starting vertices with approximate outer radius s^{1/d_F}, inside which the two clusters are disconnected. This in turn implies a scaling

$$\bar{t} \approx L^{d_F - x_2},$$

for the interleaved breadth-first search Sweeny implementation. To confirm our arguments, we show in table 4.1 the numerically extracted effective exponent κ_{IBFS}, which, apart from deviations for large q, attributed to strong finite size corrections, are consistent with $\kappa_{\mathrm{IBFS}} \approx d_F - x_2$. In Figure 4.5 we compare the average running time per operation, for the critical $q = 0.005$ case, of, among others, the SBFS and IBFS implementations. It clearly emerges that the interleaved is superior over the sequential breadth-first implementation. However, in both cases $\bar{t}$ suffers from computational slowing down, i.e. the size dependence is still in form of a power-law in L.

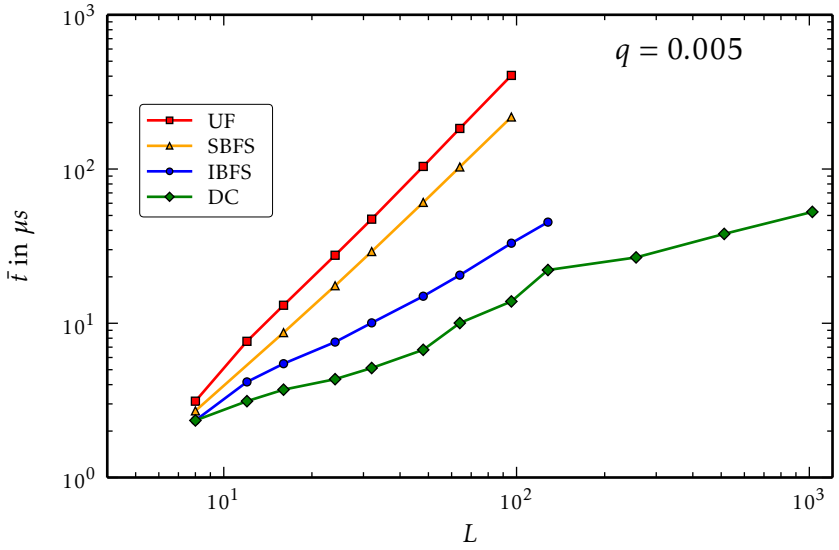


Figure 4.5: *Run-time per edge operation of simulations of the $q = 0.005$ square-lattice RCM and Sweeny's algorithm employing the SBFS, IBFS, UF and DC connectivity implementations, respectively.*

4.2.3 Union-find data structure and interleaved breadth-first search

The union-find data structure, formally introduced in Chapter 3, is an incremental data structure that has no capability to support simultaneously the deletion and insertion of edges with the same, for all practical problem instances, constant computational effort. However, we have combined it with the interleaved breadth-first search approach, with the hope to reduce the system-size scaling of $\bar{t}$ further below $L^{d_F - x_2}$. Our initial intuition was that the pseudo-constant computational complexity of inserting edges and connectivity queries in the union-find data structure pushes the overall scaling of $\bar{t}$ further down as compared to the plain interleaved approach. Unfortunately, we overlooked a crucial aspect. The problem occurs during the re-building phase of the union-find data structure following a bridge deletion. To understand this, consider the deletion of a bridge $e = (x, y)$. Removing e splits, by definition, the connected component C, to which x, y belong, into two components C_x and C_y. This demands that the tree in the union-find data structure, representing the connected component C, needs to be split, such that the resulting two trees correctly encode the graph connectivity. Cutting (and linking)

of trees can be done with logarithmic computational effort, e.g. with the Euler-tour representation, or the recent DFS-Tour tree representation [84]. Ideally, one would cut the tree at a position that corresponds uniquely to e, and the resulting two trees correspond to the connected components, induced by the original bridge-removal in the represented graph. However, the path-compression heuristics, underlying the optimal union-find algorithm data structure, destroys any structural information in the tree and merely encodes the equivalence classes corresponding to connected components. For instance, an edge between two nodes in the union-find tree does not necessarily correspond to an edge in the actual represented graph. In fact, typically most of all vertices are direct descendants of the unique representative element of the connected component.
Thus, in order to rebuild the union-find data structure, we have to re-analyse the cluster structure by means of a graph traversal that determines for all vertices, previously part of C, whether they belong to C_x or C_y, and construct the corresponding UF-tree. Note, there is no significant difference here between the sequential and interleaved BFS traversal approach, in the sense that both of them show the same asymptotic size scaling of the expected running time devoted to this re-building process: In both cases all vertices of a randomly chosen component have to be visited, and one can *not* benefit from the diversity in cluster sizes attached to nearby vertices. Therefore, one again obtains the $L^{\gamma/\nu}$ contribution to $\bar{t}$ for the deletion of bridges.

Now considering the different variants of the union-find data structure, one realises that it is possible to dispense the path-compression heuristics; This worsens the amortised cost of connectivity and union operations to a logarithm in L, as shown in Lemma 3.4.1. However, a moment of thought shows that it is now necessary to cut the representing tree in both cases, the deletion of a bridge and non-bridge (in this case the tree is later merged again by a replacement edge). However, the required graph-traversal can now exploit the fact that nearby clusters are typically very unequal in size, and hence circumvent the $L^{\gamma/\nu}$ scaling. Overall, one obtains in such an approach a contribution asymptotically proportional to $L^{d_F - x_2}$ for the removal of edges, and $\log(L)$ for the insertion of edges.
In fact, this version brings us closer to the dynamic connectivity approach, which solves the overall problem with poly-logarithmic computational effort. We therefore did not consider this modified union-find approach further. Yet, it is an interesting aspect to consider for a future study, possibly in combination with a study of related phenomena in three dimensions.
Returning to the path-compression based union-find data structure, observe that when deleting a non-bridge, one clearly benefits from the interleaved BFS, as no rebuild of the union-find data structure is necessary. Recall, that the running time bound for the union-find data structure with path compression and union-by-rank heuristics is amortised. Introducing $L^{\gamma/\nu}$-expensive operations to rebuild the data structure, clearly destroys any amortisation argument. In order to confirm these heuristic scaling arguments we numerically estimated the typical running time $\bar{t}$ for the hybrid union-find and interleaved

BFS approach. In figure 4.5 we compare the average running time of the hybrid union-find variant to all other alternatives for $q = 0.005$. Besides indicating that the asymptotic scaling of the sequential BFS and union-find approach are equivalent, Figure 4.5 reveals that the involved constants are largest for the union-find approach. We attribute this to the additional overhead imposed by maintaining the union-find data structure. Furthermore, we obtained a more quantitative measure of the system size scaling of $\bar{t}$ by fitting our numerical estimates to a power law in L with exponent κ_{UF}. The corresponding estimates for some values of q are summarised in table 4.1. Similar to the SBFS implementation, we do not exactly recover the predicted γ/ν exponent, but an effective intermediate exponent. It is also interesting to consider how the computational efficiency of the UF variant varies with q (at criticality). Figure 4.6 reveals that the involved amplitudes decrease with q. We attribute this to the decreasing density of bridges with increasing cluster weight, which makes bridge removals less likely.

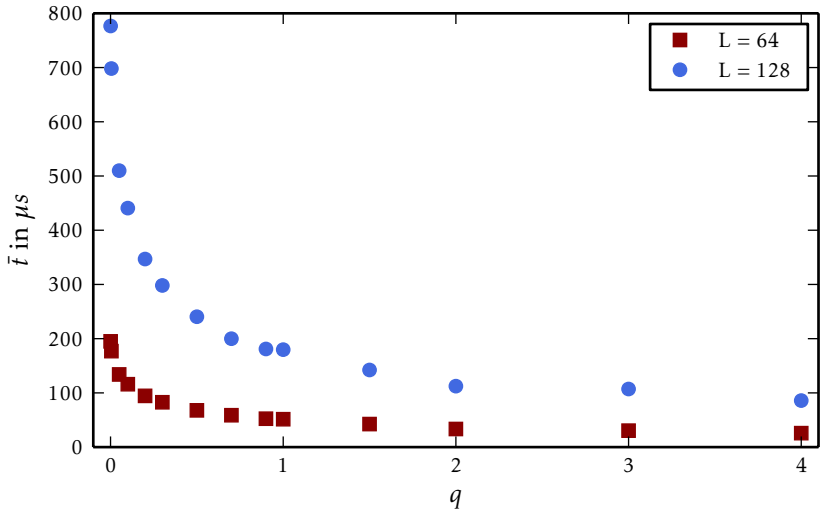


Figure 4.6: *Cluster-weight dependence of the mean running time per operation for the UF Sweeny algorithm for two characteristic system sizes.*

4.2.4 Computational efficiency of heat-bath and Metropolis variants

An interesting question is in how far the computational efficiency is influenced by the actual variant of Sweeny's algorithm, i.e. do we expect a difference between Metropolis and heat-bath variant of Sweeny's algorithm? The preceding discussion was general enough such that we can conclude that the corresponding leading asymptotic scaling is in both cases $L^{\gamma/\nu}$, due to the argument relying on the interplay between the union-find data structure rebuild, based on a breadth-first search traversal, and the fractal cluster structure. However both implementations differ in how the different contributions to the running time $\bar{t}$ are balanced. Let us first consider the Metropolis case. The following is not meant to be an exact finite-size expression for $\bar{t}$, but is intended to summarise in a

concise way the different running time contributions,

$$\begin{aligned}
\bar{t} \quad \text{"}\sim\text{"} \quad & \mathbb{P}[e \in B]\Big(c_1 \min\{1,\sqrt{q}\}L^{\gamma/\nu} + c_2(1-\min\{1,\sqrt{q}\})L^{d_F-x_2}\Big) \\
+ \quad & \mathbb{P}[e \in C]c_3 L^{d_F-x_2} + \mathbb{P}[e \notin A]c_4 L^0 \\
\sim \quad & \frac{1}{2(1+\sqrt{q})}\Big(c_1 \min\{1,\sqrt{q}\}L^{\gamma/\nu} + c_2(1-\min\{1,\sqrt{q}\})L^{d_F-x_2}\Big) \\
+ \quad & \frac{c_3\sqrt{q}}{2(1+\sqrt{q})}L^{d_F-x_2} + \frac{c_4}{2}L^0 \\
= \quad & L^{\gamma/\nu}\ \frac{c_1 \min\{1,\sqrt{q}\}}{2(1+\sqrt{q})}\Big) + L^{d_F-x_2}\left(\frac{c_2\Big(1-\min\{1,\sqrt{q}\}\Big)+c_3\sqrt{q}}{2(1+\sqrt{q})}\right)\Big(\\
+ \quad & \frac{c_4}{2}L^0. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.2.3)
\end{aligned}$$

Here c_1, c_2, c_3, c_4 are appropriate q dependent constants. We used the exact asymptotic results for the density of bridges, candidate-bridges, non-bridges and candidate-non-bridges, which we discuss in Chapter 5. Moreover, we assume that the UF Sweeny algorithm is implemented in such a way that in case of an attempted removal of an edge, the breadth-first search carries a random bit indicating whether the removal of a bridge would be rejected. This allows the BFS to terminate whenever the smaller cluster is exhausted instead of exhausting both clusters, which then will be re-joined due to the rejection of the bridge-removal. The corresponding expression for the heat-bath variant of Sweeny's algorithm can also be directly obtained and reads

$$\bar{t} \quad \text{"}\sim\text{"} \quad L^{\gamma/\nu}\ \frac{c_1\sqrt{q}/(1+\sqrt{q})}{2(1+\sqrt{q})}\Big) + L^{d_F-x_2}\left(\frac{\frac{c_2}{1+\sqrt{q}}+\sqrt{q}c_3}{2(1+\sqrt{q})}\right) + \frac{c_4}{2}L^0, \qquad (4.2.4)$$

Furthermore, we show in Figure 4.7 the amplitude for the $L^{\gamma/\nu}$ term (in units of c_1) for both Metropolis and heat-bath. One can directly observe that the heat-bath chain avoids operations $L^{\gamma/\nu}$ in average more often than the Metropolis variant and is thus, from a purely computational point of view, the preferred option.

We further note that the optimisation trick, outlined at the beginning of this chapter, does not further reduce the amplitude in front of the $L^{\gamma/\nu}$ term for the Metropolis algorithm. This is because, albeit the acceptance of an edge removal can be done in $\min\{\sqrt{q}, 1/\sqrt{q}\}$ of all cases without determining whether the edge is a bridge or non-bridge, one still needs to re-build the union-find data structure following a bridge-removal. Thus in both cases unconditional and conditional acceptance of a deletion of a bridge, we have $c_1 L^{\gamma/\nu}$ computational effort. This happens in precisely a fraction $\min\{1,\sqrt{q}\}/[2(1+\sqrt{q})]$ of all cases, where the numerator is the acceptance probability of a bridge removal, the denominator the (asymptotic) density of bridges at the self-dual point. On the other hand, the heat-bath variant can gain from this optimisation trick, due to the unconditional rejection, which happens in a fraction $1 - p_{\mathrm{sd}}(q) = 1/(1+\sqrt{q})$ of all cases. We omit the corresponding extension of (4.2.4). We note that, all Sweeny implementations in this section use the Metropolis scheme.

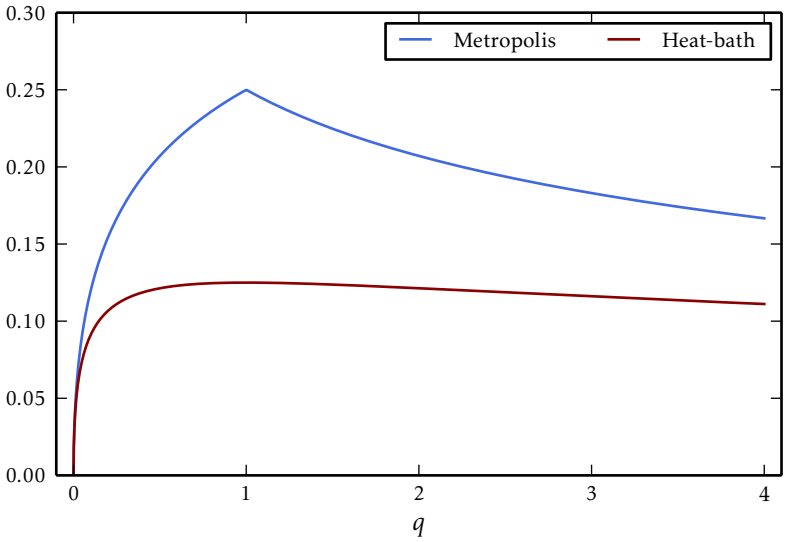


Figure 4.7: *Likelihood of an $L^{\gamma/\nu}$ operation (assuming $c_1 = 1$) for the Metropolis and heat-bath UF Sweeny algorithm. Note that the corresponding expressions for Metropolis and heat-bath are given by* (4.2.3) *and* (4.2.4)*, respectively.*

4.2.5 Dynamic connectivity algorithm

The idea of maintaining a collection of trees, that represents the current spanning subgraph (spanning forest) forms the basis of recent fully dynamic connectivity algorithms such as [24, 67, 85]. We described in Chapter 3 a slight modification of the particular poly-logarithmic fully dynamic connectivity algorithm presented in [24]. In what follows we refer to this algorithm as DC-algorithm. The property that makes the algorithm interesting for our MCMC application is Theorem 3.5.1, stating that any insertion and deletion operation, as well as connectivity query, can be done with $O((\log(L))^2)$ amortised computational complexity. This clearly outperforms any of the previously mentioned traversal based variants and reduces the computational slowing down to a poly-logarithm in the system size. Before we numerically investigate how the DC-algorithm performs, we remark that albeit the bound is amortised poly-logarithmic, the involved constants might be very large (any O statement is strictly speaking only valid for a sufficiently large system size). Secondly, due to the maintenance of $O(\log(L))$ levels, one expects a large memory footprint, which might hinder reaching system sizes for which the asymptotic poly-logarithmic dependence becomes significantly visible.
Firstly, consider for the last time Figure 4.5, which compares the DC-algorithm variant to all other traversal-based implementations of Sweeny's algorithm for the particular choice $q = 0.005$. For this choice of parameters, the DC-algorithm is the fastest approach to simulate Sweeny's algorithm. Whereas one clearly recognises a power-law scaling for the UF, SBFS and IBFS implementation, one cannot see such linearity in the log-log plot for the DC-algorithm. The kink for system-sizes around 10^2 is caused by caching effects, i.e. for the system sizes in the first regime most of the data structures can be efficiently cached

on the system used for the running time estimation. However, for larger system sizes this optimisation does not work and longer memory access times are imposed, resulting in a change of constants.
Secondly, in Figure 4.8 we show the mean running time $\bar{t}$ for the DC-algorithm for various values of q covering the range from tree-like configurations, $q = 0.0005$, over percolation, $q = 1$, to more "compact" clusters for the Ising $q = 2$ or $q = 3$. As Figure 4.8 reveals, the overall running time increases with decreasing q. We know that for $q > 1$ there are more non-bridges than for $q < 1$, and hence the DC-algorithm can exploit the level-hierarchy slightly better for large q, i.e. distributing non-tree edges over several levels enhances the amortisation of replacement-edge searches. However, the edge-level hierarchy is, for practical purposes, too "heavy" in the regime $q < 1$, as we will show in the section devoted to algorithmic optimisations. In order to obtain a quantitative estimate for the constants

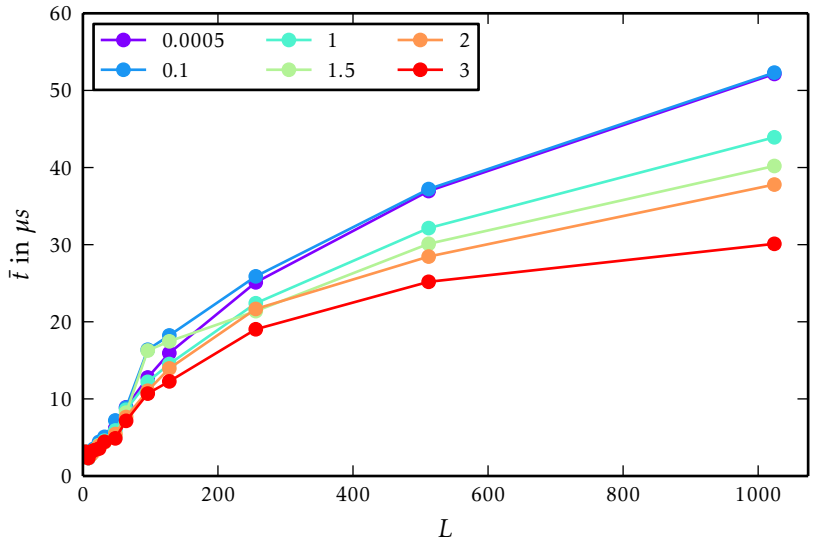


Figure 4.8: *Average run-time $\bar{t}$ for several values of q and L for the Sweeny update using a dynamic connectivity (DC) algorithm based on splay trees. The lines are only visual aid.*

involved in $\bar{t}$ we fitted the functional form

$$\bar{t}(L) = a \log^2 L + b \log L + c$$

to the data. We made the observation that the estimates for c were consistent with 0, such that we fixed $c = 0$ for the subsequent analysis. Somewhat surprisingly, our fits yield $b < 0$; we interpret this as a result of the presence of correction terms and the amortized nature of the run-time bounds leading to the asymptotic scaling only being visible for very large system sizes. Similar observations have been reported for general sets of inputs in Ref. [67]. Considering the ratio a/b, we find that its modulus increases with q, yielding a value of about 0.3 for $q = 0.0005$ and about 0.71 for $q = 2$. This corresponds to the increasing fraction of non-tree edges for increasing q, resulting in an increase of traversals

of the edge level hierarchy with the associated $\mathcal{O}(\log^2 L)$ complexity. Irrespective of this, as a consequence of the larger number of cluster-splitting operations, the total run-time is found to be largest for small q, as can be seen in Figure 4.8.
One main obstacle with the DC algorithm is the involved space complexity. The algorithm maintains $\log(n)$ instances of the graph, however with hierarchically distributed edges. This large memory consumption hindered us from reaching larger system sizes and exploring the asymptotics further. More precisely, on the system environment used for the analysis, a problem instance corresponding to the simulation of the critical point on $\mathbb{Z}^2_{1024}$ consumed 3.5GB of memory, whereas the other implementations had memory consumptions in the MB regime, c.f. [57] for more details.

4.3 Dynamical critical behaviour

Our previous discussion considered computational aspects of Sweeny's algorithm, which turned out to be a non-trivial problem and demonstrated a rich interplay between geometric properties and computational slowing down. Yet, one of the main motivations in studying algorithmic aspects of Sweeny's algorithm is its very interesting and under some circumstances peculiar dynamical critical behaviour. In this section we elaborate on certain related aspects of its dynamical critical behaviour, both in Metropolis and heat-bath variant. However, as already outlined, both the heat-bath and Metropolis version of the Sweeny dynamics are in the same dynamic universality class. We remark that the study [21] of Deng et al. considers the heat-bath dynamics, whereas our numerical study in [53] is concerned with the Metropolis version.
Markov chain Monte-Carlo methods close to a critical point suffer in many instances from *critical slowing-down*: the intrinsic time scales, such as autocorrelation times $\tau_{\text{exp}}, \tau_{\text{int},f}$ for an observable f, or even t_{mix} [86], show power laws in the system size. One expects for second order phase transitions a scaling of the form $\approx L^z$, where $z > 0$ is an appropriate dynamical critical exponent [44]. In this section we adopt the standard and define time scales in units of a *sweep*, that is m elementary Markov steps compose one elementary sweep. As described in section 2.4, the Li-Sokal bound for the Sweeny dynamics establishes a rigorous lower bound on the exponential autocorrelation time τ_{exp} as well as for the integrated autocorrelation time of energy-like observables such as, the density of open edges, denoted by $\mathcal{N}$. In the setting of this chapter, it roughly states (ignoring possible multiplicative corrections) that

$$z_{\text{exp}} \geq z_{\text{exp},\mathcal{N}} \geq z_{\text{int},\mathcal{N}} \geq \frac{\alpha}{\nu}. \tag{4.3.1}$$

This establishes a theoretical connection between a divergent (normalised) heat-capacity, a property of stationary model, and slow stationary dynamics, a dynamical property of the Markov chain. Remarkably, for the random-cluster model in two dimensions and in particular for $\mathbb{Z}^2$, one can even obtain exact, though not rigorously established, expressions for the ratio α/ν by exploiting the corresponding Coulomb-gas formulation [51]. For a

graphical illustration consider Figure 4.3. More quantitatively, we remark that $\alpha/\nu \leq 0$ for $q \leq 2$, [51]. Thus from a theoretical point of view, there is no enforced critical slowing down $q \leq 2$. To emphasise this, we can contrast this with the spin heat-bath algorithm for the Ising model, where one has the analogue to the Li-Sokal bound for the magnetisation $\mathcal{M}$: $z_{\text{int},\mathcal{M}} \geq \gamma/\nu = 2 - \eta$ at criticality, where $\eta = 0.25$ for the Ising model on $\mathbb{Z}^2$ [51]. Hence the theoretical restrictions are more severe and impose a drastic computational slowing down. Numerical results are even more "drastic" and suggest $z_{\text{int},\mathcal{M}} > 2$ or in fact $z_{\text{int},\mathcal{M}} = 2.172(6)$ [87]. Therefore, having a rather weak theoretical constraint for the Sweeny dynamics it is not a-priori clear whether and in how far the dynamics suffers from critical slowing-down (at least for $q \leq 2$).

Indeed, in [21, 53] it was shown that for certain "global" observables one has a negative dynamical critical exponent for the corresponding integrated autocorrelation time, a phenomenon termed critical speeding-up in [21], notationally opposed to the typical critical slowing down. Note that critical slowing down and critical speeding up can however coexists, as pointed out in [21]. This is because critical slowing-down refers to the slowest mode, whereas critical speeding-up describes the effect that certain observables suppress projections onto slow modes with a shift towards faster modes with increasing system size.

For instance, the observables considered in [53] are C_1, the number of vertices in the largest connected component, and $\mathcal{S}_2$, the second cluster-size moment, i.e. the sum over the squares of cluster sizes [50]. These quantities arise naturally in the Fortuin-Kasteleyn or random-cluster model representation of standard statistical physics quantities such as the susceptibility or magnetisation.

Here we show in Figure 4.9 the estimated normalised autocorrelation time $\rho_{C_1}(t)$, where we used the standard cut-off based estimation procedure, described in e.g. [44]. Firstly, note that the time t in this plot is measured in units of L single steps, that is $2L$ of such single-steps, in other words $t = 2L$, correspond to one sweep. Secondly, as can be clearly seen in the figure, the correlations decay[3] completely before the first sweep, except for $q = 2$, where apparently the critical speeding-up effect disappears. One reason for this could be the marginality of $q = 2$, where $\alpha/\nu = 0$ in the sense of a logarithmic divergence of the normalised heat capacity, see Chapter 5. Furthermore, recall that $\alpha/\nu < 0$ for $q < 2$ as well as $\alpha/\nu > 0$ for $q > 2$). The speeding-up for $q < 2$ reflects itself also in the system-size scaling of τ_{int,C_1} (now measured again in sweeps), as shown in Figure 4.10. A completely analogous behaviour was observed for the second-cluster size moment observable $\mathcal{S}_2$, which can be linked to the susceptibility of the corresponding Potts model whenever $q \in \{2, 3, \dots\}$ [57, 50]. In table 4.2 we summarise the corresponding dynamical exponents for all values of q considered. Remarkably, as first noted in [21], where comparable, the exponents for $\mathcal{S}_2$ and C_1 (and $\mathcal{N}$) are smaller for Sweeny dynamics than for the Swendsen-Wang-Chayes-Machta algorithm [88]. Furthermore the critical speeding-up phenomenon is necessarily absent for the Swendsen-Wang-Chayes-Machta

[3] Here we mean by decay, that the normalised autocorrelation function is below $1/e$.

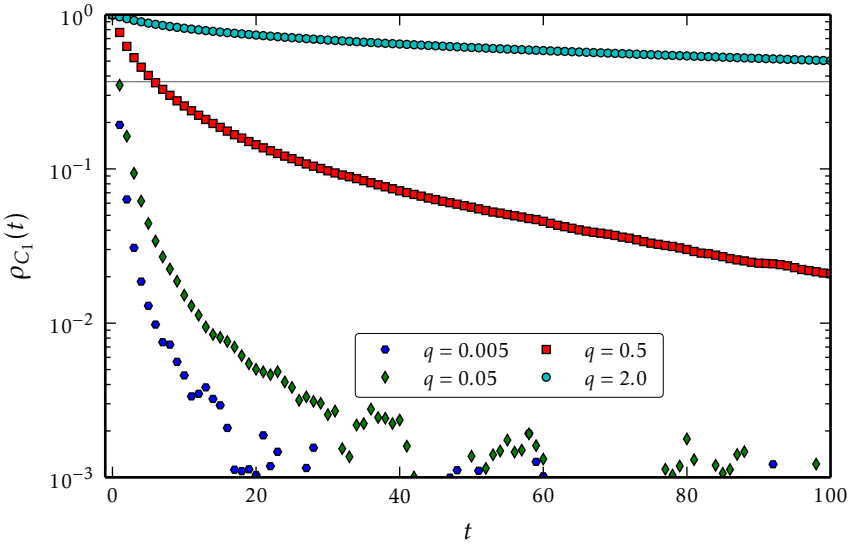


Figure 4.9: *Estimated normalised autocorrelation function for the observable measuring the size of the largest connected component, measured in units of L MC-steps. The horizontal line shows the value of* $1/e$.

dynamics. This is not surprising, as the intrinsic time scale in the latter dynamics consists of one combined (sequential) colour-bond update. We emphasise however, that the computational complexity of both approaches differ significantly and as we will show below, a combined efficiency measure has to be considered in order to conduct meaningful comparisons. However, before doing so, let us consider the bond-density, which turns out to be a very good probe for the slowness of the stationary dynamics. We emphasise, that the Li-Sokal bound in the form (4.3.1) does *not* apply to global observables such as C_1 or $\mathcal{S}_2$, thus in this case there are no apparent non-trivial theoretical restrictions.

In contrast, the bond-density $\mathcal{N}$ is ruled by the Li-Sokal bound, which forbids critical speeding-up for $q \geq 2$ but leaves in principle the freedom for a negative dynamical critical exponents for $q < 2$. However our subsequent analysis of the case $q = 1$ shows that critical speeding-up is absent for $\mathcal{N}$ hence it is plausible that this is also the case for $q \geq 1$. Still, in the absence of (rigorous) theoretical arguments, it is therefore interesting to investigate whether a similar speeding-up effect happens in this regime as for C_1 and $\mathcal{S}_2$. The numerical studies [21, 53] showed that this is not the case and one merely has $z_{\text{int},\mathcal{N}} = 0$ for $q < 2$, possibly also valid for $q = 2$, however with logarithmic divergence in L [21]. In order to understand what causes this slow relaxation it is illuminating to calculate $\tau_{\text{int},\mathcal{N}}$ exactly for $q = 1$.

4.3.1 Exact calculations for percolation

The solvability of the case $q = 1$ stems from the independence of edge updates or equivalently from the factorisation of the probability distribution over independent Bernoulli trials with parameter p. For the time being, denote the transition matrix of the heat-bath

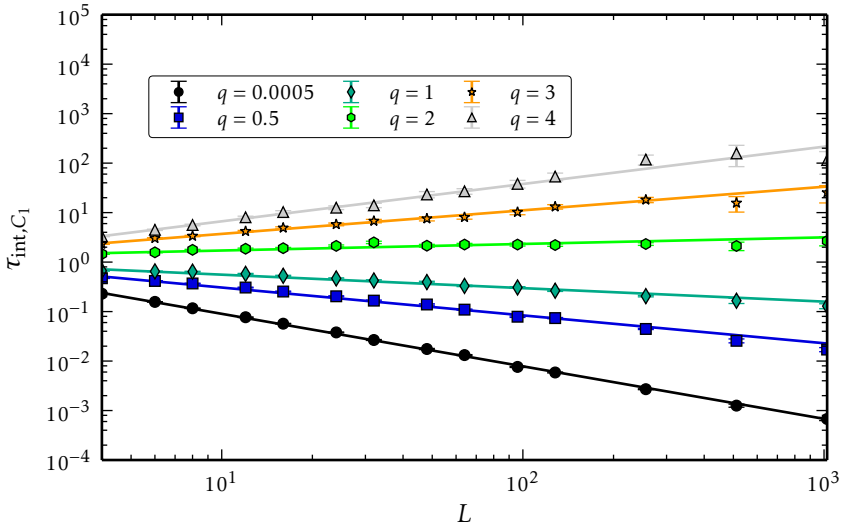


Figure 4.10: *Integrated autocorrelation time of the C_1 observable for Sweeny-Metropolis dynamics. The negative slope for $q < 2$ in the log-log plots indicates $z_{\mathrm{int},C_1} < 0$, see text below. The solid lines corrspond to least squares fits.*

q	$z_{\mathrm{int},\mathcal{S}_2}$	z_{int,C_1}	$z_{\mathrm{int},\mathcal{N}}$	$z^{\mathrm{SW}}_{\mathrm{int},\mathcal{S}_2}$ [88]	α/ν
0.0005	-1.12(1)	-1.11(1)	0.01(1)	–	-1.958
0.005	-1.09(1)	-1.09(1)	0.01(1)	–	-1.868
0.05	-1.04(1)	-1.05(2)	0.01(1)	–	-1.601
0.2	-0.86(1)	-0.88(1)	-0.01(1)	–	-1.247
0.5	-0.63(1)	-0.64(2)	0.00(1)	–	-0.878
1.0	-0.33(1)	-0.35(1)	0.00(1)	0	-0.500
1.5	-0.11(2)	-0.11(2)	0.06(2)	0	-0.227
2.0	0.03(3)	0.02(3)	0.13(2)	0.14(1)	0.000
3.0	0.44(4)	0.43(4)	0.45(4)	0.49(1)	0.400
4.0	0.75(6)	0.73(10)	0.73(6)	0.93(2)	1.000

Table 4.2: *Estimated dynamical critical exponents $z_{\mathrm{int},\mathcal{O}}$ for the two-dimensional RCM at criticality and $\mathcal{O} = \mathcal{S}_2, C_1, \mathcal{N}$ for a range of q values. The values shown for $z^{\mathrm{SW}}_{\mathrm{int},\mathcal{S}_2}$ for the Swendsen-Wang-Chayes-Machta algorithm are actually related to the nearest-neighbour connectivity observable $\mathcal{E}'$, but it was reported in Refs. [88, 50] that the observables $\mathcal{E}'$, $\mathcal{N}$ and $\mathcal{S}_2$ share the same dynamical critical exponent for this algorithm.*

chain with parameters $0 < p < 1$ and $q = 1$ by P, the probability measure over Ω_m by $\mathbb{P}_p$, and the corresponding expectation by $\mathbb{E}_p[\cdot]$. Furthermore we adopt the standard notation and write $[m]$ for the set of integers $\{1, 2, \ldots, m\}$, where m is the number of edges of the underlying graph. Moreover, the analysis is somewhat more convenient when we consider the state space as the hypercube $\Omega_m = \{0,1\}^m$ together with the convention that $E = \{e_1, e_2, \ldots, e_m\}$. This is no restriction as clearly any configuration $\mathbf{x} \in \Omega_m$, can be uniquely related to a sub-set of edges (or spanning sub-graph) via the prescription $x_i = 1$ if and only if edge e_i is open. In order to derive a set of eigenfunctions of P, recall that we

can write $P = 1/m \sum_{i\in[m]} P_i$, where

$$P_i(\mathbf{x},\mathbf{y}) = \mathbb{1}_{\{\mathbf{y}\in\Omega_{\mathbf{x},i}\}} \frac{\mathbb{P}_p[\mathbf{y}]}{\mathbb{P}_p[\Omega_{\mathbf{x},i}]} = \mathbb{P}_p[\mathbf{x}|\mathbf{y}\in\Omega_{\mathbf{x},i}] \quad \mathbf{x},\mathbf{y}\in\Omega_m,$$

where we used the notation introduced in section 2.2, that is $\Omega_{\mathbf{x},i}$ is the set of configurations that agrees with $\mathbf{x}$ at all coordinates, except possibly i. We therefore obtain the useful relation for any observable $f : \Omega_m \to \mathbb{R}$:

$$(P_i f)(\mathbf{x}) = \mathbb{E}_p[f|\Omega_{\mathbf{x},i}] = \mathbb{E}_p[f|\{x_j\}_{\{j\neq i\}}],$$

that is the expectation of f under the condition that all edges besides e_i have status (open or closed) as specified by $\mathbf{x}$. Now, consider for $i \in [m]$

$$\chi_i(\mathbf{x}) \equiv \chi_{\{i\}}(\mathbf{x}) \equiv \frac{x_i - p}{\sqrt{p(1-p)}}. \tag{4.3.2}$$

One can easily verify that $P_i\chi_j = \mathbb{1}_{\{i\neq j\}}\chi_j$ for $i,j\in[m]$, that is χ_j is an eigenfunction of P_i with eigenvalue $\mathbb{1}_{\{i\neq j\}}$. Hence we have

$$P\chi_j = \left(1 - \frac{1}{m}\right)\chi_j.$$

We can furthermore generalise and define $\chi_S \equiv \prod_{i\in S}\chi_{\{i\}}$ for $S \subseteq [m]$ with the convention $\chi_\varnothing \equiv 1$, and obtain,

$$P\chi_S = \left(1 - \frac{|S|}{m}\right)\chi_S.$$

We conclude that $\{\chi_S\}_{\{S\subseteq[m]\}}$ is a set of eigenfunctions of P. In fact, this set of functions is a orthogonal and normal (orthonormal) basis of the inner product space $\left(\mathbb{R}^{\Omega_m}, \langle\cdot,\cdot\rangle_p\right)$, i.e., one can additionally verify that $\langle\chi_S,\chi_T\rangle_p \equiv \mathbb{E}_p[\chi_S\chi_T] = \mathbb{1}_{\{S=T\}}$. This in particular allows us to expand any function $f : \Omega_m \to \mathbb{R}$:

$$f = \sum_{S\subseteq[m]} \chi_S \hat{f}(S), \tag{4.3.3}$$

where $\hat{f}(S) \equiv \mathbb{E}_p[f\chi_S]$ is known as a *Fourier-Walsh coefficient*, c.f. [89, 90]. Note that in what follows we suppress the p dependence. Indeed, the particular choice of $\{\chi_S\}_{\{S\subseteq[m]\}}$ is closely related to the discrete Fourier analysis on the hypercube. For instance, we have $\mathbb{E}[f^2] = \sum_{S\subseteq[m]} \hat{f}(S)^2$, $\mathbb{E}[f] = \hat{f}(\varnothing)$ as well as

$$\mathrm{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S\subseteq[m], S\neq\varnothing} \hat{f}(S)^2, \tag{4.3.4}$$

which is known as Parseval's theorem. One can also express the autocorrelation function at time lag $t \geq 1$ in terms of the Fourier-Walsh coefficients, or equivalently in terms of the

energy spectrum of the function f:

$$\begin{aligned} C_f(t) &= \langle (\mathrm{I}-\Pi) f, P^k (\mathrm{I}-\Pi) f \rangle \\ &= \langle f, (P-\Pi)^{|t|} f \rangle \\ &= \sum_{S \subseteq [m], S \neq \varnothing} \left(1 - \frac{|S|}{m}\right)^t \hat{f}(S)^2, \\ &= \sum_{k=1}^{m} E_f(k) \left(1 - \frac{k}{m}\right)^t, \end{aligned} \tag{4.3.5}$$

where I is the identity matrix and we used $(\mathrm{I}-\Pi)^2 = (\mathrm{I}-\Pi)$ as well as $\Pi P = P\Pi = \Pi$. Recall that $(\Pi f)(\mathbf{x}) = \mathbb{E}[f]$ for all $\mathbf{x}$, hence Π is the orthogonal projection in $(\mathbb{R}^{|\Omega|}, \langle \cdot, \cdot \rangle_p)$ onto the constant functions and consequently $\mathrm{I} - \Pi$ the orthogonal projection in the same inner product space onto the orthogonal complement of constant functions, which correspond to functions f that have $\mathbb{E}[f] = 0$. Moreover, we defined the *energy spectrum* E_f, following [90], by

$$E_f(k) \equiv \sum_{S \subseteq [m], |S|=k} \hat{f}(S)^2, \quad k \in [m].$$

We remark that the above expression for $C_f(t)$ reduces to Parseval's theorem for $t = 0$. Finally, we can express the integrated autocorrelation time $\tau_{\mathrm{int},f}$ of a function f in terms of the Fourier-Walsh coefficients or equivalently the energy spectrum:

$$\begin{aligned} \tau_{\mathrm{int},f} &= \frac{1}{2} + \sum_{t=1}^{\infty} \frac{C_f(t)}{C_f(0)} \\ &= \frac{1}{2} + \sum_{t=1}^{\infty} \sum_{k=1}^{m} \left(\frac{E_f(k)}{\sum_{\ell=1}^{m} E_f(\ell)} \left(1 - \frac{k}{m}\right)^t\right), \\ &= \frac{1}{2} + \sum_{k=1}^{m} \left(\frac{E_f(k)}{\sum_{\ell=1}^{m} E_f(\ell)} \sum_{t=1}^{\infty} \left(1 - \frac{k}{m}\right)^t\right), \\ &= \frac{1}{2} + \sum_{k=1}^{m} \left(\frac{E_f(k)}{\sum_{\ell=1}^{m} E_f(\ell)} \left(\left(\frac{m}{k} - 1\right)\right)\right) \\ &= m \sum_{k=1}^{m} \left(\frac{E_f(k)}{\sum_{\ell=1}^{m} E_f(\ell)} \frac{1}{k}\right) \left(- \frac{1}{2}\right). \end{aligned}$$

Let us now illustrate how the above spectral-formalism can be applied to the number of open edges, $N(\mathbf{x}) = \sum_{i=1}^{m} x_i$. Besides being an illustrative example, it also provides us with an understanding why $\tau_{\mathrm{int},N} = \tau_{\mathrm{exp},N} = \tau_{\mathrm{exp}}$. To begin with, we need to determine the Fourier-Walsh coefficients $\hat{N}(S)$:

$$\hat{N}(S) = \sum_{i=1}^{m} \mathbb{E}[x_i \chi_S] = \mathbb{1}_{\{|S|=1\}} \sqrt{p(1-p)} + pm \mathbb{1}_{\{S=\varnothing\}}.$$

This immediately yields

$$\begin{aligned} N &= pm + \sqrt{p(1-p)} \sum_{i=1}^{m} \chi_{\{i\}}, \\ \mathrm{Var}_p[N] &= mp(1-p), \\ E_N(k) &= \mathbb{1}_{\{k=1\}} mp(1-p), \\ \rho_N(t) &= \left(1 - \frac{1}{m}\right)^t = e^{-t/\tau_{\mathrm{exp}}}, \\ \tau_{\mathrm{int},N} &= m - \frac{1}{2}. \end{aligned}$$

A few comments are in order. Firstly, the expectation is trivially $\mathbb{E}[N] = pm$, and the variance follows also directly from the independence of the model. Secondly, the energy spectrum is localised on $k = 1$, which is precisely on eigenfunctions of the form $\chi_{\{i\}}$ and corresponds to the second largest eigenvalue, and hence to τ_{exp}. A direct consequence is that the normalised autocorrelation function $\rho_N(t)$ is a pure exponential with "scale" τ_{exp}. This in turn implies that $\tau_{\mathrm{int},N} \sim \tau_{\mathrm{exp},N} = \tau_{\mathrm{exp}}$. In general, we expect an observable with an energy spectrum strongly localised on small values of k around $k = 1$, to have a normalised autocorrelation function very close to a pure exponential and an integrated autocorrelation time approximately equal to the corresponding exponential autocorrelation time. Lastly, we emphasise that this result is independent of p. When we consider $p = 1/2$ and $\mathbb{Z}_L^2$, we can equivalently summarise the above in terms of the associated dynamical critical exponents

$$z_{\mathrm{exp},\mathcal{N}} = z_{\mathrm{int},\mathcal{N}} = z_{\mathrm{exp}} = 0. \tag{4.3.6}$$

Now returning to $q > 1$, it seems plausible to expect that the correlations have to be at least as large as for $q = 1$ (the contrary seems unphysical). Hence, one naturally generalises that one has $z_{\mathrm{exp},\mathcal{N}}, z_{\mathrm{int},\mathcal{N}} \geq 0$ for $q \geq 1$.

Furthermore, our observation that $\mathcal{N}$ is a "good" probe for the slowest mode when $q = 1$, should also generalise to $q > 1$, albeit not necessarily as "perfect" as for $q = 1$. Indeed, the authors of [21] numerically studied the Sweeny dynamics on $\mathbb{Z}_L^2$ for $q \neq 1$, and found in particular for all values of q studied in the regime $(0, 3.5]$ that the autocorrelation function is very close to a pure exponential, and conclude that $z_{\mathrm{int},\mathcal{N}} \approx z_{\mathrm{exp}}$ (the former statement was also observed by us in [53] for the Metropolis dynamics and additionally for $q = 4$). Moreover, it is conceivable that at least in the regime $1 < q < 2$ (on $\mathbb{Z}_L^2$) the eigenvalue spectrum is only weakly perturbed from the case $q = 1$, possibly only causing poly-logarithmic deviations in the spectrum. However, albeit being physically plausible, to our knowledge there has been no rigorous underpinning of similar observations, to date.

Here we establish a new rigorous result and prove that $\tau_{\mathrm{exp},\mathcal{N}} = \tau_{\mathrm{exp}}$ for the heat-bath chain for the random-cluster model in the regime $q \geq 1$. In fact we prove a more general statement that extends to monotone heat-bath chains such as the Glauber dynamics for the Ising model [48]. We note already here, that the regime $q < 1$ seems notoriously

hard to treat analytically/rigorously. This mainly follows from the non-validity of the Fortuin-Kasteleyn-Ginibre (FKG) inequality. Further, it is conjectured that certain aspects in the regime $q < 1$ show negative association, opposed to the positive association property, which is closely related to the FKG inequality [27, 91]. However, the numerical studies [21, 53] indicate that $z_{\text{int},\mathcal{N}} = 0$ also for $q < 1$.

4.4 Projection Lemma and proof of $\tau_{\exp,\mathcal{N}} = \tau_{\exp}$ for the heat-bath chain for $q \geq 1$.

In this section we prove that the bond-density observable $\mathcal{N}$ is indeed a good probe for the slowest mode of the heat-bath chain, i.e. we show that $\mathcal{N}$ has positive projection onto an increasing eigenfunction corresponding to the second-largest eigenvalue, of the transition matrix P of the heat-bath chain, for $q \geq 1$. We actually prove the slightly more general statement for strictly increasing functions. By projection, we refer, as in (4.3.5), to a spectral expansion of the normalised autocorrelation function, that is one has

$$\rho_{\mathcal{N}}(t) = \sum_{j=2}^{|\Omega|} \left(\frac{\mathbb{E}_{p,q}[\mathcal{N} f_j]^2}{\sum_{k=2}^{|\Omega|} \mathbb{E}_{p,q}[\mathcal{N} f_k]^2} \right) \lambda_j^t, \tag{4.4.1}$$

where $\mathbb{E}_{p,q}[\cdot]$ denotes the expectation with respect to the (stationary) random-cluster model measure for bond-density p, cluster weight q, and in this section $\{f_j\}_{j=1}^{|\Omega|}$ is a complete and orthonormal set of eigenfunctions of the transition matrix of the corresponding heat-bath chain. Furthermore, the eigenfunctions are sorted in decreasing order with respect to their eigenvalues[4] $\{\lambda_j\}_{j=1}^{|\Omega|}$.

Before we prove our Projection Lemma, recall that function $f : \Omega \to \mathbb{R}$ is increasing, if $f(A) \leq f(B)$ whenever $A \subseteq B$. Analogously, we call a function *strictly* increasing if $f(A) < f(B)$ whenever $A \subset B$. In order to prove the desired Projection Lemma we need a result originally used in [92] for the Ising heat-bath dynamics, which states that the second largest eigenvalue has an increasing eigenfunction for $q \geq 1$. The generalisation to the heat-bath dynamics for the monotone random-cluster model ($q \geq 1$) is straightforward, however we provide the details of the proof for completeness.

Lemma 4.4.1. *(Peres in [92], Lemma 3). The second largest eigenvalue λ_2 of the transition matrix P of the heat-bath chain for random-cluster model on a finite graph with cluster weight $q \geq 1$ and bond-density $0 \leq p \leq 1$ has an increasing eigenfunction.*

Proof. Firstly, we need the result that when f is increasing then Pf is also increasing. To show this, note for $A \subseteq B$ we have

$$\left(Pf\right)\left(A\right) \quad = \quad \sum_{C \in \Omega} \left(P(A, C) f(C) \right.$$

[4] In what follows we sometimes use the parametrisation $\lambda_j = e^{-1/\tau_j}$. Recall, it can be shown that all eigenvalues of heat-bath chains are non-negative, c.f. [45].

$$\begin{aligned} &= \mathbb{E}_{P(A,\cdot)}[f] \\ &\leq \mathbb{E}_{P(B,\cdot)}[f] \\ &= \big(Pf\big)(B). \end{aligned}$$

The inequality $\mathbb{E}_{P(A,\cdot)}[f] \leq \mathbb{E}_{P(B,\cdot)}[f]$ for any increasing function f can be easily shown via a coupling [27]. The monotonicity for $q \geq 1$ allows for the construction of a coupling (C_1, C_2) of $P(A,\cdot)$ and $P(B,\cdot)$ for any $A \subseteq B$ such that $\mathbb{P}[C_1 \leq C_2] = 1$, i.e. it preserves the order between A and B, see for instance [27]. Denote the corresponding probability measure of (C_1, C_2) by $Q_{A;B}(C_1, C_2)$. We therefore obtain

$$\begin{aligned} \mathbb{E}_{P(A,\cdot)}[f] &= \sum_{C\in\Omega} P(A,C) f(C) \\ &= \sum_{C\in\Omega} \Big(\sum_{D\in\Omega} Q_{A;B}(C,D) \Big) f(C) \\ &= \sum_{\substack{C,D\in\Omega \\ C\subseteq D}} Q_{A;B}(C,D) f(C) \\ &\leq \sum_{\substack{C,D\in\Omega \\ C\subseteq D}} Q_{A;B}(C,D) f(D) \\ &= \sum_{D\in\Omega} f(D) \Big(\sum_{C\in\Omega} Q_{A;B}(C,D) \Big) \\ &= \mathbb{E}_{P(B,\cdot)}[f]. \end{aligned}$$

The above demonstrates that Pf is increasing whenever f is increasing, for $q \geq 1$. Now, let f be any increasing function with $\mathbb{E}_{p,q}[f] = 0$. Furthermore, recall that our spectral discussion in Chapter 2 showed that we can write $f = \sum_{j=2}^{|\Omega|} q_j f_j$, where $q_j \equiv \mathbb{E}_{p,q}[f_j f]$ is the projection of f onto the j'th eigenfunction f_j (this is valid for any function $f : \Omega \to \mathbb{R}$, and not restricted to increasing functions). Iterating this yields

$$\frac{1}{\lambda_2^t} P^t f = \sum_{j=2}^{|\Omega|} q_j \left(\frac{\lambda_j}{\lambda_2}\right)^t f_j \to g \equiv \sum_{j=2}^{|\Omega|} q_j f_j \mathbb{1}_{\{\lambda_j = \lambda_2\}}, \tag{4.4.2}$$

for $t \to \infty$, where we explicitly used that the heat-bath chain for the random-cluster model has no non-negative eigenvalues, hence $1 \geq \lambda_j/\lambda_2 \geq 0$ for $j \geq 2$, c.f. [45] or Chapter 2. The crucial point is that because f increasing implies Pf increasing, one also has, by induction, that $P^t f$ is increasing for $t \geq 1$, which in particular also holds in the limit $t \to \infty$, thus g is increasing. Now the limit $P^t f$ being a linear combination of eigenfunctions with corresponding eigenvalue λ_2, implies that it is also an eigenfunction of eigenvalue λ_2, as long as g is not zero. This can be assured by choosing f non-orthogonal to at least one eigenfunction of λ_2. To see that it is indeed possible to find such an increasing function f, consider $f = f_2 + C\big(N - \mathbb{E}_{p,q}[N]\big)$, where N is the number of open edges, i.e. $N(A) = \sum_{e\in E} \mathbb{1}_{\{e\in A\}}$ for $A \subseteq E$. It is clear that $\mathbb{E}_{p,q}[f] = 0$ and $q_2 \neq 0$, because the f_j's are an

orthonormal basis in the inner product space $(\mathbb{R}^{|\Omega|}, \langle\cdot,\cdot\rangle_{p,q})$. Additionally we claim that we can find $C > 0$ sufficiently large, such that f is increasing, independent of whether f_2 is increasing or not. To begin with, define

$$\Delta \equiv \Delta_f \equiv \min_{A,B\in\Omega, A\subset B} f_2(B) - f_2(A),$$

and consider first the case $\Delta \geq 0$ with arbitrary $A, B \in \Omega$ such that $A \subset B$:

$$f(B) - f(A) = f_2(B) - f_2(A) + C\left(\sum_{e\in E}\left(\mathbb{1}_{\{e\in B\}} - \mathbb{1}_{\{e\in A\}}\right)\right) \geq \Delta + C > 0,$$

which is valid for any $C > 0$. The first inequality follows from the definition of Δ and the fact that $\sum_e \mathbb{1}_{\{e\in\cdot\}}$ is a strictly increasing function, that is it differs by at least 1 for any $A \subset B$. For the case $\Delta < 0$, we choose $C > -\Delta > 0$, and conclude likewise that f is increasing. ■

The next Lemma states our main result, which is based on ideas of [92]:

Lemma 4.4.2. *(Projection Lemma). Let P be the transition matrix of the heat-bath Markov chain for the random-cluster model on a finite graph with $0 \leq p \leq 1$ and $q \geq 1$, and let g be an increasing eigenfunction corresponding to the second largest eigenvalue. Then for any strictly increasing function f*

$$\mathbb{E}_{p,q}[fg] > 0.$$

Proof. To start with, we note that if f is strictly increasing and g is increasing, one has for some $\alpha > 0$ sufficiently small, that also $f - \alpha g$ is strictly increasing. To see this, note that for finite Ω, one can simply choose $\alpha > 0$ such that $f(B) - f(A) > \alpha\big(g(B) - g(A)\big)$ for all $A, B \in \Omega$ such that $A \subset B$, see also [92].
Additionally we need the Fortuin-Kasteleyn-Ginibre (FKG) inequality [27]: It is well known that for $q \geq 1$ the random-cluster model the FKG inequality holds, stating that for two increasing functions f, g one has $\mathbb{E}_{p,q}[fg] \geq \mathbb{E}_{p,q}[f]\mathbb{E}_{p,q}[g]$. The proof of the above Lemma yields a construction of g which in particular shows that $\mathbb{E}_{p,q}[g] = 0$, which in turn implies

$$\begin{aligned}
\mathbb{E}_{p,q}[(f-\alpha g)g] &\geq \mathbb{E}_{p,q}[f-\alpha g]\mathbb{E}_{p,q}[g] \\
\mathbb{E}_{p,q}[fg] - \alpha\mathbb{E}_{p,q}[g^2] &\geq \mathbb{E}_{p,q}[f]\mathbb{E}_{p,q}[g] - \alpha\mathbb{E}_{p,q}[g]^2 \\
\mathbb{E}_{p,q}[fg] - \alpha\mathbb{E}_{p,q}[g^2] &\geq 0
\end{aligned}$$

which further implies

$$\mathbb{E}_{p,q}[fg] \geq \mathbb{E}_{p,q}[g^2]\alpha > 0,$$

note $\mathbb{E}_{p,q}[g^2] = \sum_j \mathbb{E}_{p,q}\mathbb{1}_{\{\lambda_j=\lambda_2\}} q_j^2 > 0$, as it projects on at least one (normalised) eigenfunction of eigenvalue λ_2, by construction. ■

This clearly applies to the bond-density $\mathcal{N}$, being manifestly strictly increasing. This allows us to show that the (normalised) autocorrelation function of $\mathcal{N}$ is for large t equivalent to a constant times $e^{-t/\tau_{\exp}}$, more precisely:

Corollary 4.4.3. *The normalised autocorrelation function for the bond-density $\mathcal{N}$ has the following asymptotics*

$$\rho_{\mathcal{N}}(t) \sim W_2(\mathcal{N}) e^{-t/\tau_{\exp}} \quad \text{for } t \to \infty,$$

where $W_2(\mathcal{N}) > 0$ is a constant.

Further,

$$\tau_{\exp,\mathcal{N}} = \tau_{\exp}.$$

Proof. Denote by $\mathcal{B} = \{f_j\}_{j=1}^{|\Omega|}$ the above orthonormal eigenbasis of P. Pick an eigenfunction of $\mathcal{B}$ that corresponds to the second-largest eigenvalue λ_2 and for which g has non-zero projection, denote it by f_k (hence $q_k > 0$). Write $\mathcal{J} \equiv [m] \setminus \{k\}$, where m is the number of edges in G, as well as $\mathcal{S}_2 \equiv \{i \in [m] : \lambda_i = \lambda_2\}$ and construct $\mathcal{B}' \equiv \{f_j | j \in \mathcal{J}\} \cup \{\tilde{g}\}$, where $\tilde{g} = g/\sqrt{\mathbb{E}_{p,q}[g^2]}$. Now, one can verify that $\mathcal{B}'$ is a normalised eigenbasis of P (not necessarily orthogonal), hence we have

$$\mathcal{N} = \mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]\tilde{g} + \sum_{j\in\mathcal{J}} \mathbb{E}_{p,q}[f_j\mathcal{N}]f_j.$$

Recall that we have for $t \geq 1$

$$\rho_{\mathcal{N}}(t) = \frac{C_{\mathcal{N}}(t)}{\mathrm{Var}_{p,q}[\mathcal{N}]} = \frac{\langle \mathcal{N}, (P-\Pi)^t \mathcal{N}\rangle_{p,q}}{\langle \mathcal{N}, (\mathrm{I}-\Pi)\mathcal{N}\rangle_{p,q}} \tag{4.4.3}$$

Furthermore one has

$$\begin{aligned}
\langle \mathcal{N}, (P-\Pi)^t \mathcal{N}\rangle_{p,q} &= \lambda_2^t\, \mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 + \sum_{j\in\mathcal{J}\setminus\{1\}} \lambda_j^t\, \mathbb{E}_{p,q}[f_j\mathcal{N}]^2. \\
\langle \mathcal{N}, (\mathrm{I}-\Pi)\mathcal{N}\rangle_{p,q} &= \mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 + \sum_{j\in\mathcal{J}\setminus\{1\}} \mathbb{E}_{p,q}[f_j\mathcal{N}]^2.
\end{aligned}$$

Now, inserting this into (4.4.3) and using $\lambda_j = e^{-t/\tau_j}$ and $\lambda_2 = e^{-t/\tau_{\exp}}$ we obtain

$$\begin{aligned}
\rho_{\mathcal{N}}(t) &= \frac{\mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 e^{-t/\tau_{\exp}} + \sum_{j\in\mathcal{J}\setminus\{1\}} \mathbb{E}_{p,q}[f_j\mathcal{N}]^2 e^{-t/\tau_j}}{\mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 + \sum_{j\in\mathcal{J}\setminus\{1\}} \mathbb{E}_{p,q}[f_j\mathcal{N}]^2} \\
&\sim e^{-t/\tau_{\exp}} \left(\frac{\mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 + \sum_{j\in\mathcal{S}_2\setminus\{k\}} \mathbb{E}_{p,q}[f_j\mathcal{N}]^2}{\mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]^2 + \sum_{j\in\mathcal{J}\setminus\{1\}} \mathbb{E}_{p,q}[f_j\mathcal{N}]^2} \right),
\end{aligned} \tag{4.4.4}$$

where the term in the parenthesis equals $W_2(\mathcal{N})$, which is positive because all terms are non-negative and at least $\mathbb{E}_{p,q}[\tilde{g}\mathcal{N}]$ is positive, by the Projection Lemma. The second part

follows directly from (4.4.4) in combination with the standard relation, c.f. [44],

$$\tau_{\mathrm{exp},\mathcal{N}} = \limsup_{t\to\infty} \frac{t}{-\log|\rho_{\mathcal{N}}(t)|}.$$

■

As mentioned previously, this result was implicitly assumed in [15, 53], but albeit being a plausible assumption has to our knowledge not been proven so far. In [15] it was actually observed that $\tau_{\mathrm{int},\mathcal{N}} \sim \tau_{\mathrm{exp},\mathcal{N}}$ and hence $\tau_{\mathrm{int}\mathcal{N}} \sim \tau_{\mathrm{exp}}$, which is supported by the observation that the normalised autocorrelation function of $\mathcal{N}$ is very close to a pure exponential. This suggests that the projections on eigenfunctions with smaller eigenvalue, or equivalently, smaller exponential autocorrelation time, are strongly suppressed in comparison to the projection on eigenfunctions with exponential autocorrelation time τ_{exp}. Further support for this picture comes from the preceding exact analysis of the percolation case, $q = 1$, which concluded with (4.3.6). Thus for the case $q = 1$ one has that $\rho_{\mathcal{N}}(t)$ is precisely a pure exponential. We believe this is an interesting theoretical problem with practical impact and we intend to approach it in a future project.
We remark that the above arguments are not restricted to the monotone random-cluster model, but can naturally be extended to monotone heat-bath chains, for instance the heat-bath chain for the ferromagnetic Ising model (with and without magnetic field). Here the magnetisation $\mathcal{M} = \sum_{i=1}^{n} \sigma_i/n$ can be used as the strictly increasing function, hence $\tau_{\mathrm{exp},\mathcal{M}} = \tau_{\mathrm{exp}}$. We conclude the discussion on the dynamical critical behaviour of Sweeny's algorithm with Figure 4.11, which summarises all our z_{int} estimates of observables $\mathcal{S}_2, C_1, \mathcal{N}$ for some representative values of q in the entire second-order regime $0 < q < 4$. The sharpness of Li-Sokal's bound for $q \geq 1$ is clearly distinctive, where the apparent violation for $q = 4$ is a consequence of a combination of imposing a power-law on the numerical data and known multiplicative logarithmic corrections for $q = 4$ (at least for stationary quantities). A similar "violation" was observed in [93, 94] for Swendsen-Wang dynamics.

4.5 Overall efficiency

The preceding discussion shows that from a purely statistical point of view, i.e. considering correlations of certain standard observables, Sweeny's algorithm seems slightly more favourable than the Swendsen-Wang-Chayes-Machta approach. However, because an elementary update-step in the Sweeny dynamics has computational complexity that increases, in one or the other form, with the system size, it is not enough to compare m Sweeny steps, or one sweep, to one complete Chayes-Machta iteration (activating connected components with subsequent percolation on the active-vertex-induced spanning sub-graph.) Instead, one should consider a measure that quantifies, say, autocorrelation times measured in computer time. This will give MCMC practitioners a way of judging which method produces statistically-independent samples of the desired observable f

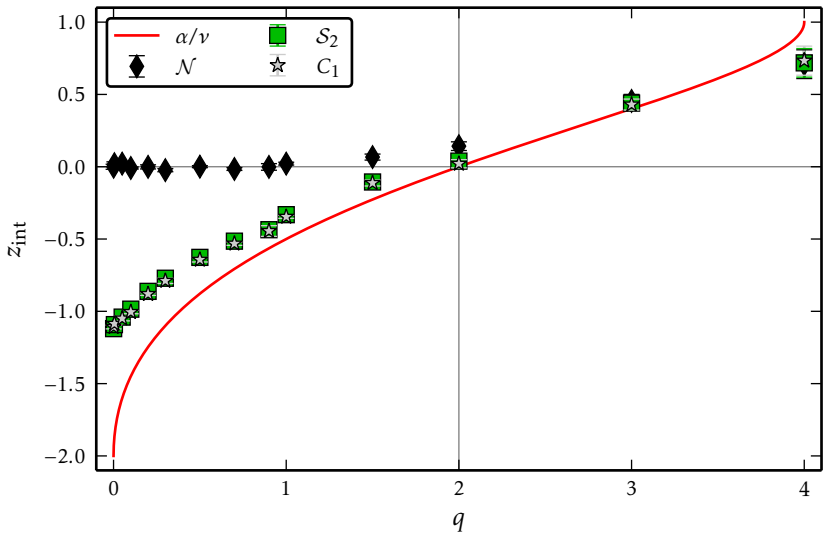


Figure 4.11: *Dynamical critical exponents for the integrated autocorrelation time of observables $\mathcal{S}_2, C_1, \mathcal{N}$ together with the exact value of α/ν, corresponding to the Li-Sokal bound for the Sweeny dynamics. Both, the vertical and horizontal gray line, are only meant to be a visual aid.*

in shorter time. A reasonable efficiency-measure was analysed by us for the Sweeny dynamics (in Metropolis variant) in [53], where we considered, for a given observable f,

$$T_f \equiv \tau_{\text{int},f}\, \bar{t}, \tag{4.5.1}$$

which is the integrated autocorrelation time of f in units of computing time on a specific system. Now the (asymptotically) smaller critical correlations for the Sweeny dynamics on $\mathbb{Z}_L^2$ will only extend to T_f when the computational effort scales sufficiently weak with L. In particular, the poly-logarithmic fully dynamic connectivity algorithm yields, for sufficiently large L, a running time dependence of L that is weak enough to extend the better dynamical critical behaviour to the overall efficiency measure. However, the involved constants might forbid the visibility of this effect for practical system sizes. To check whether such a practical limitation actually occurs, we estimated T and summarised the results in Figure 4.12. Here we show the ratio of T_f for Sweeny dynamics to T_f for Swendsen-Wang-Chayes-Machta (SWCM) dynamics, for the observables $\mathcal{N}$ and $\mathcal{S}_2$. The reason for the decision to consider ratios, is that we expect specific system dependencies of the computing device, used to perform the test, to be asymptotically insignificant, when considering ratios. This follows from the expectation that the involved constants might cancel or at least become asymptotically irrelevant when considering ratios of running times. A few comments to Figure 4.12 are in order. Firstly, as expected, the union-find approach does not yield a more efficient variant than the SWCM algorithm for any value of q considered. Secondly, the involved constants in the DC implementation seem to be too large to observe an decrease of the ratio with L, which should, at least from a theoretical (asymptotic) point of view, be the case for all choices of q considered. Finally and most

surprisingly, the interleaved breadth-first search seems to be the most efficient Sweeny variant for $q = 1, 2, 3, 4$ and shows even an actual decrease of the ratio for $q = 4$, where the exponent $d_F - x_2 = -1/8 < 0$, thus the expected computational effort saturates to a constant thus becoming, in average, asymptotically equivalent to the SWCM approach. Actually, we have $d_F - x_2 \leq 0$ for $q \gtrapprox 3.84$, however the difference between dynamical critical exponents for the relevant observables in Sweeny and SWCM dynamics are too small to see this effect clearly.

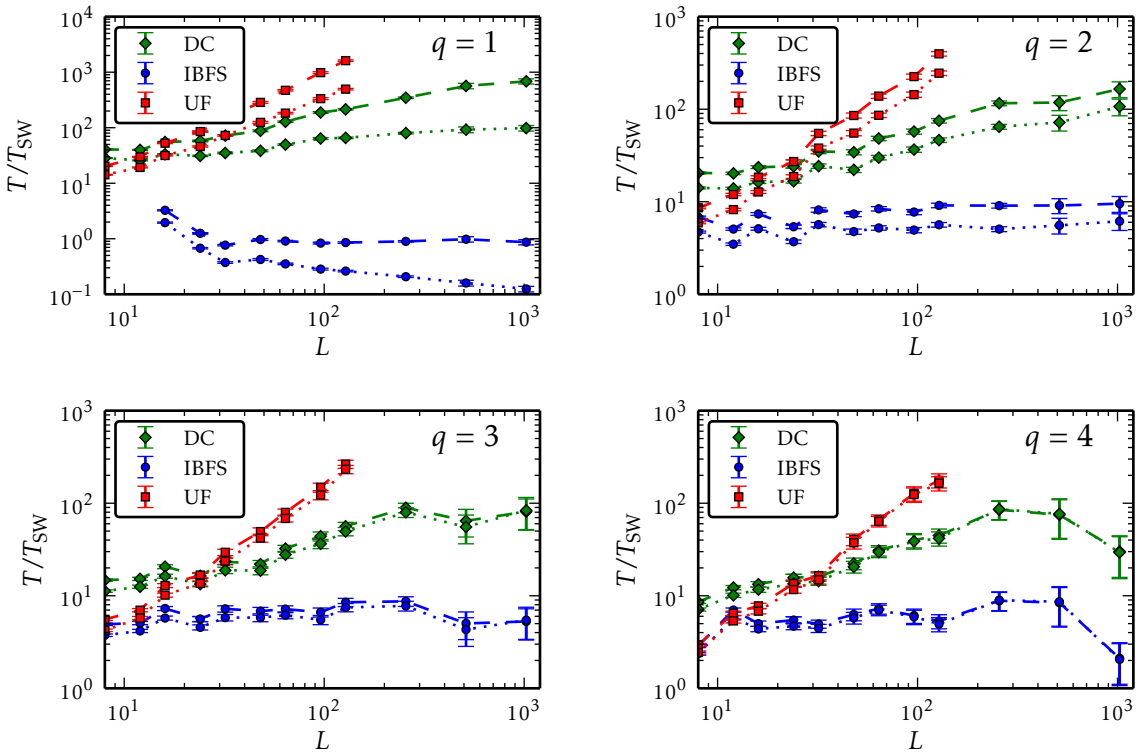


Figure 4.12: *Effective run-times T according to Eq. (4.5.1) for the different implementations relative to the time T_{SW} of the Swendsen-Wang algorithm. Dashed lines correspond to the running time to generate an independent sample of the observable $\mathcal{N}$ and dotted lines to samples of $\mathcal{S}_2$.*

4.6 Algorithmic optimisations at criticality

In this section we describe two heuristic optimisations that turned out to be particularly beneficial for simulating critical models. However, we stress, that neither improves any asymptotic running time scaling, that is the leading size dependence remains in both instances the same. Nevertheless, we show how the involved constants, both regarding to space and time complexity, can be reduced. This in turn allows one to advance to larger system sizes, previously hindered by memory or running time restrictions. We will use the optimisations in this section in a future study of the Sweeny dynamics in three dimensions. Further, we remark that the results are preliminary and a detailed study will be presented elsewhere.

4.6.1 Cluster identifiers and interleaved breadth-first search

We first consider the hybrid union-find and interleaved breadth-first search approach, which suffered from the need of rebuilds of the tree-data structure representing the connected components in the union-find data structure. We provided scaling arguments that showed that at the self-dual point the bridge-deletion operation has a typical running time of order $L^{\gamma/\nu}$. This destroys asymptotically the improved $L^{d_F-x_2}$ running time scaling of all other operations. In order to circumvent this obstacle, we replace the union-find data structure by a data structure, say an array, that assigns a cluster-identifier (CI) to any vertex. Further, we assume the availability of a data structure that is capable of returning the cluster mass, associated to a given CI, in constant time. It is easy to see, that connectivity queries can be answered by comparing CI's, that is two vertices are connected if and only if their CI's equal. Thus in case of edge-insertions, we can efficiently, without any data structure modification, decide whether the edge is pivotal or not. However, in case we decided to insert the edge, which we assume to be also pivotal, that is it merges two clusters, we need to update the CI data structure. More precisely, the components, now joined by the inserted bridge, have differing CI's. Due to the fact that we can determine the cluster mass of a given CI in constant time, we can directly, and only, start an exhaustive breadth-first search at the vertex incident to the smaller cluster that sets the CI's of all vertices encountered to the CI of the larger cluster. Additionaly, we need to update appropriately the data structure mapping CI's to cluster masses. This can be easily done with constant computational effort.
Now, compared to the union-find data structure, we deteriorated the running time of insertions, however compared to the pure interleaved breadth-first search variant, we firstly, improved the determination of the type of an edge to a constant running time which previously demanded a interleaved breadth-first search. Secondly, we halved the running time of any BFS involved, because we did not need an interleaved approach for this to work. Yet, the main advantage comes into the play, when considering deletions. As for the union-find approach, we perform two *simultaneous* breadth-first searches in order to decide whether the edge under consideration is pivotal to the current configuration or not. Both BFS's terminate as soon as their shells overlap, or the smaller is exhausted. In addition to that, both BFS's maintain a list of visited vertices. In the case one BFS terminates without "touching" the other, we conclude that the edge we try to remove is pivotal. In case this deletion is accepted, we simply iterate over all vertices in the list collected by the BFS corresponding to the smaller cluster and assign to each of the vertices the new CI of the newly created cluster. Of course, we also need to update the cluster masses associated to the old CI and add an entry for the newly created CI. This method takes, in both cases, deletion of a bridge and non-bridge, typically $L^{d_F-x_2}$ running time, and is thus perfectly exploits the fractal cluster structure.
We conclude that the above interleaved BFS in combination with cluster-identifiers has typically running time per operation $\bar{t} \approx L^{d_F-x_2}$. In Figure 4.13 we show results from a

numerical comparison of $\bar{t}$ for both, the optimised and plain, versions in 2D. We obtain roughly a factor of two improvement when using the optimised version.

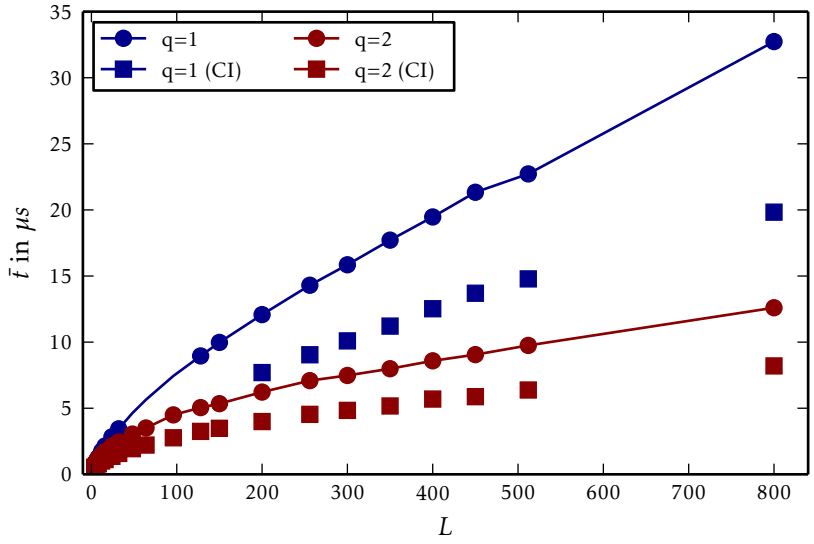


Figure 4.13: *Comparison of expected running times per operation between the cluster-id optimised and plain interleaved breadth-first search variants of Sweeny's algorithm. The running time was averaged over 10^4 steps, and an appropriate relaxation phase was discarded. The parameters p and q were fixed to the self-dual point, and the underlying graph is $\mathbb{Z}_L^2$. Note that the measurements were done on a different system than the previous running time analysis. Thus the overall constants might have changed and a direct comparison to the other results is not possible. The lines are only for visual aid.*

We can also roughly quantify how the different runtime contributions contribute to $\bar{t}$ at the self-dual point, on $\mathbb{Z}_L^2$, in other words we have for sufficiently large L

$$\begin{aligned} \bar{t} \quad &\text{"}\approx\text{"} \quad \mathbb{P}[e \in B] c_1 L^{d_F - x_2} + \mathbb{P}[e \in C] c_2 L^{d_F - x_2} \\ &+ \quad \mathbb{P}[e \in \bar{B}] \frac{c_1}{2} L^{d_F - x_2} + \mathbb{P}[e \in \bar{C}] c_3 L^0 \\ &\sim \quad \left(\frac{3c_1}{4} + \frac{c_2\sqrt{q}}{2(1+\sqrt{q})} \right) L^{d_F - x_2} + \frac{c_3}{2(1+\sqrt{q})} L^0 . \end{aligned}$$

A few comments are in order: The first term, corresponding to $e \in B$, has no dependence on whether the deletion is accepted or not, because in both cases we terminate whenever the smaller cluster is exhausted, and both BFS's collected a list of visited vertices, which has to be emptied in any case (deletion or not). The non-bridge term ($e \in C$) has a different constant c_2, which corresponds to the case where two interleaved BFS's merge, which indeed has $L^{d_F - x_2}$ scaling but not necessarily the same constant. The third term, corresponding to candidate-bridges ($e \in \bar{B}$), has a factor of $c_1/2$, as only one exhaustive BFS on the smaller cluster is needed, opposed to the case $e \in B$. Finally, the case of candidate-non-bridges, $e \in \bar{C}$ can be done with constant computational effort, as the query is constant and inserting e does not need any change in the cluster id data structure.

We also performed a numerical check of the running time improvement in $3D$ at the estimated location of a second-order phase transition. More precisely we considered bond-percolation ($q = 1$) with bond-density $p = 0.24881182$, recently estimated in [95]. Additionally we analysed $q = 2.2$ with $p = 0.4677$, corresponding to a second-order phase transition in $3D$ [96]. Figure 4.14 shows the corresponding numerical comparison. As for the $2D$ case these results indicate a speed-up of a factor of roughly 2, which is remarkable keeping in mind the simplicity of the approach.

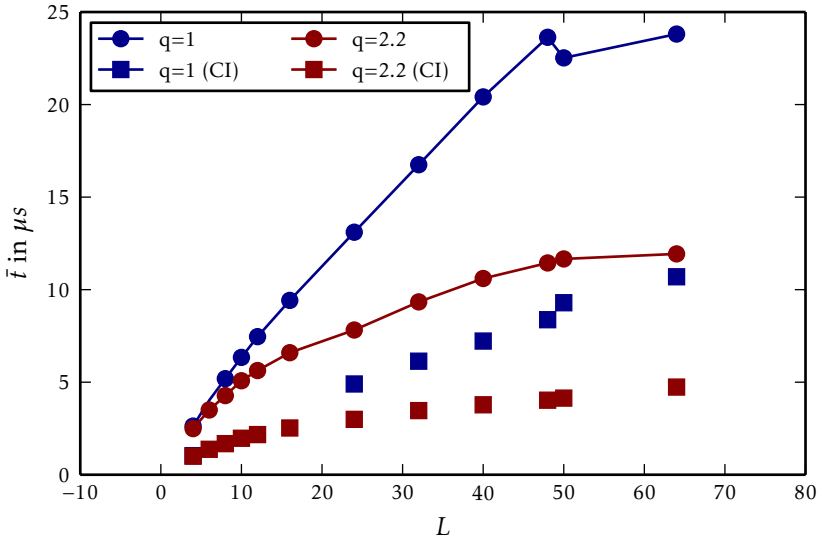


Figure 4.14: *Expected running time comparison for the BFS implementation, with and without cluster-id data structure. The underlying graph is $\mathbb{Z}_L^3$ and the value of p, for the two choices of q, was fixed to the corresponding conjectured (estimated) location of a second-order phase transition. See main text.*

4.6.2 One-level dynamic connectivity algorithm

Next, we consider the dynamic-connectivity algorithm approach to Sweeny's algorithm. From the general discussion in section 3.5 it is clear that the level hierarchy is essential in that it precisely assures that a given edge does not carry too much computational cost when amortising the cost of the entire operation sequence. However, having a memory requirement that, in addition to the typical linear scaling in the volume (number of edges), increases logarithmically with the volume, clearly complicates the advance towards large system sizes, in practise $L \geq 2048$.

Now, one might wonder whether the fractal structure of critical Fortuin-Kasteleyn cluster is dense enough for the connectivity algorithm to be able to hierarchically distribute the connected components over the entire $O(\log(n))$ levels. Recall that, following the deletion of a tree edge, all tree edges incident to the smaller of the two (temporarily) different trees are also added to the next higher level. This intuitively moves edges in denser regions to higher levels. Additionally we find that typically the smaller fragment has asymptotically a vanishing fraction of the original cluster mass.

We therefore studied how edges are typically distributed over the level hierarchy and find that for all system sizes considered here ($L \leq 1024$), that the majority of all edges resides in levels weakly increasing with the number of vertices. We hence set the number of levels to 1, that is the algorithm operates completely without the edge level hierarchy. We show in Figure 4.15 a running time comparison between the "one-level" heuristic and standard version of the DC algorithm.
Remarkably, our numerical results confirm that for all three test cases, corresponding to the regimes $q < 1$ and $q > 1$ as well as the marginal case $q = 1$, the heuristics is useful and efficient, that is it speeds up the algorithm at least by a factor of 2. We emphasise that this comes in addition to the reduction of the space complexity, i.e. the one-level version maintains only one instance of the data structure and not $\log(n)$. However, we are not able to exclude the possibility that the optimised variant still has a worse than poly-logarithmic computational complexity in the setting of critical random-cluster models. In other words, it is also conceivable that the overall effect of operating with only one level is a drastic reduction of the involved constants in addition to a possibly polynomial computational complexity. We therefore revert final conclusions to a more careful (analytical) future study in the mean-field setting, that is for the random-cluster model on the complete graph [97].

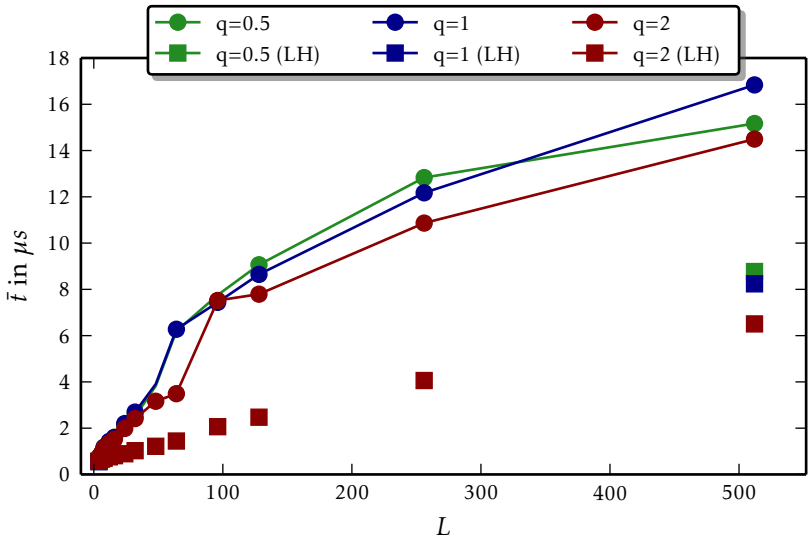


Figure 4.15: *Running time comparison of the DC implementation with and without the "one-level" heuristics.*

Our study of algorithmic optimisations revealed a further interesting structural property of critical random-cluster models (presumably also beyond $\mathbb{Z}_L^2$) and concerns the number of non-tree edges in the smaller of two disconnected clusters, attached to two neighbouring vertices. One reason for considering this quantity is the difference in approaches to finding a reconnecting edge between a BFS or DFS approach and the DC algorithm. We already mentioned that the DC algorithm can iterate between non-tree edges with at most

logarithmic (in n) computational effort, independent of their graph-theoretic distance in the actual spanning sub-graph. On the other hand, we do not have this guarantee in traversal based methods, which in the worst-case need to iterate over a huge part of the edges belonging to the cluster. Now, following a cluster break-up, the fragments are typically very unequal in size, in other words the mass of the smaller cluster is strongly suppressed with respect to the mass of the larger cluster. A natural question concerns the non-tree edges in the smaller fragment. Intuitively, we expect this to be very likely a branch or dangling end [55] with few cycles. In such cases the DC algorithm setup would clearly benefit, even without the level hierarchy, as it can efficiently iterate over those few non-tree edges with logarithmic cost, whereas the BFS method in average will still need time linear in the smaller fragment size, i.e. $L^{d_F-x_2}$. In order to confirm this geometric picture, we estimated the expected number of non-tree edges incident to the smaller of the two clusters attached to two neighbouring vertices. More precisely, let $T(A)$ be the number of tree-edges of A and $\overline{T}(A)$ the number of non-tree edges, thus $T(A)+\overline{T}(A)=|A|$ as well as $|C(A)|\geq\overline{T}(A)$ (recall that $C(A)\subseteq A$ is the set of non-bridges in A). We note that $T(A)=n-K(A)$ and $\overline{T}(A)=|A|-n+K(A)$, where $\overline{T}$ is also known as the *circuit rank* or *cyclomatic number* of A [69]. Further, we naturally extend $T,\overline{T}$ to connected components in (V,A), that is we write $T(C),\overline{T}(C)$ for the number of tree and non-tree edges in the connected component C of the spanning sub-graph (V,A), respectively. Now, for two disconnected vertices $x,y\in V$ in (A,V), such that $(x,y)\in E$, we estimated

$$\Theta\equiv\frac{1}{m}\sum_{(x,y)\in E}\mathbb{1}_{x\nleftrightarrow y}\overline{T}\Big(C_{\text{argmin}\{|C_x|,|C_y|\}}\Big), \tag{4.6.1}$$

where $C_{\text{argmin}\{|C_x|,|C_y|\}}$ is the smaller of the two components C_x and C_y, i.e. Θ is an improved estimator for the number of non-tree edges in the smaller of two clusters attached to two neighbouring, disconnected, vertices. In Figure 4.16 we show our numerical estimates of the size dependence of $\mathbb{E}[\Theta]$ (the expectation for the self-dual random-cluster model on $\mathbb{Z}_L^2$ with cluster weight q). Indeed, as expected, the number of non-tree edges, albeit increasing with L, is very small, less than 10 in expectation, for system sizes up to $L=1024$. Note for decreasing q the expected number of non-tree edges in the smaller component becomes larger. On a first view this seems paradox having in mind the smaller number of non-bridges for small q. Yet one needs to remember that the expected size of the smaller of two neighbouring vertices, ruled by a d_F-x_2 scaling, becomes actually larger with decreasing q, because d_F-x_2 is increasing with decreasing cluster weight. However, we expect the relative density of non-tree edges to decrease with q. Indeed, consider Figure 4.17 which shows the ratio of expected number of non-tree edges to the expected number of vertices in the smaller cluster, and confirms the expected decrease with q. The ratio further approaches a constant, consistent with a $L^{d_F-x_2}$ scaling of $\mathbb{E}[\Theta]$, that is $\mathbb{E}[\theta]\approx L^{d_F-x_2}$, however with a strongly suppressed constant compared to $\mathbb{E}[C_{\min,2}]$ [15]. We will conduct a careful analysis of the optimised DC and BFS approach in detail elsewhere, and report on a preliminary result. The general picture that emerges is that

the optimised DC variant is superior for $q < 2$, whereas for $q \approx 2$ the two variants become comparable and for q beyond 2 the traversal based method becomes more efficient. The reason for this transition is so far not clear to us, and we hope that in a "simplified" setting, such as e.g. the complete graph, the analysis becomes more transparent.

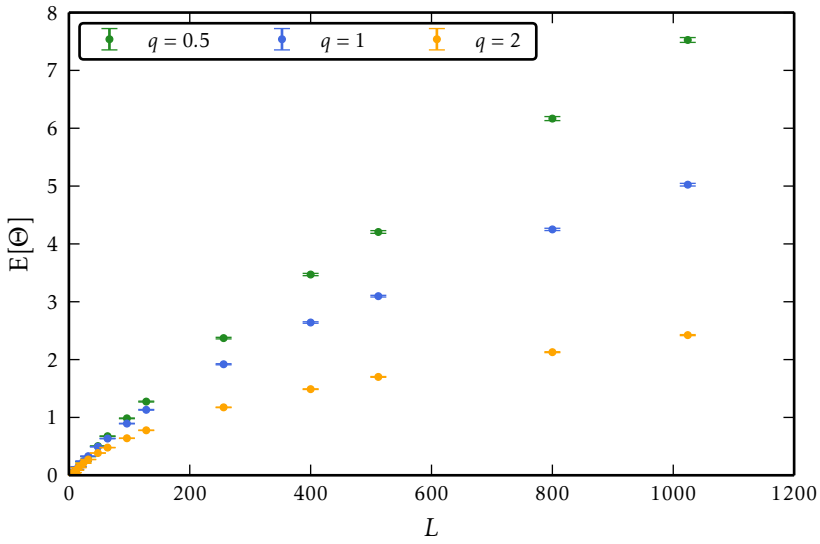


Figure 4.16: *Expected number of non-tree edges in the smaller cluster, $\mathbb{E}[\Theta]$, for three characteristic values of q. Remarkably, there are only a few non-tree edges, supporting the picture of smaller fragments being almost-trees.*

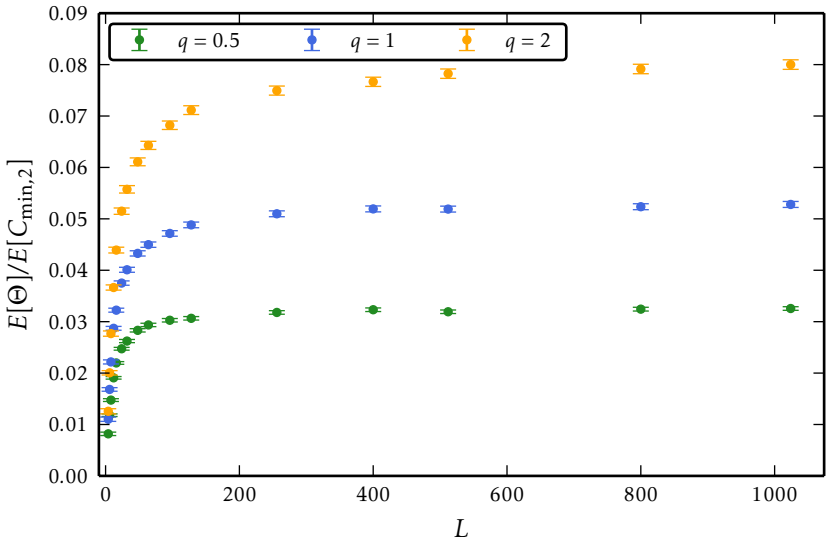


Figure 4.17: *Ratio of expected number of non-tree edges and the expected number of vertices in the smaller cluster. Note that expected total number of edges in the smaller cluster equals $\mathbb{E}[C_{\min,2} - 1] + \mathbb{E}[\Theta]$.*

Chapter 5

Fragmentation of Fortuin-Kasteleyn clusters

We have seen in chapters 2 and 4 how stationary structural properties of the random-cluster model, such as the density of bridges, the typical fragment size, induced by the removal of a bridge, or the specific heat of the Potts model, crucially influence, both, stationary dynamical and computational efficiencies. For instance, the Li-Sokal bound relates specific-heat divergences to the critical slowing-down of energy-like observables which in turn lifts the lower bound of the overall exponential autocorrelation time. On the computational side we have seen that the exponent $d_F - x_2$ determines the leading system size scaling of optimised traversal based algorithms. We hence investigate in this chapter related structural properties and derive a new[1] relationship between the (expected) bridge- and edge-density for the random-cluster model on any graph. We discuss in detail both asymptotic and size dependent effects and establish some interesting generic symmetries of the random-cluster model, which we in turn exploit to relate previously unrelated quantities. Moreover we study the fluctuations in the number of bridges by means of the variance, and reveal a surprising finite size scaling for two dimensional critical anti-monotone ($q < 1$) random-cluster models, as a consequence of the interplay between bridges and non-bridges. We continue with a detailed analysis of one-step, or stationary, fragmentation effects, both for edge and vertex fragmentation, which provides a crucial foundation for the analysis of dynamic or iterated fragmentation in Chapter 7. Further, the results in this chapter are also of independent interest, as in many cases our results do not only extend previous studies, but also provide an alternative explanation for various observed phenomena, such as the absence of finite-size corrections of particular quantities in the percolation model.

[1] **Note added in proof**: The author found out about the work [98], which contains implicitly the bridge-edge identity, however derived by different means.

5.1 Bridge-edge identity

We derive the relationship between the density of bridges and edges by utilising a powerful tool, independently discovered by G. Margulis and L. Russo, and commonly referred to as Russo-Margulis formula. As the formula is rather unknown in the statistical physics community, and also, as we believe, of use in other studies, we first devote some time to deriving it. The Russo-Margulis formula has been useful analytical tool in the study of "threshold phenomena". Such phenomena, besides their apparent connection to phase transitions in physics, occur in general in many probabilistic models reaching from pure probability theory, over computer science, to seemingly unrelated fields as economics and political sciences [99]. In essence, a threshold phenomenon occurs when the probability of an event changes from being close to 0 to close to 1 in a very narrow window. The probably most popular example is the phase transition of the bond percolation model on the square lattice $\mathbb{Z}^2$ for $p_c = 1/2$. Here one has exponentially decaying correlations for any $p < 1/2$ and for $p > 1/2$ the configuration percolates, that is $\mathbb{P}[0 \leftrightarrow \infty] \geq (p - p_c)/[p(1 - p_c)]$. See [100] for a particular short and elegant proof of this sharp threshold result (which also utilises explicitly the Russo-Margulis formula). Furthermore an adapted variant of the Russo-Margulis formula has played an important role in the recent rigorous study of the random-cluster model phase-transition on $\mathbb{Z}^2$ [28]. Here the authors proved the long-standing (well supported) conjecture that the self-dual point $p_{\mathrm{sd}}(q) = \sqrt{q}/(1 + \sqrt{q})$ is indeed the critical point for the random-cluster model on the square-lattice for $q \geq 1$.
Yet, our application of Russo's formula is more moderate and not directly related to a threshold phenomenon, as we use it merely to extract the desired relationship between the expected densities of bridges and edges in the random-cluster model. We first derive Russo's Formula in the setting of independent bond-percolation on a finite graph $G = (V, E)$ with $m \equiv |E|$ and show then how it can also be applied to the random-cluster model. The following derivation of the Russo-Margulis formula resembles [25]. To start with, we need some notation. Firstly, write $E = \{e_1, e_2, \ldots, e_m\}$ and denote by $\mathbb{P}_p[\cdot], \mathbb{E}_p[\cdot]$ the probability measure and expectation, respectively, of the bond-percolation model with parameter p. In the following we will also consider a percolation model where each edge e_j has a designated "activation" probability p_j, and we write $\mathbf{p}$ for corresponding m-dimensional vector. Now, the Russo-Margulis formula states that for any function $X : \Omega \to \mathbb{R}$,

$$\frac{d}{dp}\mathbb{E}_p[X] = \sum_{e \in E} \mathbb{E}_p[\delta_e X], \tag{5.1.1}$$

where $(\delta_e X)(A) \equiv X(A^e) - X(A_e)$ for $A \in \Omega$ is called the *derivative* of X at e and we define $A^e \equiv A \cup \{e\}$ and $A_e \equiv A \setminus \{e\}$. Further, $\mathbb{E}_p[\delta_e X]$ is called the *influence* of e on X [25, 27]. The idea of the proof of (5.1.1) is the construction of a coupling of two percolation models with parameters $\mathbf{p}$ and $\mathbf{q}$, where $q_i = p_i$ for all $i \neq j$ and $q_j = p_j + \epsilon$ with $\epsilon \geq 0$. This defines the two percolation models on a joint probability space, which in turn allows us to consider the difference of X, evaluated at the two corresponding random-percolation

configurations, as *one* random variable. Then we analyse its expectation and use linearity of the expectation to obtain the difference between expectations of X for the two models associated to $\mathbf{p}$ and $\mathbf{q}$. Finally, we take the limit $\epsilon \to 0$, which yields, in the marginal case $p_j = p$ for all j, the desired derivative. Now, to start with, we construct the desired coupling based on a collection of m independently and identically distributed uniform random numbers in $[0,1]$, denoted by $U_1, U_2, \ldots, U_m$. For a given instance of uniform random numbers construct

$$\begin{aligned} A_{\mathbf{p}} &= \{e_j | 1 \le j \le m, U_j \le p_j\}, \\ A_{\mathbf{q}} &= \{e_j | 1 \le j \le m, U_j \le q_j\}. \end{aligned}$$

It is easy to see that both random sets $A_{\mathbf{p}}$ and $A_{\mathbf{q}}$, in isolation, have precisely the desired $\mathbb{P}_{\mathbf{p}}$ and $\mathbb{P}_{\mathbf{q}}$ laws. Yet, due to the shared underlying uniform random numbers, they exist on the same probability space. Hence we have constructed a coupling of $\mathbb{P}_{\mathbf{p}}$ and $\mathbb{P}_{\mathbf{q}}$ This in particular allows us to analyse the random difference $X(A_{\mathbf{q}}) - X(A_{\mathbf{p}})$ on one probability space. Now, one has by the coupling property and linearity of the expectation

$$\mathbb{E}[X(A_{\mathbf{q}}) - X(A_{\mathbf{p}})] = \mathbb{E}_{\mathbf{q}}[X] - \mathbb{E}_{\mathbf{p}}[X].$$

Now the crucial observation is that in the coupling, by construction of $\mathbf{q}$, $X(A_{\mathbf{q}}) - X(A_{\mathbf{p}})$ can only be non-zero when $p_j < U_j \le q_j \equiv p_j + \epsilon$. Thus we have

$$\begin{aligned} \mathbb{E}\Big[X(A_{\mathbf{q}}) - X(A_{\mathbf{p}})\Big] &= \mathbb{E}\Big[X(A_{\mathbf{q}}) - X(A_{\mathbf{p}}) | U_j \in (p_j, q_j]\Big]\mathbb{P}\Big[U_j \in (p_j, q_j]\Big] \\ &+ \mathbb{E}\Big[X(A_{\mathbf{q}}) - X(A_{\mathbf{p}}) | U_j \notin (p_j, q_j]\Big]\mathbb{P}\Big[U_j \notin (p_j, q_j]\Big] \\ &= \epsilon\mathbb{E}\Big[X(A_{\mathbf{q}}) - X(A_{\mathbf{p}}) | U_j \in (p_j, q_j]\Big] \\ &= \epsilon\left\{\mathbb{E}\Big[X\big((A_{\mathbf{q}})^{e_j}\big)\Big] - \mathbb{E}\Big[X\big((A_{\mathbf{p}})_{e_j}\big)\Big]\right\} \\ &= \epsilon\left\{\mathbb{E}_{\mathbf{q}}\Big[X(A^{e_j})\Big] - \mathbb{E}_{\mathbf{p}}\Big[X(A_{e_j})\Big]\right\} \end{aligned}$$

Thus we obtain

$$\frac{\partial}{\partial p_j}\mathbb{E}_{\mathbf{p}}[X] = \lim_{\epsilon \to 0} \frac{\mathbb{E}_{\mathbf{q}}[X] - \mathbb{E}_{\mathbf{p}}[X]}{\epsilon} = \mathbb{E}_{\mathbf{p}}[\delta_{e_j} X]. \tag{5.1.2}$$

Then, finally setting $p_j = p$ for $1 \le j \le m$ we obtain, by means of the chain-rule, the Russo-Margulis formula (5.1.1). We remark, that relation (5.1.2) is of use in its own right, such as for applications to the "anisotropic" or quenched disorder version of the percolation model, and with modifications also for corresponding random-cluster models.

Before we proceed with the derivation of the announced bridge-edge formula for the random-cluster model let us consider two general observations. Firstly, the derivative of X at e is defined in terms of A^e and A_e, one of which necessarily equals A. For the following discussion it turns out to be more convenient to eliminate one of the two in favour of A

itself. To do so note

$$\begin{aligned}\mathbb{E}_p[\delta_e X] &= \sum_{A\subseteq E}\Big(\mathbb{P}_p[A]\{X(A^e)-X(A_e)\},\\ &= \sum_{\substack{A\subseteq E:\\ e\in A}}\mathbb{P}_p[A]\{X(A)-X(A_e)\}+\sum_{\substack{A\subseteq E:\\ e\notin A}}\Big(\mathbb{P}_p[A]\{X(A^e)-X(A_e)\},\\ &= \sum_{\substack{A\subseteq E:\\ e\in A}}\Big(\mathbb{P}_p[A]\{X(A)-X(A_e)\}+\frac{1-p}{p}\sum_{\substack{A\subseteq E:\\ e\in A}}\Big(\mathbb{P}_p[A]\{X(A)-X(A_e)\},\\ &= \frac{1}{p}\sum_{\substack{A\subseteq E:\\ e\in A}}\mathbb{P}_p[A]\{X(A)-X(A_e)\}. \qquad (5.1.3)\end{aligned}$$

The second general aspect we mention is how one can derive a Russo-Margulis-like formula for the random-cluster model. Firstly let us generalise and write $\mathbb{P}_{p,q}[\cdot]$ and $\mathbb{E}_{p,q}[\cdot]$ for the probability law and expectation in the random-cluster model with cluster weight q and bond-density p. Clearly we have

$$\mathbb{E}_{p,q}[X]=\frac{\mathbb{E}_p[q^K X]}{\mathbb{E}_p[q^K]},$$

and hence any expectation in the random-cluster model can be expressed in terms of expectations in the percolation model (with $q=1$). It follows, that we can apply (5.1.1), i.e.

$$\frac{\partial}{\partial p}\mathbb{E}_{p,q}[X]=\frac{\frac{d}{dp}\mathbb{E}_p[q^K X]}{\mathbb{E}_p[q^K]}-\mathbb{E}_{p,q}[X]\frac{\frac{d}{dp}\mathbb{E}_p[q^K]}{\mathbb{E}_p[q^K]}. \qquad (5.1.4)$$

thus we see that we generically need $d/dp\,\mathbb{E}_p[q^K]$, which is related to the expected number of bridges, which follows from

$$\begin{aligned}\mathbb{E}_p[\delta_e q^K] &= \frac{1}{p}\sum_{\substack{A\subseteq E:\\ e\in A}}\mathbb{P}_p[A]\Big[\Big(q^{K(A)}-q^{K(A_e)}\Big]\Big(\\ &= \frac{1-q}{p}\sum_{\substack{A\subseteq E:\\ e\in A}}\Big(\mathbb{P}_p[A]q^{K(A)}\mathbb{1}_{\{e\in B(A)\}},\\ &= \frac{1-q}{p}\mathbb{E}_p[q^K]\mathbb{P}_{p,q}[e\in B]. \qquad (5.1.5)\end{aligned}$$

Here $B(A)$ denotes the set of edges in A that are bridges in (V,A), in other words $e\in A$ is a bridge if and only if $K(A)=K(A_e)-1$. Therefore (5.1.4) becomes

$$\frac{\partial}{\partial p}\mathbb{E}_{p,q}[X]=\frac{\frac{d}{dp}\mathbb{E}_p[q^K X]}{\mathbb{E}_p[q^K]}+\frac{q-1}{p}\mathbb{E}_{p,q}[X]\mathbb{E}_{p,q}[|B|].$$

The derivative in the first term can be re-written

$$\frac{d}{dp}\mathbb{E}_p[q^K X] \;=\; \frac{1}{p}\sum_{e\in E}\sum_{\substack{A\subseteq E:\\ e\in A}}\mathbb{P}_p[A]\Big\{\Big(X(A)q^{K(A)}-X(A_e)q^{K(A_e)}\Big\},$$

$$
\begin{aligned}
&= \frac{1}{p}\sum_{e\in E}\sum_{\substack{A\subseteq E:\\ e\in B(A)}} \mathbb{P}_p[A]q^{K(A)}\{X(A)-qX(A_e)\}\\
&+ \frac{1}{p}\sum_{e\in E}\sum_{\substack{A\subseteq E:\\ e\in C(A)}} \mathbb{P}_p[A]q^{K(A)}\{X(A)-X(A_e)\}.
\end{aligned}
$$

Here we split the sum into two contributions, corresponding to either $e \in B(A)$ or $e \in C(A)$, where the latter is the set of non-bridges that is $A \setminus B(A)$ or equivalently edges in A that leave K invariant upon removal, i.e. $e \in A$ such that $K(A) - K(A_e) = 0$. We remark, that alternatively one has [27]

$$
\frac{\partial}{\partial p}\mathbb{E}_{p,q}[X] = \frac{1}{p(1-p)}\operatorname{Cov}_{p,q}[N,X].
$$

In order to derive the bridge-edge formula note that (5.1.5) implies

$$
\frac{\partial}{\partial p}Z_{p,q} = \frac{Z_{p,q}}{p}(1-q)\mathbb{E}_{p,q}[|B|], \tag{5.1.6}
$$

where we introduced the partition function of the random-cluster model $Z_{p,q} \equiv \mathbb{E}_p[q^K]$. On the other hand it is straightforward to show, by explicit differentiation, that

$$
\frac{\partial}{\partial p}Z_{p,q} = \frac{Z_{p,q}}{p}\quad\frac{\mathbb{E}_{p,q}[N]-mp}{1-p}\Big)\Big(\tag{5.1.7}
$$

Finally, by equating (5.1.6) and (5.1.7) we obtain for $q \neq 1$

$$
\mathbb{E}_{p,q}[\mathcal{B}] = \frac{\mathbb{E}_{p,q}[\mathcal{N}]-p}{(1-p)(1-q)}. \tag{5.1.8}
$$

Here we defined the density of bridges $\mathcal{B} = |B|/m$. In order to obtain an expression for the bridge-density for $q = 1$ in a general setting we can to take the limit $q \to 1$ in (5.1.8)

$$
\begin{aligned}
\lim_{q\to 1}\mathbb{E}_{p,q}[\mathcal{B}] &= \frac{1}{1-p}\lim_{q\to 1}\frac{p-\mathbb{E}_{p,q}[\mathcal{N}]}{q-1}\\
&= -\frac{1}{1-p}\lim_{q\to 1}\frac{\partial}{\partial q}\mathbb{E}_{p,q}[\mathcal{N}]\\
&= \frac{1}{p-1}\operatorname{Cov}_p[K,\mathcal{N}].
\end{aligned} \tag{5.1.9}
$$

A few comments are in order; The second equality follows from the fact that $\mathbb{E}_{p,q}[\mathcal{N}] \to p$ for $q \to 1$, followed by applying L'Hôpital's rule. The last step follows from the generic identity $\partial/\partial q\mathbb{E}_{p,q}[X] = \operatorname{Cov}_{p,q}[K,X]/q$ for any random variable $X : \Omega \to \mathbb{R}$, c.f. e.g. [27] or by explicit differentiation. Both, identity (5.1.8) and (5.1.9), are valid for any value of p and any (finite) graph. We also remark that the idea of using the Russo-Margulis formula to establish a connection to the density of bridges has already been mentioned in [101], however with a different application in mind and solely restricted to the framework of

bond percolation, without mentioning the link to the covariance for percolation and the generalisation to the random-cluster model.
It is of course also possible to obtain an analogous expression for the expected density of non-bridges $\mathbb{E}_{p,q}[\mathcal{C}]$, to this end note $\mathbb{E}_{p,q}[\mathcal{N}] = \mathbb{E}_{p,q}[\mathcal{B}] + \mathbb{E}_{p,q}[\mathcal{C}]$ and therefore

$$\mathbb{E}_{p,q}[\mathcal{C}] = \frac{p + \big[p(q-1) - q\big]\mathbb{E}_{p,q}[\mathcal{N}]}{(1-p)(1-q)}. \tag{5.1.10}$$

Before we consider special instances of (5.1.8) and (5.1.10), for which we evaluate, both exactly and numerically, we elaborate on two general, and often useful, observations.
Firstly, there is an upper threshold, or bound, for the density of bridges, valid on any graph whenever $q \geq 1$. This follows from the following general comparison inequality for the random-cluster model with $q \geq 1$, for any increasing[2] function $f : \Omega \to \mathbb{R}$, c.f. [102, 27]:

$$\mathbb{E}_{p,q}[f] \geq \mathbb{E}_{p',q'}[f] \text{ if } q \geq q', q \geq 1, \frac{p}{q(1-p)} \geq \frac{p'}{q'(1-p')}.$$

For the particular choice $q' = 1$, $q \geq 1$ and $p' = \tilde{p}(p,q) = p/[(1-p)q + p]$ one can verify the above conditions, and hence with $f = \mathcal{N}$ (which is clearly increasing) we obtain

$$\mathbb{E}_{p,q}[\mathcal{N}] \geq \mathbb{E}_{\tilde{p}(p,q),1}[\mathcal{N}] = \tilde{p}(p,q),$$

whence with (5.1.1) we obtain the desired upper bound for $\mathbb{E}_{p,q}[\mathcal{B}]$:

$$\begin{aligned} \mathbb{E}_{p,q}[\mathcal{B}] &= \frac{p - \mathbb{E}_{p,q}[\mathcal{N}]}{(1-p)(q-1)}, \\ &\leq \frac{p - \tilde{p}(p,q)}{(1-p)(q-1)}, \\ &= \tilde{p}(p,q). \end{aligned} \tag{5.1.11}$$

The above choice of $\tilde{p}(p,q)$ might seem somewhat arbitrary, however as we will see in Chapter 6, where we discuss perfect sampling for the random-cluster model by means of the coupling from the past algorithm, the percolation model with $\tilde{p}(p,q)$ provides a natural "lower" reference model for the random-cluster model with p, q.
The second observation we make is related to the expected relative densities, i.e. $\mathbb{E}_{p,q}[\mathcal{B}/\mathcal{N}]$ which measures the mean fraction of bridges among all *open* edges. More precisely, let us consider $\mathbb{P}_{p,q}[e \in B | e \in N]$, that is the probability that e is a bridge, given e is open (active). Thus this conditional probability formalises the idea of a relative density of bridges (for transitive graphs). Due to the hierarchy between open edges and bridges, that is e is a bridge $\Rightarrow$ e is open (in other words $B(A) \subseteq A$), we have

$$\mathbb{P}_{p,q}[e \in B | e \in A] \quad = \quad \frac{\mathbb{P}_{p,q}[e \in B, e \in A]}{\mathbb{P}_{p,q}[e \in N]},$$

[2] Recall f is increasing if $f(A) \leq f(B)$ whenever $A \subseteq B$.

$$
\begin{aligned}
&= \mathbb{P}_{p,q}[e \in A | e \in B] \frac{\mathbb{P}_{p,q}[e \in B]}{\mathbb{P}_{p,q}[e \in A]}, \\
&= \frac{\mathbb{P}_{p,q}[e \in B]}{\mathbb{P}_{p,q}[e \in A]}.
\end{aligned}
$$

Thus, it suffices to consider ratio of expected bridge- to expected edge density. Further the argument clearly applies as well to non-bridges. For instance for transitive graphs, such as $\mathbb{Z}^2_L$, we have

$$\mathbb{P}_{p,q}[e \in B | e \in A] = \frac{1 - \frac{p}{\mathbb{E}_{p,q}[\mathcal{N}]}}{(1-p)(1-q)}.$$

After these general observations, we now proceed with the confirmation and analysis of the above theoretical predictions in various instances of the random-cluster model.

5.2 Exact asymptotic densities for the square lattice

5.2.1 Complete solution for the Ising model

The Ising model corresponds to the choice of $q = 2$ in the random-cluster model and plays a special role, as it allows for an exact calculation of its normalised partition function (or free energy) on $\mathbb{Z}^2$ [22], as well as for various finite and semi-infinite geometries, see for instance [75] for a method to calculate the free energy for $\mathbb{Z}^2_L$. This evaluation is valid for the entire temperature regime, and is consistent with the complexity results for the Tutte polynomial, outlined in Chapter 2, i.e. the polynomial time solvability for bipartite planar graphs whenever $q = 2$. It therefore provides an intriguing test case to study the variation of the above quantities with p, which in turn provides further insight into the change of the cluster structure with p. In what follows we concern ourselves with the infinite case on $\mathbb{Z}^2$, and in order to establish a link to the random-cluster model we exploit the relationship between the internal energy of the Ising model, at inverse temperature $\beta \geq 0$ and coupling $J \geq 0$, denoted by $u(\beta, J)$, and the edge-density of the corresponding random-cluster model, i.e.

$$\mathbb{E}_{p,2}[\mathcal{N}] = [1 - u(-\log(1-p), 1/2)] \frac{p}{2} \tag{5.2.1}$$

where we explicitly assumed that the graph is $\mathbb{Z}^2$. This identity can be derived by means of the Edwards-Sokal coupling [11] between the Potts model and the random-cluster model, and the equivalence between 2-state Potts and Ising model. For the internal energy one obtains, c.f. e.g [22],

$$u(\beta, 1/2) = -\frac{1}{2} \coth\beta \left\{ 1 + \frac{2}{\pi} \left(2\tanh^2\beta - 1 \right) \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{1 - \left[\frac{4/\sinh^2\beta}{(1+1/\sinh^2\beta)^2} \right] \sin^2 t}} \mathrm{d}t \right\}$$

which has to be evaluated with the parametrisation $\beta = \beta(p) = -\log(1-p)$ in order to calculate $\mathbb{E}_{p,2}[\mathcal{N}]$ using (5.2.1). We show in Figure 5.1 the expected edge-, bridge- and non-bridge densities derived from the above expressions. First, let us consider the

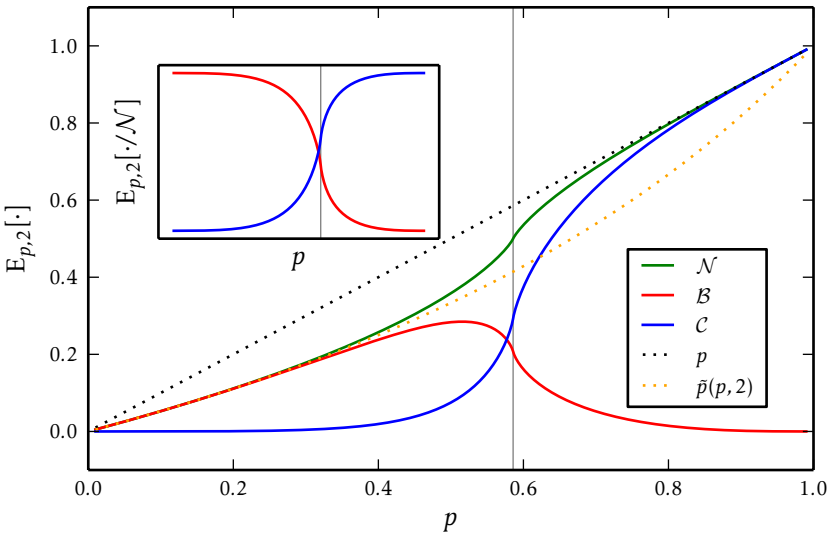


Figure 5.1: *Exact densities of edges, bridges and non-bridges for the $q = 2$ random-cluster model on $\mathbb{Z}^2$. The solid grey vertical lines indicate the location of the critical point $\sqrt{2}/(1+\sqrt{2})$.*

absolute densities, where we observe that for small values of p, both the bridge- and edge-density are very well described by $\tilde{p}(p,2)$. This is easily explained by considering the heat-bath dynamics for the random-cluster model, in particular (2.2.4), where the value $\tilde{p}(p,q)$ is precisely the corresponding insertion probability of pivotal edges (bridges). Clearly, for sufficiently low values of p, a typical configuration consists of small components with very few and short cycles. We remark, that the random-cluster model with parameters p and q on a forest (or tree) is equivalent to percolation with parameter $\tilde{p}(p,q)$. In contrast, for values of p close to 1, where most of the edges are part of cycles, the cluster-weight does not have a very strong influence, and edges, mostly non-bridges, increase in expectation roughly linearly with p. This can, again, be explained in terms of the heat-bath dynamics, which operates on non-pivotal edges with parameter p. Independently of the particular value of p, the results shown in Figure 5.1 are consistent with the upper bound (5.1.11). Equivalently, $\mathbb{E}_{p,2}[\mathcal{N}]$ is entirely bounded from below by $\tilde{p}(p,2)$. Further, the relative densities, shown in the inset of Figure 5.1, nicely reflect how active edges are more and more deposited in cycles. Interestingly, the expected density of bridges shows a maximum at a value of p smaller than $p_{\mathrm{sd}}(2)$. There is also an analytical result of relevance to the density of bridges, which originates in a singularity of the second β derivative of the free energy density of the Ising model. More precisely it can be shown, [22], that the normalised free energy f has a singular contribution $f_s \approx (\beta-\beta_c)^2 \log(\beta-\beta_c)$. Which in turn implies that the first thermal derivative of the internal energy, related to the second thermal derivative of f, diverges logarithmically for $\beta \to \beta_c$ (from below),

i.e. we have $\partial_\beta u \approx \log(\beta - \beta_c)$ and hence by means of (5.2.1) also a singularity of the p derivative of $\mathbb{E}_{p,2}[\mathcal{N}]$ at $p = p_{\mathrm{sd}}(2)$. We will have more to say when we discuss the finite size effects in the critical random-cluster model in section 5.3. We analysed numerically

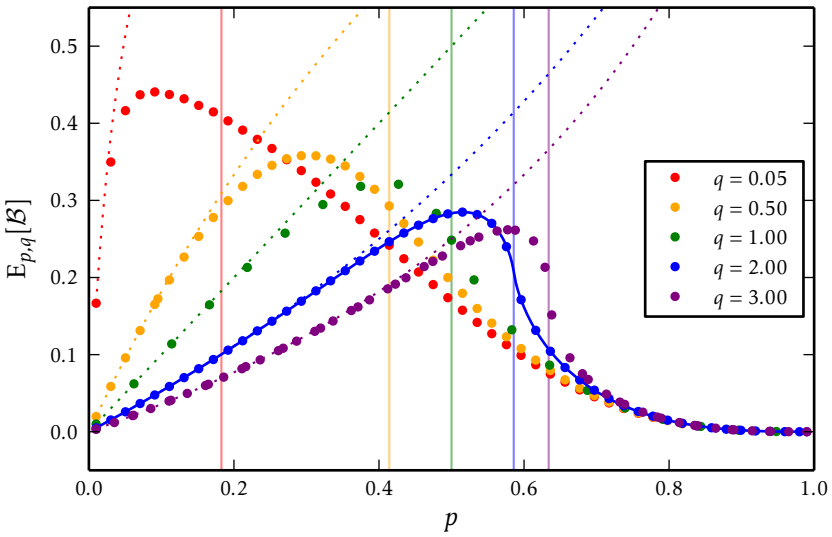


Figure 5.2: *Estimated densities of bridges for the random-cluster model on $\mathbb{Z}^2_{64}$ for some values of q. The solid blue lines shows the exact solution of Figure 5.1, and the dashed lines show the respective $\tilde{p}$ curve. Lastly, the solid vertical lines locate the corresponding self-dual points.*

the p-dependence of $\mathbb{E}_{p,q}[\mathcal{B}]$ for some other values of q for $\mathbb{Z}^2_L$, in order to check whether the above characteristics extend naturally beyond $q = 2$. As shown in Figure 5.2, the maximum of $\mathbb{E}_{p,q}[\mathcal{B}]$ is located at a value below the corresponding self-dual bond density, and the slope at p_{sd} appears to become steeper with increasing q. In order to understand better what happens at the self-dual point, let us fix $p = p_{\mathrm{sd}}(q)$ and consider the self-dual bridge- and non-bridge density, denoted by $\mathbb{E}_q[\mathcal{B}] \equiv \mathbb{E}_{p_{\mathrm{sd}}(q),q}[\mathcal{B}], \mathbb{E}_q[\mathcal{C}] \equiv \mathbb{E}_{p_{\mathrm{sd}}(q),q}[\mathcal{C}]$.

5.2.2 Critical densities for general cluster weights

A first natural question relates to the asymptotic values of $\mathbb{E}_q[\mathcal{B}]$ and $\mathbb{E}_q[\mathcal{C}]$ for $L \to \infty$. If we ignore the boundary conditions and consider the infinite square lattice $\mathbb{Z}^2$, which is self-dual and planar (in contrast to $\mathbb{Z}^2_L$ which is not planar), then, by self-duality, one has $\mathbb{E}_q[\mathcal{N}] = 1/2$. Hence we can insert this together with $p = p_{\mathrm{sd}}(q) = \sqrt{q}/(1+\sqrt{q})$ into (5.1.8) and (5.1.10) to obtain

$$\mathbb{E}_q[\mathcal{B}] = \frac{1}{2(1+\sqrt{q})}, \quad \mathbb{E}_q[\mathcal{C}] = \frac{\sqrt{q}}{2\left(1+\sqrt{q}\right)}. \tag{5.2.2}$$

See also Figure 5.3, which compares these exact values to extrapolated values from a careful finite size scaling analysis, described below. In fact, equations (5.2.2) can be

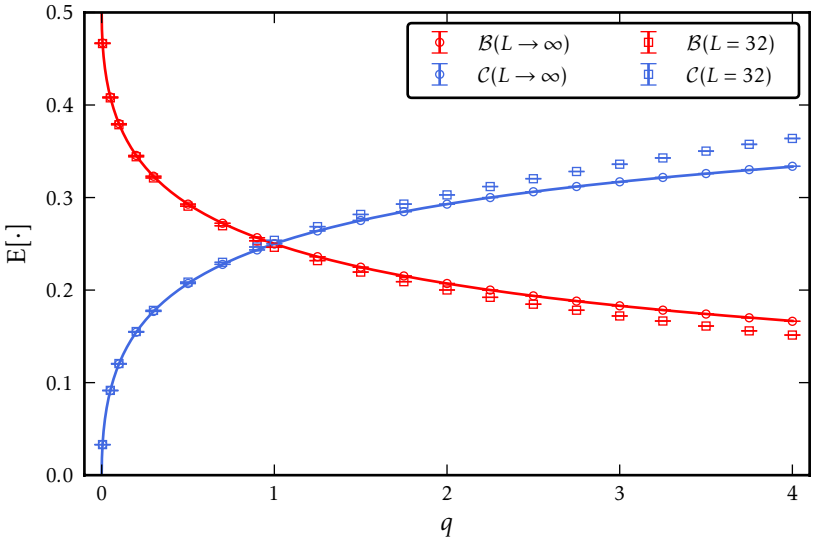


Figure 5.3: *Critical bridge- and non-bridge density. The circles correspond to the asymptotic densities extracted from the finite-size analysis described in section 5.3. Squares show the numerical estimates for a finite $L = 32$ system. The solid lines depict the exact expressions in* (5.2.2)*. The size dependent deviations, in paticular for "large" q motivate a careful analysis of the involved finite size corrections.*

written in terms of $p_{\mathrm{sd}}(q)$ and $\mathbb{E}_q[\mathcal{N}]$, as

$$\mathbb{E}_q[\mathcal{B}] = \big(1 - p_{\mathrm{sd}}(q)\big)\mathbb{E}_q[\mathcal{N}], \quad \mathbb{E}_q[\mathcal{C}] = p_{\mathrm{sd}}(q)\mathbb{E}_q[\mathcal{N}].$$

The above expressions allow us interpret $p_{\mathrm{sd}}(q)$ as the expected fraction of non-bridges among all active (occupied) edges. Clearly, the remaining fraction $1-p_{\mathrm{sd}}(q)$ of active edges have to be bridges, i.e. the discussion on relative densities implies $\mathbb{P}_q[e \in B | e \in A] = 1-p_{\mathrm{sd}}(q)$ and similarly $\mathbb{P}_q[e \in C | e \in A] = p_{\mathrm{sd}}(q)$. As $p_{\mathrm{sd}}(q)$ is increasing with q, this implies that non-bridges become more and more likely among active edges. Further, an immediate consequence of (5.2.2) is that $\mathbb{E}_q[\mathcal{B}], \mathbb{E}_q[\mathcal{C}] \to 1/4$ for $q \to 1$, which recovers a result recently shown in [55]. The result

$$\mathbb{P}_1[e \in B | e \in A] = \mathbb{P}_1[e \in C | e \in A] = \frac{1}{2},$$

reflects nicely the fact that the case $q = 1$ occupies edges independently, and for the choice $p = 1/2$ there is no bias towards either bridges or non-bridges. However, in contrast to the case $q = 2$, we are not able to obtain a general p-dependent solution for $\mathbb{E}_{p,1}[\mathcal{B}]$ on $\mathbb{Z}^2$. Interestingly, we can establish a symmetry of $\mathbb{E}_{p,1}[\mathcal{B}]$ for $\mathbb{Z}^2$, imposed by self-duality and planarity.

5.2.3 Symmetry for the bridge density for percolation

In order to establish the desired result we remark that, based on (5.1.5), one can show

$$\frac{d}{dp}\mathbb{E}_{p,1}[k] = -\frac{1}{p}\mathbb{E}_{p,1}[\mathcal{B}], \tag{5.2.3}$$

where we defined $k = K/m$. Now, Euler's formula for any finite planar graph (V, A) states [27]

$$K(A) = |V| - |A| + F(A) - 1, \tag{5.2.4}$$

where $F(A)$ is the number of faces in *any* planar embedding of (V, A). It is further well known, see for instance [27], that

$$F(A^\star) = K(A). \tag{5.2.5}$$

Here $A^\star$ is the dual configuration of A. Now, dividing both sides of (5.2.4) by m and taking expectation $\mathbb{E}_{p,1}[\cdot]$ we obtain with $|V| = n$

$$\mathbb{E}_{p,1}[k] = \frac{n}{m} - p + \frac{\mathbb{E}_{p,1}[F]}{m} - \frac{1}{m}, \tag{5.2.6}$$

when we take the limit $m \to \infty$ and, informally, assume the above holds in the limit also for $\mathbb{Z}^2$, we obtain by duality and (5.2.3):

$$\mathbb{E}_{p,1}[k] = \mathbb{E}^\star_{1-p,1}[k] + \frac{1}{2} - p. \tag{5.2.7}$$

Here we used the identity (5.2.5) and $\mathbb{E}^\star_{1-p,1}$ is the expectation of the percolation model with bond density $1-p$ on the dual graph $G^\star$. Now in combination with the self-duality of $\mathbb{Z}^2$ (it is isomorphic to its dual) and the duality of measures, i.e. $\mathbb{P}_{1,p}[A] = \mathbb{P}_{1,1-p}[A^\star]$ for any $A \subseteq E$, a consequence of the duality of the percolation measures for $q = 1$ whenever $p^\star(p, 1) = 1 - p$, we obtain the result. Lastly, by differentiating both sites of (5.2.7) with respect to p we obtain by means of (5.2.5) (discarding the q dependence):

$$\mathbb{E}_p[\mathcal{B}] = p - \frac{p}{1-p}\mathbb{E}_{1-p}[\mathcal{B}]. \tag{5.2.8}$$

When we evaluate the above for $p = 1/2$ we obtain the asymptotic result $\mathbb{E}_{1/2}[\mathcal{B}] = 1/4$, as mentioned before. What (5.2.8) states is that, due to duality, one only needs to consider $\mathbb{E}_p[\mathcal{B}]$ for $p \in [0, 1/2]$ or $p \in [1/2, 1]$, and can recover the alternative interval.

5.3 Finite-size effects

5.3.1 Percolation and pseudo-bridges

In the preceding sections we considered the asymptotic behaviour for the square lattice, but it is also intriguing to analyse the non-asymptotic or finite case, which we do here by considering $\mathbb{Z}^2_L$, the $L \times L$ square lattice with periodic boundary conditions. A related non-asymptotic study was recently presented by Xu et al. in [55] for critical bond percolation on $\mathbb{Z}^2_L$. There the authors investigated a partition of edges induced by a natural medial graph[3] property. More precisely, the authors divided open edges into two classes, corresponding to whether the two medial graph loop segments (or loop arcs) associated to a given edge belong to the same loop, or two different loops. We denote the corresponding densities by $\mathbb{E}_p[\ell_1]$ and $\mathbb{E}_p[\ell_2]$, respectively. We can also provide an alternative definition based on an application of the Russo-Margulis formula (5.1.1) to the expected number of loops on the medial graph, denoted by $\mathbb{E}_p[N_{\mathcal{L}}]$:

$$\begin{aligned}\frac{1}{m}\frac{d}{dp}\mathbb{E}_p[N_{\mathcal{L}}] &= \frac{1}{m}\sum_{e\in E}\mathbb{E}_p[\delta_e N_{\mathcal{L}}],\\ &= \frac{1}{m}\sum_{e\in E}\frac{1}{p}\sum_{\substack{A\subseteq E:\\ e\in A}}\Big(\mathbb{P}_p[A]\{N_{\mathcal{L}}(A)-N_{\mathcal{L}}(A_e)\},\\ &= \frac{1}{p}\Big(\mathbb{E}_p[\ell_1]-\mathbb{E}_p[\ell_2]\Big).\end{aligned}$$

In words, edges contributing to ℓ_2, in what follows referred to as *type-2 edges*, are edges that, upon removal, merge two previously disconnected loops and hence decrease the overall number of loops by one. Equivalently, *type-1 edges* increase the number of loops by one upon removal. The authors of [55] established rigorously that for the marginal case $q = 1$ and $p = 1/2$, one obtains for any L,

$$\mathbb{E}_{1/2}[\ell_1] = \mathbb{E}_{1/2}[\ell_2] = \frac{1}{4}. \tag{5.3.1}$$

We emphasise, that both quantities therefore show no finite-size corrections. This in particular implies (the to our knowledge so far unknown result) that the expected number of loops in the medial graph is extremal at criticality, that is we have uniformly in L

$$\frac{d}{dp}\mathbb{E}_p[N_{\mathcal{L}}]\Big|_{p=1/2} = 0.$$

Further, for planar graphs, it is straightforward to verify that a type-2 edge decreases the number of faces upon removal by one, whereas type-1 edges leave the number of faces invariant upon removal: We have Euler's formula (5.2.4) and one has for an edge $e \in A$:

$$K(A) - K(A_e) = [|V| - |A| + F(A) - 1] - [|V| - |A_e| + F(A_e) - 1] = F(A) - F(A_e) - 1, \tag{5.3.2}$$

[3] For a definition of the medial graph and a related efficient algorithm see section 3.4.1.

in other words $K(A) = K(A_e)$ if and only if $F(A) \neq F(A_e)$, see also [55]. Moreover an edge is a bridge if and only if it is a type-1-edge and a non-bridge, if and only if it is of type 2. This follows from an alternative Euler relation involving the number of loops [22, 20]:

$$|A| + 2K(A) = |V| + N_{\mathcal{L}}(A),$$

which leads to the equivalence between type-1 and type-2 edges and bridges and non-bridges, respectively for planar graphs:

$$K(A) - K(A_e) = \frac{1}{2}[N_{\mathcal{L}}(A) - N_{\mathcal{L}}(A_e) - 1],$$

where we assumed w.l.o.g. $e \in A$. However the graph $\mathbb{Z}_L^2$ is not planar and, more importantly, the set of non-planar spanning sub-graphs in the associated configuration space Ω has, asymptotically in the continuum limit, non-vanishing probability at criticality, [103, 104, 105]. Such non-planar configurations must in particular contain at least one cluster that includes a *non-trivial cycle* that wraps around both main directions of the torus simulatenously. In order to define precisely a non-trivial cycle, we use the definition of $\mathbb{Z}_L^d$ provided in [106], i.e. $\mathbb{Z}_L^d$ is the graph with vertex set corresponding to $[0, L)^d$, in other words we identify the vertices of $\mathbb{Z}_L^d$ with d-tuples $(x_1, x_2, \ldots, x_d)$ where $0 \leq x_j < L$. Two vertices in $\mathbb{Z}_L^d$ are neighbours, or adjacent, if they are equal in all coordinates but one, in which they are adjacent in the (one-dimensional) cycle graph on L vertices, denoted by C_L. Now, a *non-trivial cycle* in $\mathbb{Z}_L^d$ is a cycle with the property, that the projection along at least one coordinate (direction) induces C_L. In other words it is possible to find a specific coordinate, that when considered in isolation, is C_L. For the case $d = 2$ considered here, one has clearly only two coordinates. In what follows we consider $d = 2$ until otherwise stated. We denote by a *cross* cluster a connected component that contains a non-trivial cycle which wraps around both directions and has the topology as shown in the right panel in Figure 5.4. Note that a cross cluster is not the only way a cluster can wrap around the torus simultaneously in both directions, as illustrated in the left panel of Figure 5.4. Clearly the topology of the two clusters are different. We refer to clusters of the topology shown in the left panel of Figure 5.4 as *winding clusters*, that is a winding cluster is a connected component that includes a non-trivial cycle that wraps around both directions *but* is not a cross cluster. Observe that the existence of one cross cluster implies the absence of any other cluster that induces a non-trivial cycle of any type. The probability of a cross cluster is known in the continuum limit, thanks to the (conjectured) conformal invariance of critical bond percolation in two dimensions. Ziff et al. [103] give a closed form expression, based on earlier work of Pinson [104], that evaluates to 0.309526275… for the probability of a cross cluster for critical bond percolation on $\mathbb{Z}_L^2$ in the limit $L \to \infty$. For practical purposes, one can use the deviation from the Euler formula (5.2.4) to test for a cross cluster [103, 28], in other words the spanning sub-graph (V, A) contains one cross cluster if and only if $F(A) - K(A) + |V| - |A| - 1$ equals -2, and at least one winding cluster exists whenever the expression evaluates to -1. Now in [55] it was argued that a

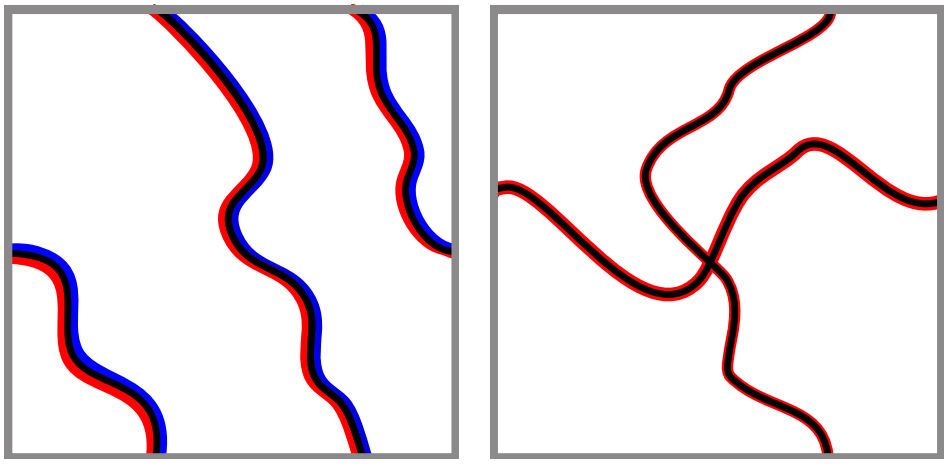


Figure 5.4: *Illustration of a winding cluster (left panel) and a cross cluster (right panel). The solid black line corresponds to open edges, the two remaining colours red and blue represent the loops in the medial graph. Observe that in both examples all edges are non-bridges. In the winding cluster no pseudo-bridges can exists as both loop arcs of a given non-bridge belong to two different loops. On the other hand for the cross cluster in the right panel we find that all non-bridges have both of its loop arcs in the medial graph in the same loop.*

cluster that induces a non-trivial cycle in both coordinates can contain non-bridges that violate the "planar identity" (5.3.2) stating that type-1 edges equal bridges. This occurs precisely for non-bridges that have both of its associated loop-arcs in the medial graph in the same loop. In fact a careful observation (using the illustrations in Figure 5.4) shows that such non-bridges can only exist on a cross cluster. More precisely, following the terminology of [55], we call such non-bridges *pseudo-bridges*. An important observation is that a pseudo-bridge is a non-bridge on a cross cluster with the distinguishing property that it is pivotal to the existence of the cross cluster. In other words, let $\chi(A)$ be the indicator function of the event that the configuration A contains a cross cluster. Then $e \in E$ is a pseudo-bridge if and only if $e \in A$ and $\chi(A) \neq \chi(A_e)$. One can verify that this happens precisely for non-bridges on a cross-cluster for which not more than one cycle "overlaps" at it (for an illustration see Figure 5.5 or Figure 1 in [55]). Summarising the previous arguments we have:

$$\begin{aligned}
\frac{1}{4} &= \mathbb{P}_{1/2}[e \in L_1], \\
&= \mathbb{P}_{1/2}[e \in L_1, e \in B] + \mathbb{P}_{1/2}[e \in L_1, e \in C], \\
&= \mathbb{P}_{1/2}[e \in L_1 | e \in B]\mathbb{P}_{1/2}[e \in B] + \mathbb{P}_{1/2}[e \in L_1, e \in C], \\
&= \mathbb{P}_{1/2}[e \in B] + \mathbb{P}_{1/2}[e \in L_1, e \in C], \\
&= \mathbb{P}_{1/2}[e \in B] + \frac{1}{2}\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi].
\end{aligned}$$

where $L_1(A), L_2(A)$ are the sets of type-1 and type-2 edges in A, respectively and $\mathcal{P}_\chi(A)$ is the set of edges that are pivotal to the event χ in A. Hence rearranging yields

$$\mathbb{P}_{1/2}[e \in B] = \frac{1}{4} - \frac{1}{2}\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi], \tag{5.3.3}$$

$$\mathbb{P}_{1/2}[e \in C] = \frac{1}{4} + \frac{1}{2}\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi]. \tag{5.3.4}$$

Note that $\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi]$ is in fact equal to the influence of e on χ, that is $\mathbb{E}_{1/2}[\delta_e \chi]$, as defined in section 5.1, a consequence of the monotonicity and Boolean nature of χ. Before we provide scaling arguments for the form of $\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi]$ we recall the numerical results found by Xu et al. in [55] for the density of bridges and non-bridges, which due to translational invariance equal the likelihood of an edge to be a bridge or non-bridge, respectively. The authors find that the difference $\mathbb{E}_{1/2}[\ell_1] - \mathbb{E}_{1/2}[\mathcal{B}]$ vanishes with increasing L to leading order as L^{-x_2}, where $x_2 = 5/4$ is the two-arm exponent for critical percolation on the square lattice. This exponent is defined in terms of the large R asymptotics of the probability of the existence of two distinct clusters in an annulus of inner radius $O(1)$ and outer radius R [74]. In other words the inner boundary of the annulus is connected to the outer boundary by two distinct clusters. Coming back to the density of (non-) bridges, the numerical findings of Xu et al. suggest, due to the absence of finite-size corrections of the expected type-1 and type-2 densities, that we have the following asymptotics:

$$\mathbb{P}_{1/2}[e \in B] = \mathbb{E}_{1/2}[\mathcal{B}] \sim \frac{1}{4} - cL^{-x_2} \tag{5.3.5}$$

$$\mathbb{P}_{1/2}[e \in C] = \mathbb{E}_{1/2}[\mathcal{C}] \sim \frac{1}{4} + cL^{-x_2}, \tag{5.3.6}$$

where $c > 0$ [55] because of pseudo-bridges. Moreover, the constraint $\mathbb{E}_{1/2}[\mathcal{B} + \mathcal{C}] = 1/2$, demands the use of the same constant c in both expressions. Comparing this to our above calculations, this suggests

$$\mathbb{P}_{1/2}[e \in \mathcal{P}_\chi] \approx L^{-x_2}.$$

We will now provide scaling arguments which establish a link between the two-arm event above and the probability of an edge being pivotal for a cross cluster event, which in turn explains the appearance of the two-arm exponent as a finite-size correction exponent for the density of bridges for critical bond percolation on $\mathbb{Z}_L^2$. The arguments are based on rigorous results and ideas of [107] for left-to-right crossings in the planar case of a $L \times L$ box (without periodic boundary conditions). Consider Figure 5.5 which shows a cross cluster and a construction of a particular arm event: There exists four paths in an annulus with inner radius $O(1)$ and outer radius $L/2$, of alternating dual and primal type, emanating from the inner to the outer boundary. This event in an annulus of the above specification centred at a particular edge happens in particular when the edge is pivotal for a cross cluster[4]. The probability of such a four-arm event is (expected to be) asymptotically equivalent to $CL^{-5/4}$. Note that Xu et al. actually state that the relevant

[4]Strictly speaking this only shows that $\mathbb{P}_p[e \in \mathcal{P}_\chi] \leq \alpha_4(L/2)$ where $\alpha_4(L/2)$ is the probability of the above four arm event in an annulus of inner radius $O(1)$ and outer radius $L/2$. Unfortunately we are not able to

exponent is the two arm exponent of the existence of at least two clusters in an annulus. However it can be verified that the above four arm event of alternating primal and dual paths is close to the event that there are two distinct (primal) clusters within the annulus, touching the two boundaries [74]. This then explain the appearance of the exponent $-5/4$ above in the finite size correction for the density of bridges for critical percolation on $\mathbb{Z}_L^2$.

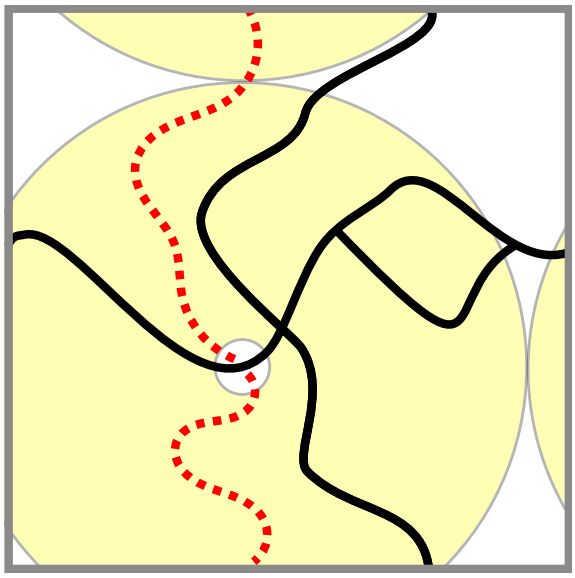


Figure 5.5: *A sketch of a cross cluster on $\mathbb{Z}_L^2$. The solid black lines are open edges composing the cross cluster, that induces two non-trivial cycles along both main directions. The dashed red line is a path in the dual graph, and the yellow annulus "around" a particular pseudo-bridge is used for the four arm event of alternating dual and primal paths.*

We remark that for $q = 1$ in any dimension, it is well known that $x_2 = d - 1/\nu$ [55, 56, 109, 110], hence the above expressions can also be written in terms of $d - 1/\nu$ using the thermal exponent $1/\nu$. However, this is only true for $q = 1$, and is therefore not clear what happens for $q \neq 1$, i.e. has the bridge-density leading finite-size correction of $-x_2$ or $1/\nu - d$? Before we answer this question, we briefly elaborate on an interesting link to the work [111], where the authors use scaling arguments and a numerical analysis to show that for critical percolation on $\mathbb{Z}_L^2$ and $\mathbb{Z}_L^3$ one has at the corresponding critical bond-density p_c

$$\mathrm{Cov}_{p_c}[K\mathcal{N}] \approx -d_1 + d_2 L^{1/\nu - d}, \tag{5.3.7}$$

where $d_1, d_2 > 0$ and d is the spatial dimension. Now, by (5.1.9), the above covariance equals, up to a factor of -2, the critical bridge density for bond-percolation, valid on any graph and any value of p. For the case $d = 2$, and hence $p_c = 1/2$, we thus have that $d_1 = 1/8$ and $2d_2 = c > 0$. This is in line to the numerical results reported in [111][5]. Besides

show the corresponding lower bound, which probably can be derived by means of the "separation of arms" technique [108].

[5] We remark that Deng et al. used a normalisation by the number of vertices, hence the results differ by a factor of $d = 2$ here. More precisely the authors show that the $d_1 = -1/4$, which with our normalisation by the number of edges corresponds to $d_1 = -1/8$.

confirming previous results, our work provides a novel geometric interpretation of the observed finite size corrections of the covariance (5.3.7).

5.3.2 Beyond $q = 1$ percolation

Let us now investigate the size dependence of $\mathbb{E}_q[\mathcal{B}]$ and $\mathbb{E}_q[\mathcal{C}]$ for general values of $q \in [0, 4]$. The absence of finite-size corrections for $\mathbb{E}_1[\ell_1]$ and $\mathbb{E}_1[\ell_2]$ seems rather special and unlikely to extend to $q \neq 1$. Indeed the proof of the corresponding "Loop Duality Lemma" in [55] explicitly used the fact that $p = 1/2$ and $q = 1$, by utilising the duality of the random-cluster model probability distribution for $\mathbb{Z}_L^2$ and $q = 1$ with $p = 1/2$, i.e. one has for a configuration A, and the corresponding dual configuration $A^\star$, $\mathbb{P}_{1/2}[A] = \mathbb{P}_{1/2}[A^\star]$. In order to get a better understanding of the size dependencies, we estimated $\mathbb{E}_q[\ell_1]$ and $\mathbb{E}_q[\ell_2]$ for a large selection of q in the entire interval $[0, 4]$, with increasing system size L. In what follows we consider only $\mathbb{E}_q[\ell_1]$, as the results for $\mathbb{E}_q[\ell_2]$ are completely analogous. Firstly, we fitted a finite scaling ansatz $\mathbb{E}_q[\ell_1] = c + aL^{-e_1} + bL^{-e_2}$ to our Monte-Carlo estimates for $\mathbb{E}_q[\ell_1]$ using the method of least squares fitting. The resulting estimates are consistent with $c = 1/[2(1 + \sqrt{q})]$ and $e_1 = 2 - 1/\nu$ as well as $e_2 = x_2$. Thus, we continued by fixing c, e_1, e_2 to the corresponding exact values, and fitted the data again, hence this time with only a and b being free parameters. We remark that this has the advantage that our fitting function now depends only linearly on the fitting parameters a and b, which we expect to yield a better numerical stability of the fitting procedure. Indeed, the fits were good with respect to the standard goodness of fit heuristics [112], even including the smallest system sizes $L = 4, 8, 16, 32$, strongly suggesting the following asymptotics

$$\mathbb{E}_q[\ell_1] \sim \frac{1}{2(1 + \sqrt{q})} + aL^{1/\nu - d} + bL^{-x_2}. \tag{5.3.8}$$

In Table 5.1 we show some of the estimates of a and b together with goodness of fit characteristics and Figure 5.6 shows our numerical data for $q = 2$ together with the best fit.

We find for all values of q analysed, that $a < 0$ and $b > 0$, with the obvious constraint that $a + b = 0$ for $q = 1$. This in turn, in combination with $-x_2 = 1/\nu - 2$, explains the absence of finite-size corrections for $q = 1$ as a cancelling of amplitudes in the random-cluster model, and hence recovers the size-independent result (5.3.1). Further, it is known that for $q \to 0$ with $p = p_{\mathrm{sd}}(q)$, one recovers the uniform spanning tree model, for which clearly no pseudo-bridges exists, and hence one has

$$\mathbb{E}_q[\ell_1] \to \frac{1}{2} - \frac{1}{2}L^{-2}, \tag{5.3.9}$$

which is in agreement with (5.3.8), due to $1/\nu \to 0$ for $q \to 0$, and the numerical observation that $a \to -1/2$ and $b \to 0$ for $q \to 0$, as it clearly emerges from the data listed in Table 5.1. We remark that the fit estimates for q close to one need to be treated with

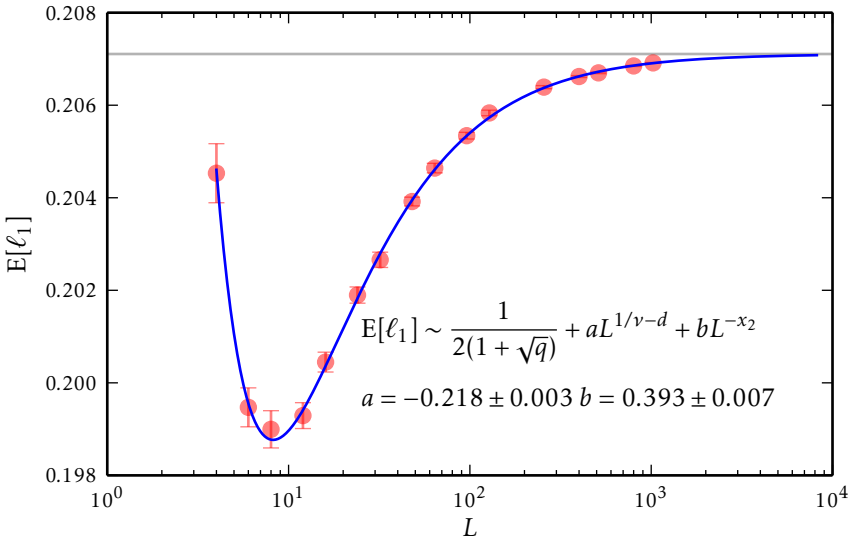


Figure 5.6: *Estimated size dependence of $\mathbb{E}_2[\ell_1]$ together with the best fit to the finite-scaling ansatz outlined in the text. All the data points starting from $L = 4$ were used in the fit and yielded $\chi^2/N_{\text{pdf}} = 0.5554$, where $N_{\text{pdf}} = 14$, is the number of degrees of freedom. The horizontal line corresponds to the density of bridges of $\mathbb{Z}^2$ in* (5.2.2).

caution, as here the two exponents $1/\nu - 2$ and $-x_2$ become very close, and hence it is numerically very difficult to distinguish the two contributions and the corresponding constants a, b. However, outside a suitable "safety"-window around $q = 1$, it appears that both a, b are increasing with q. The validity of (5.3.8) can be even further supported, by separately analysing the size-dependencies of $\mathbb{E}_q[\ell_1 - \mathcal{B}]$ and $\mathbb{E}_q[\mathcal{B}]$. Firstly, we considered the difference $\mathbb{E}_q[\ell_1 - \mathcal{B}]$ and find, as for $q = 1$, that the difference vanishes, in leading order, with the exponent $-x_2$, i.e.

$$\mathbb{E}_q[\ell_1 - \mathcal{B}] \sim b' L^{-x_2}. \tag{5.3.10}$$

This was already mentioned in [55], however based on preliminary results. We show in Figure 5.7 the numerically extracted exponent of an effective power law decay (in L) of $\mathbb{E}_q[\ell_1 - \mathcal{B}]$, and the results clearly confirm the claim (5.3.10). Additionally, we find that the involved constant b' in (5.3.8) is numerically consistent, within statistical accuracy, with $b' = b$. This suggests that the $aL^{1/\nu-d}$ term in (5.3.8) is, in leading order, cancelled out in the difference $\mathbb{E}_q[\ell_1 - \mathcal{B}]$, which in turn indicates that

$$\mathbb{E}_q[\mathcal{B}] \sim \frac{1}{2(1+\sqrt{q})} + aL^{1/\nu-d}, \tag{5.3.11}$$

where we emphasise that the constant $a < 0$ is the same as in (5.3.8). Indeed, we find, by a careful finite-size analysis of our Monte-Carlo data for $\mathbb{E}_q[\mathcal{B}]$, that the leading finite size correction is proportional to $L^{1/\nu-d}$. We summarise in Figure 5.7, the extracted exponents, and compare them to corresponding exact Coulomb gas values for $1/\nu - d$. Our analysis

q	a	b	χ^2/N_{pdf}	$L_{\min}$	Q
3.5	−0.1410(7)	0.504(7)	1.0354	4	0.4135
3.5	−0.1405(8)	0.49(2)	0.9722	8	0.4730
3.5	−0.1406(9)	0.50(4)	1.0155	16	0.4270
3.5	−0.140(1)	0.4(1)	1.1116	32	0.3514
2	−0.218(3)	0.393(7)	0.5554	4	0.9007
2	−0.219(3)	0.40(1)	0.6303	8	0.8183
2	−0.219(4)	0.40(2)	0.7086	16	0.7173
2	−0.218(6)	0.38(4)	0.85043	32	0.5580
1.25	−0.27(1)	0.32(2)	0.8451	4	0.6199
1.25	−0.26(2)	0.31(3)	0.9229	8	0.9229
1.25	−0.25(3)	0.30(4)	0.8482	16	0.5819
1.25	−0.31(4)	0.41(7)	0.4951	32	0.8607
0.9	−0.21(3)	0.19(3)	0.4983	4	0.9356
0.9	−0.22(5)	0.20(4)	0.5010	8	0.9155
0.9	−0.23(7)	0.21(6)	0.5237	16	0.8748
0.9	−0.3(1)	0.2(9)	0.6113	32	0.7692
0.5	−0.336(8)	0.192(4)	1.0286	4	0.4203
0.5	−0.32(1)	0.188(5)	1.0466	8	0.4019
0.5	−0.32(2)	0.188(7)	1.2546	16	0.2502
0.5	−0.33(4)	0.190(9)	1.5533	32	0.1332
0.05	−0.429(6)	0.033(1)	0.7662	4	0.7073
0.05	−0.41(1)	0.033(1)	0.6955	8	0.7575
0.05	−0.43(3)	0.033(1)	0.6882	16	0.7365
0.05	−0.43(7)	0.033(1)	0.8571	32	0.5522
0.005	−0.466(5)	0.0040(4)	0.9548	4	0.4979
0.005	−0.49(1)	0.0045(5)	0.6865	8	0.7663
0.005	−0.49(3)	0.0046(6)	0.6857	16	0.7389
0.005	−0.38(8)	0.0041(7)	0.5806	32	0.7948

Table 5.1: *Estimates of the parameters a and b in the finite-size scaling ansatz of the density of type-1 edges,* (5.3.8)*, obtained from least-square fits. The value Q is the "confidence-level", i.e. the probability that χ^2 would exceed the observed value, assuming that the underlying statistical model is correct.*

suggests therefore that both $\mathbb{E}_q[\ell_1]$ and $\mathbb{E}_q[\mathcal{B}]$ contain a term $aL^{1/\nu-d}$ suppressing the density with respect to the limiting value (5.2.2), however in addition to that, $\mathbb{E}_q[\ell_1]$ is ruled by another effect yielding an overestimation of bL^{-x_2} of the density for finite lattices. Now precisely for $q = 1$, it happens that these effects cancel in $\mathbb{E}_q[\ell_1]$, which in turn becomes size independent, however the global q-dependent picture actually suggests that the leading finite-size correction of the critical bridge density for percolation is actually thermal. In the remaining part of this section we show how the bridge-identity (5.1.8) imposes that the finite-size corrections for $q \neq 1$ must be, in leading order, described by $L^{1/\nu-d}$. To start with, we note that the generalisation of (5.2.1) to the q-state Potts model is well known and reads [50, 11]

$$\mathbb{E}_{p,q}[\mathcal{N}] = \frac{mp}{n} u_q\Big(-\log(1-p)\Big)\Big(\tag{5.3.12}$$

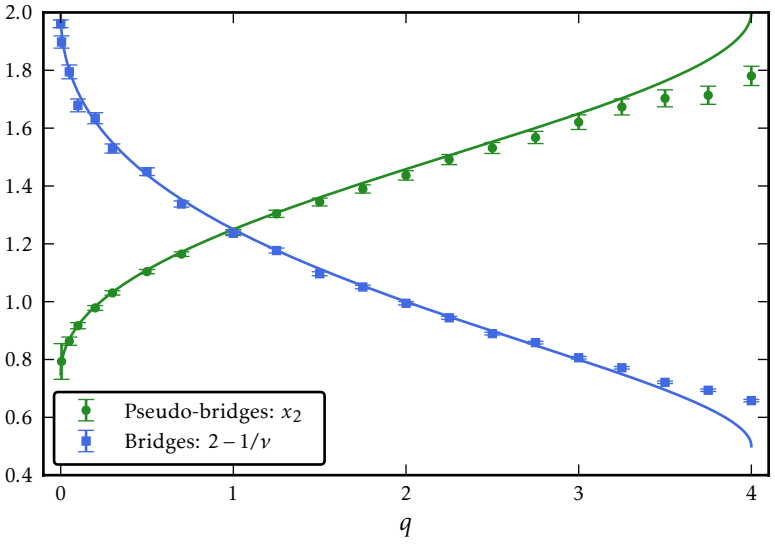


Figure 5.7: *Comparison of numerically extracted exponents of the leading finite-size correction for the density of bridges and pseudo-bridges. The solid lines show the exact Coulomb-gas value. The deviations for values of q around 4 are caused by strong sub-leading finite size corrections, which can not be incorporated into a stable fit.*

where $u_q(\beta)$ is the internal energy of the q-state Potts model with inverse temperature β [50, 44], and as before, m and n are the number of edges and vertices, respectively. The finite-size scaling for standard observables such as $u_q(\beta)$ close to a point of a second-order phase transition, is well established, c.f. [113]. We note that $\beta_c = \log(1+\sqrt{q})$ for the Potts model on $\mathbb{Z}^2$. Now, in the vicinity of β_c, the finite size corrections for $u_q(\beta_c)$ can be extracted from a scaling ansatz for the normalised dimensionless free-energy $f = -\log Z_P(\beta, q)/(\beta n)$. More generally, for a system without surfaces, such as $\mathbb{Z}_L^2$, the singular part f_s of f as a function of the thermal field $t = \beta - \beta_c$, the ordering field h, the leading irrelevant field v, and the systems size L, is expected to be of the following form [113]

$$f_s(t,h,L) = L^{-d} F\left(tL^{1/\nu}, hL^{d-\beta/\nu}, vL^{-\theta/\nu}\right) \tag{5.3.13}$$

The internal energy $u_q(\beta_c)$ can be expressed in terms of the first derivative of f with respect to t, evaluated at $t = h = 0$. Furthermore it is plausible to assume [93, 113] that the non-singular part of f has no size-dependence and directly yields the value obtained in the infinite-volume limit. These observations allow us to conclude that, to leading order in L, the size dependent deviations of $\mathbb{E}_q[\mathcal{N}]$ from its asymptotic value 1/2 are ruled by the exponent $1/\nu - d$, which in turn implies, by (5.1.8), that the expression (5.3.11) is indeed the right asymptotic form for $\mathbb{Z}_L^2$ at the self-dual point. We remark that, firstly the above arguments naturally extend to higher dimensions and other graphs, as we confirm below for $\mathbb{Z}_L^3$. Secondly, in a strict sense, assuming the correctness of the scaling ansatz (5.3.13), the above arguments are only valid for integer q. However it is plausible to assume the validity of the above arguments for $q \in (0, 4]$. A further direct consequence of (5.3.13) is that the p-derivative of $\mathbb{E}_{p,q}[\mathcal{B}]$ has a contribution proportional to $L^{\alpha/\nu}$, which implies a

singularity for $q \geq 2$, as already outlined in the discussion of the Ising case; consider also Figure 4.3.
So far we have only considered the density of bridges, and completely ignored non-bridges. This is reasonable as no new effects appear and the involved finite-size exponents are the same. However, what changes are of course the constants a, b. In particular, we find, due to pseudo-bridges, an underestimation of $\mathbb{E}_q[\mathcal{C}]$ by $\mathbb{E}_q[\ell_2]$ and the finite non-bridge density approaches the asymptotic result from above.

5.3.3 Three dimensions

We also investigated the three dimensional case $\mathbb{Z}_L^3$, in order to confirm the validity of the above arguments. In contrast to the two-dimensional case, we have no exact expressions for the standard critical exponents nor do we know the exact asymptotic densities. We therefore need to revert to appropriate numerical studies in the literature. Let us first consider critical bond percolation on $\mathbb{Z}_L^3$. One recent numerical determination is $p_c(1) = 0.24881182(10)$ [95]. We fixed the bond-density at this value and estimated the density of bridges for system sizes up to $L = 256$, and show in Figure 5.8 the corresponding data together with a fit to the basic finite-size scaling ansatz

$$\mathbb{E}_{p_c(q)}[\mathcal{B}] = c + aL^{-e}. \tag{5.3.14}$$

We extract ν via $\nu = 1/(3+e)$. Our numerical value of $\nu = 0.87(1)$ is in good agreement with the recent literature estimates $\nu = 0.8764(12)$ [95] and $\nu = 0.8751(11)$ [114]. Further we can also compare to the covariance study of Deng et al. [111], which concluded $\mathrm{Cov}_{p_c(1)}[K, N/L^3] \to -0.4783(2)$ which compares very well with our estimate $\mathbb{E}_{p_c(1)}[\mathcal{B}] \to 0.212177(3)$ which in turn implies $\mathrm{Cov}_{p_c}[K, N/L^3] = 3\mathbb{E}_{p_c}[\mathcal{B}](p_c - 1) \to -0.478155(1)$. For the cluster-weight $q = 1.5$ we are not aware of any published values for $p_c(1.5)$ to date, but Deng et al. find [96] that $p_c(1.5) = 0.311574973(43)$. In the same work the authors also estimate $1/\nu = 1.34 \pm 0.01 \pm 0.02$, where the first error bar represents statistical errors and the second error bar systematic errors due to neglected terms in the finite-size scaling analysis. This result is consistent with our estimate $1/\nu = 3 + e = 1.32(5)$. Moreover we estimate $\mathbb{E}_{p_c(1.5)}[\mathcal{B}] \to 0.19789(2)$, which as expected, is smaller than the corresponding bridge density for percolation. Lastly, we also considered the cluster-weight $q = 0.5$, for which, to our knowledge, no literature values are known. We emphasise that the purpose of the present study is not a high precision estimation of critical exponents and asymptotic densities, but merely seeks to confirm the validity of the above theoretical arguments. We estimated the location of the phase-transition, which is supposed to be of second order or continuous [88], by conducting an appropriate Binder-cumulant study, to be $p_c(0.5) = 0.154(1)$ and extract $\nu = 1.19(14)$ as well as $\mathbb{E}_{p_c(0.5)}[\mathcal{B}] \to 0.231(1)$.

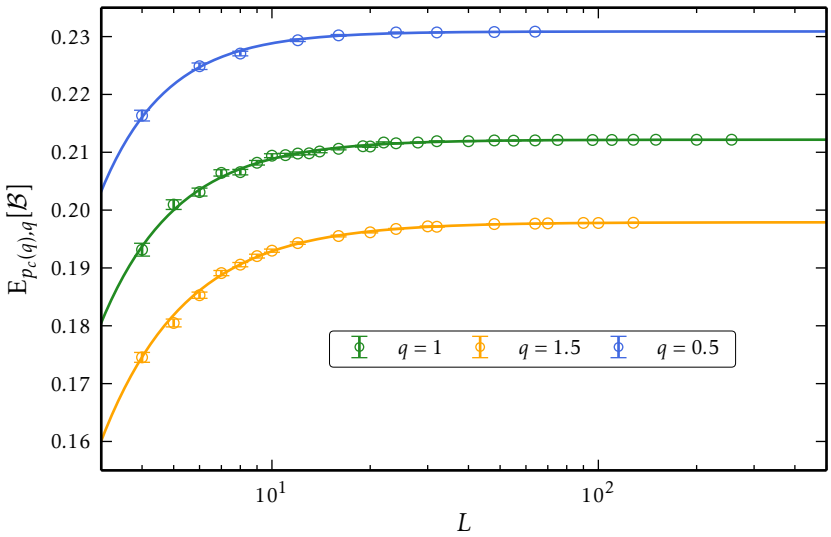


Figure 5.8: *Size dependencies of bridge densities for the random-cluster model on $\mathbb{Z}_L^3$ with $q = 0.5, 1, 1.5$ at the respective predicted location $p_c(0.5)$ of a second order phase transition. The solid lines show the corresponding best finite size ansatz* (5.3.14).

5.4 Bridges and their relation to other quantities

So far our analysis has focused on open (active) edges, that is bridges and non-bridges. Somewhat surprising, the analysis of closed (inactive) edges turns out to be very interesting too, in that it allows us to link several previously unrelated quantities, studied in the literature, to the density of bridges. This provides a unified framework for the analysis for, e.g., finite size corrections and, in particular, yields a geometric interpretation of those. Moreover, the study of closed edges provides a probe of the cluster structure. For instance, a large density of candidate-non-bridges suggests that clusters are more likely to self-entangle than to overlap with other clusters. On the other hand, a large candidate-bridge density indicates that clusters are typically disconnected, and closed edges are mostly inter cluster links.

Now, in order to analyse the density of candidate-(non-)bridges, let us consider a given closed edge $e = (x, y)$, that is e is not part of the current spanning sub-graph. Clearly we have that when x is not connected to y by a path of open edges, in what follows abbreviated by $x \nleftrightarrow y$, then by inserting e, we merge two previously disconnected clusters. Thus we conclude that a candidate-bridge is nothing else than a pair of disconnected neighbouring[6] vertices. This observation allows us to write down the following, almost trivial but useful, identity for arbitrary $e = (x, y) \in E$:

$$\mathbb{P}_{p,q}[e \in \overline{B}] = \mathbb{P}_{p,q}[x \nleftrightarrow y] = 1 - \mathbb{P}_{p,q}[x \leftrightarrow y]. \tag{5.4.1}$$

[6] A pair of vertices x, y is neighbouring if and only if $(x, y) \in E$.

We remark that for integer q one can relate (5.4.1) to $\langle \mathbb{1}_{\{\sigma_x=\sigma_y\}}\rangle_{q;-\log(1-p)}$, where $\langle\cdot\rangle_{q;\beta}$ is the expectation of the Potts model with q states and inverse temperature β and $\sigma_x, \sigma_y \in \{0,1,\ldots,q-1\}$. This follows from

$$\langle \mathbb{1}_{\{\sigma_x=\sigma_y\}}\rangle_{q;-\log(1-p)} = \frac{1}{q} + \left(1-\frac{1}{q}\right)\mathbb{P}_{p,q}[x \leftrightarrow y]$$

which can be easily verified by utilising the Edwards-Sokal coupling [11]. We therefore obtain

$$\mathbb{P}_{p,q}[e \in \overline{B}] = \frac{1-\langle \mathbb{1}_{\{\sigma_x=\sigma_y\}}\rangle_{q;-\log(1-p)}}{1-\frac{1}{q}}. \tag{5.4.2}$$

Remarkably, it is possible to relate $\mathbb{P}_{p,q}[x \nleftrightarrow y]$ to $\mathbb{P}_{p,q}[(x,y) \in B]$, the probability that $e=(x,y)$ is a bridge. This follows from a natural bijection between the set of configurations contributing to the events $\{x \nleftrightarrow y\}$ and $\{e \in B\}$. More precisely let

$$\Omega_{x,y} \equiv \{A \subseteq E | x \nleftrightarrow y \text{ in } (V,A)\}, \tag{5.4.3}$$

$$\Omega_e \equiv \{A \subseteq E | e \in B(A)\}, \tag{5.4.4}$$

where $x,y \in V$ and $e \in E$. Now simply consider the mapping between $\Omega_{x,y}$ and $\Omega_{(x,y)}$ for $x,y \in V$ such that $(x,y) \in E$

$$A \in \Omega_{x,y} \mapsto A^e \in \Omega_{(x,y)}. \tag{5.4.5}$$

We recall that $A^e \equiv A + \{e\}$. Now by the nature of the probability measure of the random-cluster model one has

$$\begin{aligned}
\mathbb{P}_{p,q}[A] &= \frac{1}{Z_{p,q}} q^{K(A)} p^{|A|} (1-p)^{m-|A|}, \\
&= \frac{1}{Z_{p,q}} q^{K(A^e)+1} p^{|A^e|-1} (1-p)^{m-(|A^e|-1)}, \\
&= \frac{q(1-p)}{p} \mathbb{P}_{p,q}[A^e].
\end{aligned} \tag{5.4.6}$$

valid for any $A \in \Omega_{x,y}$ with $e=(x,y)$. This in turn allows us to conclude

$$\begin{aligned}
\mathbb{P}_{p,q}[x \nleftrightarrow y] &= \sum_{A \subseteq E} \mathbb{P}_{p,q}[A] \mathbb{1}_{\{x \nleftrightarrow y\}}(A), \\
&= \sum_{A \in \Omega_{x,y}} \mathbb{P}_{p,q}[A], \\
&= \frac{q(1-p)}{p} \sum_{A \in \Omega_{x,y}} \mathbb{P}_{p,q}[A^e], \\
&= \frac{q(1-p)}{p} \sum_{A \in \Omega_{(x,y)}} \mathbb{P}_{p,q}[A], \\
&= \frac{q(1-p)}{p} \sum_{A \subseteq E} \mathbb{P}_{p,q}[A] \mathbb{1}_{\{e \in B(A)\}},
\end{aligned}$$

$$= \frac{q(1-p)}{p}\mathbb{P}_{p,q}[e \in B]. \tag{5.4.7}$$

The above shows that $\mathbb{P}_{p,q}[e \in \overline{\mathcal{B}}] = \frac{q(1-p)}{p}\mathbb{P}_{p,q}[e \in \mathcal{B}]$, and therefore, using (5.1.8), we obtain the general result

$$\mathbb{E}_{p,q}[\overline{\mathcal{B}}] = \frac{q}{1-q}\left(\frac{\mathbb{E}_{p,q}[\mathcal{N}]}{p} - 1\right). \tag{5.4.8}$$

This has direct consequences. Firstly, the finite size corrections of $\mathbb{E}_{p,q}[\overline{\mathcal{B}}]$ are completely determined by the ones of $\mathbb{E}_{p,q}[\mathcal{B}]$. Secondly, the same conclusion applies to $\mathbb{E}_{p,q}[\overline{\mathcal{C}}]$, due to

$$\mathbb{E}_{p,q}[\overline{\mathcal{C}}] = 1 - \mathbb{E}_{p,q}[\mathcal{N}] - \mathbb{E}_{p,q}[\overline{\mathcal{B}}].$$

Lastly, one can infer from (5.2.2) the corresponding asymptotic values for the square lattice:

$$\mathbb{E}_q[\overline{\mathcal{B}}] = \frac{\sqrt{q}}{2(1+\frac{1}{\sqrt{q}})}, \quad \mathbb{E}_q[\overline{\mathcal{C}}] = \frac{1}{2(1+\sqrt{q})}. \tag{5.4.9}$$

We pause to note that these results, together with the bridge and non-bridge results (5.2.2), have a very intuitive basis. Firstly, we remark that the density of both, open and closed, edges are, asymptotically, independent of q for $\mathbb{Z}^2$, and equal to 1/2. Thus we can equivalently study the relative densities of $\mathcal{B},\mathcal{C},\overline{\mathcal{B}}$ and $\overline{\mathcal{C}}$, which equal $1-p_{\mathrm{sd}}(q)$, $p_{\mathrm{sd}}(q)$, $p_{\mathrm{sd}}(q)$ and $1-p_{\mathrm{sd}}(q)$, respectively. Now, by increasing q, open edges are typically more and more likely to be non-bridges in the same manner as closed edges become more and more likely candidate-bridges. This clearly implies the opposite effect for bridges and candidate-non-bridges. Thus, by increasing q, the overlap between clusters is increased, which relates to a larger number of candidate-bridges. In other words, the increase of q strengthens the connection within clusters, at the "cost" of increasing overlaps between clusters (increasing the (relative) density of candidate-bridges). This reflects the interplay of K and $|A|$ in the probability weight of A, equal to $q^{K+|A|/2}$, for $p_{\mathrm{sd}}(q)$. In order to increase the weight, either K needs to be increased, going typically along with a decrease of $|A|$, or $|A|$ increases, possibly decreasing K. However increasing $|A|$ by relocating bridges to positions of candidate-non-bridges, one can increase K by conserving $|A|$. Increasing q enforces this particular relocating effect more and more.

The connection between bridge-density and connectivity further allows us to recover a recent result of Hu et al., [114], where the authors determine that for $L \to \infty$ on $\mathbb{Z}_L^2$ one has, due to translational invariance, for any $(x,y) \in E$ $\mathbb{P}_q[x \leftrightarrow y] \to 3/4$. Recall, the authors of [55] established that the asymptotic bridge density for critical percolation on $\mathbb{Z}_L^2$ is 1/4 (or consider (5.2.2)). This is clearly consistent with (5.4.7). Moreover in [114] it was shown that for the random-cluster model on $\mathbb{Z}_L^2$, with integer $q \geq 1$, one has:

$$\mathbb{P}_q[x \leftrightarrow y] \to \frac{2+\sqrt{q}}{2\left(1+\sqrt{q}\right)} \text{ for } L \to \infty \tag{5.4.10}$$

for arbitrary $(x,y) \in E$. It is straightforward to confirm that this result can be derived from (5.4.8) and (5.4.9).
As another application of (5.4.7) we can derive the expected bridge-density for the random-cluster model on the cycle graph with n vertices, denoted by C_n. Here, one can show, after some straightforward algebra [69] (see also the Appendix B), that

$$\mathbb{P}_{p,q}[x \nleftrightarrow y] = \frac{\tilde{p}(p,q)^n - \tilde{p}(p,q)^{n-1} - \tilde{p}(p,q) + 1}{1 + (q-1)\tilde{p}(p,q)^n} \to 1 - \tilde{p}(p,q) \tag{5.4.11}$$

for $n \to \infty$ and any pair x, y of neighbouring vertices. We therefore obtain for C_n in the limit $n \to \infty$:

$$\mathbb{P}[e \in B] \to \tilde{p}(p,q) = \frac{p}{p + (1-p)q}.$$

This reflects the intuition that for $n \to \infty$ the cycle graph becomes effectively a path and thus a tree, for which all edges are bridges. Recall, that the random-cluster model on a tree with parameters p and q is equivalent to independent bond percolation with parameter $\tilde{p}(p,q)$. Further the above limit matches the upper bound (5.1.11). Moreover, the asymptotic value is approached, for any choice of p and q, from below. This reflects the intuition that the finite cycle is more strongly connected than the infinite one, because it can still "sense" the topology of the cycle. However for $n \to \infty$ the influence of the boundary condition becomes irrelevant when concerned with the density of bridges. This is also consistent with what we observed for $\mathbb{Z}_L^2$ in particular (5.3.11). Now, in order to see that the asymptotic bridge-density is approached from below for C_n, it is helpful to distinguish the two cases $q \geq 1$ and $q < 1$. Showing the claim for $q \geq 1$ is straightforward and we omit the details. For the case $q < 1$ we note that one can re-write (5.4.11) as

$$\mathbb{P}_{p,q}[x \nleftrightarrow y] = \frac{\big[1 - \tilde{p}(p,q)\big]\big[\tilde{p}(p,q) - \tilde{p}(p,q)^n\big]}{(q-1)\tilde{p}(p,q)^n + \tilde{p}(p,q)}$$

which in turn with $0 \leq \tilde{p}(p,q) - \tilde{p}(p,q)^n \leq 1$ and $(q-1)\tilde{p}(p,q)^n + \tilde{p}(p,q) \geq \tilde{p}(p,q) - \tilde{p}(p,q)^n \geq 0$ yields

$$\mathbb{P}_{p,q}[x \nleftrightarrow y] \leq 1 - \tilde{p}(p,q)$$

and hence eventually

$$\mathbb{P}_{p,q}[e \in B] = \frac{p}{q(1-p)}\mathbb{P}_{p,q}[x \nleftrightarrow y] = \frac{p}{q(1-p)}\big[1 - \tilde{p}(p,q)\big] \leq \tilde{p}(p,q).$$

5.5 Bridge fluctuations

In combination with results from the Coulomb gas and the solution of the critical Potts model in two dimensions, the bridge-edge formula (5.1.8) provides a rather complete understanding of the expected behaviour of the critical bridge density. Additionally, higher moments of the bridge distribution can be discussed with similar techniques (however with more involved combinatorial/geometric interpretations of the involved

terms) as we will show now for the example of the variance. This allows us to study the scale of fluctuations in $|B|$. Somewhat naively, knowing that the first moments of the bridge- and edge- distribution are linearly related, one might expect that the variance $\mathrm{Var}_{p,q}[|B|]$ is governed by the fluctuations of N. Going further, one might then anticipate in this case that the critical variance would follow $\mathrm{Var}_q[|B|]/m \approx L^{\alpha/\nu}$, which in turn would imply a divergence with L for $q \geq 2$ and a "saturation" to a constant for $q < 2$ [15]. However, as we will show now the story is not quite as simple.

To work out the scale of fluctuations in the number of bridges $|B|$, we apply the Russo-Margulis formula to the second derivative of the partition function $Z_{p,q}$, which we then in turn equate with the expression one obtains by explicit differentiation. To start, with we have, using (5.1.6),

$$\partial_p^2 Z_{p,q} = (1-q)\partial_p \left(\frac{Z_{p,q}}{p}\mathbb{E}_{p,q}[|B|]\right) = -Z_{p,q}\frac{1-q}{p^2}\mathbb{E}_{p,q}[|B|] + \frac{1-q}{p}\partial_p \mathbb{E}_p\left[q^K|B|\right]$$

Let us now focus on the second term:

$$\begin{aligned}\frac{1-q}{p}\partial_p \mathbb{E}_p\left[q^K|B|\right] &= \frac{1-q}{p}\sum_{e,f\in E}\mathbb{E}_p\left[\delta_e\left(q^K \mathbb{1}_{\{f\in B\}}\right)\right], \\ &= \frac{1-q}{p^2}\sum_{e,f\in E}\sum_{\substack{A\subseteq E:\\ e\in A}}\mathbb{P}_p[A]\left[q^{K(A)}\mathbb{1}_{\{f\in B(A)\}} - q^{K(A_e)}\mathbb{1}_{\{f\in B(A_e)\}}\right] \qquad (5.5.1)\end{aligned}$$

Here we used the fact that the derivative δ_e for $e \in E$ considered as an operator from $\mathbb{R}^{|\Omega|} \to \mathbb{R}^{|\Omega|}$ is linear. We can split the inner sum into two sums corresponding to $e \in B(A)$ and $e \in C(A)$, of which we consider the former first

$$\frac{1-q}{p^2}\sum_{e,f\in E}\sum_{\substack{A\subseteq E:\\ e\in B(A)}}\mathbb{P}_p[A]\left[\mathbb{1}_{\{e=f\}}q^{K(A)}\mathbb{1}_{\{f\in B(A)\}} + \mathbb{1}_{\{e\neq f\}}q^{K(A)}(1-q)\mathbb{1}_{\{f\in B(A)\}}\right].$$

A few comments are in order. If $e = f$, we clearly have that $e = f \notin B(A_e)$, hence we recover the expected bridge density. For $e \neq f$ it is important to observe that removing a bridge e cannot influence the pivotality of an occupied edge f, in particular when $f \in B(A)$ then also $f \in B(A_e)$, whenever e is a bridge in A. The above can therefore be re-written as

$$\frac{1-q}{p^2}Z_{p,q}\mathbb{E}_{p,q}[|B|] + \frac{(1-q)^2}{p^2}Z_{p,q}\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e\in B, f\in B].$$

For the non-bridge contribution to the inner sum in (5.5.1) note that only summands with $e \neq f$ contribute, because otherwise if $e \in C(A)$ then trivially $f = e \notin B(A), B(A_e)$. Furthermore, if $e \neq f$ such that $e \in C(A)$ and $f \in B(A)$ then we have also a vanishing contribution because deleting a non-bridge cannot change the fact that f is a bridge. Thus, there can only be a contribution for a configuration A such that $e \in C(A)$ and $f \in C(A)$ as well as both edges are in one cycle and deleting e will destroy the cycle and hence cast f into a bridge in A_e. Moreover, both edges e and f must only be in one "linearly

independent" cycle (imagine two clusters glued together in parallel by the edges e and f, hence in particular there are no additional links between the two clusters). Write $e \overset{A}{\Leftrightarrow} f$ for the above event involving edge e and f. We obtain for the second term:

$$-Z_{p,q}\frac{1-q}{p^2}\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f].$$

We therefore obtain eventually for $\partial_p^2 Z_{p,q}$:

$$\begin{aligned}\partial_p^2 Z_{p,q} &= Z_{p,q}\frac{1-q}{p^2}\left((1-q)\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e\in B, f\in B]-\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f]\right)\\ &= Z_{p,q}\frac{1-q}{p^2}\left((1-q)\mathrm{Var}_{p,q}[|B|]+(1-q)\mathbb{E}_{p,q}[|B|]^2-(1-q)\mathbb{E}_{p,q}[|B|]-\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f]\right)\end{aligned}$$

On the other hand one can show by explicit differentiation and after some straightforward but tedious algebra that:

$$\frac{1}{Z_{p,q}}\partial_p^2 Z_{p,q} = \frac{\mathbb{E}_{p,q}[N](2p-1-2pm)-mp^2+m^2p^2+\mathrm{Var}_{p,q}[N]+\mathbb{E}_{p,q}[N]^2}{p^2(1-p)^2}$$

Equating both expressions for $\partial_p^2 Z_{p,q}$ with subsequent rearranging yields

$$\begin{aligned}\mathrm{Var}_{p,q}[|B|] = {}& \frac{\mathbb{E}_{p,q}[N](2p-1-2pm)-mp^2+\mathrm{Var}_{p,q}[N]+\left(\mathbb{E}_{p,q}[N]-mp\right)^2+2mp\mathbb{E}_{p,q}[N]}{(1-q)^2(1-p)^2}\\ &-\mathbb{E}_{p,q}[|B|]^2+\mathbb{E}_{p,q}[|B|]+\frac{1}{1-q}\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f]\end{aligned}$$

This can be further simplified by using the bridge edge identity (5.1.8), which leads to

$$\frac{\mathrm{Var}_{p,q}[|B|]}{m} = \frac{\mathbb{E}_{p,q}[\mathcal{N}](2p-1)-p^2+\mathrm{Var}_{p,q}[N]/m}{(1-q)^2(1-p)^2}+\mathbb{E}_{p,q}[\mathcal{B}]+\frac{1}{1-q}\frac{1}{m}\sum_{e\neq f\in E}\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f]. \tag{5.5.2}$$

We emphasise that (5.5.2) is an exact result valid for any p and q as well as any graph, thus it has the same range of applicability as the bridge-edge identity (5.1.8). Furthermore, we remark that for $e=(u,v)$ and $f=(x,y)$ we have by a bijection argument, similar to the one used in Sec. 5.4 to derive the relationship between the bridge density and nearest neighbour connectivity:

$$\mathbb{P}_{p,q}[e \overset{A}{\Leftrightarrow} f] = \frac{p^2}{(1-p)^2}\frac{1}{q}\left\{1-\mathbb{P}_{p,q}[(u,v)\nleftrightarrow(x,y)]\right\}$$

Here $1-\mathbb{P}_{p,q}[(u,v)\nleftrightarrow(x,y)]$ for $(u,v),(x,y)\in E$ is the probability that the two nearest neighbour pairs belong to two different clusters, such that the two distinct clusters each contain one vertex from $\{u,v\}$ and one vertex from $\{x,y\}$ (see Fig. 1(*c*) in Ref. [110]). In what

follows we focus on the self-dual line for the RC model on $\mathbb{Z}_L^2$ and study the continuum limit. As shown by Vasseur *et al.* [110] one has for two pairs of neighbouring vertices (u,v) and (x,y) at distance r the following asymptotics for large r:

$$1-\mathbb{P}_q[(u,v)\leftrightsquigarrow(x,y)]\sim A(q)r^{-2x_2(q)}.$$

This follows from the construction of a four-leg watermelon event, due to the four hulls propagating from the neighbourhood of (u,v) to the neighbourhood of (x,y), which are associated to the two clusters involved. We hence expect the following asymptotic behavior

$$\begin{aligned}\frac{1}{m}\sum_{e\neq f\in E}\mathbb{P}_q[e\Leftrightarrow f]&=\frac{1}{m}\sum_{(u,v)\neq(x,y)\in E}\Big\{1-\mathbb{P}_q[(u,v)\leftrightsquigarrow(x,y)]\Big\},\\&\approx A(q)\iint_1^{\frac{L}{2}}\mathrm{d}r2\pi r r^{-2x_2(q)},\\&=\alpha(q)L^{2-2x_2(q)}+\beta(q),\end{aligned}\tag{5.5.3}$$

where $\alpha(q)$ and $\beta(q)$ are two q dependent constants.

Inspecting the form (5.5.2), we hence see that the normalised variance of the bridge density has two contributions, one proportional to the normalised variance of the number of edges that scales as $L^{\alpha/\nu}$ at criticality, and another term related to the above mentioned watermelon event with scaling proportional to L^{2-2x_2}. Whenever $L^{\alpha/\nu}$ is the dominant contribution, the situation hence follows the naive expectation outlined above. Inspection of the Coulomb gas values of the exponents shows that this is the case for $q\geq 1$. For $q>2$, this leads to a divergence of $\mathrm{Var}_{p,q}[|B|]/m$. For $q<1$, however, the leading term is proportional to L^{2-2x_2}. Remarkably, our result[7] shows that for

$$q<\tilde{q}=4\cos^2\left(\pi/\sqrt{3}\right)=0.2315891\cdots$$

the normalised variance of $|B|$ becomes unbounded, that is it diverges with the exponent $2-2x_2$. We emphasise that this is not only quantitatively different from what the heat-capacity based arguments would predict ($\alpha/\nu\neq 2-2x_2$), but also a qualitatively different result. The situation is summarised in the lines of Fig. 5.9 showing the Coulomb gas values of α/ν and $2-2x_2$. We also analysed the variance of the bridge density numerically. The fit functions and resulting parameters are summarised in Table 5.2, and the corresponding parameter estimates are indicated by the symbols in Fig. 5.9. Clearly, we find excellent agreement with the predictions from Eq. (5.5.2) discussed above. For the marginal value $\tilde{q}=0.2315891\cdots$ we expect a logarithmic divergence, and we indeed find the corresponding form to yield the best fit to our simulation data.

[7] This value $\tilde{q}$ follows from the condition that $x_2(\tilde{q})=1$. Using $x_2=g/2-(g-4)^2/(8g)$ [15] where g is the Coulomb gas coupling, we find that this happens for $\tilde{g}=\pm 4/\sqrt{3}$. Thus taking only the positive solution we find, because of $q=2+2\cos(g\pi/2)$, that the solution is $\tilde{q}=2+2\cos(2\pi/\sqrt{3})=4\cos(\pi/\sqrt{3})^2$.

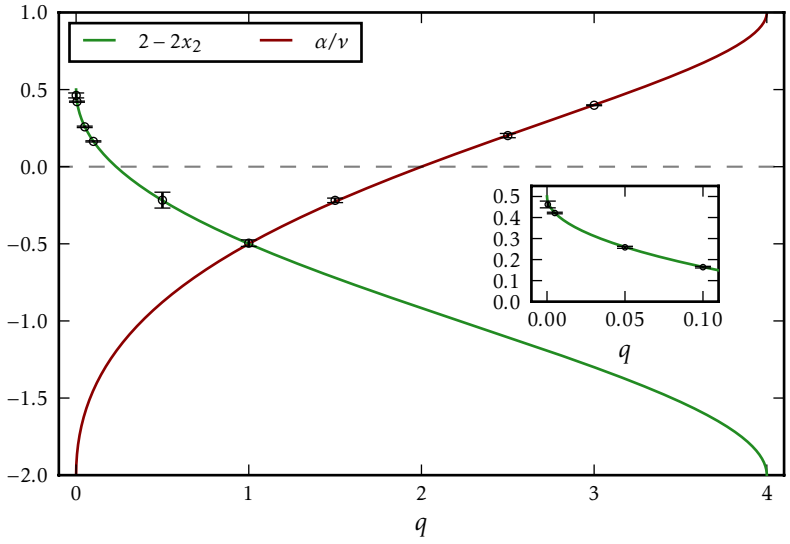


Figure 5.9: *The exponents α/ν and $2-2x_2$ appearing in the system size scaling of $\mathrm{Var}_q[|B|]/m$ at criticality. The lines show the exact values following from the Coulomb gas mapping. The symbols denote our numerical estimates from fitting to the variance of the bridge density, cf. the data collected in Table 5.2.*

It remains to discuss the percolation case $q = 1$ where the bridge-edge identity (5.1.8) becomes singular and the above derivation hence needs to be revisited. Here we derive the singular behaviour based on another bijection argument that allows us to harness recent results on logarithmic observables emerging from a careful analysis of the appropriate logarithmic conformal field theory description of critical percolation [110, 114]. Before we do so, note that in order to extract the asymptotic scaling of the variance it suffices to study the covariance $\eta_{e,f} \equiv \mathbb{P}_{p,q}[e \in B, f \in B] - \mathbb{P}_{p,q}[e \in B]\mathbb{P}_{p,q}[f \in B]$, which relates to $\mathrm{Var}_{p,q}[|B|]$ via the well known identity

$$\mathrm{Var}_{p,q}[|B|] = \sum_{e,f \in E} \eta_{e,f}.$$

To start with, recall that we consider critical bond percolation on $\mathbb{Z}_L^2$, i.e. $q = 1$ and $p = 1/2$ and write $\mathbb{P}[\cdot]$ for $\mathbb{P}_{1/2,1}[\cdot]$. Furthermore, note that $\mathbb{Z}_L^2$ is a transitive graph, and hence none of the following events depend on the explicit edge or vertex, used in the arguments. Now, fix two edges $e = (x_1, y_1), f = (x_2, y_2)$ that are distance $r \ll L$ apart. Consider the event $\{e \in B \wedge f \in B\}$. All configurations contributing to this event can be further sub-divided into two events, depending on whether e and f belong to the same connected component in (V, A) or not. Denote the two events by Ω_1 and Ω_2, respectively. Choose a configuration A that belongs to Ω_2, i.e. the two edges e and f are bridges in (V, A) and belong to two different connected components. The crucial point is that we can relate A one-to-one to $A - \{e, f\}$, a configuration where x_1, y_1, x_2, y_2 belong to four different clusters. Denote all configurations in which the four vertices belong to four different components by $\tilde{\Omega}_2$ (this is a event). Due to the choice of $q = 1$ and $p = 1/2$ we have that $\mathbb{P}[\Omega_2] = \mathbb{P}\left[\tilde{\Omega}_2\right]$. Let us now

q	Exponent (C)	χ^2_{pdf}	$L_{\min}$	Model	α/ν	$2-2x_2$
0.0005	0.462(16)	1.01	12	$A+BL^C$	−1.9576	**0.4752**
0.005	0.422(3)	1.15	4	$A+BL^C$	−1.8679	**0.4222**
0.05	0.258(5)	0.58	16	$A+BL^C$	−1.6005	**0.2599**
0.1	0.164(4)	0.76	8	$A+BL^C$	−1.4492	**0.1648**
0.231589	–	1.01	8	$A+B\log(L)$	−1.1962	0
0.5	−0.217(51)	1.4	12	$A+BL^C+DL^{\alpha/\nu}$	−0.8778	**−0.2191**
1	−0.496(21)	0.7	6	$A+BL^C[1+D\log(L)]$	−0.5	−0.5
1.5	−0.218(15)	1.05	4	$A+BL^C+DL^{2-2x_2}$	**−0.2266**	−0.7205
2.5	0.202(14)	0.94	48	$A+BL^C$	**0.2036**	−1.1052
3	0.398(3)	1.14	6	$A+BL^C$	**0.4**	−1.3

Table 5.2: *Numerical results for the leading exponent in the finite-size scaling of* $\mathrm{Var}_q[|B|]$. *The two rightmost columns show the exact values obtained from the Coulomb gas mapping. For the two cluster weights* $q = 1.5$ *and* $q = 0.5$ *we also performed a fit to the form* $A + BL^C$, *which yielded slightly worse results, due to the proximity of the two (negative) exponents. In all cases the quality-of-fit* Q *was at least 5%.*

consider the event Ω_1, i.e. the set of configurations for which e and f belong to the same component and both e and f are bridges. Now, because both edges are pivotal we can relate any such configuration $A \in \Omega_1$ one-to-one to a configuration where x_1 and y_1 as well as x_2 and y_2 are disconnected, and for which x_1, y_1, x_2, y_2 belong to three different clusters, of which one cluster contains one vertex of $\{x_1, y_1\}$ and one vertex of $\{x_2, y_2\}$. Denote the corresponding event $\tilde{\Omega}_1$. Note that any configuration $A \in \Omega_1$, where e and f are bridges belonging to the same cluster, must yield 3 disconnected clusters in $A' = A - \{e, f\}$. This is because the alternative case of 2 disconnected clusters in A' would imply that e and f are in a cycle in A, which is obviously a contradiction. As before we have $\mathbb{P}[\Omega_1] = \mathbb{P}\left[\tilde{\Omega}_1\right]$. Finally, the probabilities $\mathbb{P}\left[\tilde{\Omega}_1\right]$ and $\mathbb{P}\left[\tilde{\Omega}_2\right]$ were studied[8] in [110] in the framework of a corresponding logarithmic conformal field theory. We note that in the continuum limit both probabilities $\mathbb{P}[\Omega_1]$ and $\mathbb{P}[\Omega_2]$ only depend, to leading order, on r. Now because we have $\mathbb{P}[e \in B, f \in B] = \mathbb{P}[\Omega_1] + \mathbb{P}[\Omega_2]$ we obtain by falling back to [110]:

$$\eta_{e,f} \sim (a + b\log(r))\, r^{-2x_2} \Rightarrow \frac{\mathrm{Var}_{1/2,1}[|B|]}{m} \sim (a' + b'\log(L))\, L^{2-2x_2}, \tag{5.5.4}$$

where $x_2 = 5/4$, the two-arm exponent for critical percolation, and a, b, a', b' are constants. Our numerical analysis confirmed this scaling, as shown in Table 5.2 and Figure 5.9. Thus considering the variance of bridges for critical percolation yields yet another manifestation of the underlying logarithmic conformal field theory [110].

Lastly, it is an interesting open problem to check whether the observation that the $q < 1$ regime is governed by an exponent different from the α/ν heat-capacity value extends

[8] Vasseur et al. denote $\mathbb{P}[\tilde{\Omega}_1]$ by $\mathbb{P}_1(r)$ and $\mathbb{P}[\tilde{\Omega}_2]$ by $\mathbb{P}_0(r)$.

to higher dimensions, i.e. is this a inherent property of the $q < 1$ random-cluster model or rather a pecularity of (short-range) two dimensional systems, to which the above arguments naturally extend by means of universality[9].

5.6 Fragment sizes

In this section we focus on criticality and investigate how a cluster fragments upon deletion of a bridge, that is we study the statistics of the fragment size distribution associated to the deletion of a bridge. Such fragmentation phenomena were analysed for critical bond percolation in e.g. [29, 115, 30], and for continuum percolation clusters in [116]. Here we go beyond uncorrelated bond-percolation and study, for the critical random-cluster model, the size of the fragments formed by the removal of a bridge. More precisely we study the statistics of the quantity $\mathcal{D}_e(A)$, defined as the size of the smaller of the two clusters formed by the removal of edge e, in case it is a bridge, and 0 otherwise, i.e:

$$\mathcal{D}_e(A) \equiv \mathbb{1}_{\{A\in\Omega_e\}} \min\{C_x(A_e), C_y(A_e)\},$$

where Ω_e is defined in (5.4.4) as the set of configurations for which e is a bridge, and $C_x(A)$ generically denotes the size of the connected component in the graph (V, A), to which vertex x belongs. Our first step was to investigate the expectation $\mathbb{E}_q[\mathcal{D}_e]$, which can also be written as

$$\mathbb{E}_q[\mathcal{D}_e(A)] = \mathbb{P}_q[e \in B]\mathbb{E}_q[\min\{C_x(A_e), C_y(A_e)\}|e \in B],$$

that is we consider a conditional expectation, where we condition on e being a bridge. Further, because we used Monte-Carlo simulations to estimate the above expectation, we exploited the translational invariance of $\mathbb{Z}_L^2$ to improve the estimates. More specifically, we have that $\mathbb{E}_q[\mathcal{D}_e]$ is independent of e and hence equal to $\mathbb{E}_q[\overline{\mathcal{D}}]$, where $\overline{\mathcal{D}}(\cdot) = \frac{1}{m}\sum_{e\in E}\mathcal{D}_e(\cdot)$. The fractal, self-similar, structure at criticality suggests that this quantity scales as a power law in L, the linear dimension, i.e. we expect, in leading order, $\mathbb{E}_q[\overline{\mathcal{D}}] \approx L^y$, with a yet to be determined exponent y. We remark that y is however constrained by $y \leq \gamma/\nu$. This follows from the observation, that the expected cluster size scales at criticality, for a system with linear dimension L, as $L^{\gamma/\nu}$, c.f. e.g. [46, 15, 26]. Moreover, as already discussed in Chapter 4, this γ/ν scaling causes a drastic computational slowing down of, both, the sequential breadth/depth-frist search and interleaved breadth/depth-first-search-union-find based implementations of Sweeny's algorithm.

To start with, we fitted various scaling ansätze to our estimates for $\mathbb{E}_q[\overline{\mathcal{D}}]$, in order to extract the leading size dependence. We find that the best fit results were obtained by fitting a simple power law, L^y, to the data, with an appropriate lower system size cut-off, to account for possible unconsidered corrections. Interestingly, our data is completely consistent with $y = d_F - x_2$, which involves the cluster fractal dimension and the two-arm

[9] We expect the value of $\tilde{q}$ or its mere existence ($\tilde{q} > 0$) to be a dimensionality dependent question, which of course also deserves a further study.

exponent. For both, d_F and x_2, one has exact expressions [51], to which we compare our numerical estimates for y in Figure 5.10. The occurrence of the combination $d_F - x_2$ for

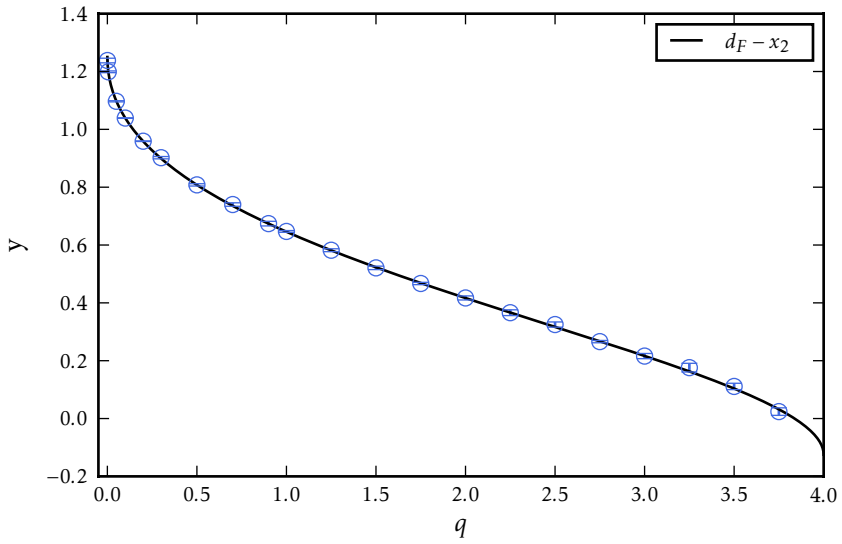


Figure 5.10: *Extracted exponents for $\mathbb{E}_q[\mathcal{D}]$, the expected smaller fragment size. The numerical estimates are compared to the predicted values for $d_F - x_2$ from Coulomb gas arguments.*

the smaller fragment size, closely resembles the finite-size dependence of the quantity $C_{\min;e}$, studied in [15], and which is defined by

$$C_{\min;e}(A) \equiv \mathbb{1}_{\{A \in \Omega_{x,y}\}} \min\{C_x(A), C_y(A)\}, \tag{5.6.1}$$

where $e = (x, y) \in E$ and $\Omega_{x,y}$, defined in (5.4.3), is the set of configurations for which vertex x and y are disconnected. The authors of [15] provided scaling arguments, supported by extensive simulations, that show that $\mathbb{E}_q[\overline{C}_{\min}] \approx L^{d_F - x_2}$, where $\overline{C}_{\min}(A) = \sum_{e \in E} C_{\min;e}(A)/m$. The identical scaling of quantities is not a coincidence, as we can show that both $\mathcal{D}$ and $C_{\min}$ have, up to a q and p dependent factor, the same expectation for any p and q. The reasoning is completely analogous to the observations that led to (5.4.7). In order to show this, note that we have $C_{\min;e}(A) = \mathcal{D}_e(A + e)$ for any $e = (x, y) \in E$ and $A \in \Omega_{x,y}$. Further we have relation (5.4.6) for our choice of A and x, y, such that we can conclude

$$\mathbb{E}_{p,q}[\overline{C}_{\min}] = \frac{q(1-p)}{p} \mathbb{E}_{p,q}[\overline{\mathcal{D}}]. \tag{5.6.2}$$

Being valid for all values of p and q, it holds in particular also at the critical point $p_{\mathrm{sd}}(q)$, and hence supports our numerical data that shows $\mathbb{E}_q[\overline{\mathcal{D}}] \approx L^{d_F - x_2}$. We emphasise that $d_F - x_2 < \gamma/\nu$ and hence $L^{d_F - x_2} \ll L^{\gamma/\nu}$. This suggests that the removal of a bridge predominantly creates very unequally sized fragments.

5.6.1 Scaling relations for the fragmentation break-up kernel

In order to understand the suppression of the smaller fragment size better, we analysed the probability of separating a cluster of mass s' from a cluster of mass s upon removal of a bridge, denoted by $b_{s'|s}$. This quantity was already analysed for critical bond-percolation in references [29, 115], where the authors, based on exact results in one dimension and for the Bethe lattice as well as exact enumeration methods and extensive simulation, proposed the following scaling form:

$$b_{s'|s} \sim s^{-\phi} \mathcal{G}\left(\frac{s'}{s}, \frac{s}{L^{d_F}}\right), \tag{5.6.3}$$

where various scaling relations for ϕ have been proposed, e.g.

$$\phi = \begin{cases} 2 - \sigma & [115, 29], \\ 1 + \left(d_H - \frac{1}{\nu}\right) d_F & [117], \\ \tau + \sigma - \frac{d_H}{d_F} & [118], \end{cases}$$

where σ is the crossover exponent of the cluster-size distribution [46] and d_H the hull fractal dimension [119]. The authors of [30] performed an extensive numerical analysis of fragmentation of percolation clusters for dimensions two to nine, and conclude that the original Edwards et al. relation [115]

$$\phi = 2 - \sigma = 2 - \frac{1}{\nu d_F}, \tag{5.6.4}$$

describes the data best. We remark that (5.6.4) is based on the implicit assumption that a cluster of mass s has typically a constant density of bridges, that is one has $a_s \approx s^{\lambda}$, with $\lambda = 1$, where a_s is the expected number of bridges on a cluster with masses s. Indeed, we recently showed in our letter [120], based on the bridge-edge formula (5.1.8) in combination with scaling-arguments, that $\lambda = 1$ is the only consistent possibility for the random-cluster model on any (finite) graph.

The (critical) random-cluster model setting provides us with another non-trivial testing ground of the validity of (5.6.4) (with suitable replacements of ν and d_F by q-dependent expression). Yet, somewhat surprisingly, our numerical analysis shows a deviation of (5.6.4) for $q \neq 1$. In what follows we will resolve this and derive the generically valid scaling-relation (5.6.9) for ϕ, which, in particular, reduces to (5.6.4) for $q = 1$. To start with, we, again, exploit the connection between $C_{\min,e}$ and $\mathcal{D}_e$. The authors of [15] analysed, in addition to moments of $C_{\min,e}$, also $\mathbb{P}[C_{\min,e} = s']$, the probability that $C_{\min,e}(A) = s'$, which due to translational invariance, does not depend on the particular edge $e \in E$, and we simply write $C_{\min}$ in what follows. The authors proposed the following scaling form, at criticality, for large but finite box, with dimension L:

$$\mathbb{P}[C_{\min} = s'] \sim s'^{-\psi_{C_{\min}}} \mathcal{F}\left(\frac{s'}{L^{d_F}}\right). \tag{5.6.5}$$

Further clever scaling arguments, supported by a careful numerical analysis, led the authors to the conclusion that $\psi_{C_{\min}} = \frac{x_2}{d_F} + 1$. We remark that this relation for $\psi_{C_{\min}}$ is consistent with $\mathbb{E}_{p,q}[\overline{C}_{\min}] \approx L^{d_F - x_2}$. Now, with identity (5.4.7) in mind, it is not hard to see that $\mathbb{P}[C_{\min} = s']$ must be related to the probability that $\mathcal{D}_e(A) = s'$, where e is an arbitrary edge in E. Indeed, we can show that valid for any transitive graph one must have,

$$\mathbb{P}[C_{\min} = s'] = \frac{q(1-p)}{q}\mathbb{P}[\mathcal{D} = s'], \tag{5.6.6}$$

which is the analogue of (5.6.2), now only on the level of distributions, and clearly more general than a relation for the first moment. Moreover we have, now for $L \to \infty$, and any $e = (x, y) \in E$:

$$\begin{aligned}
\mathbb{P}[\mathcal{D} = s'] &= \sum_{s=2s'}^{\infty} \Big(\mathbb{P}_q[\mathcal{D}_e(A) = s', C_x = s] \\
&= \sum_{s=2s'}^{\infty} \Big(\mathbb{P}_q[\mathcal{D}_e(A) = s' | C_x = s] \mathbb{P}_q[C_x = s] \\
&\approx \sum_{2s'}^{\infty} \Big(d_{s'|s} s^{1-\tau} \\
&\approx \int_{2s'}^{\infty} ds\, s^{1-\tau-\phi} \mathcal{H}\left(\frac{s'}{s}\right), \\
&\approx s'^{2-\tau-\phi}
\end{aligned} \tag{5.6.7}$$

A few comments are in order. Firstly, the sum starts from $2s'$, because in order to produce a *smaller* fragment of size s', the "mother" cluster must have size at least $2s'$. Secondly, we used the standard scaling ansatz $\mathbb{P}_q[C_x = s] \approx s^{1-\tau}$, c.f. e.g. [15, 26]. Thirdly, we replaced $d_{s'|s}$, the probability of a smaller fragment of size s' induced by a bridge-removal on a cluster of size s, by the scaling form, resembling (5.6.3), $d_{s'|s} \sim s^{-\phi}\mathcal{H}(s'/s)$. As we will show now, this form is a direct consequence of the scaling form (5.6.3).

Assume both s' and s are continuous and the random variable s' has probability density function $b_{s'|s} = s^{-\phi}\mathcal{G}(s'/s)$ for any $s \geq s'$. Now, clearly the fragmentation of a mother cluster with mass s results in two daughter fragments with masses s' and $s - s'$ (assuming no mass loss). Define $M = \min\{s', s - s'\}$ and note that for $m \leq s/2$:

$$\begin{aligned}
\mathbb{P}[M \geq m] &= \mathbb{P}[s' \geq m, s - s' \geq m], \\
&= \mathbb{P}[m \leq s' \leq s - m], \\
&= s^{-\phi} \int_{m}^{s-m} ds' \mathcal{G}(s'/s), \\
&= -s^{-\phi+1}\hat{\mathcal{G}}(m/s).
\end{aligned} \tag{5.6.8}$$

For the last step we defined $\hat{\mathcal{G}}(x) = -\int_x^{1-x} dy \mathcal{G}(y)$. It follows that we obtain for the probability density function $d_{m|s}$ of M, using a standard relation, c.f. e.g. [91]:

$$\begin{aligned} d_{m|s} &= -\frac{d}{dm}\mathbb{P}[M \geq m], \\ &= s^{-\phi}\mathcal{H}(m/s), \end{aligned}$$

with $\mathcal{H}(x) = -d/dx\hat{\mathcal{G}}(x)$, which shows the claim. Now, the intended scaling relation of ϕ follows directly by combining (5.6.5) and (5.6.7) in combination with the standard relation $\tau = 1 + d/d_F$, c.f. e.g. [46, 15]:

$$\phi = 2 - \frac{d - x_2}{d_F} = 2 - \frac{d_R}{d_F}. \tag{5.6.9}$$

We can also directly estimate ϕ from the system size scaling of $\mathbb{E}_q[\overline{\mathcal{D}}]$, because it is possible to derive a finite-size scaling relation for $\mathbb{E}_q[\overline{\mathcal{D}}]$ without referring to $b_{s'|s}$ or $d_{s'|s}$, that is in the scaling framework of e.g. [15], we obtain

$$\mathbb{E}_q[\overline{\mathcal{D}}] \approx L^y, \tag{5.6.10}$$

with $y = (3 - d/d_F - \phi)$, which allows us to extract ϕ by studying the size dependence of the numerically determined expected daughter cluster size $\mathbb{E}_q[\overline{\mathcal{D}}]$. In Figure 5.11 we compare our estimates for ϕ to the exact value predicted by (5.6.9), obtained from Coulomb-gas expressions for d_F and x_2 [15]. Our data is in perfect agreement with the

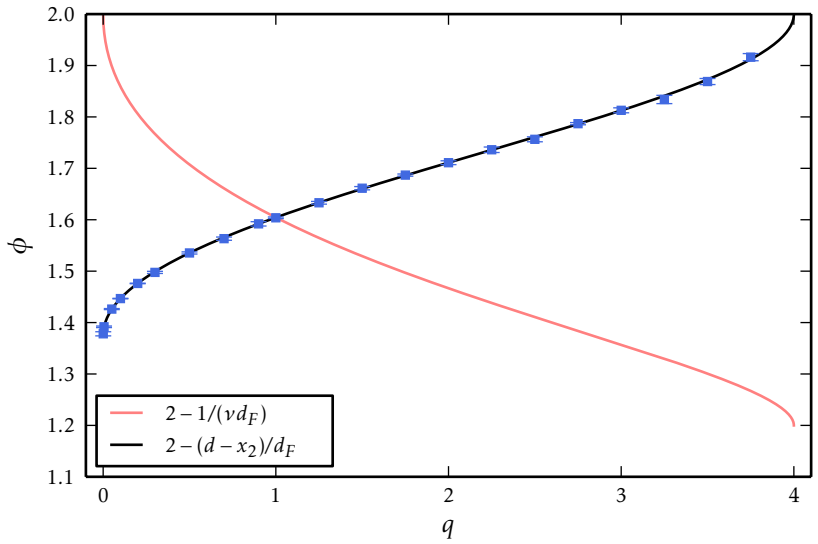


Figure 5.11: *Extracted exponent ϕ for the conditional break-up probabilities $b_{s'|s}$ and $d_{s'|s}$. The solid lines correspond to the scaling relations* (5.6.4) *and* (5.6.9)*, where the latter is clearly supported by our data.*

scaling relationship (5.6.9). Furthermore in the special case of $q \to 0$, on the square lattice, the exact value of ϕ is known to be 11/8 [121], which follows directly from (5.6.9) and

$d_F \to d = 2$ and $x_2 \to 3/4$ for $q \to 0$. For comparison, (5.6.4) states $\phi \to 2$, because of $1/\nu \to 0$ for $q \to 0$. This clearly contradicts [121]. As stated earlier, our result (5.6.9) recovers the $q = 1$ result (5.6.4) due to the following observation. It is well known, [12, 109, 15], that the two-arm exponent x_2 equals the scaling dimension of the red bonds, that is $d_R = d - x_2$, and one has for critical *percolation*, in any dimension,

$$x_2 = d - \frac{1}{\nu} \Leftrightarrow d_R = \frac{1}{\nu}. \tag{5.6.11}$$

This allows us to recover (5.6.4) from (5.6.9) for $q \to 1$. We emphasise that (5.6.11) is not valid for general values of $q \neq 1$, and hence both expressions (5.6.4) and (5.6.9) yield different exponents for for $q \neq 1$.

To gain some intuition for the scaling relation (5.6.9) it is useful to realise that s^{1/d_F} is in leading order the linear dimension of a percolation cluster of mass s. Furthermore, for a cluster of dimension ℓ one has, in average, roughly a number of ℓ^{d_R} red bonds. Therefore s^{d_R/d_F} is the typical number of red bonds on a percolation cluster of mass s. It is well known that that the red bonds are precisely the bridges that influence the cluster structure on a large scale [12]. This suggests that albeit there are bridges with a "small" impact on the cluster structure, they have no relevance for the exponent ϕ. The ratio d_R/d_F in a certain sense is an entropic factor in $d_{s'|s} \sim s^{d_R/d_F} s^{-2} \mathcal{H}(s'/s)$. The remaining exponent -2 in the power law in s corresponds, in a perfectly recursive mean field model of fragmentation, described in section 7.2.1, to a totally biased fragmentation or the surface erosion of completely dense objects.

As it is evident from (5.6.3), the distribution of daughter fragments $b_{s'|s}$ is not solely characterised by ϕ but also depends on the ratio s'/s of fragment (daughter) to mother cluster mass. This dependence is encoded into the function $\mathcal{G}$, which we now briefly analyse. The strong suppression of the average daughter cluster size relative to the size of the mother cluster suggests that fragmentation events are typically "abrasive" [120], which indicates that $\mathcal{G}(x)$ should be minimised around $x = 1/2$. We remark that the binary nature of bridge induced fragmentations enforces the symmetry $\mathcal{G}(x) = \mathcal{G}(1-x)$ [29]. Thus, it is sufficient to consider the interval $[0, 1/2]$. In a previous study, concerned solely with critical percolation ($q = 1$), in [29], the authors find that $\mathcal{G}(x)$ indeed encodes fragmentation of an abrasive nature. To understand how, and whether, the variation of q changes this effect, we numerically analysed $b_{s'|s}$ and hence inferred characteristics of $\mathcal{G}(x)$. Figure 5.12 shows the rescaled break-up kernel $b_{s'|s}$ for a system with $L = 1024$ and different values of q. For a given value of q, several regimes of mother cluster sizes s are shown in Figure 5.12. As for the $q = 1$ case, the data collapses, for fixed q, approximately to one master curve, which by (5.6.3) coincides with $\mathcal{G}(x)$. We directly observe that the creation of daughter cluster sizes comparable to the original mother cluster mass, $x \approx 1/2$, is suppressed in comparison to the production of unequally sized fragments, $x \approx 0$ (or $x \approx 1$). We remark that the plot, due to the log-scale in the horizontal direction, does not appear symmetric. Further, the abrasive nature, in other words the suppression of balanced fragmentations, becomes more pronounced for smaller cluster weights q. The

authors of [115] proposed a connection between the deepness of $G(x)$ and the branches, mainly based on the characteristics of $\mathcal{G}(x)$ for the z-coordinated Bethe tree, for which an exact solution exists. More precisely the model shows a deepening of $\mathcal{G}$ with increasing coordination number, which in turn is equivalent to an increased branching. This picture is supported by our numerical data that shows a deepening of the minimum of $\mathcal{G}$ with decreasing q. A consequence of the analysis of vertex fragmentation in the next section is that higher order fragmentation events become less and less likely. This suggests a reduced branching effect for increasing q, as long as $p = p_{\mathrm{sd}}(q)$. Unfortunately, we do not understand whether the branching argument is actually the only reason for the decreased likelihood of equally sized fragmentation events, with decreasing q, or if there are other effect causing it.

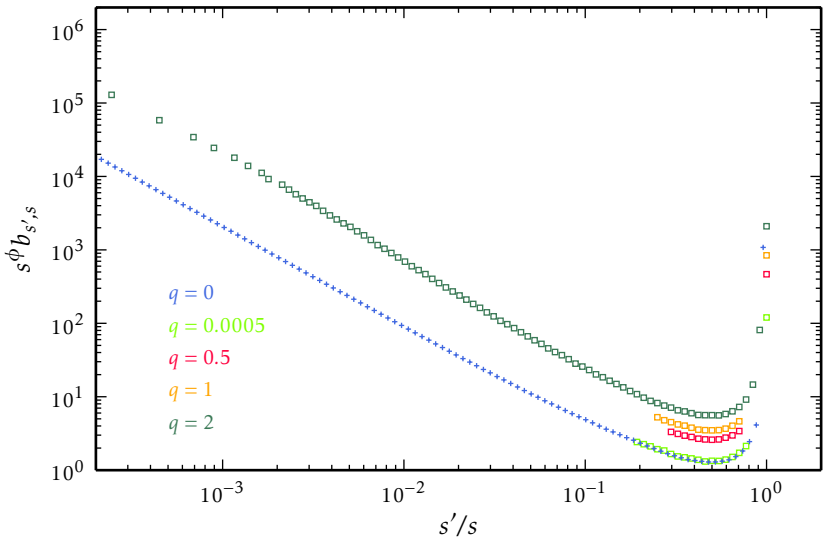


Figure 5.12: *Rescaled fragment size distribution for different mother cluster sizes s, corresponding to different symbols. The data for fixed q clearly support the scaling ansatz* (5.6.3). *Assuming the validity of* (5.6.3) *the curves approach the scaling function* $\mathcal{G}$..

5.7 Vertex fragmentation

A graph can be fragmented by removing a single edge, which happens precisely whenever the edge is a bridge. The above analysis of edge-fragmentation was substantially facilitated by the formulation of the random-cluster model as a correlated *bond*-percolation model, that is the fragmentation acts on the elementary degrees of freedom of the model. Yet, it is apparent that the removal of edges is not the only way a graph can fragment, as, for instance, the removal of a single vertex can also lead to a fragmentation. We therefore close this chapter with an analysis of certain aspects of vertex induced fragmentation. Let us mention already now, that we have (so far) not achieved a similar analytical clarity as for the bridge analysis, which is due to the bond formulation of the random-cluster model.

We therefore revert (mainly) to a purely numerical analysis, based on a sample generation by the Sweeny and Chayes-Machta-Swendsen-Wang algorithm followed by an algorithmic analysis using the cut vertex/bridge-detection algorithm presented in section 3.6.
To start with, we precisely define the vertices under study and adopt to this end graph theoretic terminology, see e.g. [47]. Formally, a *cut vertex* or *articulation point* is a vertex that upon removal increases the number of connected components. Clearly, any vertex that is incident to a bridge and has degree of at least 2 is a cut vertex. In addition to bridge-induced cut vertices, one can also have cut vertices not incident to any bridge, but at which two or more cycles overlap (and the cycles share no edge and vertex anywhere else). For an illustration consider Figure 5.13, where vertices e, f are bridge-induced cut vertices, and vertex c is a cycle-induced cut vertex. Further, we refer to a vertex incident with degree 1, as *dangling vertex*. Note that any vertex with degree 1 must be incident to a bridge. Considering again Figure 5.13, we have that only vertex g is dangling. Lastly, we refer to a vertex that is either a dangling or a cut vertex, as a *fragmenting vertex*. In Figure 5.14 we show the asymptotic densities[10] of cut and fragmenting vertices,

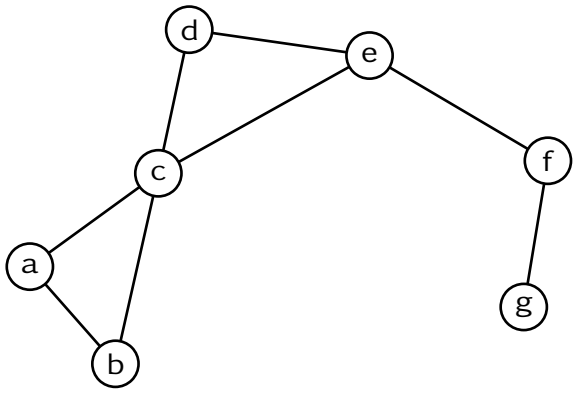


Figure 5.13: *Illustration of the cut vertex and fragmenting vertex definition. Vertices c, e, f are cut vertices, and g is a dangling vertex. Further, vertices e and f are bridge-induced cut vertices.*

obtained from a finite-size scaling analysis as described below. Analogous to our results for the bridge density, we observe that the density of both, cut or fragmenting vertices, is asymptotically constant. Indeed, because any bridge yields at least one fragmenting vertex, and we already established that the density of bridges is asymptotically constant, it follows in turn that the density of fragmenting vertices must also be asymptotically non-vanishing. Clearly, the difference of fragmenting and cut vertex densities corresponds to the density of dangling vertices, that is vertices with degree 1. For instance, for $q = 1$ it is easy to show that the density of dangling vertices must be constant, because one has,

[10] Here we normalise by the total number of vertices, e.g. $n = L^d$ for $\mathbb{Z}_L^d$.

fixing $p = 1/2$ and $G = \mathbb{Z}_L^2$, that a vertex has degree 0, 1, 2, 3, 4 with probability 1/16, 1/4, 3/8, 1/4, 1/16, respectively. Thus by translational invariance, this shows that, for any L, the density of dangling vertices is 1/4 for $q = 1$, at criticality. The same qualitative statement holds, albeit with different asymptotic constants and the existence of finite-size corrections, for $q \neq 1$ and $p \neq 1/2$, as we will show below. In other words, a constant fraction of all vertices is dangling. Our results are also consistent with exact results available for the spanning tree limit [121], where the authors, exploiting an equivalence to the Abelian sandpile model, show that the fraction of vertices with degree 1 equals $f_1 = (1 - 2/\pi)8/\pi^2 \approx 0.29454$. Any spanning tree has, by definition, the property that it contains all vertices (connected/spanning), and contains no cycles (tree). Thus all vertices with degree larger than one are bridge-induced cut vertices. This in turn implies that the fraction of cut vertices is asymptotically equal to

$$1 - f_1 = 1 - \frac{8}{\pi^2}\left(1 - \frac{2}{\pi}\right) \approx 0.705455,$$

which is highlighted by the dashed line in Figure 5.14. Unfortunately we have not been able to obtain an exact expression for $q > 0$. Our analysis of the size dependent corrections

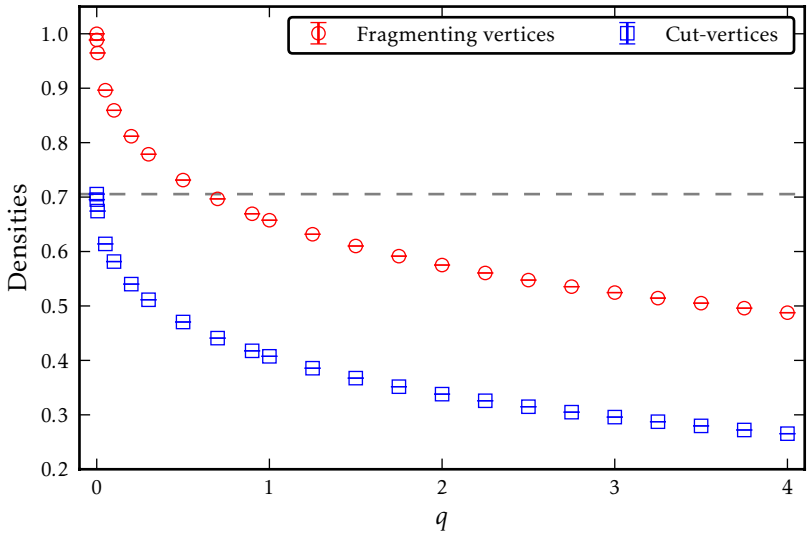


Figure 5.14: *Asymptotic density of fragmenting- and cut vertices. The horizontal dashed line shows the exact value of the cut vertex density for the spanning tree model.*

of the cut and fragmenting vertex densities at the self-dual point revealed the interesting observation that again the exponent $1/\nu - d$ describes the leading finite size correction of both densities. In Figure 5.15 we show the estimated exponents, together with the conjectured exact value. The previous discussion showed that the density of cut vertices consists of two contributions, namely bridge- and cycle-induced cut vertices. Now, we know that the leading finite size correction of the bridge density is $L^{1/\nu-2}$. Thus this picture suggests that the main effect is captured in the size-dependence of the bridge-density, and no other dominant size-dependent effect appears in the density of dangling

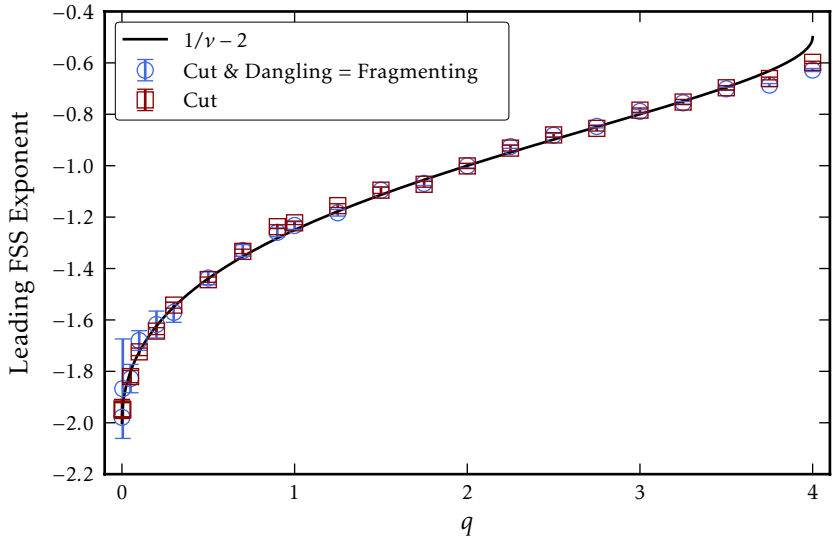


Figure 5.15: *Leading finite size correction exponent for the density of cut- and fragmenting vertices.*

vertices.

Let us now have a closer look and sub-divide the density of fragmenting vertices according to the following natural classification. To this end, note that in contrast to the edge case one is for the square lattice not restricted to binary break-up events, but rather has the possibility of *single*, *binary*, *ternary* and *quaternary* break-ups upon removal of a cut vertex. In general we say that a vertex is a type-ℓ cut vertex when, upon removal, ℓ daughter clusters (counting isolated vertices) replace the former mother cluster to which v belonged. We emphasise that not only all edges incident to v are removed but also v itself. Correspondingly, we call a break-up event ℓ-ary if it is induced by the removal of a type-ℓ cut vertex. Note that type-1 cut vertices are necessarily dangling vertices. Clearly, when one goes beyond $\mathbb{Z}_L^2$, to graphs with maximum degree $\Delta \geq 1$, one has $1 \leq \ell \leq \Delta$. However, as before, we focus in the following discussion on $\mathbb{Z}_L^2$, and have $\Delta = 4$. Now, we subdivide fragmenting vertices into four categories, corresponding to the degree of the break-up event. For a given configuration A, denote the set of type-ℓ cut vertices by F_ℓ and write $f_\ell = |F_\ell|/L^2$. We performed another finite-size scaling analysis, and extracted the leading exponent as well as the asymptotic densities. We observe that all densities have unsurprisingly a non-vanishing asymptotic mean fraction, as shown in Figure 5.16. As a reference, we can again use spanning tree results [121],

$$\begin{aligned} f_2 &= \frac{8}{\pi} - \frac{36}{\pi^2} + \frac{48}{\pi^3} \approx 0.44699, \\ f_3 &= 2 - \frac{16}{\pi} + \frac{48}{\pi^2} - \frac{48}{\pi^3} \approx 0.22239, \\ f_4 &= 1 - f_1 - f_2 - f_3 \approx 0.03608, \end{aligned}$$

where f_2, f_3, f_4 are the (expected) densities of vertices with degree 2, 3, 4, respectively, and hence for the uniform spanning tree model, equal precisely the density of binary, ternary and quaternary fragmenting-vertices. The corresponding values are clearly recovered by our numerical results for $q = 0$. Further it emerges from Figure 5.16, that the corresponding densities for $q > 0$ are below the $q \to$ limit and decrease with increasing q. Recall that the uniform spanning tree $q \to 0$ limit has the constraint that all vertices have to be connected, and changing q to a non-zero value, whilst remaining on the critical manifold $p_{\mathrm{sd}}(q)$, relaxes this connectedness constraint. Moreover, the bridge density decreases also

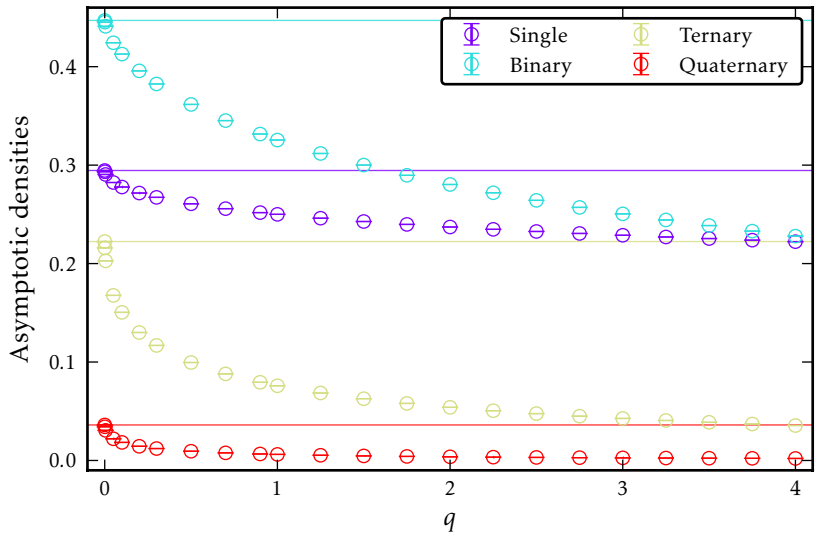


Figure 5.16: *Asymptotic densities of the different fragmenting-vertex classes, extracted from a finite-size scaling ansatz of the form $A + BL^{-C}$. For the corresponding value of C see Figure 5.17. The solid horizontal lines correspond to the known exact densities for the uniform-spanning tree model the square-lattice, following [121].*

with q, c.f. (5.2.2), implying a decrease of the density of bridge-induced cut vertices. The second feature we want to point out, is that the binary "branch" is dominating, which resembles the break-up induced by the removal of bridges. However, we remark that a binary cut vertex does not necessarily correspond to the removal of a bridge, as for instance one could have a vertex with degree 4, where the incident vertices are divided into two cycles, overlapping precisely at the corresponding cut vertex, which in turn is clearly a cycle-induced cut vertex. In general, the binary branch has the highest diversity, with cut vertices of degrees 2 to 4. For comparison, quaternary cut vertices uniquely correspond to vertices with degree 4, incident to 4 bridges, and ternary cut vertices have only two different possibilities (up to permutations of the vertices). Interestingly, the second largest contribution comes from dangling vertices. The more interesting observation, however, is that all four different fragmenting-vertex classes share the same finite-size correction exponent $1/\nu - d$ exponent, as summarised in Figure 5.17. The only, somewhat artificial, exception happens for $q = 1$, were we already showed, that the density of vertices of degree 1 has no finite-size corrections. This in turn implies that fragmenting vertices corresponding to single-break-up events, have also no size-dependent corrections. However, it is

plausible to assume, that this is a consequence of a vanishing amplitude, $B \to 0$, in front of the corresponding $L^{1/\nu-d}$ term, for $q = 1$.

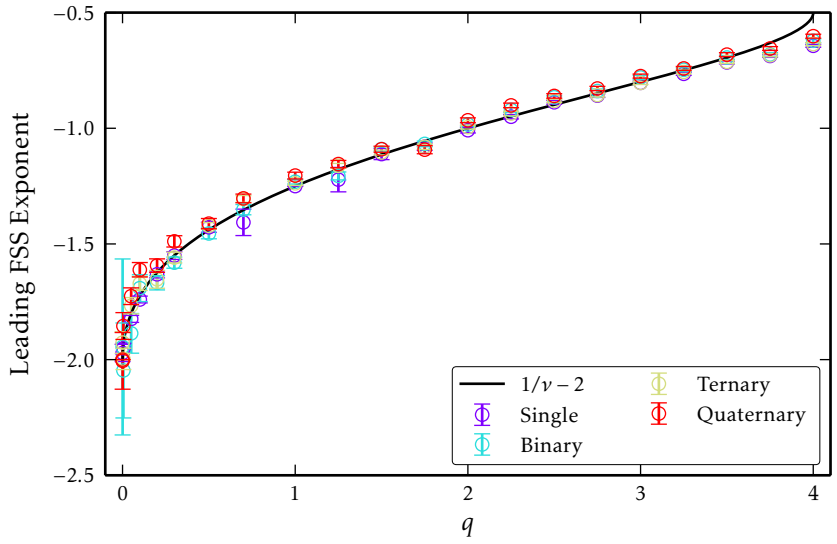


Figure 5.17: *Leading finite size correction exponent for each of the four fragmenting-vertex classes, extracted from a finite-size scaling ansatz of the form $A + BL^{-C}$. The solid line corresponds to the Coulomb-gas value of $1/\nu - d$, [51].*

5.7.1 Joint influence of multiple edges and cut vertices

In the remaining part of this section we provide theoretical arguments for the generic $L^{1/\nu-d}$ leading correction for all cut vertex type average densities. Our argument adopts some ideas used in the derivation of the bridge-edge identity (5.1.8) in section 5.1. Let us for the moment consider the simplest case $q = 1$. The Russo-Margulis formula (5.1.1) expressed the p-derivative of $\mathbb{E}_p[X]$ in terms of the expectation of $\delta_e X(A) = X(A^e) - X(A_e)$. Further (5.1.3) allows us to consider instead only the difference $X(A) - X(A_e)$. A direct consequence, easily verified and related to (5.2.3), is $\mathbb{P}_p[e \in B]/p = \mathbb{E}_p[K(A_e) - K(A)]$. This merely reflects the definition of a bridge, in other words e is a bridge in (V, A) if and only if $K(A_e) - K(A) = 1$. Equivalently, we have seen that a cut vertex v is defined as a vertex for which

$$K(A) - K(A_{E_v}) + 1 < 0.$$

A few comments are in order. Firstly we naturally generalised $A_e \equiv A - \{e\}$ to $A_{E'} \equiv A - E'$ for any $E' \subseteq E$. Secondly we denote by E_v the set of edges incident to vertex v. Thirdly, the increment of 1 incorporates the removal of vertex v in addition to all edges in E_v. In other words the number of components in the graph (V_v, A_{E_v}) by removing v together with all edges in E_v has one component less than the graph (V, A_{E_v}), obtained by removing only all edges in E_v.

Analogous to the analysis of bridges this naturally leads to the question whether

$$\mathbb{E}_p\Big[K(A) - K(A_{E_v}) + 1\Big] \tag{5.7.1}$$

can be related to the probability that v is a cut vertex. Unfortunately, this is not the case due to the different types of cut vertices, however we have

$$\mathbb{E}_p\Big[K(A) - K(A_{E_v}) + 1\Big] = -\sum_{l=2}^{\Delta}(\ell - 1)\mathbb{P}_p[v \in F_\ell] = -\sum_{\ell=2}^{\Delta}(\ell - 1)\mathbb{E}_p[f_\ell], \tag{5.7.2}$$

which involves the mean densities (probabilities) of the different types of cut vertices. Further, it is reasonable to assume that at the location of a critical point, say $p = 1/2$ on $\mathbb{Z}_L^2$, one has, by means of universality, that $\mathbb{E}_{p=1/2}[K(A) - K(A_{E'})]$ is insensitive (when concerned with scaling exponents) to whether E' is a singleton, e.g. $E' = \{e\}$ or $E' = E_v$ whenever the graph has only short-range bonds. This in particular is a valid assumption for $\mathbb{Z}_L^d$. Thus it is reasonable to assume that the leading finite size scaling exponent in (5.7.1) is inherited from $\mathbb{P}_{p=1/2}[e \in B]$, which in turn has leading correction proportional to $L^{1/\nu - d}$. Further expression (5.7.2) relates this correction to the corrections of the (weighted) sum of $\mathbb{E}_{p=1/2}[f_\ell]$ for $2 \leq \ell \leq \Delta$, whose coefficients have all the same sign. Finally, when we deem an artificial cancellation of the amplitudes of $\mathbb{E}_{p=1/2}[f_\ell]$ in the above sum as unlikely, it follows that each of the separate densities inherit the $1/\nu - d$ exponent from $\mathbb{P}_{p=1/2}[f_\ell]$. Now, we also briefly remark how the argument generalises to $q \neq 1$. Firstly the analogue to (5.7.2) reads

$$\mathbb{E}_p\Big[q^{K(A)} - q^{K(A_{E_v})+1}\Big] = Z_{p,q}\sum_{\ell=2}^{\Delta}\Big(1 - q^{\ell-1}\Big)\mathbb{E}_{p,q}[f_\ell]. \tag{5.7.3}$$

On the other hand we clearly have (5.1.6) which, is equivalent to

$$\mathbb{E}_p[q^{K(A)} - q^{K(A_e)}] = Z_{p,q}(1 - q)\mathbb{P}_{p,q}[e \in B].$$

Now we can apply the same reasonable universality arguments (assumptions) as before and conclude that

$$\mathbb{E}_q[f_\ell] \approx c_1 + c_2 L^{1/\nu - d}.$$

It is an interesting open question whether the above ideas can be further exploited or made exact, and if it is possible to extract the asymptotic densities $\mathbb{E}_q[f_\ell]$ based on duality results.

Lastly, there is an interesting adaption of the bridge-edge formula (5.1.8) to the site-random-cluster model [27] recently studied in [122]. This is a natural adaption of the bond random-cluster model. More precisely, configurations correspond to subsets of vertices, and the associated sub-graphs are constructed by adding all edges between open vertices (vertices in the configuration). One then weights each configuration with a Bernoulli factor with parameter p, that is each open vertex contributes p, and each closed vertex $1 - p$ to the weight of the configuration. Additionally, each configuration has an

additional cluster weight, as in the bond-random cluster model. Arguments completely analogous to the bond-case can be used to derive the following relation

$$\mathbb{E}^{(s)}_{p,q}[\mathcal{V}] = p + (1-p)\sum_{\ell=2}^{\Delta}\left(1-q^{\ell-1}\right)\mathbb{E}^{(s)}_{p,q}[f_\ell],$$

where the (s) stands for the *site*-random-cluster model, and $\mathcal{V}$ is the density (with respect to n) of occupied vertices.

Chapter 6

Efficiency of the coupling from the past algorithm for the random-cluster model

In this chapter we describe a numerical analysis of the coupling from the past (CFTP) algorithm applied to the monotone[1] ($q \geq 1$) random-cluster model in two and three dimensions which is complemented by a rigorous analysis in one dimension. The reason for studying the CFTP algorithm for the random-cluster model is two-fold. Firstly, the approach provides a perfect-sampling algorithm, that is it is capable of producing samples drawn precisely from the random-cluster model probability measure $\pi_{p,q}$ that are assured to be free of any initialisation bias. Additionally, samples generated by this method can be guaranteed to be independent. In this sense the CFTP algorithm circumvents both potential issues of correlation and relaxation, that normally affect MCMC methods. Secondly, certain efficiency characteristics of the CFTP algorithm, such as the expected running time, are closely related to statistical efficiency measures of the heat-bath chain, for instance in the form of the corresponding exponential autocorrelation time τ_{exp}. We have seen in chapter 4 that the dynamical critical behaviour of Sweeny's algorithm (in both its heat-bath and Metropolis flavours) showed interesting and peculiar characteristics, e.g., a speeding up effect or the possible complete absence of critical slowing down for $q < 2$. Most of these aspects are not yet, or at most partially heuristically, understood and hence by studying the efficiency of the CFTP algorithm one can, for instance, obtain upper bounds for the exponential autocorrelation time. We remark that this is usually done by estimating the autocorrelation times of a carefully chosen representative set of observables, which one believes to cover all dynamic scales. Combined with heuristic arguments one can then often derive estimates for τ_{exp}. Yet, in many cases there is no absolute confidence that the considered set of observables projects onto the slowest mode.

[1] For $q < 1$ the random-cluster model is in fact anti-monotone (which is not to be confused with negative association), for which one also has an exact sampling algorithm [32]. We will conduct a practical efficiency study elsewhere.

There are excellent resources in the literature that explain the basic and more detailed aspects of coupling from the past, such as [33, 2, 27, 123, 124]. We will therefore introduce the basic ideas rather briefly and tailor the exposition to the random-cluster model, see also in particular [124].

6.1 Introduction and theoretical background

The coupling from the past algorithm, presented by Jimm Propp and David Wilson in [33] is probably best described in terms of the *random mapping* representation of a Markov chain (more precisely its transition matrix) [2]. This representation emerges naturally in the algorithmic study of Markov chains, where the corresponding algorithmic procedure is fed with a stream of random numbers, based on which the next state is chosen. In other words the random mapping representation reflects the idea of considering a Markov chain as an iterated decomposition of random maps from and to the state space. More precisely, given a transition matrix $P(A,B)$ of a Markov chain on a state-space Ω, one can always [2] find $f : \Omega \times \Lambda \to \Omega$ such that

$$\mathbb{P}[f(A,\mathcal{R}) = B] = P(A,B), \tag{6.1.1}$$

where $A, B \in \Omega$ and $\mathcal{R}$ is an appropriate Λ-valued random number. Here Λ is often the unit-interval $[0,1]$ but can also me more general the sample space of a random vector, as we will see below in case of the heat-bath chain. Intuitively, given f, the current configuration A and a random variable or vector $\mathcal{R}$, we can iterate the chain one step in a truthful way, that is the corresponding one-step distribution is described by $P(A,\cdot)$. Moreover, one can then also compose several independent random maps, which in turn allows one to construct the sequence $A_n = f(A_{n-1},\mathcal{R}_n)$ with $A_0 = A$ and $n \geq 1$, which is fed with a sequence $(\mathcal{R}_1,\mathcal{R}_2,\cdots)$ of independent Λ-valued random numbers. It is not hard to see that the sequence $\{A_n\}_{n\geq 1}$ is a Markov chain with transition matrix P and initial state A [2], which is successively updated by f and the underlying stream of random variables. In what follows we write $f_n(\cdot) \equiv f(\cdot,\mathcal{R}_n)$ for $n \geq 1$. Furthermore in order to denote the composition of a consecutive sequence of f_n's we write for $t \geq 1$

$$F_t^{\uparrow} \equiv f_t \circ f_{t-1} \circ \cdots \circ f_2 \circ f_1 \tag{6.1.2}$$

$$F_t^{\downarrow} \equiv f_1 \circ f_2 \circ \cdots \circ f_{t-1} \circ f_t. \tag{6.1.3}$$

It is now straightforward to show that $\mathbb{P}\Big[F_t^{\uparrow}(A) = B\Big] = P^t(A,B)$ for any $t > 0$ and $A, B \in \Omega$. We remark further that also the order of the f_i's above does not matter in the following sense

$$\mathbb{P}\Big[F_t^{\uparrow}(A) = B\Big] = \mathbb{P}\Big[F_t^{\downarrow}(A) = B\Big] \quad \text{valid for any } t > 0 \text{ and } A, B \in \Omega. \tag{6.1.4}$$

At this point it might seem somewhat artificial to emphasise the difference between the two compositions $\uparrow$ and $\downarrow$, however as we will see in a moment, it is the most important

observation underlying the coupling from the past algorithm by Propp & Wilson. The reason for this difference is that the $\uparrow$-scheme does not guarantee that the output state follows the stationary distribution. We will show this precisely below. The above mapping view point also explains the notion of "coupling from the past": In order to iterate the mapping one step further in the $\downarrow$ scheme, one needs to *prepend* the new one-step mapping to the previous composition. Equivalently, one has to start at an earlier time in the past and evaluate the output at the present. Before we proceed, let us consider as an explicit example the heat-bath chain for the random-cluster model with cluster weight q and bond-density p, on a graph $G = (V, E)$. The corresponding random variable $\mathcal{R}$ is actually a tuple or vector (u, e), where u is an uniform random number in $[0, 1]$, and e is chosen uniformly in E (hence $\Lambda = [0, 1] \times E$). Let us denote by $A \subseteq E$ the current configuration. Now, in line with (2.2.3), one can choose the mapping

$$f\Big(A, (u, e)\Big) \equiv \begin{cases} A^e & \text{if } u \leq a(A; e), \\ A_e & \text{if } u > a(A; e), \end{cases} \tag{6.1.5}$$

where $a(A; e)$, defined in (2.2.4), is clearly a deterministic function of A and e. Recall, we have shown in Lemma (2.2.1) that the heat-bath chain for the random-cluster model is ergodic (aperiodic and irreducible). Hence we have, by means of the convergence Theorem (2.1.2), the following asymptotic results

$$\lim_{t\to\infty} \mathbb{P}\Big[F_t^{\downarrow}(A) = B\Big] = \lim_{t\to\infty} \mathbb{P}\Big[F_t^{\uparrow}(A) = B\Big] = \pi(B). \tag{6.1.6}$$

Here the first equality follows from (6.1.4), and the above is valid for all A, B (we suppressed the p, q-dependence on both sides). This states the fact that, in this random-map setting, the sequence of random maps $\{f_i\}_{\{i\geq 1\}}$, induced by the independent random-sequence $\{(u_i, e_i)\}_{\{i\geq 1\}}$, has the property that the distribution of the value of $F_t^{\uparrow}$ and $F_t^{\downarrow}$ becomes closer and closer to π for sufficiently large t. Furthermore the mapping looses asymptotically any dependence on the actual point in Ω for which the mapping is evaluated. This suggests that the random maps $F_\infty^{\uparrow}$, $F_\infty^{\downarrow}$ are actually constant, with the corresponding asymptotic value having distribution precisely equal to π. This has a nice intuitive basis, as it roughly means that the Markov chain (iteration of random-maps) has forgotten any dependence on the initial configuration or distribution/configuration. Let us consider this possible constancy of the two iterated mappings $\uparrow$ and $\downarrow$ more carefully. To start with define the following two random random times $\tau^{\uparrow}, \tau^{\downarrow}$ as

$$\tau^{\uparrow} \equiv \inf\Big\{t \geq 1 : F_t^{\uparrow} \text{ is constant}\Big\} \tag{6.1.7}$$

$$\tau^{\downarrow} \equiv \inf\Big\{t \geq 1 : F_t^{\downarrow} \text{ is constant}\Big\} \tag{6.1.8}$$

The first crucial observation is that both $\tau^{\uparrow}$ and $\tau_{\downarrow}$ have the same distribution. This follows from the fact that the probability that $F_t^{\downarrow}(\cdot)$ is not constant, equals the probability that $F_t^{\uparrow}(\cdot)$ is not constant. The two events correspond to $\tau^{\downarrow} > t$ and $\tau_t^{\uparrow} > t$, hence we have

$\mathbb{P}\left[\tau^{\downarrow} > t\right] = \mathbb{P}\left[\tau^{\uparrow} > t\right]$ [33]. When the difference between $\tau^{\uparrow}$ and $\tau_{\downarrow}$ does not matter, we simply write τ in what follows. More importantly, it turns out that in many cases τ is finite almost surely, that is the mappings $F_t^{\uparrow}$ and $F_t^{\downarrow}$ become, with probability one, constant after a finite time. We postpone the proof of this statement for the choice of f specified in (6.1.5) for a moment. Further note that the study of τ is closely related to the mathematical concept of coupling of probability distributions or Markov chains, see for instance [2]; we already encountered the coupling idea in the derivation of the Russo-Margulis formula in section 5.1.

Now, the original idea of Propp & Wilson is to operate in the $\downarrow$-scheme, that is successively construct $F_t^{\downarrow}$ for increasing values of t until constancy is achieved. How can constancy of $F_t^{\downarrow}$ be determined? In order to see this note the random map representation allows for a clear separation of the randomness, the stream of random numbers, from the actual algorithmic procedure f. In fact one can consider multiple starting states and iterate them "simultaneously", based on the same underlying stream of random numbers. Moreover one can, at least theoretically, simultaneously iterate all $|\Omega|$ random-maps (for fixed random number stream), that start in all possible states of Ω. Hence by simultaneously iterating all Ω configurations based on the same composition of random maps one can check for constancy. In practise this means one uses the same random numbers for all trajectories, which in fact implements what is known as a grand-coupling [2] of the corresponding Markov chain. Now having understood how in principle constancy can be determined, one further needs a protocol that describes how the value of t is increased when the current map $F_t^{\downarrow}$ is not constant. A natural way is simply to increment t by one. However, this can be shown to incur a computational cost roughly proportional to $(\tau^{\downarrow})^2$, whereas an almost optimal choice is to double t, in which case the computational load increases only linearly[2] with $\tau^{\downarrow}$ [33]. Care has to be taken in how the composition $F_t^{\downarrow}$ is extended by the new random map f_{t+1}. By construction of $F_t^{\downarrow}$ one has to *prepend* the "new" mapping f_{t+1} to the previous composition $F_t^{\downarrow}$ of mappings $\{f_i\}_{1\le i\le t}$. For instance suppose we have determined that $F_t^{\downarrow}$ is not constant and decided to continue with $t+1$, thus we have to consider

$$F_{t+1}^{\downarrow} = f_1 \circ f_2 \circ \cdots \circ f_t \circ f_{t+1} = F_t^{\downarrow} \circ f_{t+1}. \tag{6.1.9}$$

Practically this means that one has to re-use the random numbers $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t$, previously used to construct $F_t^{\downarrow}$, also for the corresponding composition from 1 to t in the construction of $F_{t+1}^{\downarrow}$. We remark that there are alternative exact sampling approaches that avoid the need of re-using random numbers [125], however we focus on the CFTP approach as it is probably the conceptually simplest approach (which is still far from being completely understood for $\mathbb{Z}_L^2$). A crucial observation is that the output of the algorithm that terminates as soon as it encountered that the map $F_t^{\downarrow}$ is constant, equals the output of any extended run of the algorithm to times beyond t. This follows from the

[2] Here we ignored the computational complexity of each operation. Thus the above is strictly speaking measured in units of the computational cost of a single step of the underlying Markov chain.

observation that for any $A \in \Omega$

$$F_{t+1}^{\downarrow}(A) = F_t^{\downarrow}(f_{t+1}(A)) = F_t^{\downarrow}(A), \tag{6.1.10}$$

by constancy of $F_t^{\downarrow}$. As already mentioned before, the $\uparrow$-scheme or "coupling to the future" algorithm (or forward coupling) does not guarantee a truthfully distributed output state. The obstacle stems from the way the compositions are extended. Here one successively builds up $F_t^{\uparrow}$ and terminates with t for which $F_t^{\uparrow}$ is constant. In this case one has in general

$$F_{t+1}^{\uparrow}(A) = f_{t+1}\Big(F_t^{\uparrow}(A)\Big)\Big(\tag{6.1.11}$$

which can (and will in most cases) differ from $F_t^{\uparrow}(A)$. Therefore in contrast to the coupling from the past algorithm the forward coupling algorithm does not leave the coalesced state invariant. In other words extending the run beyond $\tau^{\uparrow}$ can alter the coalesced state, so it is not a priori clear at what time to "pick up" the state. Crucially this ambiguity is absent in the coupling from the past algorithm of Propp & Wilson.

6.1.1 Practical CFTP: Monotonicity

It is apparent that the idea of iterating all $|\Omega|$ configurations simultaneously is not tractable for most of the interesting models, such as the random-cluster model heat-bath chain. However in some instances one does not need to consider all configurations and it suffices to consider only two extremal configurations. In order to have a notion of extreme configurations, we need a partial order in the state space. Indeed, the state space $\Omega = \{A \subseteq E\}$ of the random-cluster model is actually a poset, i.e. has a partial order, which is induced by set inclusion: We write $A \preceq B$ whenever $A \subseteq B$, which can be verified to be a partial order on Ω. Now, for the random-cluster model with $q \geq 1$ we can verify that the heat-bath dynamics is *monotone* in the sense that it preserves the partial order, [33, 27]. The monotonicity follows from the fact that for two configurations $A \preceq B$ it holds that any edge that is pivotal to B must also be pivotal to A. In other words, adding the remaining edges in $B \setminus A$ to A cannot create any pivotal edges. Note that pivotality, as opposed to being a bridge, is independent of the actual occupation of the edge under consideration. Specifically, using the random-map representation monotonicity means

$$A \preceq B \Rightarrow f(A,(u,e)) \preceq f(B,(u,e)) \tag{6.1.12}$$

for any value of $u \in [0,1]$ and $e \in E$. This can be iterated to yield $A \preceq B \Rightarrow F_t^{\downarrow}(A) \preceq F_t^{\downarrow}(B)$ for any $t \geq 1$, provided the same sequence of random numbers is used (the same holds with $\uparrow$). It is evident that this partial order has clearly two extreme elements $\varnothing$ and E. Now, in the monotone setting it is straightforward to determine whether coalescence ($F_t^{\downarrow}$ is constant) has happened: Simply check whether $F_t^{\downarrow}(\varnothing) = F_t^{\downarrow}(E)$. Before we proceed with our analysis, we provide the remaining proof that the coupling from the past algorithm indeed produces bias-free samples from π. We do not claim novelty of the proof, and the standard

references mentioned above contain proofs, however mostly with the aim of maintaining generality. Yet, we still believe studying the proof promotes the understanding of the algorithm and more importantly, we made an attempt to construct the proof with an emphasis on the difference between the forward coupling and coupling from the past variants.

6.1.2 Proof of correctness of CFTP for the random-cluster model with $q \geq 1$

For the following arguments we need another notation, generalising the composition $F_t^{\downarrow}$. We define define for $k \leq l$

$$F_l^k = f_k \circ f_{k+1} \circ \cdots \circ f_{l-1} \circ f_l, \tag{6.1.13}$$

which is the composition of random maps f_k to f_l, and we have $F_t^{\downarrow} = F_t^1$, as well as $F_t^t = f_t$. Further we write $|F_k^t|$ for the cardinality of the image of F_k^t, i.e. for the number of different output states, more precisely $|F_k^t| \equiv |\{F_k^t(A) : A \in \Omega\}|$. We therefore have that $|F_k^t| = 1$ corresponds to the case that F_k^t is a constant map. Now comes the main argument: We know that the heat-bath chain is irreducible and aperiodic, which allows us to find an integer L such that $P^L(A,B) \geq \epsilon > 0$ for any pair of states $A, B \in \Omega$. In words, there is a non-vanishing probability for an L-step transition between any two states. Moreover in the particular case when F_L^1 maps all states to state E we clearly have that F_L^1 is a constant map (coalescence occurred). Now, the event that E is the F_L^1-image of all states happens in particular when $F_L^1(\varnothing) = E$, which is a direct consequence of the monotonicity of the chain whenever $q \geq 1$. Whence by ergodicity we have that $\mathbb{P}[F_L^1(\varnothing) = E] \geq \epsilon$ and therefore

$$\epsilon \leq \mathbb{P}\Big[F_L^1(\varnothing) = E\Big] \leq \mathbb{P}\Big[F_L^1(A) = E \quad \forall A \in \Omega\Big] \leq \mathbb{P}\Big[|F_L^1| = 1\Big]. \tag{6.1.14}$$

Let us remark that by construction of the coupling from the past algorithm, we have that the same statement actually applies to $F_{(k+1)L}^{kL+1}$ for any $k \geq 1$. Moreover the events $|F_{(k+1)L}^{kL+1}| = 1$ (or > 1) for different values of k are mutually independent, as the constancy of any such mapping depends on disjoint and independent sets of random numbers. Because constancy of either $F_{(k+1)L}^{kL+1}$ or $F_{kL}^{\downarrow}$ implies constancy of $F_{(k+1)L}^{\downarrow}$ we have

$$\mathbb{P}\Big[\Big\{\Big|F_{(k+1)L}^{kL+1}\Big| = 1\Big\} \vee \Big\{\Big|F_{kL}^{\downarrow}\Big| = 1\Big\}\Big] \leq \mathbb{P}\Big[\Big|F_{(k+1)L}^{\downarrow}\Big| = 1\Big]$$

It follows by De Morgan's laws and independence of the events $\left|F_{kL}^{\downarrow}\right| > 1$ and $\left|F_{(k+1)L}^{kL+1}\right| > 1$ that

$$\mathbb{P}\Big[\Big|F_{(k+1)L}^{\downarrow}\Big| > 1\Big] \leq \mathbb{P}\Big[\Big\{\Big|F_{(k+1)L}^{kL+1}\Big| > 1\Big\} \wedge \Big\{\Big|F_{kL}^{\downarrow}\Big| > 1\Big\}\Big] = \mathbb{P}\Big[\Big|F_{(k+1)L}^{kL+1}\Big| > 1\Big]\mathbb{P}\Big[\Big|F_{kL}^{\downarrow}\Big| > 1\Big].$$

We therefore obtain

$$\mathbb{P}\Big[\Big|F_{(k+1)L}^{\downarrow}\Big| > 1\Big] \leq (1-\epsilon)\mathbb{P}\Big[\Big|F_{kL}^{\downarrow}\Big| > 1\Big] \tag{6.1.15}$$

which by induction (over k) leads to

$$\mathbb{P}\Big[\tau^{\downarrow} > kL\Big] = \mathbb{P}\Big[\Big|F^{\downarrow}_{kL}\Big| > 1\Big] \leq (1-\epsilon)^k. \tag{6.1.16}$$

This result is clearly not so surprising in light of ergodicity of the underlying chain. However we remark that it has an important theoretical consequence: For any choice of $\delta > 0$, we can find a corresponding value of $k(\delta)$, such that the probability of $\tau^{\downarrow}$ exceeding $k(\delta)L$ is not larger than δ. From a practical point of view, we remark operating with the doubling protocol, i.e. the algorithm operates with times $T_i = 2^{i-1}$ for $i \geq 1$, will in general not directly "hit" τ. However, it can be easily verified that we can find also a $i(\delta)$ such that the probability of not coalescing before/equal $T_{i(\delta)}$, is not larger than δ. Now, we have that the following implication is true

$$\tau^{\downarrow} \leq T_i \Rightarrow F^{\downarrow}_{\tau^{\downarrow}}(A) = F^{\downarrow}_{T_i}(A) \quad \forall A \in \Omega, \tag{6.1.17}$$

which entails by contraposition and (6.1.16) that for any $A \in \Omega$

$$\mathbb{P}\Big[F^{\downarrow}_{\tau^{\downarrow}}(A) \neq F^{\downarrow}_{T_{i(\delta)}}(A)\Big] \leq \delta. \tag{6.1.18}$$

We emphasise that this logical implication is the crucial difference between the coupling from the past algorithm and a forward coupling scheme. In other words we do *not* have that generically the images of any A in $F^{\uparrow}_{\tau^{\uparrow}}$ and $F^{\uparrow}_{T_i}$ equal whenever $T_i \geq \tau^{\uparrow}$. The last argument we need for the proof is the convergence theorem (2.1.2) in a somewhat more general form, as given for instance in [2]. The basic statement, adapted to our notation, states that for an ergodic chain we can find $0 < \alpha < 1$ and $C > 0$ such that

$$\max_{A\in\Omega} \Big\| \mathbb{P}\Big[F^{\downarrow}_{t}(A) = \cdot\Big] - \pi \Big\|_{TV} \leq C\alpha^t. \tag{6.1.19}$$

It can be verified that this in particular implies the convergence statement in (2.1.2). Moreover it allows us (theoretically) to choose a $t(\delta)$ such that the above total variation distance does not exceed δ. We now have all the necessary pieces to conclude the proof. In what follows we show that for any error margin $\eta > 0$ we impose, the coupling from the past algorithm will produce an output state that has distribution closer to π than η. More precisely, we have for any $A, B \in \Omega$ and any $\eta > 0$

$$\Big|\mathbb{P}\Big[F^{\downarrow}_{\tau}(A) = B\Big] - \pi(B)\Big| \leq \eta, \tag{6.1.20}$$

which is equivalent (c.f. [123]) to $\mathbb{P}[F^{\downarrow}_{\tau}(A) = B] = \pi(B)$ for all $A, B \in \Omega$. We achieve this by bounding the total variation distance between the law of $F^{\downarrow}_{\tau^{\downarrow}}(A)$ and π for any A. To start with, let $\eta = 2\delta$ and choose $j(\delta) = \max\{i(\delta), \lceil \log(t(\delta))/\log(2) \rceil\}$. Recall that the total variation distance is the maximal absolute difference of probabilities the two distributions can assign to any event. This in particular applies also to a particular elementary outcome

B. We conclude

$$\begin{aligned}\left\|\mathbb{P}\left[F^{\downarrow}_{\tau^{\downarrow}}(A)=\cdot\right]-\pi\right\|_{TV} &\leq \left\|\mathbb{P}\left[F^{\downarrow}_{\tau^{\downarrow}}(A)=\cdot\right]-\mathbb{P}\left[F^{\downarrow}_{T_{j(\delta)}}(A)=\cdot\right]\right\|_{TV}+\left\|\mathbb{P}\left[F^{\downarrow}_{T_{j(\delta)}}(A)=\cdot\right]-\pi\right\|_{TV}\\ &\leq \mathbb{P}\left[F^{\downarrow}_{\tau^{\downarrow}}(A)\neq F^{\downarrow}_{T_{j(\delta)}}(A)\right]+C\alpha^{T_{j(\delta)}}\\ &\leq \delta+\delta\leq\eta,\end{aligned}$$

which establishes the claim. Here we used the triangle inequality for the total variation distance, which allows us to use the convergence result (6.1.19) and (6.1.18). The latter is utilised by bounding the total variation distance between the distributions of $F^{\downarrow}_{\tau^{\downarrow}}(A)$ and $F^{\downarrow}_{T_{j(\delta)}}(A)$ in terms of the natural coupling of the two using the same stream of random numbers (see [2] for bounding the total variation distance using the coupling method).

We note that (6.1.16) can also be used to actually prove that $\mathbb{P}[\tau^{\downarrow}<\infty]=1$, that is the algorithm terminates *almost surely*. The argument uses the fact that the coupling from the past algorithm is ruled by a $0-1$ law [123], which states that $\mathbb{P}[\tau^{\downarrow}<\infty]$ must be either 0 or 1, that is the algorithm either terminates almost surely, or does almost surely run forever. However as we have established that $\mathbb{P}[\tau^{\downarrow}<\infty]>0$, we must have almost surely $\tau^{\downarrow}<\infty$. Likewise, this applies also to $\tau^{\uparrow}$, i.e. coalescence will in both cases almost surely happen. We will have more to stay about the actual distribution of $\tau^{\uparrow}$ ($\tau^{\downarrow}$) in section 3.

After these general, somewhat theoretical but revealing, considerations we can now formulate the heat-bath chain induced coupling from the past algorithm for the random-cluster model as shown in Algorithm 4. A few comments are in line. Firstly, note that we assume the availability of a routine $RAND(t)$ that returns the t'th pair of (pseudo)-random numbers. In particular, for fixed t, successive calls to $RAND(t)$ (within one invocation of $CFTP\_HB$) must return the same tuple (u_t,e_t), which itself is random and independent of all other tuples returned by $RAND$. Secondly, we invoked routine $a(X,e)$, already defined in (2.2.4), returns, depending on the pivotality of edge e in X, either $\tilde{p}$ or p. This implicitly assumes the availability of a sub-routine to determine this pivotality, which can be provided by any of the methods described in chapter 3. Lastly, we used the particular doubling-protocol mentioned above.

Before we proceed with the presentation of our results, let us remark that the coupling from the past procedure has the following important property

$$F^{\downarrow}_{t}(E)\subseteq F^{\downarrow}_{t'}(E)\quad\text{and}\quad F^{\downarrow}_{t'}(\varnothing)\subseteq F^{\downarrow}_{t}(\varnothing) \tag{6.1.21}$$

whenever $t'\leq t$. This intuitively means that extending the run can at most push the iterated state of E further down, and at the same time can, if at all, only elevate the corresponding image of $\varnothing$. Thus the final output configuration is more and more revealed as t is increased. This is precisely shown in Figure 6.1 for the evolution of the top configuration E.

Algorithm 4 Coupling from the past algorithm for the random-cluster model

```
function CFTP_HB
    t ← 1
    while X ≠ Y do
        X, Y ← ∅, E
        t' ← t
        while t' > 0 do
            u_{t'}, e_{t'} ← RAND(t')
            if u_{t'} ≤ a(X, e_{t'}) then
                X ← X^{e_{t'}}
            else
                X ← X_{e_{t'}}
            end if
            if u_{t'} ≤ a(Y, e_{t'}) then
                Y ← Y^{e_{t'}}
            else
                Y ← Y_{e_{t'}}
            end if
            t' ← t' − 1
        end while
        t ← 2t
    end while
    return X
end function
```

6.2 Expected coupling time

The efficiency of the coupling from the past algorithm depends on the statistics of τ, the random time at which all coupled trajectories coalesce. One reason for that follows from the fact that τ and the final output state are not independent and one will introduce a bias by systematically suppressing longer runs [126]. We remark that there exist alternative perfect sampling approaches commonly referred to as "interruptible" perfect sampling algorithms, such as [126], that allow for such "user impatience" or running time constraints. On the other hand we have seen in the previous section that we have the almost sure guarantee that the algorithm will terminate. We have already remarked that the perfect sampling capability of the coupling from the past algorithm is not the only reason of our study, because the time of coalescence τ allows us to study the relaxation and stationary correlations of the underlying heat-bath chain. For instance, consider the following standard relation [2]:

$$\max_{A\in\Omega} \left\| P^t(A,\cdot) - \pi \right\|_{TV} \leq \mathbb{P}[\tau > t], \tag{6.2.1}$$

which relates the tail of the distribution of τ to the total variation distance of the time dependent law of the Markov chain from the stationary distribution. In words, the more likely it is that at a given time t a grand-coupling coalesces, equivalently $F_t^{\downarrow}$ is a constant map, the smaller is the initialisation bias. It is evident that this inequality is closely related

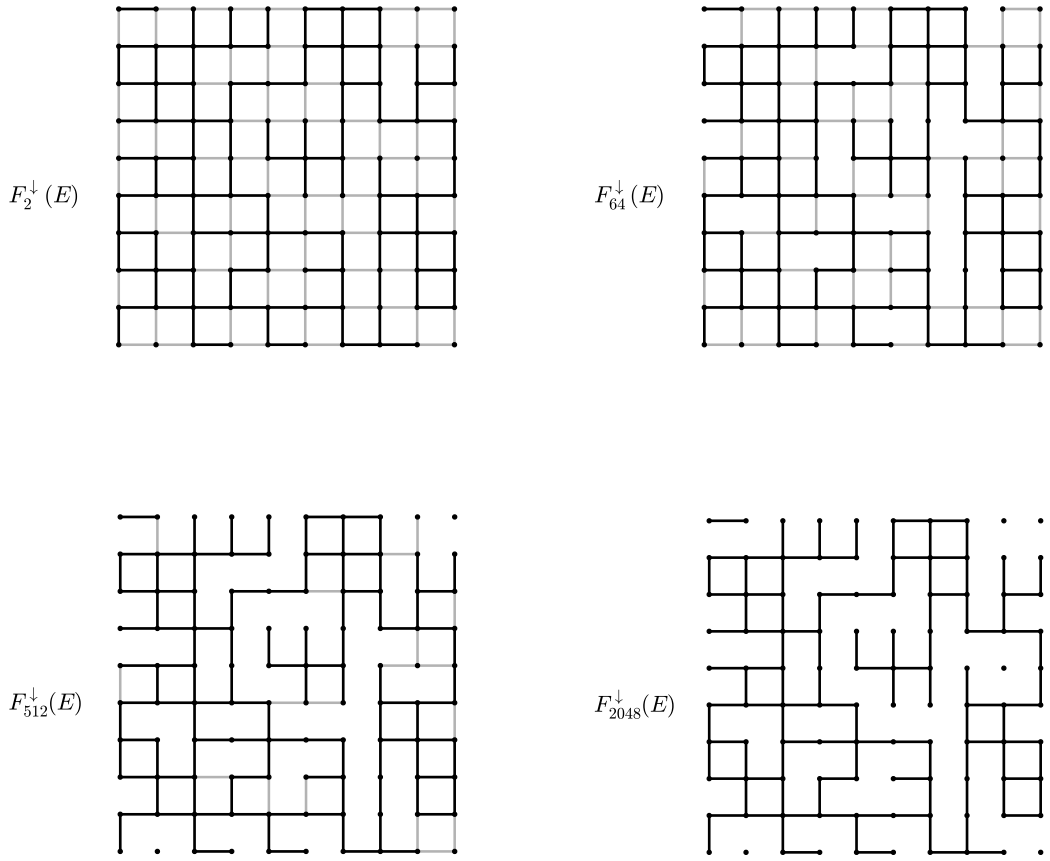


Figure 6.1: *Snapshots of the iterated value of $F_t^{\downarrow}(E)$ for some values of t, as produced by a coupling from the past run for $q = 4$ and $p = p_{\mathrm{sd}}(4) = 2/3$. The black lines in each configuration show the edges that are part of the final constant output state $F_{\tau}^{\downarrow}(E)$, whereas the grey lines correspond to the remaining edges in $F_t^{\downarrow}(E)$ (For a similar Figure consider also [124], Fig. 7.2.).*

to the mixing time (see chapter 2), and in fact provides the basis of many works that use coupling to bound the mixing time [2]. Practically, it is somewhat easier to estimate the expectation $\mathbb{E}[\tau]$ than the tail of the distribution of the coupling time τ. Now, using a probability tail bound known as Markov's inequality [38] we can indeed obtain the following relationship valid for any positive integer t:

$$\mathbb{P}[\tau > t] \leq \frac{\mathbb{E}[\tau]}{t}. \tag{6.2.2}$$

This allows us to bound the mixing time t_{mix} from above by $\mathbb{E}[\tau]$:

$$\begin{aligned} \mathbb{E}[\tau] &\geq \left\lceil \frac{t_{\mathrm{mix}}}{2} \right\rceil \mathbb{P}\left[\tau > \left\lceil \frac{t_{\mathrm{mix}}}{2} \right\rceil\right] \\ &\geq \frac{t_{\mathrm{mix}}}{2} \mathbb{P}\left[\tau > \left\lceil \frac{t_{\mathrm{mix}}}{2} \right\rceil\right] \end{aligned}$$

$$\begin{aligned} &\geq \frac{t_{\text{mix}}}{2} \max_{A\in\Omega} d_A\left(\left\lceil \frac{t_{\text{mix}}}{2} \right\rceil\right) \\ &\geq \frac{t_{\text{mix}}}{8}. \end{aligned}$$

The first inequality is a direct consequence of (6.2.2), the third follows from (6.2.1), and the last follows from the fact that t_{mix} is defined as the first time t such that $d_A(t) \leq 1/4$ for all $A \in \Omega$, hence for any time earlier t_{mix} there must be at least one A such that $d_A \geq 1/4$. Recall that we use the definition, introduced in chapter 2, $d_A(t) \equiv \left\| P^t(A,\cdot) - \pi \right\|_{TV} = \frac{1}{2}\sum_{B\in\Omega} \left| P^t(A,B) - \pi(B) \right|$. From a practical point of view we would also like to know how much larger $\mathbb{E}[\tau]$ can be. In [33] Propp & Wilson indeed show that the mixing time t_{mix} yields an upper bound for $\mathbb{E}[\tau]$, that is $\mathbb{E}[\tau] \leq 2t_{\text{mix}}(1+\log(m))$, which we tailored to the random-cluster model on a graph with m edges. We conclude

$$\frac{t_{\text{mix}}}{8} \leq \mathbb{E}[\tau] \leq 2t_{\text{mix}}\,(1+\log(m)).$$

Because we can relate the mixing time to the relaxation time or exponential autocorrelation time it is not surprising that we can also relate $\mathbb{E}[\tau]$ to τ_{exp} using precisely (2.1.17) and (2.1.16) to obtain

$$\frac{\tau_{\text{exp}}-1}{8}\log(2) \leq \mathbb{E}[\tau] \leq 2\Big(\tau_{\text{exp}}+1\Big)\log\ \frac{4}{\pi_{\min}}\Big)\Big(\log(m)+1)$$

We further need a lower bound for $\pi_{\min}$ for the random-cluster model, which is straightforward (but probably far from optimal), and follows from $\pi_{p,q}(A) \geq \Big[p(1-p)\Big]^m q^{-n}$ for $q \geq 1$. We therefore obtain for the heat-bath chain of the random-cluster model with bond-density p and cluster weight $q \geq 1$ on a graph with m edges and n vertices:

$$\frac{\tau_{\text{exp}}-1}{8}\log(2) \leq \mathbb{E}[\tau] \leq 2\Big(\tau_{\text{exp}}+1\Big)\ \log(4) + m\log\ \frac{q^{n/m}}{p(1-p)}\Big)\Big)(\log(m)+1). \qquad (6.2.3)$$

For the regular graphs we are interested in, we have in particular that n/m is constant (e.g. 2 for $\mathbb{Z}_L^2$). Here we can therefore equivalently use the O-notation and concisely conclude

$$\tau_{\text{exp}} \leq O(\mathbb{E}[\tau]) \quad \text{and} \quad \mathbb{E}[\tau] \leq O(\tau_{\text{exp}} m \log(m)). \qquad (6.2.4)$$

Let us interpret the above result, as it provides the basis for the conclusions we make at a later stage. The first bound allows us to obtain an upper bound for the exponential autocorrelation time, that is the *slowest* mode of the chain. For instance, our Projection Lemma in section 4.4 shows that this is relevant for the correlations in the time series of the bond-variable $\mathcal{N}$, an important quantity (in particular for chapter 5). In particular, this also applies at the location of a second order phase transition. Here in the study of (dynamical) critical phenomena one is particularly interested in determining dynamical critical exponents, such as z_{exp} in an expected power law scaling $\tau_{\text{exp}} \approx L^{z_{\text{exp}}}$. Similarly we

also expected a power law scaling for $\mathbb{E}[\tau]$, that is

$$\mathbb{E}[\tau]/L^d \approx L^w,$$

where we defined the *(Propp-) Wilson exponent* w. Due to (6.2.4) we have the constrain that z_{exp} cannot exceed w and $w \leq d + z_{\text{exp}}$, where d is the dimensionality of the system, such as in $\mathbb{Z}_L^d$. Thus by analysing the expected running time of the coupling from the past algorithm we can obtain a safe time scale on which all observables relax. Additionally the second bound in (6.2.3) quantifies that $\mathbb{E}[\tau]$ can not be larger than roughly $m\log(m)$ times the exponential autocorrelation time. A factor of $m\log(m)$ is in principle enough for establishing the rapid-mixing property, let alone for the purpose of obtaining upper bounds of τ_{exp}. However for the practical question, whether the coupling from the past algorithm is significantly more inefficient than a standard MCMC approach using the heat-bath chain (ignoring the residual bias in the latter), a factor of $m\log(m)$ matters. We will therefore devote the rest of this section to analyse the expected coupling time, firstly for the off-critical random-cluster model on $\mathbb{Z}_L^d$ for $d = 2, 3$, and secondly for the self-dual point, that is at criticality.

6.2.1 Off criticality

Recall that we established that one has the freedom to study either $\tau^{\downarrow}$ or $\tau^{\uparrow}$ when concerned solely with the statistics of the time of coalescence. We decided to implement the forward coupling (corresponding to $F_t^{\uparrow}$) and hence determined $\tau^{\uparrow}$. This decision was simply made because of the smaller computational overhead in the forward coupling.
Before we present our numerical results we pause and describe briefly how this randomised algorithm operates in practise. This will at the same time provide us with an intuition for the mechanism responsible for a possible slowing down. As outlined in the previous sections, the two coupled Markov chains $V_t \equiv F_t^{\uparrow}(\varnothing)$ and $\Lambda_t \equiv F_t^{\uparrow}(E)$, starting in configuration $\varnothing$ and E, respectively, use the same stream of random numbers. This has the consequence that both chains operate at any given step, say t, on the same edge e (which of course varies with t). Note that, strictly speaking, the heat-bath chain does not attempt to "flip" the status of e, but assigns an independent status to e. The probability of e being open or closed in the subsequent configuration depends only on the pivotality of the edge in the current configuration, i.e., it equals $\tilde{p}(p,q)$ for pivotal edges and p for non-pivotal edges. However this means that the two coupled chains might differ in the subsequent configuration at edge e, in spite of the fact that both chains update the same edge. Compare this to the case $q = 1$, where $\tilde{p}(p,1) = p$. Therefore for $q = 1$ both configurations agree at e after time t. In contrast for $q \neq 1$ there is a non-vanishing probability that $e \in \Lambda_{t+1}$ but $e \notin V_{t+1}$. This event happens with probability not larger than $p - \tilde{p}(p,q)$. Strictly speaking, it happens precisely with probability $p - \tilde{p}(p,q)$ when e is pivotal in V_t but not pivotal in Λ_t, and with probability 0 otherwise. It is apparent, that for small values of t, V_t is still very close to $\varnothing$ and Λ_t does not differ much from E. This

suggests that for sufficiently small times a chosen edge e has differing pivotality in V_t and Λ_t. Hence the two configurations disagree at e with probability $p - \tilde{p}(p,q)$ from time $t+1$ on, until (at least) the next time e is selected. This happens in average after m steps. Now, for later times the two chains might differ significantly from their initial configuration, which might change the cycle structure significantly, e.g. the configuration Λ_t might be sufficiently sparse due to a low bond density p, such that the two configurations V_t and Λ_t are structurally enforced to agree at e in many instances. It therefore crucially depends on the structural evolution of the two configurations, whether the coupling time is extremely delayed, reflecting a slowing down, or close to optimal (to be specified below).

Let us now consider the case where the stationary model is away from a critical point. In Figure 6.2 we show our estimates for $\mathbb{E}[\tau]/m$, both for the graph $\mathbb{Z}_L^2$ and $\mathbb{Z}_L^3$ at some off-critical values of p for several cluster weights q. Recall that for $\mathbb{Z}^2$ the critical point is precisely $p_{\mathrm{sd}}(q)$, which implies that the bond-density is sub-critical ($p < p_{\mathrm{sd}}(q)$) for both cluster weights shown. We remark that an estimation of super-critical values in $2D$ would yield qualitatively the same result, as the duality result for the heat-bath chain, in particular inequalities (2.3.7), entails that the expected coupling time for pairs $(p, p^\star(p,q))$ are tied together. In 3D for the graph $\mathbb{Z}_L^3$ we have numerical estimates of the location of the critical point (see section 5.3), which implies that the bond-density $p = 0.6$ is super-critical for both cluster weights $q = 1.5$ and $q = 2.2$. The pair $p = 0.2$ and $q = 1.5$ is sub-critical. Overall our off-critical results, both in 2D and 3D, suggest that coupling happens on average after a number of sweeps (m steps) increasing with roughly logarithmically with the number of edges m. This resembles what is known rigorously for percolation ($q = 1$), where the dependence on the pivotality of an edge is lost ($p = \tilde{p}(p,1)$) and hence the coupling time τ is equal to the random variable T, counting the number of steps until each edge in E is selected (collected) at least once (when edges are selected uniformly at random in E, with replacement). The study of T is related to a well known problem in probability theory [91, 127, 38] which is commonly known as the *coupon collector's problem*. A standard result states that the expectation of T is "almost" linear in the number of edges, in the following sense

$$m\log(m) \le \mathbb{E}[T] = m\sum_{k=1}^{m}\frac{1}{k} = m\log(m) + m\gamma + \frac{1}{2} + o(1) \le m\Big(\log(m)+1\Big), \tag{6.2.5}$$

where the last term $o(1)$ vanishes for $m \to \infty$, and $\gamma = 0.5772156649\ldots$ is the Euler-Mascheroni constant. The corresponding exact value is depicted by the solid black line in Figure 6.2. It appears that $\mathbb{E}[\tau]$ and $\mathbb{E}[T]$ differ, in leading order, by a constant factor. Indeed consider the inset of Figure 6.2, which shows the size dependence of $\mathbb{E}[\tau]/\mathbb{E}[T]$ of the corresponding values in the main Figure. Albeit a weak size dependence is still visible, it is conceivable that we have indeed $\mathbb{E}[\tau] \sim C(p,q)\mathbb{E}[T]$, where $C(p,q) \ge 1$, also depending on the dimensionality, or simply $\mathbb{E}[\tau] \approx \mathbb{E}[T]$. Support for this conjecture comes from our rigorous analysis of the coupling process in one dimension, that is for $\mathbb{Z}_L$. One particular result of analysis is stated in section 6.4 in form of Lemma 6.4.5 which proves asymptotic

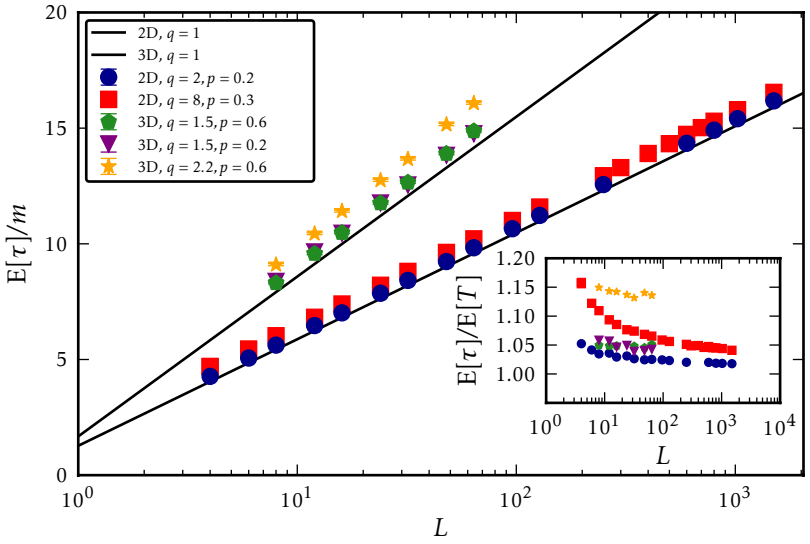


Figure 6.2: *Estimated mean coupling time $\mathbb{E}[\tau]$ per edge, for off critical random-cluster models in 2D and 3D with varying cluster weights and bond-densities. Note the log-scale in the horizontal direction, such that a straight line indicates $\mathbb{E}[\tau]/m \approx C\log(m)$ with $C > 0$ being a constant. This clearly resembles the exact result for $q = 1$ (coupon collector's problem [38]). The inset shows the ratio $\mathbb{E}[\tau]/\mathbb{E}[T]$.*

equivalence, that is

$$\mathbb{E}[\tau] \sim \mathbb{E}[T].$$

Note that, the random-cluster model on $\mathbb{Z}_L$ (the L-cycle) is exactly solvable and is in particular off-critical for any value of $q > 0$ and $p < 1$, see Appendix B for details. Our results suggest that the dominant effect for $q > 1$ away from a critical point remains the coupon collector mechanism, and the introduction of a cluster weight introduces merely a change in the overall time scale.

Partial couplings

In order to gain more insight into the coupling process we introduce the concept of *partial couplings*, for which we need some further notation and concepts. Recall that we write $\Lambda_t \equiv F_t^{\uparrow}(E)$ and $V_t \equiv F_t^{\uparrow}(\varnothing)$ for the states of the two coupled chains at time t, starting in E and $\varnothing$ at time 0, respectively. Further define $S_t \equiv \Lambda_t - V_t$, that is the set of edges that are in Λ_t but not in V_t. We clearly have $S_0 = E$ and $S_{\tau^{\uparrow}} = \varnothing$. Therefore we can alternatively define $\tau^{\uparrow}$ by $\tau^{\uparrow} \equiv \inf\{t > 0 : S_t = \varnothing\}$. Because S_t cannot be empty before all edges are visited at least once we have that $\tau \geq T$, which yields $\mathbb{E}[\tau] \geq \mathbb{E}[T] \sim m\log(m)$. We define the *local coupling time* at edge e by

$$\tau_e^{\uparrow} \equiv \sup\{t \geq 0 : e \in S_t\} + 1, \tag{6.2.6}$$

which is the first time the two chains agree ultimately at e, and we have that $\tau_e^{\uparrow} \leq \tau^{\uparrow}$ for any $e \in E$, or $\tau^{\uparrow} = \max\{\tau_e^{\uparrow} : e \in E\}$. We order the edges corresponding to their value $\tau_e^{\uparrow}$, i.e., we write $\tau_k^{\uparrow}$ for k'th smallest local coupling time ($\tau_m^{\uparrow} = \tau^{\uparrow}$). It follows that the two

configurations Λ_t and V_t agree (ultimately) on at least k edges when $\tau_k^{\uparrow} \leq t$. We therefore term $\tau_k^{\uparrow}$ the *degree k partial coupling*. Equivalently, we have:

$$\tau_k^{\uparrow} \leq t \Rightarrow |S_t| \leq m - k. \tag{6.2.7}$$

We emphasise that the reverse implication $|S_t| \leq m - k \Rightarrow \tau_k^{\uparrow} \leq t$ is *not* true. This is because an edge that is not part of S_t might become a member at a later stage, due to the two configurations disagreeing at e caused by a differing cycle structure. We remark that this obstacle can be avoided when considering partial couplings in the backward coupling, where one can analogously define $\tau_e^{\downarrow} \equiv \inf\{t > 0 : e \notin S_t\}$ with $S_t \equiv F_t^{\downarrow}(E) - F_t^{\downarrow}(\varnothing)$ and we further define $\tau_k^{\downarrow}$ analogously. Here one has $\tau_k^{\downarrow} \leq t \Leftrightarrow |S_t| \leq m - k$, because it follows from (6.1.21) that if $e \notin S_t$ then in fact $e \notin S_{t'}$ for any $t' \geq t$. In other words once the two configurations agree at e in the forward coupling they will do so for all subsequent times. Lastly, let us remark that conceptually the quantities $\tau_e^{\uparrow}$ depend on the complete evolution of the coupling process until overall coalescence has occurred. On the other hand in order to determine $\tau_e^{\downarrow}$ it suffices to wait for the two configurations to agree at e.

In practise we determined the value $\tau_e^{\uparrow}$ by associating a time variable to e, which is initially set to 0. In case e is selected at time t, we set the corresponding time variable to t *only* when the outcome of the update at time t is that the two configurations agree at edge e *and* the time variable was set to the value 0 before the current update. In case the two configurations disagree at e after the update, we reset the time variable to 0. It is now straightforward to confirm that upon termination of the forward coupling algorithm, the time variable of edge e precisely stores $\tau_e^{\uparrow}$.

Let us now come back to the forward coupling and use the analysis of the partial couplings to investigate how efficient the coupling process works at different stages, corresponding to different degrees k of partial couplings. Note, that as a lower reference we can again consider the case $q = 1$, so let $T_k \equiv \tau_k^{\uparrow}$ in case of $q = 1$. One can verify that $T_k - T_{k-1}$ is a geometric random variable with mean $m/(m-k+1)$ for $k > 1$, and that $T_1 \equiv 1$, so that we conclude

$$\mathbb{E}[T_k] = 1 + \sum_{\ell=2}^{k} \mathbb{E}[T_\ell - T_{\ell-1}] = 1 + m \sum_{\ell=2}^{k} \frac{1}{m-\ell+1} = m \sum_{\ell=m-k+1}^{m} \frac{1}{\ell} = m(H_m - H_{m-k}),$$

where $H_k = \sum_{\ell=1}^{k} 1/\ell$ is the k'th harmonic number. In order to achieve a local coupling at e one needs that e is visited at least once and therefore one has $\mathbb{E}[T_k] \leq \mathbb{E}[\tau_k]$. We show in Figure 6.3 the ratio $\mathbb{E}[\tau_k]/\mathbb{E}[T_k]$ for some representative off critical choices of p and q. The most distinguished feature is that ratio decreases with k. This can be explained in terms of a certain "recycling" mechanism: Consider first the coupon collector process and how it advances from a degree k partial coupling (k edges collected) to a degree $k+1$ partial coupling ($k+1$ edges collected). We have already seen that $T_{k+1} - T_k$ is a geometric random variable with mean $m/(m-k+1)$, which corresponds to a phase of re-visiting an expected number of $m/(m-k+1)$ previously visited edges. Now suppose that at a given instance

the coupon collector process has visited k edges and is in the process of collecting the $k+1$ edge, however the coupling process has only achieved a degree-k' partial coupling with $k' < k$. It is clear that it possible that some of the edges not-yet coupled are re-visited. If additionally the cluster structure in the two configurations V_t and Λ_t is locally sufficiently similar, such that edges are both typically pivotal (or non-pivotal), then it is conceivable that the coupling process can re-use this sequence of re-visits to advance to a degree $k'+1$ coupling. In particular at late stages, where k is close to m, where re-visit sequences become typically comparable in length to m, one expect this re-cycling or "catching-up" process to be most efficient. This is precisely what we observe in Figure 6.3. We note that a closer inspection reveals that the final ratio decreases slowly with increasing system size, consistent with what is shown in Figure 6.2. Furthermore, it appears that the size dependence in the remaining regime for values of k/m is comparably weaker than the final size dependence. It is intriguing to study the partial coupling concepts in more detail, which we plan to pursue in a future study for the complete graph.

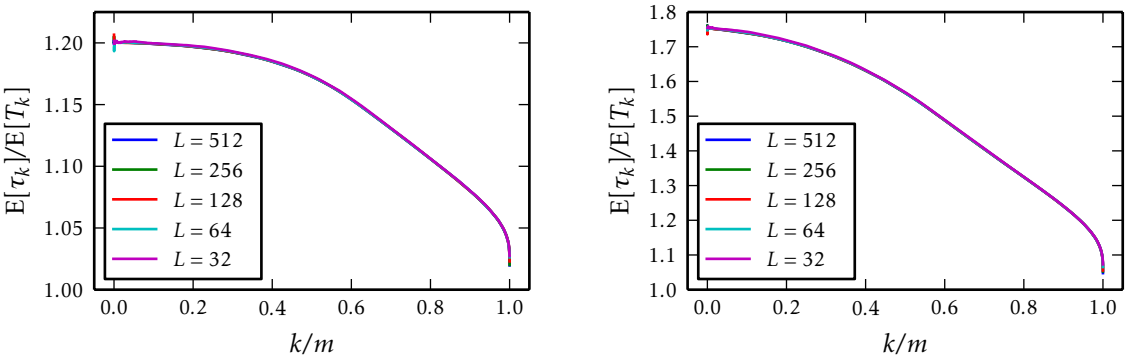


Figure 6.3: *Ratio of expected coupling time to expected coupon collector time for* $(q = 2, p = 0.2)$ *and* $(q = 8, p = 0.3)$*, corresponding to the left and right Figure, respectively.*

6.2.2 At criticality

Let us first consider again the square lattice $\mathbb{Z}_L^2$ where we fix the bond density to the corresponding critical value of $\mathbb{Z}^2$, that is to $p = p_{\mathrm{sd}}(q) = \sqrt{q}/(1+\sqrt{q})$. Recall that we only consider $q \geq 1$ due to lack of monotonicity of the heat-bath chain for $q < 1$. There are two qualitatively different regimes corresponding to first and second order phase transitions. For instance it is rigorously known [128] that for $\mathbb{Z}^2$ the phase transition is continuous or second order for $1 \leq q \leq 4$ and expected to be discontinuous or first order for $q > 4$ [22]. In particular at the location of a first-order phase transition we expect the coupling time to be exponential in L due to the coexistence of ordered and disordered phases. Strictly speaking the disordered and ordered phase refer to the Potts model: In the ordered phase ($\beta > \beta_c(q)$) one of the q spins dominates and in the disordered phase all q-different spins appear roughly with the same frequency. Furthermore the overall probability mass is roughly equally distributed among the two phases. However the Edwards-Sokal coupling allows us to relate the two coexisting phases to random-cluster

phases. Roughly, there is a coexistence between configurations consisting of typically very small (roughly logarithmically large) components, and configurations containing one macroscopic giant component among many very small components. Crucially, there exists an exponentially large probabilistic "barrier" that separates configurations from one phase to its complement, that is configurations outside the phase. Moreover it is plausible that in the coupling from the past algorithm the configuration Λ_t starting in E will very likely be "trapped" in configurations corresponding to the ordered phase, and hence produce a giant component. On the other hand the configuration V_t starting from $\varnothing$ will very likely remain in configurations part of the disordered phase. The two trajectories are hence separated by an exponentially large probabilistic barrier, as they belong to the two phases which a separated by exponentially suppressed regions in phase space. This then yields a exponentially large coupling time. It is apparent that such an exponential computational complexity impeded a study of the size dependence of the expected coupling time numerically. Apart from that, we remark that in such situations, one would, probably, revert to other well established alternative and heuristically more efficient, though not "perfect", simulation methods, such as for instance the multi-canonical algorithm [129, 130] or Wang-Landau sampling [131]; for a general introduction see also [1].

On the other hand we analysed the second order regime extensively. The analysis in the previous section showed that the size dependence of the expected coupling time is likely ruled by the coupon collector mechanism and any cluster weight induced delay causes, asymptotically, at most a change in the overall constant. Yet we do *not* expect this to be the case at criticality. One reason for this is that our numerical analysis of the exponential autocorrelation time in chapter 4 supported a power law scaling in L. By the nature of the phase transition we expect likewise that $\mathbb{E}[\tau] \approx L^w$, where w is the Propp-Wilson exponent, that clearly depends on q. Furthermore we have $w \geq z_{\text{exp}}$ as a consequence of (6.2.3). Additionally there is also a theoretical argument, independent of our numerical observations: The Li-Sokal bound, discussed in section 2.4, demands that $\tau_{\text{exp}} \geq 2\,\text{Var}[N] + 1$ and hence, by (6.2.3), the same slowing down mechanism applies to $\mathbb{E}[\tau]$(

$$\mathbb{E}[\tau] \geq \frac{\log(2)}{8}\,(2\text{Var}[N] - 1). \tag{6.2.8}$$

This enforces a critical slowing down for $q \geq 2$ in case of $\mathbb{Z}_L^2$ [the case $q = 2$ has $\alpha/\nu = 0$ with $\text{Var}[N]/L^2 \approx C \log(L)$] [22].

To start with, we numerically estimated $\mathbb{E}[\tau]$ based on at least 10^4 independent samples for each choice of q considered. Firstly, consider the size dependence of $\mathbb{E}[\tau]$ in Figure 6.4. We extracted the (effective) exponent w by fitting two standard power law model functions AL^w and $AL^w + B$ to our estimates for $\mathbb{E}[\tau]/m$. However without having an appropriate scaling theory for the expected coupling time we were not able to predict a generic ansatz with anticipated corrections. Further, it is unclear if the logarithmic multiplicative factor present for $q = 1$ persists for $q > 1$. Detecting this or distinguishing it from a power law with a small exponent is numerically very hard. Thus the extracted

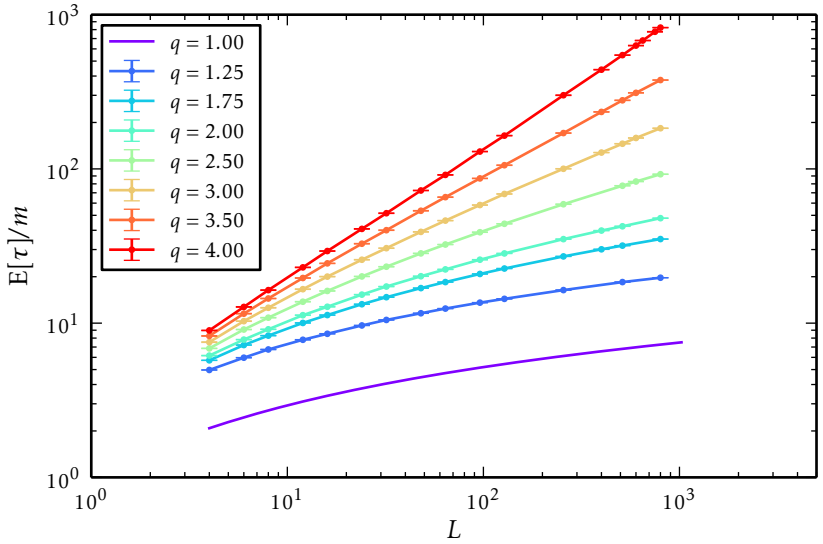


Figure 6.4: *Mean coupling time for the coupling from the past algorithm for the critical random-cluster model on $\mathbb{Z}_L^2$ ($m = 2L^2$) for some cluster weights. Note that the line corresponding to $q = 1$ represents the exact value $\mathbb{E}[T]/m = H_m \sim 2\log(2L) + \gamma$ as in* (6.2.5).

Propp-Wilson exponent w has to be interpreted as an effective exponent, with possible deviations from a true exponent, caused by, e.g., logarithmic corrections. In table 6.1 we summarise our final estimates for the leading exponent w, chosen to be the model that yielded the better goodness of fit. Note that for the system size dependence of $\mathbb{E}[\tau]$ when $q = 1.25$ we expect, if at all, a very small exponent, which is technically very hard to distinguish from a logarithm, which in turn is also a possible scenario consistent with the Li-Sokal bound and hence cannot be ruled out. In fact we are not aware of any reason why the possibility $\mathbb{E}[\tau]/m \approx C\log(m)^p$ with $p \geq 1$ should be excluded when $q \leq 2$.

We also tried to fit a functional form of $AL^B\log(L) + C$ to the data to probe for the possibility of a multiplicative logarithm as for $q = 1$. We find that for cluster weights larger than 2 such fits yield good results, which slightly decrease the estimated Propp-Wilson exponent. On the other hand for $q \leq 2$ we find that such a logarithmic ansatz does not yield reasonable fits. Yet we decided to use the estimates shown in table 6.1, as they are slightly larger and therefore provide a more pessimistic (safer) upper bound for z_{exp}.

For the tricritical case $q = 4$ we find that the before-mentioned fitting functions do not yield any reasonable fit. Motivated by the existence of multiplicative logarithmic corrections for the Potts model on the square lattice when $q = 4$ and $\beta_c = \log(3)$, [94], we tried to fit the functional form $AL^w\left(\log L\right)^{-B}$ to the data. In contrast to the previous fits, this approach yielded a reasonable fit, as shown in table 6.1. We determined the involved exponent B to $B = 0.6(1)$. This suggests that the logarithmic corrections found for equilibrium properties extend to the expected coupling time. Further it suggests that the Li-Sokal bound becomes close to sharp as one approaches $q = 4$.

Before we discuss the possible sharpness of the Li-Sokal bound, let us compare our

q	$z_{\text{int},\mathcal{N}}$	α/ν	w	χ^2/N_{DOF}	Q	$L_{\min}$	$L_{\max}$	Model	$z^{(\text{CM})}_{\text{int},\mathcal{E}'}$
1.25	–	−0.3553	0.013(3)	1.95	0.05	16	2048	$A+BL^w$	0
1.75	–	−0.1093	0.111(2)	0.65	0.59	256	1024	$A+BL^w$	0.06(1)
2	0(log)	0(log)	0.162(5)	1.10	0.36	64	1024	$A+BL^w$	0.14(1)
2.5	0.26(1)	0.2036	0.351(1)	1.01	0.42	48	800	$A+BL^w$	0.31(1)
3	0.45(1)	0.4000	0.509(9)	1.07	0.38	48	800	$A+BL^w$	0.49(1)
3.5	0.636(2)	0.6101	0.696(1)	1.01	0.38	48	600	AL^w	0.69(1)
4	0.75(6)	$1(\log^{-3/2})$	0.98(2)	0.93	0.50	32	1024	$AL^w\big(\log L\big)^{-B}$	0.93(2)

Table 6.1: *Extracted exponent w for a power law system size scaling $\mathbb{E}[\tau]/m \approx L^w$. The model estimates shown correspond to the best choice of the fitting model and system size constraints $L_{\min}$ and $L_{\max}$. Additionally the dynamical critical exponent for the integrated autocorrelation time of the bond density $\mathcal{N}$ is also shown as estimated in [21] (the value for $q = 4$ is taken from our work [53] and only seemingly in violation with the Li-Sokal bound: We simply did not take into account logarithmic corrections in the fitting procedure and enforced a power law, which also yielded a reasonably fit.). The values of $z^{(\text{CM})}_{\text{int},\mathcal{E}'}$ are taken from [88].*

estimates for w to known estimates for z_{exp} from [53, 21]. We show in Figure 6.5 some of the literature values, together with our estimates for the Propp-Wilson exponent w. We can immediately record that the theoretical constraint $\mathbb{E}[\tau]/\tau_{\text{exp}} = O(m\log(m))$ is too crude, as it allows for a difference $w - z_{\text{exp}}$ up to $d = 2$. We however observe merely a difference less than 1/2 in the entire second order phase transition regime. Thus our method together with the Li-Sokal bound leaves only an interval for z_{exp} of approximate width not larger than 1/2. We therefore conjecture that the heat-bath chain for the random-cluster model on $\mathbb{Z}_L^2$ at $p = p_{\text{sd}}(q)$ fulfills

$$\frac{\alpha}{\nu} \le z_{\text{exp}} \le \frac{\alpha}{\nu} + \frac{1}{2}, \quad 1 \le q \le 4, \tag{6.2.9}$$

where we ignore any poly-logarithmic factors in $\mathbb{E}[\tau]$ and the first inequality is of course the Li-Sokal bound. The bound (6.2.9) is supposed to be valid for any $1 \le q \le 4$, however it appears that the width for possible values of τ_{exp} becomes actually smaller as the tricritical point is approached (note that our data suggests that the width is maximised for $q = 1$, where we have $z_{\text{exp}} = 0$, $w = 0$ and $\alpha/\nu = -1/2$). In other words, for $q \to 4$ lower (Li-Sokal) and upper (CFTP) bound become very close, which in turn narrows down the interval of allowed values for z_{exp}.

Sharpness of the Li-Sokal bound

The authors of [21, 53] argued that a conceivable scenario for the critical random-cluster model heat-bath dynamics for $\mathbb{Z}_L^2$ is an "almost" sharpness of the Li-Sokal bound for $q \ge 2$, in the sense that τ_{exp} is by a factor proportional to a poly-logarithm or L^p, with $p > 0$

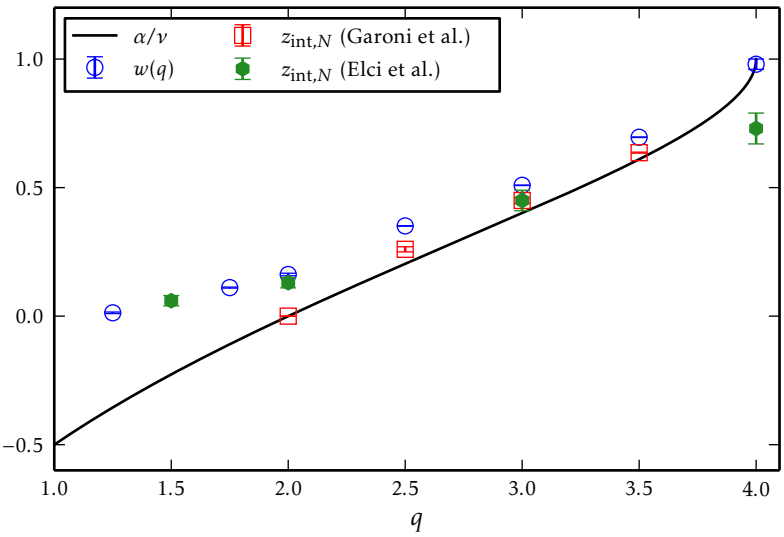


Figure 6.5: *Graphical comparison of critical exponents w, for the expected coupling time, and z_{exp} for the random-cluster model heat-bath chain, taken from [53, 21]. For comparison, the solid red-line shows the Li-Sokal bound. Compare also to Figure 4.11. The estimate for $z_{\text{int},\mathcal{N}}$ for $q = 4$ appears to violate the Li-Sokal bound, however consider the caption of Table 6.1 for an explanation.*

but small, larger than the variance of N. Now, our results shown in Figure 6.5 and Table 6.1 suggest that a similar sharpness property appears for the expected coupling time at criticality when $q = 4$. To quantify this let us compare our fit result for $\mathbb{E}[\tau]$ for $q = 4$ to the corresponding exact leading finite scaling expressions for the variance of N. To start with, note that, in general, one can show, e.g., by means of the Edwards-Sokal coupling [11] (see also [50]), that $\text{Var}[N]$ is related to the heat-capacity of the Potts model, whenever $p = 1 - e^{-\beta}$ and $q \in \{2, 3, \dots\}$. Specifically, this relation applied to the critical $q = 4$ Potts model on $\mathbb{Z}_L^2$ implies that

$$\frac{\text{Var}[N]}{L^2} \approx CL\Big(\log L\Big)^{-3/2} \tag{6.2.10}$$

where C is a constant, and we ignored further sub-leading logarithmic corrections [94]. We thus conclude, ignoring constants, that $\mathbb{E}[\tau]/m$ scales at least as $L\Big(\log L\Big)^{-3/2}$. Now, the numerical observation, that $\mathbb{E}[\tau]$ is well described by $AL^w\Big(\log L\Big)^{-B}$, where w is very close, possibly equal to 1, and $B < 3/2$, suggests that the Li-Sokal bound for $q = 4$ is sharp up to a poly-logarithmic factor in L, or possibly a small power in L, which in turn is numerically very hard to distinguish from a poly-logarithm. This observation is interesting, both in the small-exponent and poly-logarithm scenario, as it narrows down window of possible values of τ_{exp}, supporting the sharpness of the Li-Sokal bound for $q = 4$ for the heat-bath dynamics on $\mathbb{Z}_L^2$ when $p = p_{\text{sd}}(4)$. However, at the current stage we have no explanation for this observation but we believe that $z_{\text{exp}} = \alpha/\nu$ for $q = 4$ is a conceivable scenario modulo poly-logarithms. We remark, that a similar sharpness observation was made in [93] for the Swendsen-Wang-Chayes-Machta dynamics for the tri-critical Potts model on $\mathbb{Z}_L^2$.
That the tricritical point plays a special role with regards to the sharpness of the Li-Sokal

bound can also be seen in Figure 6.6, which shows the ratio $\mathbb{E}[\tau]/\mathrm{Var}_{p,q}[N]$. This method has also been used for instance in [88, 21] to probe the sharpness of the Li-Sokal bound for τ_{exp}.

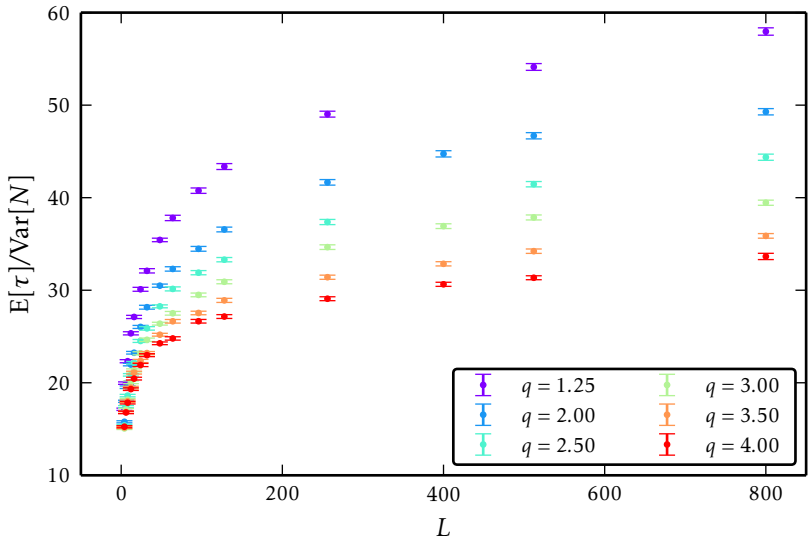


Figure 6.6: *Ratio of expected coupling time* $\mathbb{E}[\tau]$ *to variance of the number of edges* $\mathrm{Var}[N]$ *at criticality. A simple check of the sharpness of the Li-Sokal bound for the coupling time. Notice the improvement of the bound for increasing* q*.*

One can clearly recognize how the Li-Sokal bound becomes better for increasing values of q, where the ratio still shows a very weak system size dependence, as described above. It is an interesting challenging problem to discover the responsible mechanism, that leads to the improvement of the bound as $q = 4$ is approached in 2D.

Three dimensions

So far we have restricted the analysis to the square lattice, but considering the known literature values for z_{exp} for the random-cluster model heat-bath chain in $3D$ presented in [21], it appears that the Li-Sokal bound is far from being sharp for $\mathbb{Z}_L^3$. Furthermore, if we consider the dynamical critical exponent for $q = 2$ (Ising) at criticality on $\mathbb{Z}_L^3$, we find $z_{\mathrm{exp}} \approx z_{\mathrm{int},\mathcal{N}} = 0.35(1)$ for the heat-bath chain [21] and $z_{\mathrm{exp}} \approx z_{\mathrm{int},\mathcal{E}'} = 0.46(3)$ for the Chayes-Machta-Swendsen-Wang algorithm, which is clearly a considerably larger difference as observed in $2D$ for the entire divergent heat-capacity regime ($q \geq 2$). This suggests that Sweeny's algorithm becomes even more superior over the Chayes-Machta algorithm in 3D.

Motivated by this observation, we provide results confirming this superiority of the heat-bath chain assuming a poly-logarithmic dynamic connectivity implementation of the Sweeny chain. More precisely we estimated the expected coupling time on $\mathbb{Z}_L^3$ for cluster weights $q = 1.5, 1.8, 2, 2.2$, at the corresponding anticipated critical bond-densities. Like for the $2D$ case, the estimates are based on 10^4 independent samples for each cluster weight. We considered system sizes up to $L = 96$ for all cluster weights, except for $q = 2.2$ where we

only studied sizes up to $L = 64$. Let us start with $q = 1.5$, for which we set the bond-density to $p_c = 0.31157497$ [96]. We find that the best fit can be achieved with a functional model $AL^w + B$. We correspondingly extract $w = 0.05(1)$. Note, that the corresponding specific heat exponent is estimated to be $\alpha/\nu = -0.32(4)$, c.f. [15]. For comparison, a corresponding recently estimated dynamical critical exponent for the integrated autocorrelation time of the nearest-neighbour connectivity observable $\mathcal{E}'$ in the Chayes-Machta-Swendsen-Wang dynamics, short CMSW dynamics, was estimated 0.13(1) [88]. Moreover, the authors observe that the corresponding normalised autocorrelation function is very close to a pure exponential, which together with the plausible assumption that $\mathcal{E}'$ is not orthogonal to the slowest "mode", suggests that $\tau_{\text{int},\mathcal{E}'} \approx \tau_{\text{exp},\mathcal{E}'} \approx \tau_{\text{exp}}$, i.e. $z_{\text{int},\mathcal{E}'}$ roughly equals z_{exp} for the CMSW dynamics on $\mathbb{Z}_L^3$. This value is in particular larger than our upper bound for z_{exp} for the heat-bath chain obtained from the estimation of $\mathbb{E}[\tau]$.
The results for the next cluster weight $q = 1.8$ are similar (here we set $p_c = 0.34096070$, see [96]), we again obtain the best fit with a functional form $AL^w + B$ and determine $w = 0.18(2)$, whereas $\alpha/\nu = -0.15(5)$, and the CMSW exponent reads 0.29(1).
For the Ising model, i.e. $q = 2$, we find for $p_c = 0.35809123$, that the data is again best described by the same functional form, and extract $w = 0.406(6)$. The estimated value of α/ν is 0.174(1) [88], and the corresponding exponent for the CMSW chain was estimated $z_{\text{int},\mathcal{E}'} = 0.46(3)$.
Lastly, the data for $q = 2.2$, still believed to be a second order phase transition at the anticipated critical point $p_c = 0.37361401$, is best described by AL^w, and we determine $w = 0.698(3)$, whereas $\alpha/\nu = 0.50(4)$ [88], and $z_{\text{int},\mathcal{E}'} = 0.76(1)$ [88]. We summarise our 3D results in table 6.7.

q	p_c	w	α/ν [88]	$z_{\text{int},\mathcal{E}'}$ [88]
1.5	0.31157497(43)	0.05(1)	−0.32(4)	0.13(1)
1.8	0.34096070(39)	0.18(2)	−0.15(5)	0.29(1)
2.0	0.35809123(4)	0.406(6)	0.174(1)	0.46(3)
2.2	0.37361401(35)	0.698(3)	0.50(4)	0.76(1)

Figure 6.7: *Estimates for the exponent w in the system-size scaling of the expected coupling time $\mathbb{E}[\tau]/m \approx L^w$ for the heat-bath chain for the random-cluster model at criticality on $\mathbb{Z}_L^3$. The estimates for p_c for $q = 1.5$, $q = 1.8$ and $q = 2.2$ have been privately communicated to us by Garoni et al. [96]. The critical value p_c for $q = 2$ is taken from [132].*

Therefore our results on the expected coupling time at critcality in three dimensions show, using $z_{\text{exp}} \leq w$, that the Sweeny approach with dynamic connectivity algorithm is asymptotically more effcient than the CMSW approach for $1 < q \leq 2.2$. This motivates a more detailed study of the Sweeny dynamics in three dimensions that provides estimates for z_{exp} beyond $q = 2$.

We remark that the study of [21] also considered the case $q = 0$ as well as $q = 1$, however for the former the CMSW does not work and in the percolation case both algorithms become equivalent in the sense that CMSW is a sequential sweep of m heat-bath updates. It would also be desirable to see whether the Sweeny approach becomes even more superior in $D > 3$, and if so to understand the underlying mechanism.

6.3 Coupling time distribution

Let us know investigate the coupling time distribution, that is the probability that the map $F_t^{\uparrow}$ (or $F_t^{\downarrow}$) is not constant, which clearly equals, per definition, $\mathbb{P}[\tau > t]$. One reason for our interest in the distribution stems from the fact that the expectation provides only very loose information on the tail of the distribution. One standard way to obtain a tail-bound using only the first moment (expectation) of the distribution of τ is to exploit the point wise inequality $\mathbb{1}_{\{\tau \geq t\}} t \leq \tau$, which after taking expectation on both sides yields

$$\mathbb{P}[\tau \geq t] \leq \frac{\mathbb{E}[\tau]}{t}.$$

It is clear that without further study Markov's inequality does not allow for a detailed insight into the likelihood of outliers, which can drastically impair the efficiency in any application. In fact even in the case $q = 1$ where the expectation is known explicitly, it turns out that the bound is too loose, as fluctuations typically happen on a smaller scale than $\mathbb{E}[\tau]$.

As it is standard for randomised algorithms [38], one needs to understand the likelihood of such extreme events and would ideally like to obtain a high probability bound suppressing extreme deviations from the mean for increasing system sizes. Beyond this practical motivation, we can also gain insight into the coupling process itself and attempt to understand previous observations, such as the almost optimal coupling time for off-critical random-cluster models. In fact, we will study asymptotic ($m \to \infty$) properties of the distribution, and therefore need to introduce a normalisation of τ in order to obtain a non-degenerate asymptotic distribution. More precisely we consider the probability of relative deviations from the expectation $\mathbb{E}[\tau]$ by studying the re-scaled (or standardised) random-variable

$$\eta \equiv \frac{\tau - \mathbb{E}[\tau]}{\sqrt{\mathrm{Var}[\tau]}} \tag{6.3.1}$$

To start with, we consider the simplest case $q = 1$, i.e. bond percolation, where, as stated above, $\mathbb{P}[\tau > t] = \mathbb{P}[T > t]$ for any t, and T is the coupon collector time. The study of deviations of T from its mean is very well understood, both asymptotically, and in the finite case where non-asymptotic concentration results are known [38]. A fundamental result concerning the limiting distribution of a slightly different re-scaled coupling time

goes back to Erdős & Rényi in 1961 [133]:

$$\lim_{m\to\infty} \mathbb{P}\left[\frac{T - m\log(m)}{m} \leq x\right] = e^{-e^{-x}}. \tag{6.3.2}$$

The double exponential on the right hand side is the cumulative distribution function of the standard Gumbel distribution, one member of the family of generalised extreme value (GEV) distributions [134]. Before we derive the corresponding result for η, let us elaborate on two aspects. Firstly, the result (6.3.2) can be derived heuristically in a continuous time version of the coupon collector problem. Here each coupon is issued at time τ_i for $1 \leq i \leq m$ that has distribution

$$\mathbb{P}[\tau_i \leq t] = \mathbb{1}_{\{t\geq 0\}}\left(1 - e^{-\lambda t}\right), \tag{6.3.3}$$

that is an exponential waiting time with rate λ or mean $1/\lambda$. We assume that all waiting times τ_i are identically and independently distributed. Clearly, in order to collect all edges we need to wait for the last edge (coupon) to be issued. This casts the study of T into an extreme-value problem of the random variable $M = \max\{\tau_i\}_{i\leq i\leq m}$:

$$\begin{aligned}
\mathbb{P}[T \leq t] &= \mathbb{P}[M \leq t], \\
&= \prod_{i=1}^{m} \mathbb{P}[\tau_i \leq t] \\
&= \mathbb{1}_{\{t\geq 0\}}\left(1 - e^{-\lambda t}\right)^m.
\end{aligned}$$

At this point we need to find a normalisation such that the distribution of T is non-degenerate in the limit $m \to \infty$. Similar to the above arguments we obtain for two sequences $(a_m)_{m\geq 1}$ and $(b_m)_{m\geq 1}$

$$\mathbb{P}\left[\frac{T - a_m}{b_m} \leq t\right] = \left(1 - e^{-\lambda(tb_m + a_m)}\right)^m.$$

Now, having Euler's well known limit formula for the exponential in mind, i.e.,

$$\lim_{m\to\infty}\left(1 + \frac{x}{m}\right)^m = e^x,$$

it is not hard to see that with the choice $a_m = \log(m)/\lambda$ and $b_m = 1/\lambda$, we obtain

$$\lim_{m\to\infty} \mathbb{P}[\lambda T - \log(m) \leq t] = e^{-e^{-t}},$$

which coincides with the limit in (6.3.2) for $\lambda = 1$. Let us emphasise that the Gumbel distribution is recovered independent of the choice of $\lambda > 0$. The variation of λ simply changes the overall scale and location of the distribution, but the nature of the fluctuations is unaffected by the particular choice, as it is the waiting process for the last edge to be issued which characterises the fluctuations. For instance, a smaller value of λ corresponds

to typically longer waiting times for coupons, which we might imagine could mimic, in a very naive approach, the slowing down caused by weak correlations off criticality. Indeed, this expectation is supported by our numerical analysis of the coupling time distribution, as described below.
The second aspect we want to mention is that one can easily derive a non-asymptotic tail bound for the distribution of T by means of a union-bound argument [38]:

$$\begin{aligned}\mathbb{P}[T > t] &= \mathbb{P}[\cup_{e\in E}\{\tau_e > t\}],\\ &\leq m\mathbb{P}[\tau_e > t],\\ &= m\left(1-\frac{1}{m}\right)^t,\\ &\leq me^{-t/m}.\end{aligned}$$

Recall, for $q = 1$ τ_e is the first time edge e is visited. For instance, for any $\epsilon > 0$ we can choose $t = (1+\epsilon)m\log(m)$, such that

$$\mathbb{P}[T > (1+\epsilon)m\log(m)] \leq m^{-\epsilon} = o(1). \tag{6.3.4}$$

This result shows that for increasing system size it becomes more and more unlikely that T exceeds $\mathbb{E}[T]$ by a multiple of $m\log(m)$. In other words fluctuations around the mean occur at a smaller scale. We can see that according to (6.3.2) the right scale for the fluctuations is indeed m, the number of edges. Yet slightly re-phrased, one can still observe fluctuations asymptotically, when measuring T in terms of sweeps (division by m), however will suppress fluctuations when measuring in in terms of the intrinsic coupon-collector scale $m\log(m)$. Compare (6.3.4) to what one obtains from Markov's inequality: $\mathbb{P}[T > (1+\epsilon)m\log(m)] \leq c/(1+\epsilon)$, where c is a constant. Notably this does not provide us with a concentration result stating that with increasing system size deviations on this scale become negligible. Finally, let us remark that the same arguments can in principle also be applied to the coupling time τ for $q > 1$, that is

$$\mathbb{P}[\tau > t] \leq m\mathbb{P}[e \in S_t], \tag{6.3.5}$$

and note that because $e \in S_t \Rightarrow \tau_e > t$ we have $\mathbb{P}[e \in S_t] \leq \mathbb{P}[\tau_e > t]$. We remark that in order to bound the mixing time, it suffices to find a time t such that $\mathbb{P}[e \in S_t]$ or $\mathbb{P}[\tau_e > t]$ is less or equal $1/(4m)$ uniformly in e, which then would yield $t_{\text{mix}} \leq t$. Interestingly, also from a practical point of view such a time t is sufficient to obtain results suppressing large deviations with high probability, i.e. knowing that $\mathbb{P}[\tau > t] \leq 1/4$ for a particular choice of t, allows one for instance to "boost" this bound and obtain $\mathbb{P}[\tau > t\log(m)/\log(4)] \leq 1/m$. In fact we have already used this argument to derive (6.1.16). We can apply similar arguments to derive $\mathbb{P}[\tau^{\downarrow} > kt] \leq \frac{1}{4}^k$, provided $\mathbb{P}[\tau^{\downarrow} > t] \leq 1/4$. In practise however, without any particular information on the tail of the distribution of τ_e, this is merely a re-formulation of the problem. Yet, it is conceivable that under "mild" conditions such as off-criticality, the tail probability $\mathbb{P}[\tau_e > t]$ could be asymptotically equivalent to or

bounded by $Ce^{-t/(m+o(m))}$, where C is again a constant positive constant.
Now, it remains to show how we can obtain a non-degenerate limiting law for η in case of the coupon collector's problem. Firstly, it is clear that $\mathbb{E}[\eta] = 0$ as well as $\mathrm{Var}[\eta] = 1$. Thus the corresponding limiting non-degenerate distribution must also fulfil these two constraints. Note, that in general the Gumbel distribution has a scale and a location parameter, denoted by $\beta > 0$ and $\mu \in \mathbb{R}$, respectively. The particular re-scaling in (6.3.2) corresponds to $\beta = 1$ and $\mu = 0$, which is commonly known as the standard case. The following derivation of the limiting distribution of η closely follows the original ideas of Erdős & Rényi in [133] that led in particular to (6.3.2). We remark, that no conceptually new ideas are needed to derive the result for η. However, we provide the derivation, as it allows us to understand the relevant assumptions underlying it. To start with, let $\varphi_m(t) = \mathbb{E}[e^{it\eta}]$ (where i is the imaginary unit) be the characteristic function of η, we have

$$\varphi_m(t) = e^{-\frac{imH_m t}{\sqrt{\mathrm{Var}[T]}}} \rho_m\left(\frac{t}{\sqrt{\mathrm{Var}[T]}}\right),$$

where $\rho_m(t)$ is the characteristic function of a sum of m independent geometric random variables $\{\delta_k\}_{k=1}^m$ with parameters $p_k = (m-k+1)/m$. Because we have for a geometric random variable with parameter p the characteristic function equals $pe^{it}/(1-(1-p)e^{it})$ we obtain

$$\begin{aligned}\rho_m(t) &= \prod_{k=1}^{m}\left(\frac{e^{it}\frac{m-k+1}{m}}{1-\frac{k-1}{m}e^{it}}\right)\\ &= \prod_{k=1}^{m}\left(\frac{\frac{k}{m}}{e^{-it}-\frac{m-k}{m}}\right)\\ &= \frac{1}{\prod_{k=1}^{m}\frac{m}{k}\left(e^{-it}-1\right)+1},\end{aligned}$$

and we obtain after some straightforward algebra

$$\varphi_m(t) = \frac{1}{\prod_{k=1}^{m} e^{\frac{it}{\alpha_m k}}\left(\frac{m}{k}\left(e^{-\frac{it}{m\alpha_m}}-1\right)+1\right)}, \tag{6.3.6}$$

here we defined $\alpha_m \equiv \frac{\sqrt{\mathrm{Var}[T]}}{m}$. We note that $\alpha_m \to \frac{\pi}{\sqrt{6}} \equiv \alpha$ for $m \to \infty$, see for instance [38]. Taking the limit in (6.3.6) we obtain

$$\lim_{m\to\infty}\varphi_m(t) = \frac{1}{\prod_{k=1}^{\infty} e^{\frac{it}{\alpha k}}\left(1-\frac{it}{\alpha k}\right)}. \tag{6.3.7}$$

Now, using the product representation of the gamma function [133], the product in (6.3.7) can be written in terms the Gamma function,

$$\lim_{m\to\infty}\varphi_m(t) = \Gamma\left(1-\frac{it}{\alpha}\right)e^{-\frac{it\gamma}{\alpha}}, \tag{6.3.8}$$

where $\gamma \approx 0.5772156649$ is the Euler-Mascheroni constant. Lastly, the r.h.s. in (6.3.8) is the characteristic function of the Gumbel distribution with location parameter $-\gamma\sqrt{6}/\pi$ and scale parameter $\sqrt{6}/\pi$, hence we have by Levy's continuity theorem

$$\lim_{m\to\infty} \mathbb{P}\left[\frac{T-\mathbb{E}[T]}{\sqrt{\mathrm{Var}[T]}} \leq x\right] = e^{-e^{-x\frac{\pi}{\sqrt{6}}-\gamma}}. \tag{6.3.9}$$

In general, a Gumbel distribution with location parameter μ and scale parameter $\beta > 0$, has mean $\mu + \beta\gamma$ and variance $\pi^2\beta^2/6$. The particular choice of μ and β above clearly recovers the desired result, that the mean is 0 and the variance equals 1. Further, a characteristic of the Gumbel distribution is its skewness, that is the right tail is longer and heavier than the left. Compare this to, say, a normal distribution where left and right tail are symmetric. Needless to say that this asymmetry has a direct influence on the running time of the coupling from the past algorithm for $q = 1$, as running times exceeding the mean are more likely than shorter runs.

The Gumbel distribution is a ubiquitous distribution, for instance it describes the freezing-temperature fluctuations in the random-energy model [135], the distribution of pseudo-critical temperatures in the Edwards-Anderson spin glass model [136] or the distribution of the cover time of the simple random-walk on the discrete torus of dimension at least three [137], to name just a few.

The previous derivation of the Gumbel distribution (6.3.2) crucially relied on the fact that the coupon collector time can be decomposed as a sum of independent (but not identically) distributed geometric random-variables. In principle we can make an analogous decomposition ansatz for τ, based on the degree k partial-coupling times, more precisely:

$$\tau \equiv \tau_m = \sum_{k=1}^{m} (\tau_k - \tau_{k-1}) \tag{6.3.10}$$

with the convention that $\tau_0 = 0$. Thus the problem is equivalent to the study of the sum of m terms of the form $\tau_k - \tau_{k-1}$. However, the problem is that for $q > 1$ the terms for different values of k's are not independent. Clearly, when concerned with the task of obtaining tail bounds, one can dispose of the requirement of independence, and work in the dependent setting such as for example in case of negative dependence or martingales [138]. Unfortunately we were not able to cast the study of τ into such a framework.

Furthermore, let us remark that the idea of using a telescoping sum trick, as in (6.3.10), was used in the interesting work [139], that shows how extreme value statistics can be formulated as the study of the sum of a particular class of correlated random-variables. More importantly, the work [139] shows that under certain conditions a sum of correlated random-variables leads to a generalised Gumbel distribution, however without any apparent interpretation in terms of an extreme value problem. Yet, we could not confirm the relevant conditions for our problem, as it requires a specific functional form of the joint probability distribution of the terms composing the sum. Beyond that, we actually lack any knowledge about the functional form of the marginal distribution of each of the

terms in the sum, as well as the corresponding mean.
We therefore decided to fall back to a numerical analysis of η. This study provides us with valuable insight, which may lead to a future rigorous study. In fact we mention already here that the numerical observations stimulated our rigorous analysis of the asymptotics of the coupling time distribution in the case of the cycle graph in section 6.4. Now, let us continue with our numerical results, which are divided into two subsections, corresponding to the off-critical and critical situations, respectively.

6.3.1 Off critical

We analysed the coupling time distribution for off-critical parameter choices in $2D$ and $3D$ by considering the square- and simple cubic lattice with periodic boundary conditions, that is $\mathbb{Z}_L^2$ and $\mathbb{Z}_L^3$, respectively. The before-mentioned study of the expected coupling time for off critical parameters suggests that the fluctuations stem from the underlying coupon collector process, and the introduction of a cluster weight merely reflects itself in a sub-leading change of scale. It is therefore plausible to expect that η follows again (6.3.9) asymptotically.
Indeed, in the left panel of Figure 6.8 we show the empirical distribution of η in two dimensions for $q = 8$ and $p = 0.3 < p_{\mathrm{sd}}(8)$ in $2D$ and compare it to the probability density function corresponding to (6.3.2), which is shown as the solid green line. The graphical coincidence with the asymptotic result in (6.3.9) supports our initial anticipation. Likewise, the right panel in Figure (6.8) shows the empirical distribution of η for $q = 1.5$ and $p = 0.2$ in $3D$. Recall that one recent [96] estimate for the location of the second order phase transition of the random-cluster model on $\mathbb{Z}^3$ for $q = 1.5$ is $p_c = 0.31157497 \pm 0.00000059$, hence $p = 0.2$ is clearly sub-critical. In order to obtain a quantitative measure of how well

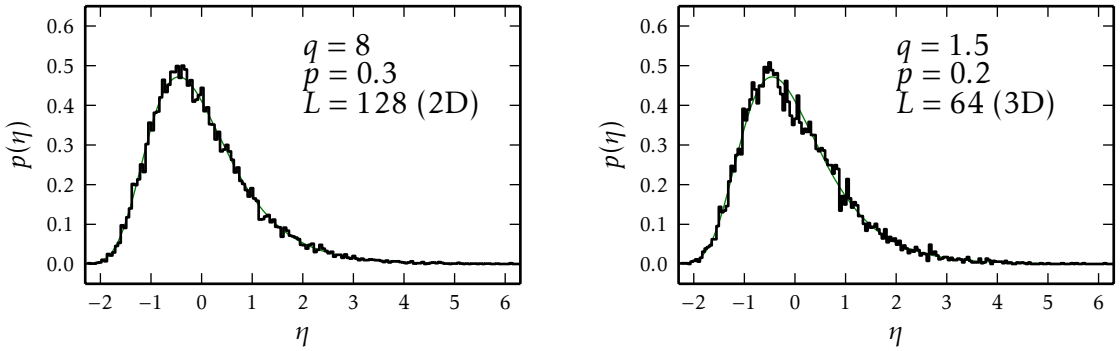


Figure 6.8: *Histogram for η for the coupling process on $\mathbb{Z}_L^2$ and $\mathbb{Z}_L^3$, for parameters as specified in the figure. The histograms are based on* 20000 *independent samples in 2D and* 10000 *in 3D. Here $p(\eta)$ denotes the corresponding probability density function of η. The solid green line shows the probability density function corresponding to* (6.3.9).

the distribution of η is described by (6.3.9), we used the method of maximum likelihood estimation [112] to extract the corresponding shape, $\xi \in \mathbb{R}$, location, $\mu \in \mathbb{R}$, and scale, $\sigma > 0$, parameters of the generalised extreme value (GEV) distribution with cumulative

distribution function:

$$F_{\mathrm{GEV}}\left(x;\xi,\mu,\sigma\right) = \begin{cases} e^{-e^{-(x-\mu)/\sigma}} & \xi = 0, \\ e^{-\left(1+\xi(x-\mu)/\sigma\right)^{-1/\xi}} & \xi \neq 0, \end{cases} \tag{6.3.11}$$

where the support is $(-\infty,\infty)$ whenever $\xi = 0$ and $[\mu - \sigma/\xi, \infty)$ when $\xi > 0$, the case $\xi < 0$ corresponding to the Weibull distribution, is not relevant to this work. One reason for the use of the GEV framework is that it nicely interpolates between the three different families of extreme value distributions as ξ is varied. In particular, for $\xi = 0$ it can be verified that $F_{\mathrm{GEV}}(x;\xi,\mu,\sigma)$ describes the Gumbel family. Speficically, for $\xi = 0$, $\mu = -\gamma\sqrt{6}/\pi$ and $\sigma = \sqrt{6}/\pi$ we can actually recover the limiting Gumbel distribution in (6.3.9). However, the variation of ξ allows us to quantitatively probe for deviations from the Gumbel scenario. From a practical point of view, ξ might fluctuate, simply because of the fact that our computational analysis is non-asymptotic and the work with empirical samples. Therefore we believe that a shape parameter ξ consistent to 0 supports the asymptotic validity of (6.3.9).
For the particular choice $p = 0.3$ and $q = 8$ in 2D, we obtain for $L = 128$ (corresponding to the left panel in Figure 6.8)

$$\xi = 0.01(1) \quad \mu = -0.45(1) \quad \sigma = 0.77(1), \tag{6.3.12}$$

which is based on 19000 independent samples, and the error bars are obtained by the bootstrap re-sampling method [112]. This result indicates that $\xi = 0$ and is further in perfect agreement with the predicted values

$$\mu = -\frac{\gamma\sqrt{6}}{\pi} = -0.45005320754\ldots \quad \sigma = \frac{\sqrt{6}}{\pi} = 0.77969780123\ldots. \tag{6.3.13}$$

Completely analogous, when we consider the heat-bath chain on $\mathbb{Z}_L^3$ for $q = 1.5, p = 0.2$, corresponding to the right panel in Figure 6.8, we obtain

$$\xi = 0.01(1) \quad \mu = -0.45(2) \quad \sigma = 0.78(1), \tag{6.3.14}$$

which is based on 11000 samples. Additionally, we have also data for the super-critical parameter choice $q = 2.2$ and $p = 0.6$ in 3D. Here we also obtained 11000 samples and estimate $\mu = -0.45(2)$, $\xi = 0.00(1)$ and $\sigma = 0.79(1)$.
Based on our findings we are confident to conjecture that the re-scaled coupling time distribution of the heat-bath chain for the random-cluster model on $\mathbb{Z}_L^d$ is asymptotically, i.e., for $L \to \infty$, equal to the right hand side of (6.3.9), whenever the equilibrium model is away from a point where it undergoes a phase transition on $\mathbb{Z}^d$. Our observations confirm that fluctuations in the coupling time stem asymptotically only from a waiting process of localised "defects", where the scale and centre of the fluctuations are determined by the particular off-critical choice of p and q. In contrast, we have observed a qualitatively

different behaviour at the point where the model on the corresponding infinite lattice undergoes a continuous phase transition. It is therefore tempting to check whether (6.3.9) holds or not.

6.3.2 At criticality

Continuous regime

In the left panel of Figure 6.9 we show the empirical distribution of η for $q = 3.5$ and $p = p_{\mathrm{sd}}(3.5)$ for the graph $\mathbb{Z}^2_{800}$. Recall, that for $q \leq 4$ the phase transition at $p = p_{\mathrm{sd}}(q)$ is continuous. Somewhat surprisingly we find again a very good graphical agreement with (6.3.9). The same holds for the three dimensional system considered, where we set $q = 1.5$ and $p = 0.311575$ for $\mathbb{Z}^3_{64}$, as shown in the right panel of Figure 6.9. These observations strongly suggest that a continuous phase transition in the random-cluster model is sufficiently "well-behaved" for the dominant effect, besides a change in the scale and location of fluctuations, to still originate from fluctuations of a waiting-type process resembling the mechanism causing fluctuations in the coupon collector's problem. A more

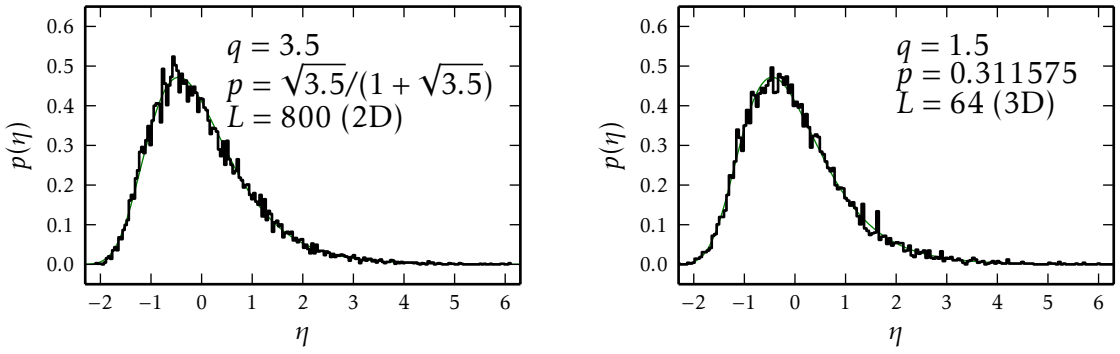


Figure 6.9: *Histogram of η for the coupling process on $\mathbb{Z}^2_L$ and $\mathbb{Z}^3_L$, for critical parameters as specified in the figure. The histograms are based on* 15000 *independent samples in 2D and* 10000 *in 3D. The solid green line shows the probability density function corresponding to* (6.3.9).

quantitative test can again be obtained by means of the maximum-likelihood method in combination with a bootstrap error estimation. We actually extended the analysis of the distribution of η in two dimensions to cluster weights in the interval $(1, 4]$ already encountered in the analysis of the system size scaling of $\mathbb{E}[\tau]$ along the self-dual line. In fact we used the same samples underlying the estimation of $\mathbb{E}[\tau]$. We find generally, that for all cluster weights strictly less than 4, the estimates for shape, location and scale parameters are consistent with $\xi = 0$ and the Gumbel distribution (6.3.9).
We summarise our results in table 6.2. In $3D$ we find, both at the corresponding (estimated) critical point for $q = 1.5$ and $q = 2.2$, that the resulting estimates are consistent with the Gumbel distribution (6.3.9).
In the marginal case $q = 4$ and $p = p_{\mathrm{sd}}(4) = 2/3$, for which the random-cluster model is still continuous, we find a peculiar behaviour, as shown in table 6.3. The data suggests a

q	L	μ	σ	ξ	N_s
1.75	1024	−0.45(3)	0.80(2)	−0.02(2)	3360
1.75	512	−0.45(2)	0.79(1)	−0.01(2)	9800
2.00	512	−0.45(2)	0.78(1)	0.00(1)	9000
2.00	800	−0.45(2)	0.80(1)	−0.02(1)	9000
2.50	512	−0.45(2)	0.79(2)	−0.01(1)	9000
3.00	512	−0.45(2)	0.78(1)	0.01(1)	9000
3.50	512	−0.46(2)	0.77(1)	0.02(2)	8990
3.50	800	−0.46(1)	0.77(1)	0.01(1)	15730

Table 6.2: *Shape, location and scale parameter estimates for choices of* $(q, p_{\mathrm{sd}}(q)$ *obtained from maximum likelihood estimators for the generalised extreme value distribution* (6.3.11).

deviation from the $q = 1$ Gumbel distribution. More precisely, we observe that the location parameter tends to settle at a value slightly in the Frèchet regime ($\xi > 0$). One immediate consequence is the smaller support of the Fréchet family, i.e., a Frèchet random variable is supported in $[\mu - \sigma/\xi, \infty)$, which implies that for sufficiently large L there exists an effective lower cut-off for relative deviations from $\mathbb{E}[\tau]$. This is an effect which is absent in the Gumbel family, where η can in principle attain arbitrary small values, as long as this value is conform to the trivial cut-off that $\tau, T \geq m$. However, based on our numerical data we cannot exclude that the observed deviation from (6.3.9) is only a finite size artifact and vanishes for $L \to \infty$, for which, strictly speaking, even the rescaled distribution for $q = 1$ only reduces to (6.3.9). It appears that $q = 4$ is in many regards the most challenging case in the entire second order critical regime for the square lattice.

L	μ	σ	ξ	N_s
128	−0.46(2)	0.73(1)	0.05(2)	8990
256	−0.47(2)	0.71(1)	0.08(2)	9000
512	−0.47(1)	0.73(1)	0.06(1)	18940
800	−0.47(1)	0.72(1)	0.07(1)	20000
1024	−0.47(2)	0.72(1)	0.07(1)	9850

Table 6.3: *Maximum-likelihood estimators for the shape-, location- and scale parameter of the generalised extreme value distribution* (6.3.11) *for different system sizes and fixed* $q = 4$ *and* $p = p_{\mathrm{sd}}(4) = 2/4$.

Discontinuous regime

Even more severe becomes the deviation from the Gumbel distribution for $q > 4$ along the self-dual line, where the phase transition is believed to be discontinuous and the expected coupling time should hence be exponential in L. We generated 10000 samples of the coupling time for the choice $q = 5$ and $p = p_{\mathrm{sd}}(5)$ on $\mathbb{Z}_{256}$, which is a comparably weak first order phase transition. We obtained the following GEV parameter:

$$\xi = 0.19(2) \quad \mu = -0.49(2) \quad \sigma = 0.62(1). \tag{6.3.15}$$

The histogram together with the GEV probability density function corresponding to the parameters (6.3.15) is shown in Figure 6.10. This is strong evidence for the validity of the Fréchet distribution with corresponding shape parameter. Further, notice that, in comparison to $q = 4$, the shape parameter ξ increased and the scale parameter decreased. This effectively increases the lower bound of the support of the Fréchet distribution and hence worsens the fluctuations around the mean, as running times below the mean become more unlikely (suppressed). The astronomically large time scales involved in the coupling

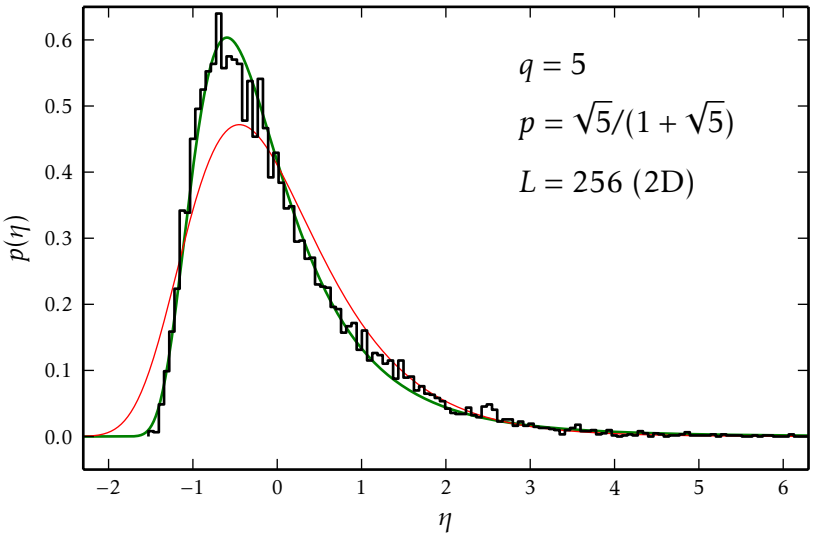


Figure 6.10: *Histogram for η for the heat-bath chain on $\mathbb{Z}^2_{256}$, with $q = 5$ and $p = p_{\mathrm{sd}}(5)$ based on 10000 samples. The red solid curve is the density function corresponding to (6.3.9) and the green corresponds to a generalised extreme value distribution with shape, location and scale parameters estimated as described in the text.*

process at the location of a first order phase transition, make it nearly impossible to reach larger system sizes and collect sufficient statistics. However, the Frèchet scenario for first order and possibly tri-critical (marginal) phase transitions seems plausible. It is an interesting and challenging problem to find out whether and under which circumstances the Frèchet distribution describes the distribution of η asymptotically. We hope that this first numerical observation motivates a (rigorous) study of related phenomena.

6.3.3 Other models and localisation

We also studied the coupling time of the Glauber dynamics, or spin-flip heat-bath dynamics, for the Ising model on $\mathbb{Z}_L^2$. In case of the Ising model one can utilise a deep relationship between spatial and temporal mixing. More precisely, the Ising model (without magnetic field) on a finite box of square lattice with arbitrary boundary conditions is known to have the *strong spatial mixing property* [140] for $0 \leq \beta < \beta_c = \log(1+\sqrt{2})/2$. This property, very roughly, states that the influence (measured in variation distance) of a change in the boundary condition at a vertex v, on the probability of an event in a distant region, decays exponentially with the distance between the region and v. Now, remarkably strong spatial mixing can be shown to imply rapid mixing for the Glauber dynamics on amenable graphs, see for instance [141] for a probabilistic proof. In fact the arguments of [141] can be used to show that the two coupled Ising-Glauber Markov chains, either in the forward or backward coupling, will very likely agree at any fixed vertex at times larger or equal $Cn\log(n)^2$ (for some constant C and where n is the number of vertices in the box). More precisely for such times the probability of disagreement at a fixed vertex is not larger than $1/4n$, independent of the boundary condition (spin assignments at the boundary of the box). Clearly, this allows us, together with the above union-bound argument, to establish rapid mixing and desired tail bounds. For a detailed and more formal treatment we refer the reader to [141].

The reason why we mention this argument is simply that it formalises the idea of locality in correlated systems and shows that this can in some instances indeed be enough to establish rapid mixing. Now, if we dare to conjecture, one might say that such a localisation, or a weaker notion of it as outlined below, must be a necessary condition for Gumbel-type fluctuations in the coupling time distribution. One is tempted to generalise the ideas of [141] to the heat-bath chain for the random-cluster model. However, to our knowledge it is not established that the random-cluster model on $\mathbb{Z}^2$ possesses the strong spatial mixing property. On the other hand, a different, weaker, notion of spatial mixing can be shown to hold under certain conditions for the random-cluster model on $\mathbb{Z}^2$ with $q \geq 1$ [142]. Remarkably, only in conjunction with very recent results [28] it is possible to confirm that this weaker notion of spatial mixing holds precisely for $p < p_{\mathrm{sd}}(q) = \sqrt{q}/(1+\sqrt{q})$ and correspondingly, by duality, also for $p > p_{\mathrm{sd}}(q)$ (whenever $q \geq 1$). For the critical point it was recently established in [128] that whenever $1 \leq q \leq 4$ one has *polynomial ratio weak mixing* which does not any more provide exponential decay but a polynomial decay. We will pursue in a future study the question whether such weaker notions of spatial mixing are sufficient to establish the rapid mixing property.

Lastly, coming back to the spin heat-bath chain for the Ising model, note that since the recent work [48] it is now rigorously established that the coupling from the past algorithm for the Ising model (in spin formulation) on a finite box of the square lattice with arbitrary boundary conditions has expected running time at most polynomial in the volume for any temperature $\beta \leq \beta_c$, that is in particular *including* criticality. However, the precise value of the exponent at criticality is not established rigorously.

Our numerical analysis of the coupling time distribution for the Ising model shows very good agreement with the Gumbel law (6.3.9) whenever $\beta < \beta_c = \log(1+\sqrt{2})/2$, and hence precisely where the strong spatial mixing holds. On the other hand, for $\beta = \beta_c$ we find that the limiting distribution seems to fall into the Frèchet class. This is possibly related to the sharpness of the phase transition (from a dynamical point of view), as it is known that for $\beta > \beta_c$ the mixing is at least exponential in L. For more details we refer to [143]. Moreover, we analysed the "single-bond chain", a Markov chain on the random-cluster model state space, that naturally appears in the study of the Chayes-Machta algorithm [144]. We find a completely analogous behaviour to the heat-bath chain for the random-cluster model, i.e., the histogram for η is best described by the Gumbel law (6.3.9) whenever $p \neq p_{\mathrm{sd}}(q)$ for all $q \geq 1$ and for $p = p_{\mathrm{sd}}(q)$ as long as $1 \leq q < 4$. Again, we refer the interested reader for more details to [143]. Finally, let us remark that the Gumbel distribution was also found in the coupling time distribution of a Markov chain on sets of two-dimensional random rhombus tilings [145].

6.4 Rigorous results in one dimension

In this section we rigorously establish that the asymptotic distribution of a properly rescaled coupling time in case of the cycle graph on m vertices (m-cycle or $\mathbb{Z}_m$) converges to the standard Gumbel distribution in (6.3.2). This is possibly the most basic case to analyse which does *not* correspond to a Bernoulli measure over edges, like for instance for forests and trees.

Before we dive into the details let us first derive a straightforward upper bound for the expected coupling time. The crucial observation is that we can bound τ by the running time $\tilde{\tau}$ of a simplified random process, which operates as follows. Run the coupling process until all edges have been visited at least once, that is up to time T_1, where T_1 is the corresponding coupon collector's time. In what follows we call the interval $[1, T_1]$ the (first) *coupon collector epoch*. Clearly, any edge e has a time $L(e)$ for which it was visited *last* in the first coupon collector epoch. Let e^* be the edge with the smallest such time, i.e., the edge that was visited last before all other edges. The important point to realise is that in $\big(L(e^*)+1, T_1\big]$ each of the remaining edges in $E \setminus \{e^*\}$ is assured to be visited at least once, while the status e^* in the two coupled chains remains unchanged. Note that if e^* was removed in the top configuration at instance $L(e^*)$ (and hence also in the bottom configuration by monotonicity), all remaining edges will be pivotal in both coupled configurations upon their last visit in the first coupon collector epoch, whence coalescence happens at time T_1 and the modified algorithm terminates at time T_1.

If instead e^* was not removed from the top configuration at time $L(e^*)$, we extend the algorithm for another coupon collector epoch, that is from $(T_1, T_1 + T_2]$. Here we check again whether e^* was removed at instance $L(e^*)$, where both e^* and $L(e^*)$ both are now defined with respect to the second coupon collector epoch. We continue for additional coupon collector epochs until e^* is removed at the corresponding instance $L(e^*)$ in one

epoch. Denote the (random-) number of such epochs by N.
Now, note that we can couple the original coupling process with this modified process, such that with probability one we have $\tau \leq \tilde{\tau}$. This simply reflects the fact that the modified algorithm will never terminate before the coupling process. Moreover, it is not hard to see that N is dominated by a geometric random variable with mean $1/(1-p)$: The removal of e^* happens with probability at least $1-p$ (remember $\tilde{p}(p,q) \leq p$ for $q \geq 1$). Finally, define

$$\hat{\tau} \equiv \sum_{i=1}^{\hat{N}} T_i,$$

where T_i are independent coupon collector times and $\hat{N}$ is a geometric random-variable with mean $1/(1-p)$. The above arguments now allows us conclude

$$\mathbb{P}[\tau > t] \leq \mathbb{P}[\tilde{\tau} > t] \leq \mathbb{P}[\hat{\tau} > t] \tag{6.4.1}$$

and hence we can use Wald's equation as well as $\mathbb{E}[Y] = \sum_{t \geq 0} \mathbb{P}[Y > t]$, which is valid for any non-negative integer-valued random-variable Y, to derive the desired upper bound for $\mathbb{E}[\tau]$

$$\mathbb{E}[\tau] = \sum_{t \geq 0} \mathbb{P}[\tau > t] \leq \sum_{t \geq 0} \mathbb{P}[\hat{\tau} > t] = \mathbb{E}[\hat{\tau}] = \frac{\mathbb{E}[T]}{1-p} \sim \frac{m \log(m)}{1-p}.$$

This shows that $\mathbb{E}[\tau] \asymp \mathbb{E}[T]$, that is the expected coupling time is of the same order of magnitude as the coupon collector time.
In order to derive the Gumbel law for the coupling time distribution, the above construction is however too wasteful, in that another epoch is used in case e^* is not removed. In fact, as we will see below, we can indeed allow for e^* not to be removed at $L(e^*)$ as long as we assure that e^* is pivotal in the top configuration at $L(e^*)$. This together with a careful analysis of a short (as specified below) sequence of subsequent updates will be enough to show that coupling happens very likely (with high probability) at the end of the first coupon collector epoch. In order to precisely state the result and provide a proof, we firstly need to formalise the quantity $L(e)$ and introduce some further notation. Firstly, let us denote the edge selection process by $\mathbf{E} \equiv \{E_t : t \geq 1\}$, where the E_t's are identically and independently distributed with marginal distribution $\mathbb{P}[E_t = e] = 1/m$. Given $\mathbf{E}$ we define

$$R_t \equiv \{e : E_u = e \text{ for some } u \leq t\}, \tag{6.4.2}$$

that is R_t is the set of distinct edges visited by the process $\mathbf{E}$ up to time t. We clearly have $T = \inf\{t > 0 : |R_t| = m\}$. Further, the times L_e, from above, can be formally defined as

$$L(e) \equiv \sup\{t \leq T : E_t = e\}. \tag{6.4.3}$$

We order the $L(e)$ from the smallest to largest, and denote them by L_i for $1 \leq i \leq m$. The crucial observation we made is that by the time the first edge is visited last within the first

coupon collector epoch, with high probability, more than $\log(\log(m))$ *distinct* edges have already been visited:

Theorem 6.4.1.

$$\mathbb{P}\Big[\big|R_{L_1}\big| \leq \log\Big(\log(m)\Big)\Big] = o(1). \tag{6.4.4}$$

We postpone the proof of Theorem 6.4.1 to sub-section 6.4.1, and show first how it leads to the fact that with high probability τ does not exceed T, and hence with high probability $\tau = T$, more precisely

Theorem 6.4.2.

$$\lim_{m\to\infty} \mathbb{P}\Big[\tau = T\Big] = 1.$$

Proof. We first recall a standard concentration result for a binomial random variable X, commonly known as Chernoff-Hoeffding bound, c.f. e.g. [138], that states:

$$\mathbb{P}\Big[X \leq (1-\epsilon)\mathbb{E}[X]\Big] \leq e^{-\frac{\epsilon^2}{2}\mathbb{E}[X]} \quad \epsilon \in (0,1).$$

Now, for a binomial random-variable X with parameters $(1-p)$ and m the above implies that we can choose any $\beta \in (0,1)$ such that we can find a $\gamma \in [0,1)$ for which the following holds:

$$\mathbb{P}\Big[X \leq \beta(1-p)m\Big] \leq \mathrm{e}^{-\gamma m}.$$

Simply observe that $\mathbb{E}[X] = (1-p)m$ and set $\epsilon = 1-\beta$ together with $\gamma = (1-\beta)^2(1-p)/2$.
Let e^* be the edge selected at time L_1, i.e. $e^* \equiv E_{L_1}$.
Define $a_m = \lfloor \log(\log(m)) \rfloor$ and denote by $P_1^{(1)} < P_2^{(1)} < \ldots < P_{a_m}^{(1)}$ the *last* visit of a_m of the edges belonging to the set $R_{L_1} \setminus \{e^*\}$ *before* time L_1. Note that $|R_{L_1}| > a_m$ implies $|R_{L_1}| \geq a_m + 1$, because both $|R_{L_1}|$ and a_m are integers. Hence there are at least a_m distinct edges in $R_{L_1} \setminus \{e^*\}$. At each of these times we flip a coin and say that the edge is removed if the coin shows head. The probability that the coin shows head is at least $(1-p)$. Note that if the coin is head, the removal happens in both chains (again by monotonicity).
Define the event A_1 that at least $\beta(1-p)a_m$ edges are removed before time L_1. The event A_1 holds with probability at least $1 - \mathrm{e}^{-\gamma a_m}$ conditioned on the event that $|R_{L_1}| > a_m$. This follows from the Chernoff-Hoeffding bound above. Moreover the edge visited by the process **E** at time $P_i^{(1)}$ is **not** visited by **E** at times contained in the interval $(P_i^{(1)}, L_1]$. Notice that if at least $\beta(1-p)a_m \geq 2$ edges are removed before time L_1, then all of edges are pivotal for the top configuration at time L_1. Hence in particular by the time L_1, the edge $e^* \equiv E_{L_1}$ is pivotal to both configurations, w.h.p.. If at time L_1 the edge e^* is removed, then the coupling will happen at T. In fact, if e^* is removed, as it is pivotal, it is removed for both chains. Moreover, e^* is not visited again in the interval $(L_1, T]$, while all the other edges are visited during this interval at least once. Hence, all of the edges visited in $(L_1, T]$ are pivotal two both configurations upon each visit.

If L_1 is not removed we focus on L_2. Set $A_0 = \{|R_{L_1}| > a_m\}$. Condition on the event $A_1 \cap A_0$, that is the existence of at least $\beta(1-p)a_m$ edges that are removed by time L_1. In order for

the edge visited at time L_2 to be not pivotal, the process must visit at least $\lfloor \beta(1-p)a_m \rfloor$ of the edges which were removed by time L_1 and must add back each of them. We denote by A_2^c a larger event, defined on $A_1 \cap A_0$, as follows. On A_1, fix a set of $a_m^{(1)} \equiv \lfloor \beta(1-p)a_m \rfloor$ edges that are deleted by time L_1 and denote it by $\mathcal{B}_2$. As before, we define $P_1^{(2)} < P_2^{(2)} \ldots < P_{|\mathcal{B}_2|}^{(2)}$ as the last visits of edges in $\mathcal{B}_2$ before time L_2. For each element in $\mathcal{B}_2$, if it is not visited in (L_1, L_2), we flip a biased auxiliary coin, that shows head with probability $1-p$, (which is independent of anything else, i.e. is an extra coin) for this edge. These auxiliary coins do not affect the behaviour of the two coupled chains, i.e., are not used to upgrade the status of these processes. We apply the following colouring.

- If an edge in $\mathcal{B}_2$ is visited in the interval (L_1, L_2) and is removed at its last visit in this interval, then colour it red;
- If an edge in $\mathcal{B}_2$ is not visited in the interval (L_1, L_2) and the auxiliary coin shows head, then colour it red;
- Do not colour it otherwise.

All the red edges are removed edges by time L_2 (flipping another coin for edges that are not visited again in (L_1, L_2) does not change the fact that they had been removed by the time L_1 and hence remain removed up to time L_2). We define A_2^c as the event that there are at most $\beta(1-p)a_m^{(1)}$ red edges. Notice that

$$\begin{aligned}\mathbb{P}\Big[A_2 \cap A_1 \mid A_0\Big] &= \mathbb{P}\Big[A_2 \mid A_1 \cap A_0\Big]\mathbb{P}\Big[A_1 \mid A_0\Big] \\ &\geq \Big(1 - \mathrm{e}^{-\gamma \lfloor \beta(1-p)a_m \rfloor}\Big)\Big(1 - \mathrm{e}^{-\gamma a_m}\Big) \\ &\geq \Big(1 - \mathrm{e}^{-\gamma[\beta(1-p)\log(\log(m)) - \beta(1-p) - 1]}\Big)\Big(1 - \mathrm{e}^{-\gamma[\log(\log(m)) - 1]}\Big)\end{aligned}$$

We emphasise that on $A_2 \cap A_1 \cap A_0$ there are at least $\beta(1-p)\lfloor \beta(1-p)a_m \rfloor$ removed edges at time L_2. On $A_2 \cap A_1 \cap A_0$, if the edge E_{L_2} is removed, as it is not visited again in the interval $(L_2, T]$, while all the other edges (with the exception of e^*) are visited during this interval. The two coupled chains already agree on e^* as at time L_1 this edge was pivotal. Moreover all other edges are pivotal edges when visited in $(L_2, T]$, and hence a coupling happens at time T.

Next A_3 is defined on the event $A_2 \cap A_1 \cap A_0$ as follows. We reset the colour of the edges, so that are all uncoloured. On $A_2 \cap A_1 \cap A_0$ fix a set $\mathcal{B}_3$ of $a_m^{(2)} \equiv \lfloor \beta(1-p)a_m^{(1)} \rfloor$ of removed edges. Colour the vertices using the method described above (the auxiliary trials are assumed to be independent of anything else). Define A_3 the event that there are at least $\beta(1-p)a_m^{(2)}$ red edges. Notice that

$$\mathbb{P}[A_3 \cap A_2 \cap A_1 \mid A_0] \geq \prod_{j=0}^{2}\Big(1 - \mathrm{e}^{-\gamma a_m^{(j)}}\Big) \geq \prod_{j=0}^{2}\Big(1 - \mathrm{e}^{-\gamma \beta^j (1-p)^j \log(\log(m)) + \gamma \phi(\infty)}\Big)$$

Note that it is easy to verify that $a_m^{(j)} \geq \beta^j (1-p)^j \log(\log(m)) - \phi(j)$ for $j \in \mathbb{N}$ with the convention that $a_m^{(0)} \equiv a_m$, where we set $\phi(j) \equiv \sum_{k=0}^{j} \beta^k (1-p)^k$, with $\phi(j) \leq \phi(\infty) = 1/(1-(1-p)\beta)$. If $A_3 \cap A_2 \cap A_1$ holds and the edge $E(L_3)$ is removed, then a coupling time happened before or at time T, for the same reason explained above. We proceed in the same fashion for $b_m \equiv \lfloor \log(\log(a_m)) \rfloor$ steps. We say that the algorithm works if $\bigcap_{j=1}^{b_m} A_j$ holds and at least one of the L_j results in a removal of the relative edge. If the algorithm works then the coupling happens at T. The probability that this algorithm works, given A_0, is at least

$$(1-p^{b_m}) \prod_{j=0}^{b_m-1} \left(1 - K \mathrm{e}^{-\gamma(1-p)^j \beta^j \log(\log(m))}\right), \tag{6.4.5}$$

where $K = e^{\gamma\phi(\infty)} = e^{\gamma/(1-(1-p)\beta)}$. Next we prove that the quantity in (6.4.5) converges to 1 as $m \to \infty$. In fact

$$\prod_{j=1}^{b_m-1} \left(1 - K \mathrm{e}^{-\gamma(1-p)^j \beta^j \log(\log(m))}\right) \geq \left(1 - K \mathrm{e}^{-\gamma(1-p)^{b_m-1} \beta^{b_m-1} \log(\log(m))}\right)^{b_m-1} = 1 - o(1),$$

where we used the fact that $b_m = \lfloor \log(\log(a_m)) \rfloor$.

This fact, together with the fact that A_0 holds w.h.p. proves that the coupling time will happen, w.h.p., at T.

■

Corollary 6.4.3.

$$\lim_{m\to\infty} \mathbb{P}\left[\frac{\tau - m\log(m)}{m} \leq x\right] = \mathrm{e}^{-\mathrm{e}^{-x}}.$$

Proof.

$$\lim_{m\to\infty} \mathbb{P}\left[\frac{\tau - m\log(m)}{m} \leq x\right] = \lim_{m\to\infty} \mathbb{P}\left[\frac{T - m\log(m)}{m} \leq x, T = \tau\right] = \lim_{m\to\infty} \mathbb{P}\left[\frac{T - m\log(m)}{m} \leq x\right] = \mathrm{e}^{-\mathrm{e}^{-x}},$$

where in the last step we used the classical result that the coupon collector time, properly rescaled, converges in distribution to a Gumbel distribution [133]. ■

In fact Theorem 6.4.2 allows us to invoke standard concentration results for the coupon-collector time T, due to the simple observation

$$\mathbb{P}[\tau > t] = \mathbb{P}[\tau > t, \tau > T] + \mathbb{P}[\tau > t, \tau = T] \leq o(1) + \mathbb{P}[T > t]. \tag{6.4.6}$$

Furthermore, recall the union-bound argument that led to (6.3.4), which states that for any $\epsilon > 0$ one has

$$\mathbb{P}[T > (1+\epsilon) m \log(m)] \leq m^{-\epsilon}. \tag{6.4.7}$$

Combining this with (6.4.6) we obtain

$$\mathbb{P}[\tau > (1+\epsilon)m\log(m)] = o(1). \tag{6.4.8}$$

Note that because $m\log(m) \leq mH_m = \mathbb{E}[T] \leq \mathbb{E}[\tau]$ for any positive integer m, we have also

$$\mathbb{P}[\tau > (1+\epsilon)\mathbb{E}[\tau]] = o(1).$$

This concentration result stating that deviations of the order of magnitude of $\mathbb{E}[\tau]$ from $\mathbb{E}[\tau]$ itself become suppressed for increasing system size, together with the generic fact that $\mathbb{E}[\tau] \geq \mathbb{E}[T]$ suggest that an asymptotic equivalence between $\mathbb{E}[T]$ and $\mathbb{E}[\tau]$ might hold. And indeed, we can use the concentration result (6.4.8) to infer the asymptotics of $\mathbb{E}[\tau]$. The argument crucially relies on the following Lemma:

Lemma 6.4.4. *(Propp & Wilson [33]) For the coupling time τ and a positive integer t we have*

$$\mathbb{P}[\tau \leq t] \leq \frac{t}{\mathbb{E}[\tau]}. \tag{6.4.9}$$

In some sense this is the reverse of Markov's inequality. However it relies crucially on the sub-multiplicativity property of τ, that is we have for two non-negative integers t_1, t_2 $\mathbb{P}[\tau > t_1 + t_2] \leq \mathbb{P}[\tau > t_1]\mathbb{P}[\tau > t_2]$, which follows from the same arguments that led to (6.1.15).

Proof. For a fixed integer $t > 0$ define $\epsilon = \mathbb{P}[\tau > t]$, assume $\epsilon < 1$ and observe

$$\mathbb{E}[\tau] = \sum_{k>0}\mathbb{P}[\tau > k] = \sum_{\ell=0}^{\infty}\sum_{k=0}^{t-1}\Big(\mathbb{P}[\tau > \ell t + k] \leq \sum_{\ell=0}^{\infty}\sum_{k=0}^{t-1}\mathbb{P}[\tau > \ell t] \leq t\sum_{\ell=0}^{\infty}\Big(\epsilon^{\ell} = \frac{t}{1-\epsilon}.$$

By definition we have $1-\epsilon = \mathbb{P}[\tau \leq t]$, which proves the claim. Note that in case $\epsilon = 1$, we do not have that the sum above converges. However the bound (6.4.9) is trivially valid as we have in such cases $\mathbb{P}[\tau \leq t] = 0$. ■

We emphasise, Equipped with this information we are now ready to establish the following result:

Lemma 6.4.5. *The expected coupling time $\mathbb{E}[\tau]$ on the m-cycle is asymptotically equivalent to the expected coupon collector time $\mathbb{E}[T]$, i.e.,*

$$\lim_{m\to\infty}\frac{\mathbb{E}[\tau]}{\mathbb{E}[T]} = 1, \tag{6.4.10}$$

or simply $\mathbb{E}[\tau] \sim \mathbb{E}[T]$.

Proof. For any $1 > \epsilon > 0$ let $t = (1+\epsilon/3)m\log(m)$. Due to (6.4.7) we have that $\mathbb{P}[\tau \leq t] \to 1$ for $m \to \infty$. Formally, this means that we can find an integer M_ϵ such that for any $m \geq M_\epsilon$

we have $1/\mathbb{P}[\tau \le t] \le 1 + \epsilon/3$. We therefore have for any $m \ge M_\epsilon$:

$$\frac{\mathbb{E}[\tau]}{\mathbb{E}[T]} - 1 \le \frac{(1+\frac{\epsilon}{3})m\log(m)}{\mathbb{E}[T]\mathbb{P}[\tau \le t]} - 1 \le \frac{1+\frac{\epsilon}{3}}{\mathbb{P}[\tau \le t]} - 1 \le (1+\frac{\epsilon}{3})(1+\frac{\epsilon}{3}) - 1 \le 1 + 2\frac{\epsilon}{3} + \frac{\epsilon^2}{9} - 1 \le 2\frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

where we used $\mathbb{E}[T] \ge m\log(m)$, see for instance Exercise 2.4 in [2]. This proves the asymptotic equivalence of $\mathbb{E}[\tau]$ and $\mathbb{E}[T]$. ■

Further, we remark that we have not been able to establish (6.3.9) for the m-cycle. However, our numerical investigation supports the claim that (6.3.9) is also valid for the m-cycle.

6.4.1 Proof of Theorem (6.4.1)

We now provide the remaining proof of Theorem 6.4.1. Before going into the details let us describe the intuition of the proof. The first step is to assure that w.h.p. all of the first $\lfloor \log(\log(m)) \rfloor$ (distinct) edges encountered by the edge selection process are sufficiently often revisited, more precisely are visited more than $c\log(m)$ times (for small enough c), by the coupon collector time T. We then can in particular assume that the edge E_{L_1} (the edge first visited last in the first coupon collector epoch) is among the first $\lfloor \log(\log(m)) \rfloor$ distinct edges encountered by the process, because it is easy to see that in the alternative case the claim holds trivially. Consequently, by the time L_1 the edge E_{L_1} is visited more than $c\log(m)$ times, w.h.p.. However, on the other hand it is intuitive that in order for the process to achieve such a large number of hits to edge E_{L_1}, a sufficiently large amount of time (visits to other edges) must have elapsed. Indeed, we prove that in such cases at time L_1 a number of distinct edges is visited by the process that in particular exceeds the considered threshold of a_m with high probability.

Proof. Recall, we need to prove that by time L_1, with high probability, more than $\log(\log(m))$ different edges have been visited by the process $\mathbf{E} = \{E_t, t \ge 1\}$, equivalently

$$\mathbb{P}\Big[\big|R_{L_1}\big| \le \log(\log(m))\Big] = o(1). \tag{6.4.11}$$

Now, recall that $a_m = \lfloor \log(\log(m)) \rfloor$ and denote by e_1 the first edge to be visited by $\mathbf{E}$ and e_i the i-th distinct edge visited by $\mathbf{E}$. Set $S_i(t) = |\{k\colon k \le t, E_k = e_i\}|$. To prove (6.4.11), we first prove that for all small enough constants c

$$\lim_{m\to\infty} \mathbb{P}\left[\bigcup_{i=1}^{a_m} \Big\{S_i(T) \le c\log(m)\Big\}\right] = 0. \tag{6.4.12}$$

We use a union-bound to prove the above, and hence first prove that for fixed integer $i \le a_m \equiv \lfloor \log(\log((m))) \rfloor$, for all small enough c

$$\mathbb{P}[S_i(T) \le c\log(m)] \le K_1 \exp\left\{-K_2\sqrt{\log(m)}\right\} \tag{6.4.13}$$

Let T_k the first time the process $\mathbf{R} = \{R_t, t \ge 1\}$ has cardinality k, that is the hitting time of e_k. Define, for $k \ge 1 + a_m$, the random variable

$$Y^{(i)}(k) \equiv \sum_{j=T_{k-1}+1}^{T_k - 1} \mathbb{1}_{\{E_j = e_i\}}.$$

Note that because $k \ge 1 + a_m$ we have that $k - 1 \ge a_m$ and hence any edge e_i with $i \le a_m$ has been visited at least once by the time $T_{k-1} + 1$. Therefore

$$\sum_{k=a_m+1}^{m} Y^{(i)}(k)$$

counts visits to edge e_i starting after the first a_m edges have been explored by E_t, and hence is a lower bound for the overall number of visits to edge e_i. Further, note that the random variables $Y^{(i)}(k)$ have different distributions for different values of k, but they are independent. Next, we use the usual Chernoff bound (which we derive, for completeness) to prove that, for all $\theta < 0$, we have

$$\begin{aligned}
\mathbb{P}[S_i(T) \le c\log(m)] &\le \mathbb{P}\left[\sum_{k=a_m+1}^{m} Y^{(i)}(k) \le c\log(m)\right] \\
&\le \mathbb{P}\left[\exp\left\{\theta \sum_{k=a_m+1}^{m} Y^{(i)}(k)\right\} \ge \exp\{\theta c\log(m)\}\right] \\
&\le e^{-\theta c\log(m)}\mathbb{E}\left[\exp\left\{\theta \sum_{k=a_m+1}^{m} Y^{(i)}(k)\right\}\right] \\
&= \exp\left\{\left(\theta c\log(m) - \sum_{k=a_m+1}^{m} \log\left(\mathbb{E}[e^{\theta Y^{(i)}(k)}]\right)\right)\right\}
\end{aligned}$$

In the second step it was crucial that $\theta < 0$. The third step follows from Markov's inequality and the last from the independence of the $Y^{(i)}(k)$'s.

Observe that conditioned on $T_k - T_{k-1} = t$ the distribution of $Y^{(i)}(k)$ for $i \le k-1$ is Binomial with parameters $1/(k-1)$ and $t-1$. We therefore obtain

$$\mathbb{E}\left[e^{\theta Y^{(i)}(k)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\theta Y^{(i)}(k)} \middle| T_k - T_{k-1} = t\right]\right]$$

$$= \mathbb{E}\left[\left(\left(e^{\theta}\frac{1}{k-1}+1-\frac{1}{k-1}\right)^{t-1}\right]. \tag{6.4.14}$$

The inner conditional expectation is in fact the moment generating function of a binomial with parameters $1/(k-1)$ and $t-1$. For the outer expectation recall that $t = T_k - T_{k-1}$ is distributed as a geometric random variable with average $m/(m-k+1)$. We therefore obtain

$$\begin{aligned}\mathbb{E}\left[e^{\theta Y^{(i)}(k)}\right] &= \frac{1}{1-\left(e^{\theta}\frac{1}{k-1}+1-\frac{1}{k-1}\right)\frac{k-1}{m}}\ \frac{m-k+1}{m}\\ &= \frac{1}{1+\frac{1-e^{\theta}}{m-k+1}}\end{aligned}$$

Hence

$$\sum_{k=a_m+1}^{m}\log\left(\mathbb{E}\left[e^{\theta Y^{(i)}(k)}\right]\right) = -\sum_{k=a_m+1}^{m}\log\ 1+\frac{1-e^{\theta}}{m-k+1}$$

We choose $\theta = -1/\sqrt{\log(m)}$. Recall that for any pair of sequences b_m and c_m, we denote by $b_m \sim c_m$ the relation $b_m/c_m \to 1$ as $m \to \infty$. Next, we show that for large m

$$\sum_{k=a_m+1}^{m}\log\ 1+\frac{1-e^{\theta}}{m-k+1} \sim \sqrt{\log(m)}. \tag{6.4.15}$$

In fact, as $\left(e^{\theta}-1\right) \to 0$ as $m \to \infty$ for our choice of θ, we have

$$\begin{aligned}\log\ 1+\frac{1-e^{\theta}}{m-k+1} &\sim \frac{1-e^{\theta}}{m-k+1},\\ &\sim \frac{1}{\sqrt{\log(m)}}\frac{1}{m-k+1}.\end{aligned}$$

In the first step, we used $\log(1+x_m) \sim x_m$, whenever x_m is a sequence satisfying $x_m \to 0$ as $m \to \infty$. The second step follows from, $1\text{-}e^{\theta} \sim \theta = 1/\sqrt{\log(m)}$. By taking the sum over k and using the fact that

$$\sum_{k=a_m+1}^{m}\frac{1}{m-k+1} \sim \log(m-\log(\log(m))) \sim \log(m),$$

we have (6.4.15). Notice that $c\theta\log(m)$ becomes $-c\sqrt{\log(m)}$, yielding (6.4.13) for all c small enough. Hence

$$\mathbb{P}\left[\bigcup_{i=1}^{a_m}\left\{S_i(T) \le c\log(m)\right\}\right] \le \log(\log(m))K_1\exp\left\{-K_2\sqrt{\log(m)}\right\} = o(1) \tag{6.4.16}$$

yielding (6.4.12). Let us briefly explain the role of c and the appearance of the constants K_1 and K_2. Note that strictly speaking, we have only shown that

$$\mathbb{P}\left[\bigcup_{i=1}^{a_m}\left\{S_i(T)\le c\log(m)\right\}\right]\le \log(\log(m))e^{c\sqrt{\log(m)}}\exp\left\{-\sum_{k=a_m+1}^{m}\log\ 1+\frac{1-\mathrm{e}^{\theta}}{m-k+1}\right)\right\}.$$

However, because we have (6.4.15), we can invoke Lemma C.0.1 proven in Appendix C, and obtain for some $K_1\ge 1$ and $K_2\in(0,1)$ that the above is bounded by:

$$K_1\log(\log(m))e^{(c-K_2)\sqrt{\log(m)}},$$

which for $c<K_2$ yields the result.

Let us now continue and show how the fact that the first a_m distinct edges encountered by the process **E**, will, with high probability, each be visited at least $\lfloor c\log(m)\rfloor+1$ times by the time T, can be used to prove that (6.4.11) holds.
Firstly, when the edge E_{L_1} does not belong to the first a_m different edges explored by the process **E**, or in other words in case of the event $E_{L_1}\notin R_{T_{a_m}}$, we have $|R_{L_1}|>\lfloor\log(\log(m))\rfloor$ and hence $|R_{L_1}|>\log(\log(m))$. Therefore (6.4.11) would follow in this case directly. Hence, in order to prove (6.4.11), it suffices to prove that

$$\mathbb{P}\left[|R_{L_1}|\le\log(\log(m)),E_{L_1}\in R_{T_{a_m}}\right]=o(1).$$

Furthermore, because of (6.4.16), it is actually enough to prove that

$$\mathbb{P}\left[|R_{L_1}|\le\log(\log(m)),\,E_{L_1}\in R_{T_{a_m}},\bigcap_{i=1}^{a_m}\left\{S_i(T)>c\log(m)\right\}\right]=o(1). \tag{6.4.17}$$

Let I be the **first** time an edge is visited by the process **E** exactly $\lfloor c\log(m)\rfloor+1$ times. We prove that

$$\mathbb{P}\left[|R_I|\le a_m\right]=o(1) \tag{6.4.18}$$

In fact, on the event $\{|R_I|\le a_m\}$ at most a_m edges are visited by the process **E** by time I. Hence, if we denote by I_k, the first time e_k is visited exactly $\lfloor c\log(m)\rfloor+1$ times, we have that $I\le I_k$, for all k. Hence

$$\mathbb{P}\left[|R_I|\le a_m\right]\le\mathbb{P}\left[\bigcup_{k=1}^{a_m}|R_{I_k}|\le a_m\right]\le\sum_{k=1}^{a_m}\mathbb{P}\left[|R_{I_k}|\le a_m\right]. \tag{6.4.19}$$

Next, we provide an upper bound for each of the following probabilities

$$\mathbb{P}\left[|R_{I_k}|\le a_m\right]\qquad k\le a_m.$$

Notice that is suffices to bound the lower tail of $\Delta_k=I_k-T_k$ as we have $\mathbb{P}\left[I_k-\lfloor c\log(m)\rfloor-1\le\gamma_1 m\right]\le\mathbb{P}\left[\Delta_k-\lfloor c\log(m)\rfloor\le\gamma_1 m\right]$. Furthermore as Δ_k is a negative

binomial with parameters $(\lfloor c\log(m)\rfloor, 1/m)$, it follows from standard (lower) tail bounds for sums of geometric random variables, see e.g. [146], that there exist $\gamma_0, \gamma_1, \gamma_2 > 0$, with $\gamma_1 < c$,such that

$$\mathbb{P}\left[\Delta_k - \lfloor c\log(m)\rfloor \leq \gamma_1 m\right] \leq \gamma_0 \mathrm{e}^{-\gamma_2 \log(m)}. \tag{6.4.20}$$

We remark that this bound is far from being sharp, however completely sufficient for our purposes. Notice that $I_k - \lfloor c\log(m)\rfloor - 1$ counts the number of times the process, during the time interval $\mathbb{N}\cap[1, I_k]$ visits an edge different from e_k. Given that $\{I_k - \lfloor c\log(m)\rfloor - 1 > \gamma_1 m\}$, the probability that process **E** visits less than a_m different edges is less than

$$\mathbb{P}\left[T_{a_m+1} > \gamma_1 m\right] = \mathbb{P}\left[\sum_{i=1}^{a_m} (T_{i+1} - T_i) > \gamma_1 m\right] \leq \gamma_3 \mathrm{e}^{-\gamma_4 m}, \tag{6.4.21}$$

for some $\gamma_3, \gamma_4 > 0$. This follows again from the fact that we consider a sum of (independent) geometric random variables with different parameters $(1 - i/m)$, and we utilize standard bounds for the upper tail of such a sum, see e.g. [146]. Combining this fact with (6.4.20), we have by the law of total probability

$$\mathbb{P}\left[|R_{I_k}| \leq a_m\right] \leq \gamma_0 \mathrm{e}^{-\gamma_2 \log(m)} + \gamma_3 \mathrm{e}^{-\gamma_4 m}.$$

This, together with (6.4.19), yields (6.4.18).

It is clear that $S_{e^*}(T) = S_{e^*}(L_1)$, as by definition L_1 is the time of the last visit to edge e^* up to time T. Notice that R_k is a non-decreasing process in k, and on the event $\{S_{e^*}(L_1) > c\log(m)\}$ we have $I \leq L_1$, because in this case edge e^* is one candidate of an edge that is visited at least $\lfloor c\log(m)\rfloor + 1$ times, so I cannot exceed L_1. Hence the event which appears in (6.4.17) is a subset of the event

$$\mathbb{P}\left[|R_{L_1}| \leq a_m, E_{L_1} \in R_{T_{a_m}}, S_{e^*}(L_1) > c\log(m)\right] \leq \mathbb{P}\left[|R_I| \leq a_m, S_{e^*}(L_1) > c\log(m)\right] \leq \mathbb{P}\left[|R_I| \leq a_m\right] = o(1)$$

proving (6.4.11).

■

In this chapter we have provided numerical evidence, heuristic arguments and rigorous results that strongly suggest that the coupling process is very efficient in the sense that $\mathbb{E}[\tau] \approx m\log(m)$ (or possibly a poly-logarithmic factor), whenever the random-cluster model is not at a location of a phase-transition (in the thermodynamic limit).
Furthermore, the closeness to the uncorrelated case $q = 1$ holds not only on the level of the first moment $\mathbb{E}[\tau]$, but also for the nature of the asymptotic fluctuations, as probed by the distribution of η.
Our surprising observation is that the limiting distribution of $\eta = (\tau - \mathbb{E}[\tau])/\sqrt{\mathrm{Var}[\tau]}$ is the Gumbel distribution (6.3.9), not only off criticality, but also at the point where the model undergoes a continuous phase transition in the thermodynamic limit. In contrast, the expected coupling time for continuous phase transitions is expected to possess a

power law system size scaling, which is clearly far off the optimal poly-logarithmic scaling observed off criticality (when measured in sweeps).
Lastly, we have seen that the nature of the phase transition reflects itself in the type of limiting distribution, as a discontinuous phase transition showed a clear deviation from the Gumbel class towards to the Frèchet family. It is an interesting question whether this sensitivity of the fluctuations of τ can be used to detect the nature of the phase-transition in situations where the precise location of a crossover from a continuous do discontinuous phase transition is not known.

Chapter 7

Iterated fragmentation with cut-off

In this section we present a, mostly phenomenological, study of *iterated*, *repeated* or *dynamic* fragmentation of fractal random structures, in form of critical Fortuin-Kasteleyn clusters. The aim of this chapter is mainly to present a novel computational fragmentation study and stimulate future research in this field, as many aspects of associated non-equilibrium phenomena remain unexplained or poorly understood. Specifically, we consider the continued (random) removal of edges or bonds from fractal random structures. In general, dynamical (stochastic) fragmentation processes play a fundamental role in science and engineering and have a variety of applications on different length and time scales, ranging from geology [147] over fracture of (brittle) solids [148] to the break-up of nuclei [149]. Part of the results presented in this chapter have recently appeared in letter form in [120]. We discuss various numerical methods to infer information about the fragmented object from a study of certain characteristics of the fragments produced in the dynamical process.

7.1 Stable fragment size distribution

The break-up kernel $b_{s'|s}$ analysed in chapter 5 is a key quantity for the equilibrium fragmentation, that is it characterises single break-up events in equilibrium. The kernel $b_{s'|s}$ encodes aspects of the equilibrium cluster structure and is therefore a priori, without further justification, only relevant for "one-step" fragmentation processes that stay close to equilibrium. However, the self-similar structure of critical Fortuin-Kasteleyn clusters suggests a "recursive" applicability of $b_{s'|s}$. In fact the main motivation of our study is to understand to which extent the break-up kernel $b_{s'|s}$, or some of its characteristics, in particular the break-up or fragmentation exponent ϕ, are relevant for the process of iterated fragmentation of fractal (self-similar) random structures. The study of fragment size distributions is a ubiquitous method to probe the structure of materials or the nature of the fragmentation process [150] and has apparent practical applications in engineering such as in milling [120]. A connected, and from a physics-point-of-view probably the most interesting, question is whether a power law size dependence can be observed in

the final distribution of fragments. The appearance of power laws in the size distribution of fragments and related scale-independent properties of fragmentation is a universal phenomenon in nature, see for instance [151, 152, 153, 154].

The main scenario we consider is the following setup. Starting with an initial bond configuration, chosen to be a critical equilibrium Fortuin-Kasteleyn configuration, edges are removed uniformly at random, or associated with an intrinsic removal rate in a continuous time setting. Firstly, notice that this process has a unique absorbing state in the sense that at some (random[1]) time, when all bonds are removed, all clusters have unit size, corresponding to isolated vertices. Here and in what follows we refer to the number of vertices in a fragment or cluster as its *mass* or *size*. This, in contrast to other definitions, such as in [29], ensures mass conservation in a bond-fragmentation process. In order to obtain a non-degenerate final distribution, we will introduce a cut-off mechanism, in the spirit of the "one-dimensional" studies in [155] and [156]. The cut-off influences the fragmentation process in the following way: Clusters are distinguished into *stable* and *unstable* fragments, where initially all fragments are unstable: Unstable fragments continue to fragment by random-edge removals, as before. More precisely, the next bond to be removed is chosen uniformly among all bonds that are part of unstable fragments. If the removal of a bond induces a fragmentation event, then any of the two "daughter" fragments with mass $s \leq s_c$ becomes stable, and will not undergo further fragmentation. Clearly we recover the unrestricted fragmentation process for $s_c = 1$, which ultimately leads to a degenerate fragment size distribution concentrated on mass 1.

Here we consider the square lattice as the base graph on which the Fortuin-Kasteleyn configurations "live". We emphasise that this choice introduces the possibility of cycles or multiple connections holding together clusters, a completely new aspect in the study of dynamic fragmentation, absent in previous one-dimensional studies such as for instance [155, 157] as well as in cycle-free bond-structures (trees or forests) such as recently studied in [158]. Cycles allow us to account for a possible material resistance against failures of bonds, which in the random setting, that is starting from a Fortuin-Kasteleyn configuration, can even be spatially varying: Recall, that any non-bridge resides in at least one cycle and hence the corresponding cluster is resistant against break-ups upon the removal of a non-bridge. Furthermore, we have seen in section 5.2.2 that the variation of the cluster weight parameter q together with $p_{\mathrm{sd}}(q)$ has a direct influence on the overall cohesion (bridge-density), while at the same time preserving self-similarity (with varying exponents). Contrast this to a variation of the bond-density in the Bernoulli bond percolation model, which does not allow for a continuous family of self-similar models for fixed graph G as p is varied.

In the following we consider the limiting distribution of stable fragments, defined by:

$$n_{s_c}(s,\infty) \equiv \frac{\langle N_{s_c}(s,\infty)\rangle}{\sum_{s=1}^{s_c}\langle N_{s_c}(s,\infty)\rangle}, \tag{7.1.1}$$

[1] The randomness stems from the randomness in the initial configuration. For fixed initial configuration it is easy to verify that this time is deterministic.

where $\langle N_{s_c}(s,\infty)\rangle$ is the expected number of stable fragments after the process terminated, and the expectation exhausts both all possible and correctly weighted initial Fortuin-Kasteleyn configurations, and any evolution of the fragmentation process. As already mentioned, here we restrict the analysis to the final ($t \to \infty$) cluster size distribution. In general, one can also consider the finite time analogue $n_{s_c}(s,t)$, which allows to study the full time dependence of the fragmentation process with cut-off. We will pursue a time dependent analysis, both for stable and unstable fragment distributions, in a future study. Yet, as it turns out, the final distribution $n_{s_c}(s,\infty)$ shows very interesting features, and is interesting in its own right.

The probably most striking feature is the existence of an approximate power law scaling in s. In Figure 7.1 we show the numerically estimated cluster size distribution $n_{s_c}(s,\infty)$, obtained from an iterated fragmentation starting from critical Fortuin-Kasteleyn configurations. Here we consider cluster weights $q = 0$, that is uniform spanning trees, bond percolation $q = 1$, and $q = 2, 3$, corresponding to the Fortuin-Kasteleyn cluster for the Ising and 3-state Potts model, respectively.

More precisely, the initial configuration was restricted to the largest component in the respective critical configuration on the square lattice. This was simply done to avoid a mixing of the stable fragment size distribution with the equilibrium cluster size distribution (with the Fisher exponent τ, see section 5.6.1 and [46]). For each value of q we used 40000 different starting configurations, and the lattice size was set to 256×256. Before we

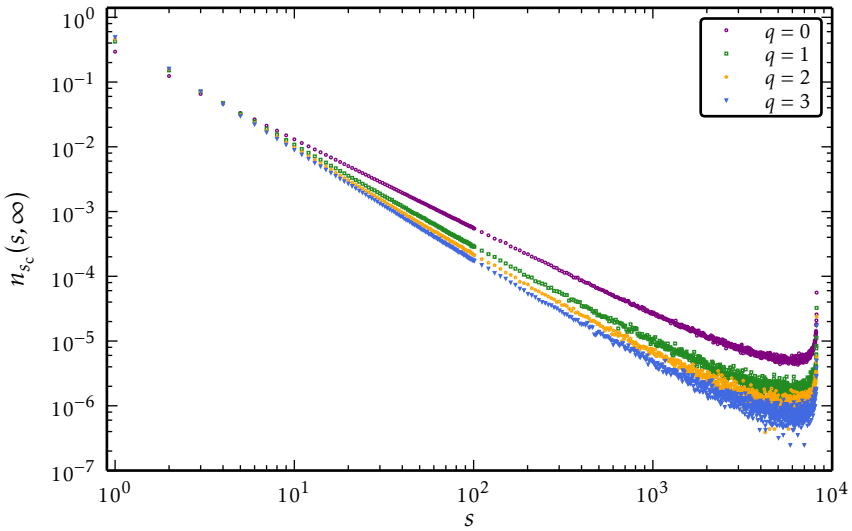


Figure 7.1: *Final fragment size distribution $n_{s_c}(\cdot,\infty)$ of stable fragments for $q = 0$ (uniform spanning trees) and $q = 1, 2, 3$, with a fixed cut-off equal to 8192. The linear dimension of the underlying square lattice was set to $L = 256$.*

.

extract the corresponding exponent, let us consider the cut-off dependence of $n_{s_c}(s,\infty)$. For instance, it is clear that in the limit of large s_c we ultimately recover the fragment size distribution kernel $b_{s'|s}$ for the equilibrium model, as only a few fragmentation events can happen before all fragments become stable. Now, consider Figure 7.2, which shows the

final stable fragment distribution for the critical percolation case for various cut-offs. It is apparent that a power law scaling persists upon variation of s_c, however with a possible change in the relevant exponent. Indeed, in Figure 7.4 we show the cut-off dependence of

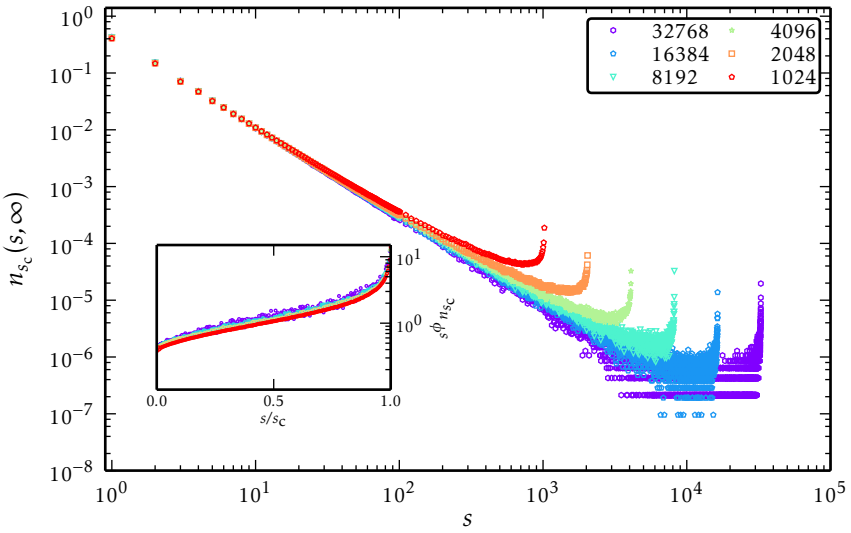


Figure 7.2: *Fragment size distribution $n_{s_c}(s,\infty)$ for critical percolation on the 256×256 square lattice for various cut-offs. The inset shows a scaling "collapse" confirming the scaling ansatz* (7.1.2).

.

a numerically extracted effective exponent χ assuming asymptotic scaling of the form

$$n_{s_c}(s,\infty) \sim s^{-\chi} \quad \text{for } s_c, s \to \infty.$$

We analysed the cluster weights already shown in Figure 7.1. More precisely we initially used a rather simplistic method to extract χ by fitting a simple power law to the empirical distribution. As expected, one observes a variation of χ with s_c, however consistent with $\chi \to \phi$ when $s_c/L^2 \to 1$, as indicated by the solid horizontal lines in Figure 7.4. The variation of χ with s_c is not surprising, when taking into account the simplicity of the numerical method used to determine the exponent. In fact, a closer analysis suggests that the extracted exponent χ is indeed only effective, in the sense that it appears to be distorted by an unconsidered cut-off dependence. Rather, we find that our data for the stable fragment size distribution $n_{s_c}(s;\infty)$ is very well described by the following scaling form

$$n_{s_c}(s,\infty) \sim s^{-\phi} \mathcal{F}\left(\frac{s}{s_c}, \frac{s_c}{L^{d_F}}\right) \tag{7.1.2}$$

valid for a large but finite system with linear dimension L, at criticality, in the limit of large s and s_c. Here the scaling function $\mathcal{F}$ precisely describes the variation of the power law regime with s_c and accounts for deviations of the exponent when $s_c/L^{d_F} \ll 1$. In the thermodynamic limit, we expect that the dependence on s_c is only via the ratio s/s_c. For a graphical confirmation of the scaling relation (7.1.2) consider the inset of Figure 7.2,

which shows a scaling collapse according to (7.1.2) for fixed system size L.
In order to have a more quantitative check of (7.1.2), we considered the cut-off dependence of the mean stable fragment size $\bar{f}$, defined as the first moment of $n_{s_c}(s,\infty)$, that is:

$$\bar{f} \equiv \sum_{s=1}^{s_c} n_{s_c}(s,\infty)s.$$

As a direct consequence of (7.1.2) for sufficiently large volumes the mean stable fragment size $\bar{f}$ should have large s_c asymptotics

$$\bar{f} \approx s_c^{d_R/d_F}.$$

Indeed, we show the cut-off dependence of $\bar{f}$ for two cluster weights in Figure 7.3. Using

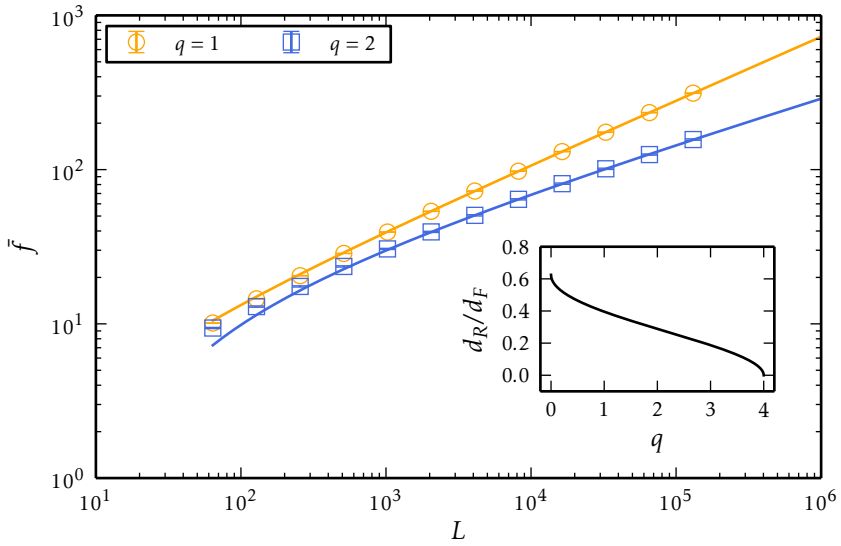


Figure 7.3: *Cut-off dependence of the mean stable fragment size $\bar{f}$ for the iterated fragmentation starting from a giant component at criticality in the random-cluster model on $\mathbb{Z}_L^2$ for cluster weights $q = 1$ and $q = 2$. The solid lines correspond to the best least squares fits of the functional form $\bar{f} = A + Bs_c^{\chi}$ to the data. The numerical results clearly support the scaling $\bar{f} \approx s_c^{d_R/d_F}$.*

the least squares fitting method with various standard power law model functions, such as $\bar{f} = A + Bs_c^{\chi}$ and $A + Bs_c^{\chi}(1 + 1/s_c)$, we find generally that the former describes our data for $\bar{f}$ best when discarding data points for small cut-offs. In particular for critical percolation, we estimate $\chi = 0.40(1)$, where the error-bar corresponds to a systematic error accounting for not considered corrections. The exact value of $d_R/d_F = 1/(\nu d_F)$ for $q = 1$ in two dimensions is $36/91 = 0.\overline{395604}$. We therefore find good agreement with (7.1.2) and further support for the related conjecture that $\chi = \phi$.
It is also helpful to consider the situation where $q \neq 1$ and seek for a confirmation of (7.1.2). We considered the Ising case, that is critical Fortuin-Kasteleyn cluster with $q = 2$ and $p = \sqrt{2}/(1 + \sqrt{2})$. Here the exact value of d_R/d_F is $13/45 = 0.2\overline{8}$, c.f. e.g. [15, 21], and

we extract $\chi = 0.28(1)$, where we discarded data points for cut-offs below 4096.

These observations suggest that the cut-off effectively preserves certain equilibrium

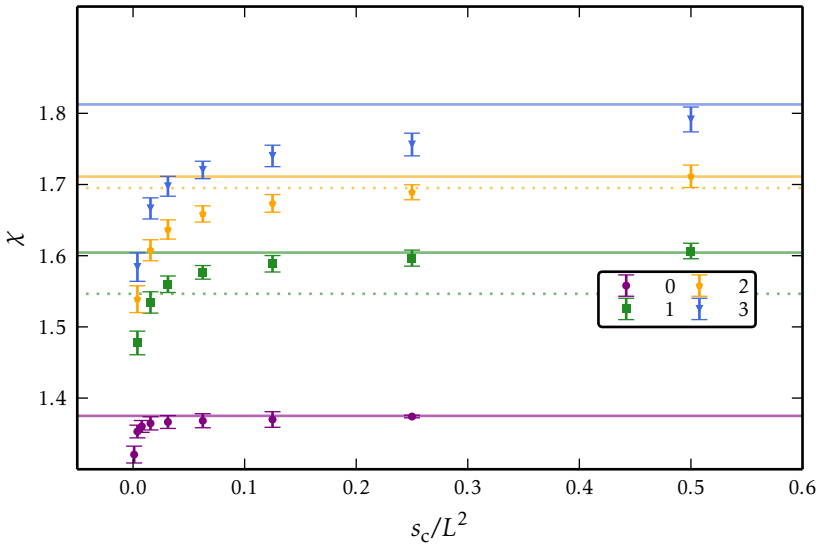


Figure 7.4: *Numerically extracted exponent χ from an imposed power law scaling of $n_{s_c}(s,\infty)$ as the cut-off is varied. The figure shows the estimation results for four cluster weights, where the solid lines show the corresponding exact value $\phi = 2 - d_R/d_F$. For comparison, the dashed horizontal lines show the value of ϕ for the same cluster weight at the respective critical point in 3D.*

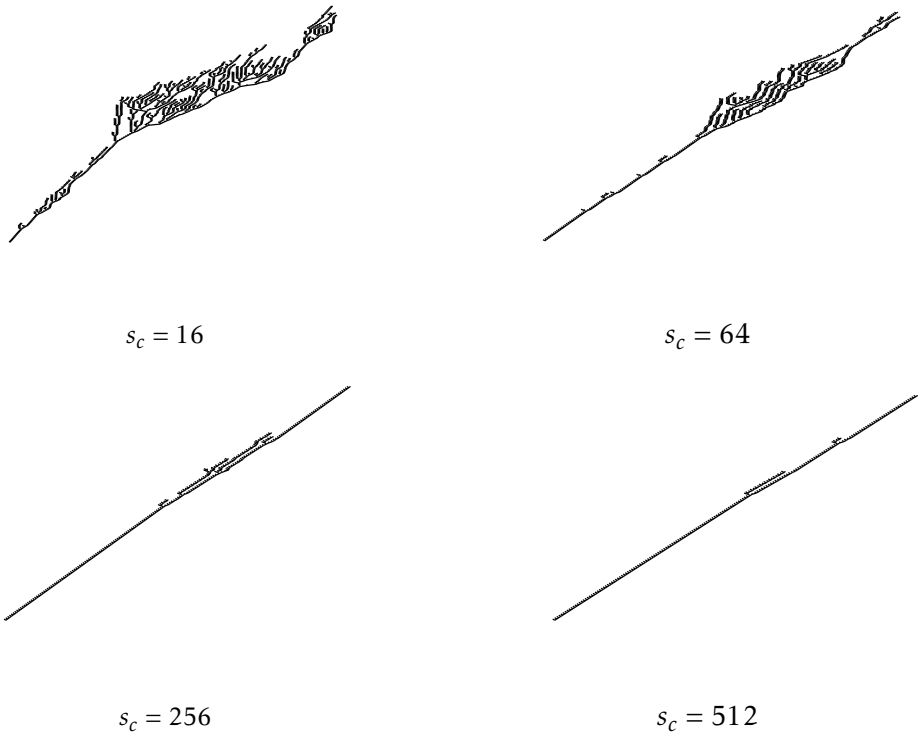
characteristics, such as in particular the exponent ϕ. Hence the intrinsic existence of a cut-off or the mechanically imposed suppression of fragmentations below a certain cut-off, provides a probe for the initial self-similar cluster structure. On the other hand, the scaling function $\mathcal{F}$ effectively introduces a "smooth" deviation from the initial power law scaling of $b_{s'|s}$ as s_c decreases. Nevertheless, we emphasise that more research is necessary to understand precisely the conditions of validity of the above scaling ansatz and more extensive data is needed to check its validity, in particular for $q \neq 1$ critical Fortuin-Kasteleyn clusters in two dimensions and beyond.

7.2 Splitting trees

A particular insightful point of view on fragmentation processes can be adopted by considering the naturally associated *splitting tree*. This is a rooted tree, with the property that each node either has two children or none. Splitting events in the fragmentation process correspond to vertices in the splitting tree with two children and stable fragments are vertices with no children (leafs). Here we adopt standard computer science notation [159], and call a vertex with two children *internal*, and one with 0 children *external*. Thus splitting events correspond to internal vertices and stable fragments to external vertices

in the splitting tree. In fact splitting trees are also known as *full* binary trees[2] in the computer science literature [159].
A similar perspective on fragmentation phenomena has been previously adopted in [160]. There the authors use the reverse idea, that is the authors utilise the analogy between recursive (continuous) fragmentation processes with a cut-off constraint and search trees in computer science, to derive asymptotic results for height and balance characteristics of the associated (search-)trees. Intuitively, a balanced tree corresponds to a fragmentation process with a break-up kernel that favours the production of daughter fragments of comparable mass. We have seen in the preceding discussion that this is in fact *not* the case for equilibrium fragmentation of Fortuin-Kasteleyn clusters. We therefore expect the splitting tree to be highly unbalanced.
Consider Figure 7.5, which shows several graphical illustrations of splitting trees for iterated fragmentation of the largest component, taken from a critical $q = 2$ Fortuin-Kasteleyn configuration, for different values of the cut-off s_c. Let us start with the cut-off value $s_c = 512$, for which the tree is characterised by a long path with only two sub-branches deeper than 1. In other words, the fragmentation process consists mainly of an "erosion" of small fragments from the giant component, until the giant component is sufficiently small to fall below the cut-off. Then, lowering the cut-off to $s_c = 256$ additionally extends a few of the sub-branches, corresponding to a further fragmentation of fragments, which were stable under the constraint of a larger cut-off. This observation extends until for relatively small cut-offs, here $s_c = 64$ or $s_c = 32$, many of the sub-branches "grow" and become deeper, corresponding to iterated fragmentations of previously stable fragments until their mass falls below the cut-off. This in turn must change the distribution of fragments $n_{s_c}(s, \infty)$ drastically, as their structure or mass distribution becomes significantly altered, by continued fragmentations, compared to the equilibrium case. This mechanism is "encoded" in the scaling function $\mathcal{F}$, which ultimately accounts for the transition from a power law distribution in s to a degenerate distribution concentrated at $s = 1$ for $s_c \to 1$. We emphasise, that more research is needed to study the precise form of $\mathcal{F}$.
With this insight from the splitting tree representation of iterated fragmentation, we can now heuristically explain the observed power law in $n_{s_c}(s, \infty)$ for sufficiently large cut-offs s_c. The continued abrasion of the giant component produces, typically, one daughter fragment of sufficiently small size to fall below the cut-off and hence to become stable. This in turn provides a mechanism which creates a fingerprint of the equilibrium cluster structure in the ensemble of stable fragments. At the same time when a fragmentation happens, the overall morphology of the unstable fragment, the giant component, does not change drastically by the separation of a, comparably, small fragment. We therefore expect that due to this erosion mechanism, the break-up kernel $b_{s'|s}$ remains applicable beyond a single-step fragmentation and acts as the kernel for an (limited) iterated fragmentation.

[2] A *full* binary tree is recursively defined as either a single vertex or two full binary trees joint at a vertex by two edges. Thus in a full binary tree each node has either 0 or 2 children. An *extended* binary tree allows also for the possibility of 1 child.



$s_c = 16$ $s_c = 64$

$s_c = 256$ $s_c = 512$

Figure 7.5: *Representation of the splitting tree of an iterated fragmentation for several values of cut-offs, starting from the giant component of a critical Ising $q = 2$ Fortuin-Kasteleyn configuration on a 128×128 square lattice. The corresponding cut-off values s_c are specified in the sub-captions.*

The above picture of an iterative applicability of $b_{s'|s}$ can indeed be verified for the iterated fragmentation of critical bond percolation on the infinite d-ary tree. The corresponding splitting kernel can be shown to equal (see [161]):

$$b_{s'|s} = \frac{s + \frac{1}{d-1}}{s-1} \frac{T_{s'} T_{s-s'}}{T_s},$$

where

$$\begin{aligned} T_s &= \frac{\left(1 - \frac{1}{d}\right)^{ds}}{s} \binom{ds}{s-1} \frac{1}{(d-1)^{s-1}}, \\ &\sim \frac{1}{\sqrt{\pi}} s^{-3/2}. \end{aligned}$$

Here we used the Stirling approximation [159] and the asymptotics are with respect to $s \to \infty$. We therefore obtain in the particular case $s' = xs$ for $x \in (0,1)$ fixed and $s \to \infty$:

$$b_{s'|s} \sim \frac{1}{\sqrt{\pi}} s^{-3/2} x^{-3/2} (1-x)^{-3/2}. \tag{7.2.1}$$

This clearly resembles the generic form (5.6.3) and has, expectedly, coinciding exponent with the Bethe lattice case [11]. Now, the relevant concept for an recursive applicability of the break-up kernel is the recently discussed "randomness-preservation property" [161],

which precisely assures that fragments are structurally equivalent to the mother cluster and hence $b_{s'|s}$ is applicable recursively. It is tempting to analyse the influence of the cut-off in this "simplified" setting. Unfortunately we have not been able to derive even the corresponding expected number of stable fragments of size s, $\langle N_{s_c}(s,\infty)\rangle$. This is a stimulating problem, we shall pursue in a future study. Here we provide merely a generally valid recursion relation for the quantity $\langle N_{s_c}(s,s_0)\rangle$, where we suppressed the asymptotics indicator ∞ and added the dependence on the initial mass of the starting fragment s_0. Now assuming a perfect recursive applicability of $b_{s'|s}$ we have

$$\langle N_{s_c}(s,s_0)\rangle = b_{s|s_0} + \sum_{s'=s_c+1}^{s_0-1} b_{s'|s_0}\langle N_{s_c}(s,s')\rangle, \tag{7.2.2}$$

whenever $s_0 > s_c$ and $s \leq s_c$. For instance, in the case of $q_{s'|s} = 1/(s-1)$, corresponding to the fragmentation of percolation on the trivial one dimensional path graph (1-ary tree) with at least s vertices and any bond-density p, we obtain $\langle N_{s_c}(s;s_0)\rangle = 2s_0/(s_c(s_c+1))$ and hence $n_{s_c}(s;\infty) = 1/s_c$, independent of s_0, corresponding to the uniform distribution over all s_c stable fragment sizes. We can cast this also into the form of the scaling ansatz proposed in (7.1.2) with $\mathcal{F}(s/s_c) = s/s_c$ and $\phi = 1$. This is equivalent to $d_R/d_F = 1$, assuming an object that is everywhere breakable, which is precisely the case for a typical cluster for bond percolation with $p > 0$ on the path graph. Interestingly, it turns out that the fragmentation of critical percolation clusters on the d-ary tree as well as the trivial example of fragmentation of the path graph, are related to a family of recursively defined fragmentation models known as the beta-splitting model, which we describe next.

7.2.1 Analogy to the beta-splitting model

It is instructive to consider mathematically amenable models of fragmentation that show similar characteristics as these found for the iterated fragmentation of self-similar (random) structures. One promising model, or actually family of fragmentation processes, is Aldous' *beta-splitting* model [31]. This one-parameter family of models was initially introduced to construct probability distributions on cladograms or evolutionary trees. It utilises the apparent relationship between binary fragmentation and binary trees (splitting trees). In somewhat different terms, the beta-splitting model is a one-parameter family of probability distributions on the space of full binary trees with a fixed, say s, number of external nodes. It is clear that each full binary tree $\mathcal{T}_s$ with s external nodes encodes a binary fragmentation process terminating with s fragments, thus it is equivalent to what we referred to as splitting tree before.
Now, the beta-splitting model is defined as follows. Fix a number of s (point) "particles", each of unit "mass", and locate them uniformly at random on the unit interval. Then split the unit interval at a position $0 < x < 1$, which is given by a random variable with probability density function $f(x)$. Now separate the set of particles into two sets, corresponding to particles to the left and to the right of x. This clearly introduces a split or fragmentation

of the original mass of s particles into a bunch or particle to the left of x, with (random) mass s', and a (right) bunch of mass $s-s'$. This process is continued for each bunch of particles (with appropriate length-rescaling for x) until each bunch or fragment consists of one particle (mass 1). To formalise, denote the probability that the left fragment has mass s' by $q_{s'|s}$ for $1 \leq s' < s$. For the particular choice of f being the probability density function of a beta random variable with parameter $-1 < \beta < \infty$ [31], that is

$$f(x) = \frac{\Gamma(2\beta+2)}{\Gamma^2(\beta+1)} x^\beta (1-x)^\beta, \quad 0 < x < 1$$

one obtains for $1 \leq s' < s$:

$$q_{s'|s} = \frac{1}{a_s(\beta)} \frac{\Gamma(\beta+s'+1)\Gamma(\beta+s-s'+1)}{\Gamma(s'+1)\Gamma(s-s'+1)}, \tag{7.2.3}$$

with $a_s(\beta)$ such that $\sum_{s'=1}^{s-1} q_{s'|s} = 1$. Note that $q_{s'|s}$ is defined as the size of the left fragment in the process, so we obtain $b_{s'|s} = q_{s-s'|s} + q_{s'|s} = 2q_{s'|s}$, by the left-right symmetry of the break-up. One important observation Aldous made in [31] is that, albeit the Beta distribution is not a probability distribution in the strict sense for $\beta \leq -1$, the expression (7.2.3) is still well defined and in particular a valid (normalisable) probability distribution on $\{1,\ldots,s-1\}$ for $-2 < \beta \leq -1$. However, the analogy to the interval splitting induced by a Beta random-variable is lost. Yet in principle, one can still construct a recursive interval splitting process by using $q_{s'|s}$ explicitly. Furthermore, note that in particular for $-2 < \beta < -1$ and $s' = xs$ with $0 < x < 1$ one obtains the following large s asymptotics

$$b_{s'|s} \sim \kappa_\beta^{-1} x^\beta (1-x)^\beta s^\beta, \tag{7.2.4}$$

with a constant κ_β. This clearly resembles the general form of break-up kernels encountered in section 5.6 for the fragmentation analysis of critical Fortuin-Kasteleyn clusters, in particular relation (5.6.3). Moreover for the particular case $\beta = -3/2$ we recover, up to a constant of proportionality, the same asymptotics as (7.2.1).

For a given fragmentation process in the beta-splitting model one can now consider the associated splitting tree and moreover study its characteristics, such as the typical and maximal depth, as β varies. For instance, for $\beta \to \infty$ it is not hard to see that the process is a deterministic interval-bisection process, which produces perfectly balanced splitting trees. The choice $\beta = -3/2$ leads to a uniform distribution on binary trees with s external nodes, when started with s particles [31]. One of the remarkable features of the beta-splitting model is that it possesses a certain phase transition in the structure of the associated splitting tree, as follows. As already outlined, the case $\beta \to \infty$ produces perfectly balanced splitting trees. On the other hand, for the case $\beta = -3/2$ the analogy to the fragmentation of critical percolation clusters on the d-ary tree, in particular the observation that a fragmentation typically produces very unbalanced fragments, suggests that a splitting tree is typically very unbalanced. Indeed Aldous established corresponding balance characteristics for the beta-splitting model in [31].

In order to formalise the notion of balance, we can consider the height statistics of the associated splitting tree. More precisely, we define the *depth* of an external node in $\mathcal{T}_s$ as the number of internal nodes along the unique path from it to the root of $\mathcal{T}_s$ (including the root). Viewed as a fragmentation process, this simply counts the number of fragmentation events a given fragment was involved in. Now, following Aldous [31], we define *the height* of the corresponding splitting tree $\mathcal{T}_s$ as the maximal depth over all external vertices. Note, that as a measure of balance we can also consider the average depth over all external nodes. The phase transition occurs in the expected height, or expected average depth, as β is varied. In particular the expected height is logarithmic in s (initial mass) when $-1 < \beta \leq \infty$, and asymptotically proportional to $s^{-\beta-1}$ when $-2 < \beta < -1$. The expected maximal depth for the marginal case $\beta = -1$ is believed to be of the order $\log(s)^2$. Concerning the expected average height the order of magnitude does not change from what is known for the expected height, however in contrast to the expected height, the expected average height is also known rigorously for $\beta = -1$.

7.2.2 Depth of fragmentation splitting trees

The fact that the beta-splitting model, being manifestly recursive, shows a power law scaling in the height of the associated splitting tree for $-2 < \beta < -1$, together with the asymptotic equivalence of the $\beta = -3/2$ model to the fragmentation of critical percolation clusters on the infinite d-ary tree, raises the question whether the iterated fragmentation of critical Fortuin-Kasteleyn clusters in general shares some of beta-splitting characteristics. Recall that we have established that the exponent ϕ equals $2 - d_R/d_F$ for fragmentation of critical Fortuin-Kasteleyn clusters in any dimension. Further, because the red bond fractal dimension can not exceed the cluster fractal dimension, we obtain the (trivial) generic constraints $1 \leq \phi \leq 2$. This observation leads directly to the regime $-2 < \beta < -1$ in the beta-splitting model, whose associated splitting trees have expected height increasing polynomially with s. A natural generalisation of the $s^{-\beta-1}$ scaling of the expected average depth and height in the beta-splitting model, to the Fortuin-Kasteleyn scenario, in a box with linear dimension L goes as follows. Notice that the fragmentation starts only on the largest component, which has typical size L^{d_F} in a system, hence, to leading order, we can substitute $s \to L^{d_F}$ in $s^{-\beta-1} = s^{\phi-1} = s^{1-d_R/d_F}$, where we identified $-\beta = \phi$. We therefore obtain, denoting the height and average depth by h and $\bar{d}$, respectively:

$$\mathbb{E}[h] \sim C_h L^{d_F - d_R} \quad \text{and} \quad \mathbb{E}\left[\bar{d}\right] \sim C_{\bar{d}} L^{d_F - d_R}, \tag{7.2.5}$$

where $\mathbb{E}[\cdot]$ denotes the corresponding expectation of the iterated fragmentation process, including initial conditions, and C_h and $C_{\bar{d}}$ are appropriate constants. In order to confirm the validity of (7.2.5), we first determined the depth and average height of the splitting trees of 10000 iterated fragmentations with cut-off $s_c = 1$, starting with the largest component of a critical bond-percolation configuration on the square lattice. The exact value of $d_F - d_R$ in this case equals $55/48 = 1.1458\overline{3}$ [46]. We find that our estimates for $\mathbb{E}[h]$ and

$\mathbb{E}\left[\tilde{d}\right]$, shown in Figure 7.6, are best described by a scaling form

$$\mathbb{E}[h], \mathbb{E}\left[\tilde{d}\right] \sim AL^{\varphi}\left(1 + CL^{-D}\right). \tag{7.2.6}$$

Specifically, we extract $\varphi = 1.16(1)$ and $\varphi = 1.147(7)$ for the height and average depth statistics, respectively. The values of φ are both statistically consistent with $\varphi = d_F - d_R$. We remark, that the involved correction exponent is comparably large in both cases, that is $D = 0.70(5)$ in both cases. As a further check of (7.2.5) we also analysed the iterated fragmentation of critical Ising Fortuin-Kasteleyn cluster. Here the estimates are again based on 10000 samples, and we again find the above scaling form to describe our data best. Our estimates read $\phi = 1.32(2)$ for the expected height, and $\phi = 1.31(1)$ for the expected average height. The correction exponent D now differs between the height and average depth estimates, i.e., $D = 0.55(3)$ for the former, and $D = 0.83(7)$ for the latter. Yet, we find perfect agreement between the estimated values of φ and the exact value $\varphi = d_F - d_R = 4/3 = 1.\overline{3}$ [22]. For reference, we show in the inset of Figure 7.6 the q-dependence of the difference $d_F - d_R$ for the critical random-cluster model in two dimensions, as obtained from the corresponding Coulomb gas formulation of two-dimensional critical random-cluster models [51]. Our analysis of the splitting trees so

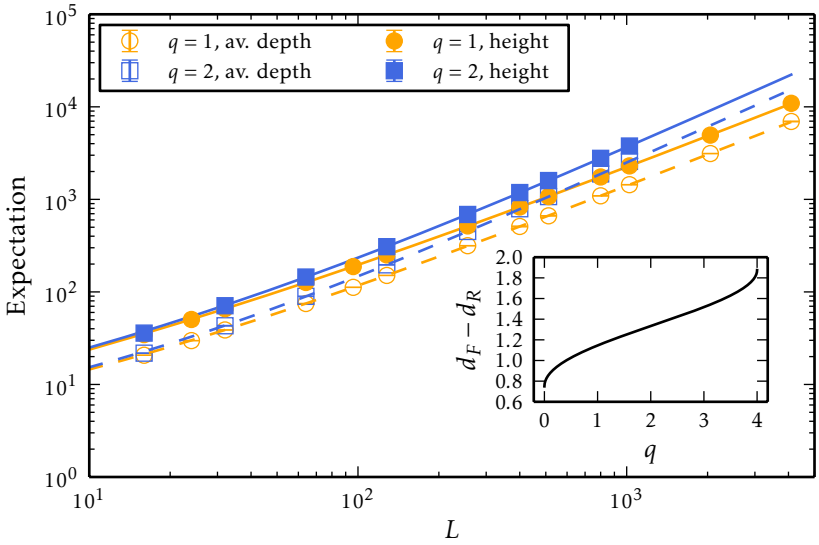


Figure 7.6: *System size dependence of the expected height and average depth of the splitting tree associated to the iterated fragmentation of critical Fortuin-Kasteleyn cluster. The solid and dashed lines correspond to the best fits to the scaling ansatz* (7.2.6). *The inset shows the exact value of $d_F - d_R$ for the two dimensional random-cluster model at criticality, as q is varied.*

.

far has only been concerned with an unconstrained iterative fragmentation, that is $s_c = 1$. However, it is clear that the splitting tree varies with s_c, and so also potentially height and depth properties. A preliminary analysis suggests however, that the variation of s_c in any non-trivial regime has a minor effect on the height of the splitting tree (consider

Figure 7.5), such that the expected height remains polynomial for a wide range of cut-offs. However we emphasise that more research is needed to understand this quantitatively. We therefore suggest that the next step in a subsequent study should consider a non-trivial cut-off $s_c > 1$, and study the dependence of $\mathbb{E}[\bar{d}]$ and $\mathbb{E}[h]$ on s_c. Another interesting question is whether there exists a (size dependent) cut-off $s_c(L)$, such that the expected height becomes poly-logarithmic. Furthermore potentially interesting are statistical quantities beyond the expectation, such as the distributional nature of the depth and height of associated splitting trees. It is apparent that the study of splitting trees corresponding to iterated fragmentation processes raises many questions, and we hope that future studies aim for a resolution of some of the posed questions.

7.3 Rate equation approaches

Being manifestly dynamic, the analysis of iterated fragmentation leads naturally to so-called *rate-equation* approaches to fragmentation, see for instance [162, 163]. Here one models the time evolution of a cluster size density $u(s,t)$, that is the density of clusters with mass in the interval $(s, s+\mathrm{d}s)$, by the following integro-differential equation

$$\frac{\partial u(s,t)}{\partial t} = -\int_0^s u(s,t)c(s,s',t)\mathrm{d}s' + 2\iint_s^\infty u(s',t)c(s',s,t)\mathrm{d}s', \tag{7.3.1}$$

where $c(s,s',t)$ is also called the "splitting kernel" and is the analogue to $b_{s'|s}$, however now it specifies the rate at which a cluster of size s breaks into two clusters of size s' and $s-s'$. Typically, one makes a factorisation ansatz $c(s,s',t) = a(s,t)b(s|s';t)$, where $a(s,t)$ is the rate at which a cluster of mass s fragments, and $b(s'|s;t)$ the probability that a cluster of mass s splits into fragments of mass s' and $s-s'$. Crucial underlying assumptions are threefold (in addition to the assumption of a sufficiently large system to justify taking a continuum limit): One assumes spatial homogeneity (fragment densities are independent of the specific location), shape independence of the fragmentation mechanism and that the break-up is caused by external drive rather than interactions between fragments [163]. Now, if we wish to introduce a cut-off s_c, it is useful to distinguish between the distribution $n(s,t;s_c)$ of stable fragments and $u(s,t;s_c)$ of unstable fragments. In this setting the rate-equations read [156]

$$\frac{\partial u(s,t;s_c)}{\partial t} = -\int_0^s u(s,t;s_c)c(s,s',t)\mathrm{d}s' + 2\iint_s^\infty u(s',t;s_c)c(s',s,t)\mathrm{d}s' \quad s > s_c, \tag{7.3.2}$$

$$\frac{\partial n(s,t;s_c)}{\partial t} = 2\int_{s_c}^\infty u(s',t;s_c)c(s',s,t)\mathrm{d}s' \quad s \le s_c. \tag{7.3.3}$$

Firstly, note that the fist equation is analogous to the previous rate equation without cut-off, however only valid for $s > s_c$. Secondly, the two equations are obviously decoupled, and it therefore suffices to solve the rate-equation for $u(s,t;s_c)$, which in turn can be used to solve for $n(s,t;s_c)$, at least in principle. It is possible for a break-up kernel that is

time-independent and homogeneous, that is $c(\lambda s, \lambda s') = \lambda^{\alpha-1} c(s, s')$ with an $\alpha > 0$ to show that (7.3.1) has a self-similar, or scaling, solution [164, 163], that is

$$u(s,t) = t^{2/\alpha} \Psi\left(s t^{1/\alpha}\right) \tag{7.3.4}$$

Inserting this into (7.3.3) we obtain a formal solution for $n(s,t;s_c)$:

$$n(s,t;s_c) = 2 \int_0^t \int_{s_c}^{\infty} c(s',s) t'^{2/\alpha} \psi\left(s' t'^{1/\alpha}\right) \mathrm{d}s' \mathrm{d}t'. \tag{7.3.5}$$

As an example, let us consider the possibly simplest case, following [165, 164], with a splitting-kernel $c(s,s') = s^{\alpha-1}$ for $\alpha > 0$. This can be shown to correspond to a system, starting with a unit-mass fragment, where each fragment of mass s breaks with a rate s^α into two fragments $s\mathcal{R}$ and $(1-\mathcal{R})s$, where $\mathcal{R}$ is chosen uniformly at random in the unit interval and independent of the break-up of all other fragments. In this case one can obtain the explicit expression $\Psi(x) = \alpha \exp(-x^\alpha)/\Gamma(2/\alpha)$, which then yields after some algebra

$$n(s;t;s_c) = \frac{2}{\Gamma\left(\frac{2}{\alpha}\right)} \frac{1}{s_c^2} \gamma(2/\alpha, s_c^\alpha t),$$

where $\gamma(s,x) = \int_0^x t^{s-1} e^{-t} \mathrm{d}t$ is the lower incomplete gamma function. Moreover, as can be easily verified we have

$$n(s;\infty;s_c) \equiv \lim_{t\to\infty} n(s;t;s_c) = \frac{2}{s_c^2}.$$

One can further directly verify that $\int_0^{s_c} s n(s;\infty;s_c) \mathrm{d}s = 1$, which is of course mass-conservation. To summarise, the stable fragment distribution is uniform in $(0, s_c)$, a direct consequence of the uniformity of the break-up process at each step. Moreover, the dependence on α is completely lost when concerned with the asymptotic distribution. This is a consequence of the fact that in the particular case of the example above, α is only related to the rate at which a cluster of a given mass fragments, which has no relevance for the asymptotic analysis.

The above example is probably the simplest case that supports the intuition that the nature of the fragmentation mechanism and hence the actual material structure can be probed by considering the distribution of stable fragments. Yet, the uniform case is very special and apparently far from the nature of the fragmentation for critical Fortuin-Kasteleyn clusters. More generally, it belongs to the class of fragmentation processes with a *separable* splitting kernel, in the sense that $c(s,s') = d(s)e(s')$, e.g. $d(s) = s^{\alpha-1}$ and $e(s') = 1$ in the uniform example. However, our analysis of the equilibrium fragmentation has shown that $c(s,s') = a(s) b_{s'|s}$, where $a(s) \propto s$ and $b_{s'|s} = s^{-\phi} g(s'/s)$, which therefore is non-separable. Yet, the corresponding splitting kernel $c(s,s')$ is in fact homogeneous with parameter $1-\phi$, that is $c(\lambda s, \lambda s') = \lambda^{1-\phi} c(s,s')$. Recall, that we derived $\phi = 2 - d_R/d_F$, which implies that the corresponding exponent α in the scaling solution (7.3.4) equals d_R/d_F, the ratio of red bond fractal and cluster fractal dimension. For instance, in two-dimensions, we can continuously vary d_R/d_F from 5/4 for $q \to 0$ to 0 for $q \to 4$. However, the analytical

analysis of the rate-equation approach with a splitting-kernel of the above form becomes considerably more involved due to the ratio-dependence in $g(x)$ which causes the kernel to be non-separable.

Unfortunately, so far we have not been able to construct an appropriate continuum analogue to the above iterated fragmentation of Fortuin-Kasteleyn clusters for the quantity $g(x)$. A first step into this direction could therefore be a careful estimation, or derivation in simplified settings, of the function $g(x)$, which then is fed into the above rate-equation formalism. We expect, that the asymptotic distribution of stable fragments should then resemble the scaling form (7.1.2), supported by our numerical analysis.

Lastly, let us mention that the role of cycles in the iterated fragmentation of fractal random structures is not yet clear. One possible scenario is that that removal of cycles merely introduces a rescaling of time, and is irrelevant for the validity of $b_{s'|s}$ in the iterated fragmentation scenario. It would be interesting to compare the iterated fragmentation of fractal random structures as done so far, with the iterated fragmentation of a (uniformly at random chosen) spanning tree of the giant component [166]. Such a spanning tree provides in some sense a "backbone" of the original giant component, as it the break-up kernel $b_{s'|s}$ remains invariant in equilibrium. However, it is not clear to us how the iterated fragmentation interacts with cycles and whether such "universal" aspects as the exponent ϕ remain unchanged between the two variants of iterated fragmentation.

Chapter 8

Conclusion and Outlook

In this thesis we established novel results for a range of aspects that reflects the structural, geometric, computational and dynamical variety of the random-cluster model. The Sweeny dynamics, being a single-bond update Markov chain, operates sufficiently close to the original cluster structure to provide a dynamical and algorithmic perspective on the model. The interplay between the dynamical and algorithmic properties of the Sweeny dynamics and the equilibrium cluster structure of the random-cluster model is one of the salient themes of this thesis. We deepened the understanding of how the cluster weight parameter in the random-cluster model influences the structure of clusters by studying their fragility and introduced a natural classification of edges, based on their relevance for the connectivity structure of clusters. We showed that the cluster weight directly affects the balance between the different types of edges by deriving an exact identity, valid for any graph and any $q \neq 1$, that allows us to express the corresponding densities of the four classes of edges (which we denoted as bridges, non-bridges, candidate bridges and candidate non-bridges, respectively) in terms of the overall density of edges. In this framework, our results reflect nicely the intrinsic correlations between edges in the random-cluster model. Interestingly, in the case $q = 1$ we were able to relate the bridge-density to a previously studied covariance [111]. For the particular case of the square lattice we derived, using self-duality, the exact asymptotic densities of the above edge types for the critical random-cluster model as a function of q. We determined the leading finite-size corrections and extended our analysis also to the three dimensional case, for which we confirmed all our theoretical arguments.

The studied densities of bridges and non-bridges are also of significance for the algorithmic efficiency of the Sweeny dynamics, and we explicitly showed how the runtime scaling of the various implementations considered in this thesis depend on the proportions of the different types of edges.

It is interesting to find further applications of our bridge-edge identity, possibly in the study of random graphs or complex networks, where a particular interest lies in resilience properties [167]. From an analytical perspective it would be useful to obtain further information on higher moments of the considered densities and study correlations

between the types of edges. In fact we have already made some partial progress in the bond percolation setting. Based on the results of [110], we are able to derive for two arbitrary edges $e, f \in E$ at distance r on the square lattice:

$$\mathbb{P}_{\frac{1}{2},1}[e \in B, f \in B] - \mathbb{P}_{\frac{1}{2},1}[e \in B]\mathbb{P}_{\frac{1}{2},1}[f \in B] \approx \frac{2\sqrt{3}}{\pi} \log(r) r^{-2x_2}.$$

Recall that B is the set of bridges in the spanning sub-graph (V, A). The corresponding details will be presented elsewhere [168].
Additionally an interesting quantity to study is what we refer to as the *bridge-load*. The bridge-load of an edge e is defined as the number of nearest neighbour pairs x, y (that is $(x, y) \in E$), for which the edge e must be part of any connecting path. It is hence a measure of the importance or load of a given bridge. Intuitively, the more nearest neighbour pairs rely on a fixed bridge, the larger is the overlap of the two clusters glued together by the bridge. It is possible to relate the bridge-load to the first p-derivative of the bridge density for percolation [168]. Studying the bridge-load could therefore be useful to understand the variation of the cluster fragility with p. Our results in Chapter 5 showed that the maximum of the density of bridges does not coincide with the critical point, but is rather located on a smaller bond density p_f. It is therefore tempting to find the exact value of $p_f(q)$, say on the square lattice, and understand what distinguishes the point $p_f(q)$ geometrically. A realistic starting point could be a careful analysis of the case $q = 1$ on $\mathbb{Z}^2$ as well as the Ising case, possibly utilising Onsager's exact solution [22] for the Ising case on the square lattice.

The fragility of random sub-graphs of a fixed graph $G = (V, E)$ can also be studied using the following model of *bridge-weighted percolation*, which we define by the following weight function w on $\Omega = \{A | A \subseteq E\}$:

$$w(A) = v^{|A|} b^{|B|}.$$

Here $v > 0, b > 0$ and clearly for $b = 1$ we recover the independent bond percolation model with the usual parametrisation $p = v/(1 + v)$. For $b > 1$ ($b < 1$) the model has a tendency to favour (disfavour) bridges. We could not find a related study in the literature nor relate it to a particular evaluation of the Tutte polynomial [69], however see [169] for the different but related "bridge-percolation" model. We believe that it is therefore desirable to study the phase diagram of the model and check whether it shares aspects with the random-cluster model beyond $b = 1$. Two immediate related questions pop up: Is it possible to determine a critical manifold $v(b)$ as for the random-cluster model in the case of the square lattice or planar graphs (using duality)? Does the model have a continuum of universality classes upon variation of b and are the universality classes relatable to the random-cluster model?

We extended our study of the fragility of Fortuin-Kasteleyn clusters by an analysis of the fragment sizes produced upon a break-up of a cluster. We confirmed a scaling ansatz,

originally introduced in the setting of independent bond percolation [115, 29, 30], for the break-up kernel $b_{s'|s}$ and derived the scaling relation $\phi = 2 - d_R/d_F$ between the fragmentation exponent and standard fractal dimensions in the random-cluster model. We find that the previously assumed scaling relation $\phi = 2 - 1/(\nu d_F)$ due to Edwards et al. [115] does not extended to the random-cluster model. In fact our relation for ϕ reduces to the above percolation relation only due to the coincidence of the thermal exponent $1/\nu$ and the red bond fractal dimension d_R for $q = 1$ [15].

There are a few open questions and interesting generalisations of our study, out of which we would like to highlight two. Firstly, the natural degrees of freedom in the Potts model are the spins located on vertices and one can therefore also consider the *spin clusters*, consisting of vertices with equal spin and all edges between them. The spin clusters do not correspond to the Fortuin-Kasteleyn clusters [170], so it is interesting to study the fragility of such spin clusters. These clusters are denser than the Fortuin-Kasteleyn clusters (the latter are obtained from the former by removing edges with probability $e^{-\beta}$). Besides studying the densities of bridges and other types of edges, one might about the value of ϕ. Does our scaling relation $\phi = 2 - d_R/d_F$ naturally extend to spin clusters, by replacing the Fortuin-Kasteleyn fractal dimensions with the spin cluster fractal dimensions [170]? Considering the two dimensional case, it is known that $d_R < 0$ for $q = 2, 3$ and $d_R = 0$ at tricriticality $q = 4$. This suggests that $\phi > 2$ for $q = 2, 3$, which appears somewhat peculiar. However, interpreting $d_R < 0$ as "asymptotically there are no red bonds" which means there are no bridges that can cause large scale break-ups, we might also conclude that effectively $d_R = 0$ for $q = 2, 3$ as it is the case for $q = 4$. Equivalently we could conclude $\phi = 2$ for all Potts spin cases. This would then be consistent with the heuristic view developed in Chapter 7, stating that the case $\phi = 2$ corresponds to a perfectly dense object that only fragments due to minimal surface erosion. Indeed, note that in [171] the authors study the fragmentation of Ising spin clusters, and find an estimate of ϕ consistent with $\phi = 2$.

The second interesting generalisation of our results is the determination of the fragmentation exponent ϕ for vertex fragmentation. We have seen that vertex fragmentation is not restricted to binary break-up events but can rather produce higher order break-ups. Here one can also study the fragmentation exponent $\phi^{(k)}$ for vertex fragmentations that produce k fragments. A natural question is whether $\phi^{(k)}$ is relatable to the k-arm exponent for $k \geq 2$. This is a natural generalisation of our result for binary fragmentation, using the k-arm exponent [15]. In particular, being true, this would then yield a practical method to measure the k-arm exponents, which are of great importance to the mathematical study of percolation [26] and the random-cluster model [27, 74].

The most striking feature that emerged from our analysis of the iterated cut-off fragmentation process of critical Fortuin-Kasteleyn clusters is the possibility of inferring structural characteristics of the fragmented object by carefully considering the produced fragments. More precisely, we analysed the size distribution of stable fragments, that is clusters sufficiently often broken apart to fall below the cut-off. In this connection we

proposed a scaling ansatz that describes the influence of the cut-off on the size distribution. Specifically, we find the size distribution to follow a power law with an exponent close to the equilibrium fragmentation exponent $\phi = 2 - d_R/d_F$. In fact our scaling ansatz predicts that the power law contribution in the size distribution is indeed ruled by the exponent ϕ but the range of validity of this power law varies with the cut-off. We confirmed that asymptotically only the ratio of stable fragment size to cut-off is required in a scaling function to describe the deviation from a power law with exponent ϕ.
We then analysed the splitting tree representing the entire history and genealogy of the fragmentation process. Remarkably we found that the morphology of the splitting tree can be used to extract the difference $d_F - d_R$ and hence an equilibrium property of the fragmented object. This is a rather intriguing observation taking into account that the splitting tree is a priori a property of the whole dynamical fragmentation process. We explained the above observations by a recursive applicability of the break-up kernel $b_{s'|s}$, due to the self similar structure of critical Fortuin-Kasteleyn clusters (in other words unstable fragments closely resemble the structure of the original fragment). This finding lends additional credibility to rate equation approaches [150] as descriptions of the iterated fragmentation of random fractal structures.
It is tempting to understand whether the ubiquitous observation of power laws in size distributions of fragmentation process of various kinds found in nature and technology can be related to some sort of self-similar cohesive structure in the original fragmented object. The insensitivity to microscopic details implied by the universality of critical phenomena indicates that our results for dynamic fragmentation should indeed be comparable to experiments. In fact we reported already in our recent letter [120] that one finds experimentally fragmentation exponents spanning a range of around 1.2 to 1.9, which are covered by our scaling relation $\phi = 2 - d_R/d_F$ for the critical random-cluster model. However a closer analysis and comparison beyond comparing power law exponents is needed, which we hope to pursue in a future study.
So far our study of the fragmentation process was mainly computational and heuristic, that is many of the aspects of the fragmentation process remain to be mathematically analysed. For instance: What is the precise influence of the cut-off on the fragment size distribution, what is the functional form of the involved scaling function? Moreover, it is desirable to have an exactly solvable fragmentation process at our disposal, for which it is possible to show that a power law in the break-up kernel persists in the stable fragment size distribution. We plan to analyse the role of the cut-off in the beta-splitting model [31], which we showed to posses surprising similarities with our iterated fragmentation process of Fortuin-Kasteleyn clusters.

The fragmentation properties of the random-cluster model established in this thesis are of independent interest, yet we showed that they naturally arise in a study of the computational efficiency of the Sweeny dynamics. For instance, the expected smaller fragment size determines the running time of an interleaved breadth-first search based implementation [15] or our cluster identifier based implementation. Albeit the dynamic

connectivity algorithm [24] is only provably poly-logarithmic when implemented with the full edge level hierarchy, we find that the typical smaller fragment size suppression at criticality allows for a sufficient "practical amortisation" of the computational costs of expensive operations without the prescribed $\log(n)$ level high forest hierarchy. The so constructed algorithm is hence a connectivity algorithm that maintains a spanning forest of the current configuration based on the Euler-tour data structure and scans the smaller tree upon a tree edge removal for a replacement edge. The crucial functionality that boosts the replacement edge search is the hierarchical information stored in the Euler-tour data structure that allows iterating from one non-tree edge to the next with only logarithmic computational effort. So far we have however not been able to prove, in a probabilistic setting, that the running time is indeed poly-logarithmic at criticality for this one-level variant of the connectivity algorithm. Hence it is possible that we are deceived by small constants in front of a polynomial $L^{d_F - x_2}$ contribution to the running time.

The combination of the poly-logarithmic dynamic connectivity algorithm with the critical-speeding up property of the Sweeny dynamics for important non-local quantities makes the single-bond dynamics a powerful tool. We compared the Swendsen-Wang dynamics for the Potts model to the Sweeny dynamics at criticality on the square lattice, and utilised a joint efficiency measure considering the statistical and computational efficiency of both chains. Albeit it is known [53, 88, 21] that the dynamical critical exponents determined in the literature are smaller for the Sweeny dynamics in two and three dimensions, we merely find in the best case an approximate equivalence of the two methods. Note however, that the poly-logarithmic approach in combination with a slightly smaller exponent will eventually, for large enough system sizes, become the more efficient and hence favourable option. Yet, because our joint analysis was conducted (in two dimensions) using the plain unoptimised poly-logarithmic connectivity algorithm, whose large memory footprint hindered the simulation of large enough system sizes to see the asymptotic running time scaling, we believe that using the optimised version will allow us to advance to larger system sizes where the superiority of the Sweeny approach becomes distinguished. We will pick up this question and a careful study of the three-dimensional case in a future study.

Deng et al. [21] observed numerically that the critical speeding-up effect is related to the fractal dimension of red bonds and becomes stronger with increasing red bond fractal dimension. In the case of the heat-bath chain for bond percolation on the square lattice we are able to prove the critical speeding up for the indicator function of certain crossing probabilities (with obvious importance for the study and location of phase transitions in the percolation model). The arguments crucially rely on recent results on the noise sensitivity of Boolean functions and percolation [90], and naturally bring the red bond fractal dimension into the discussion, which supports the heuristics of [21]. It is tempting to extend the percolation study to the variety of non-Boolean functions considered in Refs. [53, 15]. Furthermore we also plan to study the speeding up in more detail in higher

dimensions and will present the results together with the rigorous results for the critical speeding-up of crossing probabilities in due course.

Our study of the coupling from the past algorithm for the random-cluster model in two and three dimensions revealed that such an exact sampling procedure away from a phase-transition location in the stationary model is very efficient. In other words, the expected coupling time $\mathbb{E}[\tau]$ is essentially numerically not distinguishable, up to a constant, from the expected coupon collector time $\mathbb{E}[T]$, the minimal number of steps required to assure that each edge is visited at least once in a random hit setting. We supported this picture by a rigorous analysis of the coupling process on the cycle graph, which essentially is off-critical for all bond densities. Here we showed that the optimal case occurs asymptotically, that is $\mathbb{E}[\tau] \sim \mathbb{E}[T]$.
In two and three dimensions we found at criticality power laws for the expected coupling time and determined, numerically, the corresponding Propp-Wilson exponent w. Our estimates are surprisingly close to the corresponding values of z_{exp} determined in the literature. We showed that this has two immediate consequences. Firstly, in combination with the known sharpness of the (lower) Li-Sokal bound in two dimensions we can narrow down the allowed values for z_{exp}, as the expected coupling time yields an upper bound for z_{exp}. The two bounds become intriguingly close for $q \to 4$, essentially narrowing down to $z_{\text{exp}} = 1$ modulo multiplicative corrections for $q = 4$. Unfortunately, the responsible mechanism for this tightening of the lower and upper bounds for $q \to 4$ is not yet identified and remains a challenging open problem. Secondly, the closeness of z_{exp} and w is also of practical relevance, as it imposes only a minimal overhead to move from an approximate MCMC sampling scheme ruled by z_{exp} to a perfect sampling CFTP scheme ruled by w.
Our results for the coupling time in three dimensions at criticality showed the surprising fact that the Propp-Wilson exponent w is, for all considered cluster weights, in fact smaller than the anticipated value z_{exp} for the Chayes-Machta chain [88]. We conclude with the appealing fact that, for all studied cluster weights, the coupling from the past algorithm in combination with a poly-logarithmic connectivity algorithm is more efficient than the Chayes-Machta algorithm at criticality in three dimensions when considering the expected coupling time as the measure of efficiency for the former. A further, yet preliminary, study shows that the ratio of standard deviation to expectation of the coupling time decreases with the system size for all cluster weights considered. This casts the expected coupling time to the dominant scale[1]. Furthermore, knowing that the dynamical critical exponent of the Sweeny dynamics can not exceed the Propp-Wilson exponent, this applies naturally also to the Sweeny dynamics and hence promotes the use of local algorithms in three dimensions. This gives further motivation for a thorough study and comparison of the Sweeny and Chayes-Machta dynamics in three dimensions going beyond the analysis of [15, 88].

[1] On more mathematical grounds this can be used, utilising Chebyshev's inequality, to derive a concentration result for fluctuations around the expectation of the same order of magnitude.

The probably most interesting observation with regards to the coupling from the past algorithm is the appearance of universal Gumbel fluctuations in the coupling time τ. We showed that as long as the model is not close to the location of a discontinuous phase transition, the coupling time fluctuations are of Gumbel-type, in the sense that the distribution of a rescaled coupling time approaches the distribution of a standard Gumbel random variable for increasing system sizes. We provided rigorous support of this observation via the convergence of a rescaled coupling time distribution to a Gumbel distribution for the coupling process on the cycle graph. Furthermore, at a discontinuous phase transition the fluctuations seem to be of the Frèchet type, yet a more detailed analysis is needed, which should also consider the tricritical case, for which we could not make any conclusive statement. We also found the Gumbel law in the coupling distribution for the Ising-spin heat-bath dynamics as long as the model possesses the strong spatial mixing property [48, 140], which for the square lattice is equivalent to be above the critical temperature. As a future project, we plan to study whether the weaker notions of spatial mixing given for the random-cluster model on the square-lattice [142, 128, 28] can be related to rapid mixing and Gumbel fluctuations.

Another promising application of the coupling from the past algorithm is the ferromagnetic random bond Potts model in the Fortuin-Kasteleyn representation [172, 173, 174]. The standard problem which one faces when concerned with dynamics in disordered systems is the rugged probability or energy landscape. This introduces many time scales into the dynamics and the study of autocorrelation times and hence relaxation times become are a priori probabilistic, see for instance [175]. This clearly complicates the study of relaxation and correlation, of utmost importance for a thorough application of the MCMC method. We have seen that the coupling from the past algorithm self-determines the required (random) running time τ to produce an independent sample. Thus, this approach can in principle circumvent the problem of determining the autocorrelation time in order to guarantee also relaxation. Still, the study of τ in this setting is very intriguing in its own right, as the presence of disorder raises many question in particular for instances where a first order phase transition is "softened" to a continuous phase transition. Does the Gumbel law persist with disorder, even for instances where the clean model would undergo a first order phase transition? Can we generalise the Propp-Wilson inequalities relating the relaxation or autocorrelation time to the expected coupling time to the disordered setting? Do we have an analogue of the Li-Sokal bound, that is are there rigorous arguments dictating a critical slowing down? Does critical speeding up persist? In fact, we started to analyse certain aspects for the particular instance of the random bond Potts model with $q = 8$ and bivariate bond-coupling distribution on the square lattice at the respective critical temperature. One of the main open issues we would like to address refers to the universality aspects of the model, namely the value of the critical exponent ν of the correlation length and its evolution (or not) with the number of states q in the originally first order regime of the model [176, 177, 178, 179, 174, 180]. We will present detailed dynamical and static results in the future elsewhere.

Appendix A

Notation

We use standard O-notation, that is $f(n) \leq O(g(n))$ means there exists some constant $C < \infty$ such that $f(n) \leq Cg(n)$ for any $n \geq 1$. Additionally the "small-o" notation is also occasionally used in this thesis, that is write $f(n) = o(g(n))$ when $\lim_{n\to\infty} f(n)/g(n) = 0$. Furthermore, we denote by $f(n) \asymp g(n)$ that f and g have the same order of magnitude, that is there exists some constant $C < \infty$ such that $g(n)/C \leq f(n) \leq Cg(n)$ for all $n \geq 1$. The symbol $\sim$ denotes asymptotic equivalence, that is $f(n) \sim g(n)$ if $\lim_{n\to\infty} f(n)/g(n) = 1$. Lastly, we also use what is commonly referred to as "logarithmic equivalence", that is we write $f(n) \approx g(n)$ whenever $\lim_{n\to\infty} \log f(n)/\log g(n) = 1$. We use this notation in particular for quantities with power-laws, for instance for scaling ansätze at criticality, when we do not want (or can) consider constants and/or multiplicative "corrections".

Appendix B

The random-cluster model on the cycle graph

Let $\mathbb{Z}_n$ be the n-cycle, hence n corresponds to the number of vertices and edges, respectively.

B.1 Connectivity function

Let $0 \leq i < j < n$ be two distinct vertices. Our goal is to determine $\mathbb{P}[i \leftrightarrow j]$, that is the probability that i is connected to j. Recap, the partition function of the random-cluster model with edge dependent couplings $\underline{v} = (v_e)_{e \in E}$ on a finite graph $G = (V, E)$ is given by:

$$Z_G(\underline{v}, q) \equiv \sum_{A \subseteq E} q^{k(A)} \prod_{e \in E} v_e^{\omega(e,A)},$$

where $\omega(e, A) = 1$ iff $e \in A$ and $\omega(e, A) = 0$ otherwise. This can be re-written using the circuit rank $c(A)$ and $k(A) = c(A) + |V| - |A|$:

$$Z_G(\underline{v}, q) = \sum_{A \subseteq E} q^{c(A)} \prod_{e \in E} \left(v_e^{\omega(e,A)} q^{1-\omega(e,A)} \right).$$

From now on we set $G = \mathbb{Z}_n$. In this case the cycle space is 0 dimensional for any $A \neq E$ and one-dimensional for $A = E$, hence $c(A) = 1$ iff $A = E$ and $c(A) = 0$ otherwise. After straightforward algebra we obtain:

$$Z_{\mathbb{Z}_n}(\underline{v}, q) = \prod_{e \in E} (q + v_e) + (q - 1) \prod_{e \in E} v_e.$$

Let $\mathcal{P}_{i,j}$ be the set of edges of the clockwise path between i and j and $\overline{\mathcal{P}}_{i,j}$ the one in anti-clockwise orientation. Hence we have by the inclusion-exclusion principle:

$$\mathbb{P}[i \leftrightarrow j] = \mathbb{P}\left[\mathcal{P}_{i,j} \subseteq A\right] + \mathbb{P}\left[\overline{\mathcal{P}}_{i,j} \subseteq A\right] - \mathbb{P}\left[\left(\mathcal{P}_{i,j} \subseteq A\right\} \wedge \{\overline{\mathcal{P}}_{i,j} \subseteq A\}\right]\Big($$

Notice that $\mathbb{P}\Big[\{\mathcal{P}_{i,j} \subseteq A\} \wedge \{\overline{\mathcal{P}}_{i,j} \subseteq A\}\Big] = \mathbb{P}[A = E]$ and hence

$$\mathbb{P}\Big[\{\mathcal{P}_{i,j} \subseteq A\} \wedge \{\overline{\mathcal{P}}_{i,j} \subseteq A\}\Big] = \frac{q}{Z_{\mathbb{Z}_n}(\underline{v}, q)} \prod_{e \in E} v_e.$$

Furthermore we have:

$$\begin{aligned}
\mathbb{P}\Big[\mathcal{P}_{i,j} \subseteq A\Big] &= \frac{1}{Z_{\mathbb{Z}_n}(\underline{v}, q)} \left(\prod_{e \in \mathcal{P}_{i,j}} v_e \right) \left(\left(\prod_{e \in \mathcal{P}_{i,j}} \frac{\partial}{\partial v_e} \right) Z_{\mathbb{Z}_n}(\underline{v}, q) \right) \\
&= \frac{1}{Z_{\mathbb{Z}_n}(\underline{v}, q)} \left[\left(\prod_{e \in E - \mathcal{P}_{i,j}} q + v_e \right) \left(\prod_{e \in \mathcal{P}_{i,j}} v_e \right) + (q-1) \prod_{e \in E} v_e \right]
\end{aligned}$$

Analogous for $\mathbb{P}\Big[\overline{\mathcal{P}}_{i,j} \subseteq A\Big]$. Now we let $v_e = v$ for all $e \in E$ and obtain after some algebra:

$$\mathbb{P}[i \leftrightarrow j] = \frac{\rho^{j-i} + \rho^n \rho^{i-j} + (q-2)\rho^n}{1 + (q-1)\rho^n},$$

where we defined

$$\rho \equiv \rho(v, q) \equiv \frac{v}{q+v} < 1.$$

Notice that $\rho = p/(p + (1-p)q \equiv \tilde{p}(p, q)$ in the alternative parametrisation. Recall that $\tilde{p}(p, q)$ equals the probability to insert a pivotal edge in the Glauber dynamics for the random-cluster model. For $v, q > 0$ and we obtain for $n \to \infty$:

$$\mathbb{P}[i \leftrightarrow j] \to e^{-|i-j| \log\left(1 + \frac{q}{v}\right)}$$

B.2 Consistency check: Ising model

Baxter [22] shows that for the Ising model on $\mathbb{Z}_n$ in the limit $n \to \infty$ with inverse temperature $\beta > 0$:

$$\mathbb{E}[\sigma_i \sigma_j] = (\tanh(\beta))^{j-i},$$

where $\sigma_i, \sigma_j \in \{-1, +1\}$ and $i \leq j$. Notice that for $v = e^{2\beta} - 1$ one has:

$$\mathbb{E}[\sigma_i \sigma_j] = \mathbb{P}[i \leftrightarrow j].$$

It is easy to check that the $n \to \infty$ limit of $\mathbb{P}[i \leftrightarrow j]$ for $i \leq j$ and $v = e^{2\beta} - 1$ recovers the known result for the Ising model on $\mathbb{Z}_n$.

Appendix C

Asymptotic equivalence and bounds

Here we prove the following, probably well known, result for the sake of completeness:

Lemma C.0.1. *For two positive sequences a_n, b_n such that $a_n \sim b_n$, that is $a_n/b_n \to 1$ for $n \to \infty$, one has that the following implication is true uniformly in n:*

$$C \leq e^{-a_n} \Rightarrow \exists K_1 \geq 1, K_2 \in (0,1) \text{ such that } C \leq K_1 e^{-K_2 b_n}.$$

Proof. Recall, the fact that $a_n \sim b_n$ means that for any $\epsilon > 0$ there exists an integer N_ϵ such that $|a_n/b_n - 1| \leq \epsilon$ for any $n \geq N_\epsilon$. Thus in particular if we choose any $0 < \rho < 1$ we have that there is an integer N_ρ such that $a_n/b_n \geq \rho$ whenever $n \geq N_\rho$ (here we used positivity of a_n and b_n). Therefore we have for any such $n \geq N_\rho$:

$$C \leq e^{-a_n} = e^{-b_n a_n/b_n} \leq e^{-b_n \rho}.$$

It remains to prove the implication for $n < N_\rho$. Here we observe

$$C \leq e^{-a_n} = e^{-b_n \rho} e^{b_n \rho - a_n} \leq e^{-b_n \rho} K.$$

Where we defined $K \equiv \max_{n<N_\rho} \left\{ e^{b_n \rho - a_n} \right\}$(Finally, we choose $K_1 = \max\{1, K\}$ and $K_2 = \rho$. ■

Bibliography

[1] Kurt Binder and Dieter Heermann. *Monte Carlo simulation in statistical physics: an introduction*. Springer Science & Business Media, 2010.

[2] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.

[3] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Rev. Mod. Phys.*, 49:435–479, Jul 1977.

[4] F. Y. Wu. The Potts model. *Rev. Mod. Phys.*, 54:235–268, Jan 1982.

[5] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, Jan 1987.

[6] L. Chayes and J. Machta. Graphical representations and cluster algorithms I. Discrete spin systems. *Physica A: Statistical Mechanics and its Applications*, 239(4):542–601, 1997.

[7] L. Chayes and J. Machta. Graphical representations and cluster algorithms II. *Physica A: Statistical Mechanics and its Applications*, 254(3):477–516, 1998.

[8] Cornelius Marius Fortuin and Piet W. Kasteleyn. On the random-cluster model: I. Introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[9] Cornelis Marius Fortuin. On the random-cluster model II. The percolation model. *Physica*, 58(3):393–418, 1972.

[10] Cornelis Marius Fortuin. On the random-cluster model: III. The simple random-cluster model. *Physica*, 59(4):545–570, 1972.

[11] Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D*, 38(6):2009, 1988.

[12] Antonio Coniglio. Cluster structure near the percolation threshold. *Journal of Physics A: Mathematical and General*, 15(12):3829, 1982.

[13] Antonio Coniglio. Fractal structure of Ising and Potts clusters: exact results. *Physical Review Letters*, 62(26):3054, 1989.

[14] Youjin Deng, Henk W.J. Blöte, and Benard Nienhuis. Geometric properties of two-dimensional critical and tricritical Potts models. *Physical Review E*, 69(2):026123, 2004.

[15] Youjin Deng, Wei Zhang, Timothy M. Garoni, Alan D. Sokal, and Andrea Sportiello. Some geometric critical exponents for percolation and the random-cluster model. *Physical Review E*, 81(2):020102, 2010.

[16] Oded Schramm. Scaling limits of loop-erased random walks and uniform spanning trees. *Israel Journal of Mathematics*, 118(1):221–288, 2000.

[17] Steffen Rohde and Oded Schramm. Basic properties of SLE. In *Selected Works of Oded Schramm*, pages 989–1030. Springer, 2011.

[18] Wouter Kager and Bernard Nienhuis. A guide to stochastic Löwner evolution and its applications. *Journal of statistical physics*, 115(5-6):1149–1229, 2004.

[19] John Cardy. SLE for theoretical physicists. *Annals of Physics*, 318(1):81–118, 2005.

[20] Mark Sweeny. Monte carlo study of weighted percolation clusters relevant to the Potts models. *Physical Review B*, 27(7):4445, 1983.

[21] Youjin Deng, Timothy M. Garoni, and Alan D. Sokal. Critical speeding-up in the local dynamics of the random-cluster model. *Physical Review Letters*, 98(23):230602, 2007.

[22] Rodney J. Baxter. *Exactly solved models in statistical mechanics*. Dover Publications, 2008.

[23] Mark E. J. Newman and Robert M. Ziff. Fast Monte Carlo algorithm for site or bond percolation. *Physical Review E*, 64(1):016706, 2001.

[24] Jacob Holm, Kristian De Lichtenberg, and Mikkel Thorup. Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *Journal of the ACM (JACM)*, 48(4):723–760, 2001.

[25] Ariel Yadin. Percolation (Lecture notes). `http://www.math.bgu.ac.il/~yadina/teaching/percolation/perc.html`, 2013. [Online; accessed 19-March-2015].

[26] Geoffrey R. Grimmett. *Percolation (Grundlehren der mathematischen Wissenschaften)*. Springer: Berlin, Germany, 2010.

[27] Geoffrey R. Grimmett. *The random-cluster model*, volume 333. Springer Science & Business Media, 2006.

[28] 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, 2012.

[29] Mark F. Gyure and Boyd F. Edwards. Fragmentation of percolation clusters at the percolation threshold. *Physical Review Letters*, 68(17):2692, 1992.

[30] Mookyung Cheon, Muyoung Heo, Iksoo Chang, and Dietrich Stauffer. Fragmentation of percolation clusters in general dimensions. *Physical Review E*, 59(5):R4733, 1999.

[31] David Aldous. Probability distributions on cladograms. In *Random Discrete Structures*, pages 1–18. Springer, 1996.

[32] Olle Häggström and Karin Nelander. Exact sampling from anti-monotone systems. *Statistica Neerlandica*, 52(3):360–380, 1998.

[33] James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, 1996.

[34] Mario Ullrich. *Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions*. PhD thesis, Friedrich-Schiller-Universität Jena, 2012.

[35] François Jaeger, Dirk L. Vertigan, and Dominic J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 108, pages 35–53. Cambridge University Press, 1990.

[36] Christopher David Godsil, Gordon Royle, and C. D. Godsil. Algebraic graph theory. 2001.

[37] DL Vertigan and Dominic JA Welsh. The computational complexity of the tutte plane: the bipartite case. *Combinatorics, Probability and Computing*, 1(02):181–187, 1992.

[38] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[39] Neal Madras and Gordon Slade. *The self-avoiding walk*. Springer, 2012.

[40] Claude Kipnis and S. R. Srinivasa Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[41] Charles J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

[42] David Aldous and Jim Fill. Reversible Markov chains and random walks on graphs, 2002.

[43] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 2012.

[44] Alan D. Sokal. Monte Carlo methods in statistical mechanics. 1989.

[45] Martin Dyer, Catherine Greenhill, and Mario Ullrich. Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra and its Applications*, 454:57–71, 2014.

[46] D. Stauffer and A. Aharony. *Introduction to Percolation Theory*. Taylor & Francis, London, second edition, 1994.

[47] Alan Gibbons. *Algorithmic graph theory*. Cambridge University Press, 1985.

[48] Eyal Lubetzky and Allan Sly. Critical Ising on the square lattice mixes in polynomial time. *Communications in Mathematical Physics*, 313(3):815–836, 2012.

[49] Xiao-Jian Li and Alan D. Sokal. Rigorous lower bound on the dynamic critical exponents of the Swendsen-Wang algorithm. *Physical Review Letters*, 63(8):827, 1989.

[50] 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, 2011.

[51] B. Nienhuis. Coulomb gas formulation of two-dimensional phase transitions. In C. Domb and J. L. Lebowitz, editors, *Phase Transitions and Critical Phenomena*, volume 11, page 1. Academic Press, London, 1987.

[52] Antonio Blanca and Alistair Sinclair. Dynamics for the mean-field random-cluster model. *arXiv preprint arXiv:1412.6180*, 2014.

[53] Eren Metin Elçi and Martin Weigel. Efficient simulation of the random-cluster model. *Physical Review E*, 88(3):033303, 2013.

[54] Jens M. Schmidt. A simple test on 2-vertex-and 2-edge-connectivity. *Information Processing Letters*, 113(7):241–244, 2013.

[55] Xiao Xu, Junfeng Wang, Zongzheng Zhou, Timothy M. Garoni, and Youjin Deng. Geometric structure of percolation clusters. *Physical Review E*, 89(1):012120, 2014.

[56] Xuan-Wen Liu, Youjin Deng, and Jesper Lykke Jacobsen. Recursive percolation. *arXiv preprint arXiv:1410.3603*, 2014.

[57] Eren Metin Elçi and Martin Weigel. Dynamic connectivity algorithms for Monte Carlo simulations of the random-cluster model. In *Journal of Physics: Conference Series*, volume 510, page 012013. IOP Publishing, 2014.

[58] Jakub Łącki and Piotr Sankowski. Optimal decremental connectivity in planar graphs. *arXiv preprint arXiv:1409.7240*, 2014.

[59] Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: A simple proof. *Random Structures & Algorithms*, 43(2):131–138, 2013.

[60] Jeff Erickson. Algorithms, etc. http://web.engr.illinois.edu/~jeffe/teaching/algorithms/, February 2015.

[61] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.

[62] Robert Endre Tarjan. Amortized computational complexity. *SIAM Journal on Algebraic Discrete Methods*, 6(2):306–318, 1985.

[63] Daniel Dominic Sleator and Robert Endre Tarjan. Self-adjusting binary search trees. *Journal of the ACM (JACM)*, 32(3):652–686, 1985.

[64] Robert Endre Tarjan. Efficiency of a good but not linear set union algorithm. *Journal of the ACM (JACM)*, 22(2):215–225, 1975.

[65] Raimund Seidel and Micha Sharir. Top-down analysis of path compression. *SIAM Journal on Computing*, 34(3):515–525, 2005.

[66] Joanna A. Ellis-Monaghan and Iain Moffatt. *Graphs on surfaces: dualities, polynomials, and knots*. Springer, 2013.

[67] Raj Iyer, David Karger, Hariharan Rahul, and Mikkel Thorup. An experimental study of polylogarithmic, fully dynamic, connectivity algorithms. *Journal of Experimental Algorithmics (JEA)*, 6:4, 2001.

[68] Youjin Deng, Timothy M. Garoni, and Alan D. Sokal. Ferromagnetic phase transition for the spanning-forest model ($q \to 0$ limit of the Potts model) in three or more dimensions. *Physical Review Letters*, 98(3):030602, 2007.

[69] Alan D. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. *Surveys in combinatorics*, 327:173–226, 2005.

[70] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2):146–160, 1972.

[71] John Hopcroft and Robert Tarjan. Efficient planarity testing. *Journal of the ACM (JACM)*, 21(4):549–568, 1974.

[72] R. Endre Tarjan. A note on finding the bridges of a graph. *Information Processing Letters*, 2(6):160–161, 1974.

[73] Richard M. Karp and Robert Endre Tarjan. Linear expected-time algorithms for connectivity problems. In *Proceedings of the twelfth annual ACM symposium on Theory of computing*, pages 368–377. ACM, 1980.

[74] Stanislav Smirnov. Critical percolation in the plane: Conformal invariance, Cardy's formula, scaling limits. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 333(3):239–244, 2001.

[75] Arthur E. Ferdinand and Michael E. Fisher. Bounded and inhomogeneous Ising models. I. Specific-heat anomaly of a finite lattice. *Physical Review*, 185(2):832, 1969.

[76] Sergio Caracciolo, Andrea Pelissetto, and Alan D. Sokal. Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *Journal of Statistical Physics*, 60(1-2):1–53, 1990.

[77] Peter Grassberger. Numerical studies of critical percolation in three dimensions. *J. Phys. A*, 25(22):5867, 1992.

[78] Peter Grassberger. Pair connectedness and shortest path scaling in critical percolation. *J. Phys. A*, 32(35):6233, 1999.

[79] Peter Grassberger. Spreading and backbone dimensions of 2D percolation. *J. Phys. A*, 25(21):5475, 1992.

[80] Robert M. Ziff. Exact critical exponent for the shortest-path scaling function in percolation. *J. Phys. A*, 32(43):L457–L459, 1999.

[81] John Cardy. The number of incipient spanning clusters in two-dimensional percolation. *J. Phys. A*, 31(5):L105, 1998.

[82] Stanislav Smirnov and Wendelin Werner. Critical exponents for two-dimensional percolation. *Math. Res. Lett.*, 8(6):729–744, 2001.

[83] Zongzheng Zhou, Ji Yang, Youjin Deng, and Robert M. Ziff. Shortest-path fractal dimension for percolation in two and three dimensions. *Phys. Rev. E*, 86:061101, December 2012.

[84] Gabriele Farina and Luigi Laura. Dynamic subtrees queries revised: the depth first tour tree. *arXiv preprint arXiv:1502.05292*, 2015.

[85] Monika R. Henzinger and Valerie King. Randomized fully dynamic graph algorithms with polylogarithmic time per operation. *Journal of the ACM (JACM)*, 46(4):502–516, 1999.

[86] Yun Long, Asaf Nachmias, and Yuval Peres. Mixing time power laws at criticality. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 205–214. IEEE, 2007.

[87] Peter Grassberger. Damage spreading and critical exponents for "model A" Ising dynamics. *Physica A: Statistical Mechanics and its Applications*, 214(4):547–559, 1995.

[88] Youjin Deng, Timothy M. Garoni, Jonathan Machta, Giovanni Ossola, Marco Polin, and Alan D. Sokal. Critical behavior of the Chayes-Machta-Swendsen-Wang dynamics. *Physical Review Letters*, 99(5):055701, 2007.

[89] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[90] Christophe Garban and Jeffrey E. Steif. *Noise sensitivity of Boolean functions and percolation*, volume 5. Cambridge University Press, 2014.

[91] Geoffrey Grimmett. *Probability on graphs: random processes on graphs and lattices*, volume 1. Cambridge University Press, 2010.

[92] Şerban Nacu. Glauber dynamics on the cycle is monotone. *Probability theory and related fields*, 127(2):177–185, 2003.

[93] Jesús Salas and Alan D. Sokal. Dynamic critical behavior of the Swendsen-Wang algorithm: The two-dimensional three-state Potts model revisited. *Journal of Statistical Physics*, 87(1-2):1–36, 1997.

[94] Jesus Salas and Alan D. Sokal. Logarithmic corrections and finite-size scaling in the two-dimensional 4-state Potts model. *Journal of Statistical Physics*, 88(3-4):567–615, 1997.

[95] Junfeng Wang, Zongzheng Zhou, Wei Zhang, Timothy M. Garoni, and Youjin Deng. Bond and site percolation in three dimensions. *Physical Review E*, 87(5):052107, 2013.

[96] Youjin Deng, Timothy M. Garoni, and Sokal A. D. Private communication, 2014.

[97] Elçi Eren Metin. Computational complexity of a dynamic connectivity algorithm for the mean-field dynamics of the random-cluster model. In preparation.

[98] M. Caselle, F. Gliozzi, and S. Necco. Thermal operators and cluster topology in the q-state Potts model. *Journal of Physics A: Mathematical and General*, 34(3):351, 2001.

[99] Gil Kalai and Shmuel Safra. Threshold phenomena and influence: perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, pages 25–60, 2006.

[100] Hugo Duminil-Copin and Vincent Tassion. A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$. *arXiv preprint arXiv:1502.03051*, 2015.

[101] J. W. Essam. Connectedness and connectivity in percolation theory. *North-Holland Mathematics Studies*, 144:41–57, 1987.

[102] Geoffrey Grimmett. The stochastic random-cluster process and the uniqueness of random-cluster measures. *The Annals of Probability*, pages 1461–1510, 1995.

[103] Robert M. Ziff, Christian D. Lorenz, and Peter Kleban. Shape-dependent universality in percolation. *Physica A: Statistical Mechanics and its Applications*, 266(1):17–26, 1999.

[104] Haru T. Pinson. Critical percolation on the torus. *Journal of Statistical Physics*, 75(5):1167–1177, 1994.

[105] Robert P. Langlands, Claude Pichet, Phillippe Pouliot, and Yvan Saint-Aubin. On the universality of crossing probabilities in two-dimensional percolation. *Journal of Statistical Physics*, 67(3-4):553–574, 1992.

[106] Noga Alon. Economical elimination of cycles in the torus. *Combinatorics, Probability and Computing*, 18(05):619–627, 2009.

[107] Christophe Garban, Gábor Pete, and Oded Schramm. The Fourier spectrum of critical percolation. *Acta Mathematica*, 205(1):19–104, 2010.

[108] Oded Schramm and Jeffrey E. Steif. Quantitative noise sensitivity and exceptional times for percolation. In *Selected Works of Oded Schramm*, pages 391–444. Springer, 2011.

[109] Youjin Deng, Henk W. J. Blöte, and Bernard Nienhuis. Backbone exponents of the two-dimensional q-state Potts model: A Monte Carlo investigation. *Physical Review E*, 69(2):026114, 2004.

[110] Romain Vasseur, Jesper Lykke Jacobsen, and Hubert Saleur. Logarithmic observables in critical percolation. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(07):L07001, 2012.

[111] Youjin Deng and Xianqing Yang. Finite-size scaling of energylike quantities in percolation. *Physical Review E*, 73(6):066116, 2006.

[112] Peter Young. Everything you wanted to know about data analysis and fitting but were afraid to ask. *arXiv preprint arXiv:1210.3781*, 2012.

[113] Vladimir Privman. Finite-size scaling theory. *Finite Size Scaling and Numerical Simulation of Statistical Systems*, 1, 1990.

[114] Hao Hu, Henk W. J. Blöte, Robert M. Ziff, and Youjin Deng. Short-range correlations in percolation at criticality. *Physical Review E*, 90(4):042106, 2014.

[115] Boyd F. Edwards, Mark F. Gyure, and M. Ferer. Exact enumeration and scaling for fragmentation of percolation clusters. *Physical Review A*, 46(10):6252, 1992.

[116] Sang Bub Lee. Multiple fragmentation of critical continuum percolation clusters. *Physica A: Statistical Mechanics and its Applications*, 393:480–488, 2014.

[117] J. F. Gouyet. Invasion noise during drainage in porous media. *Physica A: Statistical Mechanics and its Applications*, 168(1):581–591, 1990.

[118] S. Roux and Etienne Guyon. Temporal development of invasion percolation. *Journal of Physics A: Mathematical and General*, 22(17):3693, 1989.

[119] Hubert Saleur and Bertrand Duplantier. Exact determination of the percolation hull exponent in two dimensions. *Physical Review Letters*, 58(22):2325, 1987.

[120] Eren Metin Elçi, Martin Weigel, and Nikolaos G. Fytas. Fragmentation of fractal random structures. *Phys. Rev. Lett.*, 114:115701, Mar 2015.

[121] S. S. Manna, Deepak Dhar, and S. N. Majumdar. Spanning trees in two dimensions. *Physical Review A*, 46(8):R4471, 1992.

[122] Songsong Wang, Yuan Yang, Wanzhou Zhang, and Chengxiang Ding. Percolation of the site random-cluster model by Monte Carlo method. *arXiv preprint arXiv:1411.4408*, 2014.

[123] Olle Häggström. *Finite Markov chains and algorithmic applications*, volume 52. Cambridge University Press, 2002.

[124] Mark Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. In *Probabilistic methods for algorithmic discrete mathematics*, pages 116–165. Springer, 1998.

[125] David Bruce Wilson. How to couple from the past using a read-once source of randomness. *Random Structures & Algorithms*, 16(1):85–113, 2000.

[126] James Allen Fill. An interruptible algorithm for perfect sampling via markov chains. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 688–695. ACM, 1997.

[127] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[128] Hugo Duminil-Copin, Vladas Sidoravicius, and Vincent Tassion. Continuity of the phase transition for planar potts models with $1 \leq q \leq 4$. *Preprint*, 50, 2013.

[129] Bernd A. Berg and Thomas Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68(1):9, 1992.

[130] Bernd A. Berg and Thomas Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267(2):249–253, 1991.

[131] DP Landau, Shan-Ho Tsai, and M Exler. A new approach to monte carlo simulations in statistical physics: Wang-landau sampling. *American Journal of Physics*, 72(10):1294–1302, 2004.

[132] Youjin Deng and Henk W. J. Blöte. Simultaneous analysis of several models in the three-dimensional Ising universality class. *Physical Review E*, 68(3):036125, 2003.

[133] Paul Erdős and Alfréd Rényi. On a classical problem of probability theory. *Publ. Math. Inst. Hung. Acad. Sci.*, 6:215–220, 1961.

[134] Emil Julius Gumbel. *Statistics of extremes*. Courier Corporation, 2012.

[135] Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613, 1981.

[136] Michele Castellana, Aurélien Decelle, and Elia Zarinelli. Extreme value statistics distributions in spin glasses. *Physical Review Letters*, 107(27):275701, 2011.

[137] David Belius. Gumbel fluctuations for cover times in the discrete torus. *Probability Theory and Related Fields*, 157(3-4):635–689, 2013.

[138] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[139] Eric Bertin and Maxime Clusel. Generalized extreme value statistics and sum of correlated variables. *Journal of Physics A: Mathematical and General*, 39(24):7607, 2006.

[140] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. *Communications in Mathematical Physics*, 161(3):447–486, 1994.

[141] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24(4):461–479, 2004.

[142] Kenneth S. Alexander. On weak mixing in lattice models. *Probability theory and related fields*, 110(4):441–471, 1998.

[143] Elçi E. M., Garoni T., Collevecchio A., and Weigel M. Gumbel law for the coupling time distribution of the heat-bath chain for the random-cluster model. In preparation.

[144] Mario Ullrich. Swendsen-Wang is faster than single-bond dynamics. *SIAM Journal on Discrete Mathematics*, 28(1):37–48, 2014.

[145] Nicolas Destainville. Flip dynamics in octagonal rhombus tiling sets. *Physical Review Letters*, 88(3):030601, 2002.

[146] Svante Janson. Tail bounds for sums of geometric and exponential variables. http://www2.math.uu.se/ svante/papers/sjN14.pdf.

[147] D. L. Turcotte. Fractals and fragmentation. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 91(B2):1921–1926, 1986.

[148] Brian R. Lawn. *Fracture of brittle solids*. Cambridge University Press, 1993.

[149] K. C. Chase, P. Bhattacharyya, and A. Z. Mekjian. Randomly broken nuclei and disordered systems. *Physical Review C*, 57(2):822, 1998.

[150] S. Redner. Fragmentation. *Statistical models for the fracture of disordered media*, pages 321–348, 1990.

[151] J. A. Åström. Statistical models of brittle fragmentation. *Advances in Physics*, 55(3-4):247–278, 2006.

[152] Lene Oddershede, Peter Dimon, and Jakob Bohr. Self-organized criticality in fragmenting. *Physical Review Letters*, 71(19):3107, 1993.

[153] Jan A. Åström, Finn Ouchterlony, R. P. Linna, and J. Timonen. Universal dynamic fragmentation in d dimensions. *Physical Review Letters*, 92(24):245506, 2004.

[154] Laércio Dias and Fernando Parisio. Minimal fragmentation of regular polygonal plates. *Physical Review E*, 90(3):032405, 2014.

[155] David S. Dean and Satya N. Majumdar. Phase transition in a random fragmentation problem with applications to computer science. *Journal of Physics A: Mathematical and General*, 35(32):L501, 2002.

[156] Jie Huang, Xiaopei Guo, Boyd F. Edwards, and Arnold D. Levine. Cut-off model and exact general solutions for fragmentation with mass loss. *Journal of Physics A: Mathematical and General*, 29(23):7377, 1996.

[157] P. L. Krapivsky, E. Ben-Naim, and I. Grosse. Stable distributions in stochastic fragmentation. *Journal of Physics A: Mathematical and General*, 37(8):2863, 2004.

[158] Ziya Kalay and Eli Ben-Naim. Fragmentation of random trees. *Journal of Physics A: Mathematical and Theoretical*, 48(4):045001, 2015.

[159] Robert Sedgewick and Philippe Flajolet. *An introduction to the analysis of algorithms*. Addison-Wesley, 2013.

[160] Satya N. Majumdar, David S. Dean, and Paul L. Krapivsky. Understanding search trees via statistical physics. *Pramana*, 64(6):1175–1189, 2005.

[161] Alois Panholzer. Cutting down very simple trees. *Quaestiones Mathematicae*, 29(2):211–227, 2006.

[162] Robert M. Ziff and E. D. McGrady. The kinetics of cluster fragmentation and depolymerisation. *Journal of Physics A: Mathematical and General*, 18(15):3027, 1985.

[163] Pavel L Krapivsky, Sidney Redner, and Eli Ben-Naim. A kinetic view of statistical physics. 2010.

[164] David J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, pages 3–48, 1999.

[165] Michael D. Brennan and Richard Durrett. Splitting intervals II: Limit laws for lengths. *Probability theory and related fields*, 75(1):109–127, 1987.

[166] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 296–303. ACM, 1996.

[167] Mark E. J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[168] Elçi E. M., Weigel M., and Fytas N. G. Bridges in the random-cluster model. In preparation.

[169] N. A. M. Araújo, K. J. Schrenk, J. S. Andrade Jr., and H. J. Herrmann. Bridge percolation. *arXiv preprint arXiv:1103.3256*, 2011.

[170] Wolfhard Janke and Adriaan MJ Schakel. Geometrical vs. Fortuin-Kasteleyn clusters in the two-dimensional q-state Potts model. *Nuclear Physics B*, 700(1):385–406, 2004.

[171] Muyoung Heo, Mookyung Cheon, Iksoo Chang, and Dietrich Stauffer. Ising cluster fragmentation at the critical point. *International Journal of Modern Physics C*, 10(06):1059–1063, 1999.

[172] Bertrand Berche and Christophe Chatelain. Phase transitions in two-dimensional random Potts models. *Order, Disorder and Criticality Advanced Problems of Phase Transition Theory, Yu. Holovatch (Ed.), World Scientific, Singapore*, pages 147–200, 2004.

[173] Yurij Holovatch. *Order, disorder and criticality: advanced problems of phase transition theory*. World Scientific, 2004.

[174] Christophe Chatelain. *Random and Out-of-Equilibrium Potts models*. PhD thesis, Université de Lorraine, 2012.

[175] Wolfhard Janke and Desmond A. Johnston. Non-self-averaging in autocorrelations for Potts models on quenched random gravity graphs. *Journal of Physics A: Mathematical and General*, 33(14):2653, 2000.

[176] John Cardy and Jesper Lykke Jacobsen. Critical behavior of random-bond Potts models. *Physical review letters*, 79(21):4063, 1997.

[177] Christophe Chatelain and Bertrand Berche. Finite-size scaling study of the surface and bulk critical behavior in the random-bond eight-state Potts model. *Physical review letters*, 80(8):1670, 1998.

[178] Gábor Palágyi, Christophe Chatelain, Bertrand Berche, and Ferenc Iglói. Boundary critical behaviour of two-dimensional random Potts models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(2):357–367, 2000.

[179] Jesper Lykke Jacobsen and Marco Picco. Large-q asymptotics of the random-bond Potts model. *Physical Review E*, 61(1):R13, 2000.

[180] S. Chen, Alan M. Ferrenberg, and D. P. Landau. Monte Carlo simulation of phase transitions in a two-dimensional random-bond Potts model. *Physical Review E*, 52(2):1377, 1995.

Certificate of Ethical Approval

Applicant:

Eren Elci

Project Title:

Computer simulation studies in statistical physics.

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

16 April 2015

Project Reference Number:

P32941