SHARP PHASE TRANSITIONS IN PERCOLATION MODELS

Dissertation an der Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München

---

THOMAS BEEKENKAMP

August 13, 2021

1. Gutachter: Prof. Dr. Markus Heydenreich
2. Gutachter: Prof. Dr. Daniel Rodrigues Valesin
Tag der mündlichen Prüfung: 19. November 2021.

ACKNOWLEDGMENTS

My thesis is the result of five years of life at the institute in Munich. It is impossible to say what it look like if it had not been written here. I have always felt at home at the institute, for which I can thank everyone both from the probability group as well as from other groups.

I also thank Daniel for taking the time to read my thesis and giving valuable feedback. This significanly improved several parts of the thesis. I am likewise grateful to Matija for several nice discussions on the orthant model which is the topic of Chapter 6.

Finally I thank Markus for giving me the freedom to find my own way with all faith and support.

ABSTRACT

In percolation models, vertices or edges are removed from a graph according to a particular probabilistic rule. The connectivity properties of the resulting graph are then of interest. The initial graph is typically taken to be transitive and infinite, for example $\mathbb{Z}^d$ with nearest neighbour edges. The classical example of such a model is Bernoulli bond percolation, in which an edge is removed from the graph with probability $1 - p$, and thus kept with probability p, independently for every edge. It is well known that this model exhibits a phase transition: for small values of p, there exist only finite clusters almost surely, while for large values of p, there exists an infinite cluster almost surely. In particular, there exists a critical point $p_c \in (0, 1)$ at which this transition occurs. Moreover, the phase transition is sharp: for $p < p_c$, the clusters are exponentially small.

This behaviour is not specific to Bernoulli percolation. Rather, it is a common theme in percolation models. Nevertheless, the original proofs of the sharpness of the phase transition were very specific for Bernoulli percolation, and they are not easily applied to models with dependencies. Recently however, a celebrated new proof was given by Duminil-Copin, Raoufi and Tassion, which is far more robust. It makes use of Boolean function theory, in particular the OSSS inequality for decision trees. In this thesis, we will explore this new technique, and apply it to three models: the contact process, the orthant model, and the corrupted compass model.

The application of the OSSS method to these models is far from straightforward, and some model-specific hurdles have to be overcome. As an instructive model with dependencies, we will start with the corrupted compass model, since the dependencies are relatively easily controlled in this model. This is in contrast to the contact process, where the dependencies of the model are rather elusive. We give a new proof of the sharpness of the phase transition at λ_c, the phase transition for the survival of the infection. We then investigate the percolation phase transition for the time-t-measure, and show that this transition is sharp as well. Furthermore, we investigate how this might be extended to the upper invariant measure for the contact process. Finally, we will examine the orthant model, which is quite different in nature, since it is a directed model in which there exists an infinite cluster across the entire parameter range. Still, we can speak of a phase transition in this model, and we will prove that it is sharp. The idiosyncratic nature of this model is reflected in the proof.

ZUSAMMENFASSUNG

In Perkolationsmodellen werden Knoten oder Kanten aus einem Graph nach einer Zufallsregel entfernt. Wir sind dann an der Zusammenhängingkeit des resultierenden Graphen interessiert. Der ursprüngliche Graph wird typischerweise unendlich und transitiv gewählt, zum Beispiel der Graph $\mathbb{Z}^2$ mit Nächster-Nachbar-Kanten. Das klassische Beispiel eines solchen Modells ist Bernoulli-Kantenperkolation, in dem eine Kante aus dem Graph mit Wahrscheinlichkeit $1 - p$ entfernt wird, und damit mit Wahrscheinlichkeit p behalten wird, unabhängig für jede Kante. Es ist allgemein bekannt, dass dieses Modell einen Phasenübergang aufweist: Für kleine Werte von p existieren fast sicher nur endliche Komponenten, während für große Werte von p fast sicher eine unendliche Kompentente existiert. Insbesondere gibt es einen kritischen Punkt $p_c \in (0, 1)$, an dem diesen Übergang passiert. Weiterhin ist dieser Phasenübergang scharf: Für $p < p_c$ sind alle Komponenten exponentiell klein.

Dieses Verhalten ist nicht spezifisch für Bernoulli-Perkolation, sondern ein gemeinsames Thema in Perkolationsmodellen. Nichtsdestotrotz sind die ursprüngliche Beweise für den scharfen Phasenübergang sehr spezifisch für Bernoulli-Perkolation, und lassen sich nicht leicht auf Modelle mit Abhängigkeiten übertragen. Vor kurzem aber wurde ein zelebrierter neuer Beweis durch Duminil-Copin, Raoufi und Tassion erbracht, welcher viel robuster ist. Dieser Beweis benutzt Boolesche Funktionentheorie, insbesondere die OSSS-Ungleichung für Entscheidungsbäume. In dieser Doktorarbeit werden wir diese neue Technik erfor-

schen, und sie auf drei Modelle anwenden: der Kontaktprozess, das Orthantmodell, und das Korrupter-Kompass-Modell.

Die Anwendung der OSSS-Methode auf diese Modelle ist alles andere als einfach, und einige modellspezifischen Hürden müssen überwunden werden. Zunächst werden wir als Beispiel für ein Modell mit Abhängingkeiten das Korrupter-Kompass-Modell betrachten, denn die Abhängigkeiten in diesem Modell sind relativ leicht zu beherrschen. Im Gegensatz zu dem Kontaktprozess, in dem die Abhängigkeiten des Modells schwer fassbar sind. Wir werden einen neuen Beweis für den scharfen Phasenübergang bei λ_c, den Phasenübergang für das Überleben der Infektion, präsentieren. Wir werden dann den Perkolationsphasenübergang für den Zeit-t-Maß untersuchen und zeigen, dass dieser Phasenübergang ebenso scharf ist. Außerdem werden wir untersuchen, wie dieses Resultat möglicherweise erweitert werden kann, um es für das obere invariante Maß zu zeigen. Schließlich werden wir das Orthantmodell erforschen, welches wesentlich anders beschaffen ist, da es sich dabei um ein gerichtetes Modell handelt, in dem über dem gesamten Parameterbereich eine unendliche Komponente existiert. Dennoch können wir in diesem Modell über einen Phasenübergang sprechen und werden weiterhin zeigen, dass dieser Phasenübergang scharf ist. Die Eigenart dieses Modells spiegelt sich im Beweis wider.

CONTENTS

1 INTRODUCTION

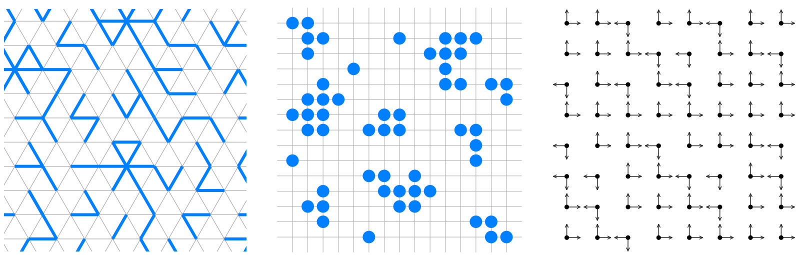
Percolation theory studies random networks, in particular their connectivity properties. The prototypical percolation model is Bernoulli percolation. In this model we take an infinite graph. (Mathematicians call networks graphs.) We then remove the edges from a graph in a random and independent fashion. If we remove enough edges, in other words if the probability that we remove an edge is high enough, the graph breaks apart in small finite components. On the other hand, if we do not remove too many edges, the structure of the original graph remains largely intact, although some holes might have appeared. This model where the original graph is the square lattice is shown in Figure 1.2 for several values of p, where p is the probability that we keep an edge in the graph.

Removing edges independently from a graph is the defining property of Bernoulli percolation, but there are many different ways to remove edges from a graph. Every different removal rule results in a different percolation model. Besides Bernoulli percolation, we focus our attention in this thesis to the corrupted compass model, the contact process, and the orthant model. These models are shown in Figure 1.1.

Seeing a realization of a particular percolation already gives some hint toward the probabilistic rule that determines which edges remain. In the corrupted compass model we choose for every vertex exactly one edge that is incident to it to remain. The model is so named, because we say that the compass at a vertex points in the direction of the edge that is chosen to remain. However, some compasses are corrupted. That means that for these vertices all incident edges are kept. Thus the higher the probability that a compass is corrupted, the more edges remain in the graph.

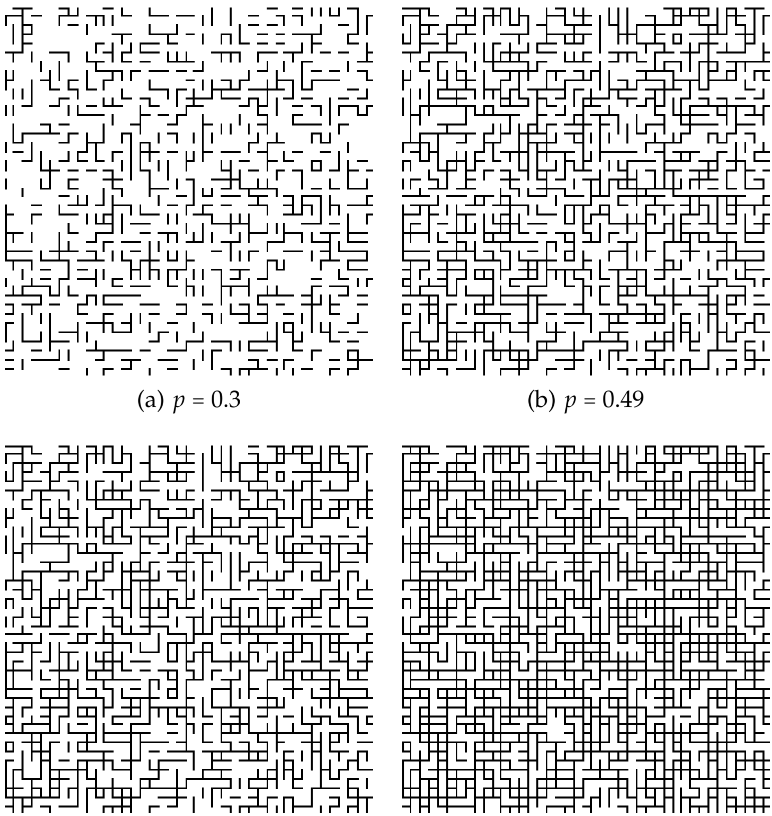
The contact process models the spread of infection on a graph. Contrary to the other models in this thesis, this model evolves over time. We start the process with some initial infected sites in the graph. These vertices can then spread the infection to their neighbours over time, or they become healthy. When a site is healthy it can no longer spread the infection, but it is possible that it gets (re-)infected by a neighbouring site. We then look at the configuration of infected sites after some fixed time has past, or even let the model evolve for an infinitely long time. Instead of removing edges from the graph, we then have healthy and infected sites, but we can ask the same question: how large are the clusters of infected sites?

The last model in this thesis is the orthant model. This model takes place on a directed graph: every edge has an orientation,



(a) The corrupted compass model. (b) The contact process. (c) The orthant model.

Figure 1.1: Several percolation models.



(a) $p = 0.3$ (b) $p = 0.49$

(c) $p = 0.51$ (d) $p = 0.7$

Figure 1.2: Bernoulli percolation on $\mathbb{Z}^2$ for different values of p [21].

which means that it can only be traversed in one direction. The model takes place on the cubic lattice with directed nearest neighbour edges. The probabilistic rule is the following: for each vertex we keep all edges pointing in the direction of the positive orthant ($\uparrow$, and $\rightarrow$ in two dimensions) with probability p, independently for each edge. Otherwise, so with probability $1 - p$, we take the edges pointing in the other direction ($\downarrow$, and $\leftarrow$ in two dimensions). An interesting feature of this model is that every vertex is in an infinite component, since the edge $\uparrow$ or $\leftarrow$ is available at every site, and hence it is not possible to get stuck while exploring the graph.

SHARP PHASE TRANSITIONS

A central theme that percolation models have in common is that of a phase transition, and moreover these transitions have a tendency to be sharp. This phenomenon is easiest explained for Bernoulli percolation. We already saw in Figure 1.2 that the graph consists of only finite clusters if p is small, whereas for large p there exists a giant infinite component. That means that at some value of p the behaviour of the model changes. We call this behavioural change the phase transition. For Bernoulli percolation on $\mathbb{Z}^2$ this transition occurs at $p_c = 1/2$. Looking closely at Figure 1.2, we can see that there is no path from top to bottom if $p = 0.49$. However, a slight increase of p pushes the model over the critical threshold, and we can see that a path from top to bottom exists for $p = 0.51$.

If the model is subcritical, that is if $p < p_c$, we can prove that the components of the graph after the removal of the edges are exponentially small. This property is called a sharp phase transition: exponentially small clusters in the subcritical parameter range, and the existence of an infinite component in the supercritical regime. Such a statement requires a mathematical proof, and in the case of Bernoulli percolation this was first given by Menshikov [45] and independently by Aizenman and Barsky [1] around the same time in the 80's. Bernoulli percolation is far from the only percolation model that exhibits a sharp phase transition. However proving this behaviour has long been out of reach for most other models, even after Menshikov and Aizenman and Barsky wrote their proofs, since these proofs use the independence of the edges in a crucial way. A notable exception is the Ising model, a model of ferromagnetism, for which Aizenman and Barsky also managed to prove a sharp phase transition [2]. However, a major breakthrough in the study of sharp phase transitions has recently been achieved.

THE OSSS INEQUALITY

In 2019 Duminil-Copin, Raoufi and Tassion published their proof of the sharp phase transition in the random-cluster model [15]. This model is a genaralization of Bernoulli percolation and the Ising model, which exhibits dependencies between edges that have long made a proof for its sharp phase transition out of reach. The proof of Duminil-Copin, Raoufi and Tassion uses a revolutionary approach using the OSSS inequality for Boolean functions.

Simply stated, a Boolean function is a function $f\colon \{-1,1\}^n \to \{-1,1\}$. The theory of Boolean functions has a long history, in particular in computer science. We call the entries of $(x_1,\ldots,x_n) \in \{-1,1\}^n$ variables or bits. An inequality from the field of Boolean functions is the OSSS inequality: it bounds the variance of a Boolean function by a sum over the influences of the variables, discounted by a factor equal to the probability that a decision tree queries that variable. A decision tree is an algorithm that determines the value of a Boolean function. It does this by sequentially revealing the values of $x_1, \ldots, x_n$, until it has gathered enough information to decide what the value of the Boolean function is. The variables $x_1, \ldots, x_n$ are typically not revealed in order. Instead, the decision tree decides which variable to reveal next based on what it has seen before. If a decision tree can determine the value of f efficiently, that is, by revealing a small amount of variables in the process, then the OSSS inequality is strong.

In the context of of percolation models, the variables are typically the state of the edges, and a useful Boolean function is the indicator of a connection event such as $0 \longleftrightarrow \partial\Lambda_n$: the event that 0 is connected to the boundary of a box of size n. The OSSS inequality can then be harvested to obtain a differential inequality from which the sharp phase transition can be deduced. However, this process can be quite involved, since a suitable decision tree needs to be constructed, which is typically rather specific to the model in question. Thus, even though we now have a new proof technique, it is still far from easy to prove the sharpness of phase transitions in percolation models.

After the paper for the random-cluster model, Duminil-Copin, Raoufi and Tassion published papers proving sharp phase transitions for Voronoi percolation and Boolean percolation [14, 16]. The proof technique was applied to the Widom-Rowlison model by Dereudre and Houdebert [12]. The level sets of the Gaussian free field were considered by [13, 46]. A different proof for the sharpness of the phase transition in the random-cluster model is given by Hutchroft [37]. This proof still uses the OSSS inequality, but applied to a different Boolean function and corresponding decision tree. We will study this proof in Chapter 3 in the case of Bernoulli percolation.

STRUCTURE OF THE THESIS

We start by introducing the Boolean function framework in Chapter 2. In this chapter we give several proofs of the OSSS inequality. Most notable is the proof of the inequality for Boolean functions depending on countably many variables, which has not previously appeared in the literature. In Chapter 3 we apply the OSSS inequality in two different ways to prove the sharp phase transition of Bernoulli percolation. We then investigate how to apply these ideas to the corrupted compass model in Chapter 4. We prove that this model undergoes a sharp phase transition, which was previously unattainable, considering the dependencies of the model. In Chapter 5 we consider the contact process. This model undergoes two phase transitions, both of which we prove to be sharp. Finally we look at the orthant model in Chapter 6. Since all components in this model are infinite, the typical definition of a sharp phase transition does not apply in this model. Nevertheless, this model undergoes a phase transition, and there is a natural notion of its sharpness, for which we will give a proof.

2 BOOLEAN FUNCTIONS

In the simplest terms, a Boolean function is a mapping from $\{-1,1\}^n$ to $\{-1,1\}$, for a fixed $n \in \mathbb{N}$. The choice of $\{-1,1\}$ as the possible values of the bits is arbitrary, but it makes the analysis more elegant. The range $\{-1,1\}$ is not an essential property of a Boolean function either, and we will often talk of real-valued Boolean functions instead. Since we are interested in probabilistic models, we want to consider a probability measure on $\{-1,1\}^n$. We start off by considering the product measure under which every bit is equal to 1 with probability ½ and -1 with probability ½, independently of each other. Even though this choice is restrictive from an application point of view, it does make the analysis more elegant as well. Once we have introduced the main objects in this setting, we will move to a more general setting, and see that these objects have natural analogues there. We will consider a general product space Ω^n for some finite set Ω, equipped with a product probability measure. In this way, we can apply the results of Boolean function theory to the models that are of interest to us.

Generally speaking, we can say the theory of Boolean functions is not so much characterized by the exact definition of the objects, but rather by the approach to analyzing them. We will have to spend some time to lay down the Boolean function ground work, so that we can reap the rewards of this work later on. A lot can be said about Boolean functions and their applications, but we will focus on the results that are relevant for percolation models. In this context, the typical Boolean function of interest is the indicator that 0 is connected to the boundary of the box of size r: $\mathbb{1}\{0 \longleftrightarrow \partial\Lambda_r\}$. The precise domain depends on the model in question, but if we take Bernoulli bond percolation as an example, we take E to be the set of edges inside Λ_r, and set $n = |E|$, so that $\mathbb{1}\{0 \longleftrightarrow \partial\Lambda_r\}$ is indeed a Boolean function. The theory of Boolean functions is then fundamental in analyzing the properties of this indicator function, which in turn gives detailed information on the behaviour of the model.

A comprehensive book on the theory of Boolean functions has been written by O'Donnel [47]. The presentation of the Boolean function framework as given in Sections 2.1 and 2.3 is largely based on this source, albeit more specific for our purposes. In Section 2.2 we state and prove the first version of the OSSS inequality. This section is based on the original paper of O'Donnell, Saks, Schramm and Servedio [48]. In Sections 2.4 and 2.5 we give proofs of two generalizations of the OSSS inequality, the first of which is based on the proof given by O'Donnel [47], while the second has not previously appeared in the literature to the best of our knowledge. In the final section of this chapter we loosen the assumption that the probability measure on Ω^n is a product measure, and focus on positively associated measures.

2.1 THE BOOLEAN FUNCTION FRAMEWORK

Let $n \in \mathbb{N}$, and write $[n] = \{1, \dots, n\}$. We start by considering functions of the form $f\colon \{-1, 1\}^n \to \mathbb{R}$. An element $x \in \{-1, 1\}^n$ consists of the variables $x_1, \dots, x_n$, which we also call bits, considering the relevance of Boolean functions in computer science. Every Boolean function can be written as a multinomial in the variables $x_1, \ldots, x_n$. To see this, we first consider the indicator function $\mathbb{1}_{\{a\}}(x)$ for $a \in \{-1,1\}^n$, which can be written as:

$$\mathbb{1}_{\{a\}}(x) = \Big(\frac{1 + a_1 x_1}{2}\Big)\Big(\frac{1 + a_2 x_2}{2}\Big) \ldots \Big(\frac{1 + a_n x_n}{2}\Big).$$

With this observation, we can write f as the multinomial given by

$$f(x) = \sum_{a \in \{0,1\}^n} f(a) \mathbb{1}_{\{a\}}(x).$$

For $S \subseteq [n]$, we define the monomial $x^S = \prod_{i \in S} x_i$. We can rearrange the terms in the above sum and write

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) x^S, \tag{2.1}$$

where $\hat{f}(S)$ is the coefficient in front of x^S, which is called the Fourier coefficient of f on S. The expansion in (2.1) is called the Fourier expansion of f. This expansion is unique, as we will now argue. The set of Boolean function defined on $\{-1,1\}^n$ form a vector space over the reals. We can define an inner product on this space as follows:

$$\langle f, g \rangle := \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x) g(x).$$

We then also have the corresponding norm on this space: $\|f\| = \sqrt{\langle f, f \rangle}$. For $S \subseteq [n]$ we define the Boolean functions $\chi_S(x) := x^S$, so that in particular $\chi_\varnothing(x) \equiv 1$. Note that for $T \subseteq [n]$, we have $\chi_S(x)\chi_T(x) = \chi_{S \triangle T}(x)$, where $S \triangle T$ is the symmetric difference between S and T, since $x_i^2 = 1$. This is one of the reasons why we have taken -1 and 1 as the values for the bits. We now have

$$\langle \chi_S, \chi_T \rangle = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} \chi_{S \triangle T}(x) = \begin{cases} 1 & \text{if } T = S, \\ 0 & \text{otherwise.} \end{cases}$$

From this observation it follows that the functions $\{\chi_S : S \subseteq [n]\}$ form an orthonormal basis for the vector space of Boolean functions, since every Boolean function can also be written as a linear combination of these function, as we saw in (2.1). It follows that the Fourier coefficients are uniquely given by $\hat{f}(S) = \langle f, \chi_S \rangle$, and

$$f(x) = \sum_{S \subseteq [n]} \langle f, \chi_S \rangle \chi_S(x).$$

Furthermore, the norm of f can be expressed in terms of the coefficients:

$$\|f\|^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2, \qquad \text{(Parseval's identity)}$$

and more generally,

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S). \tag{2.2}$$

Now that we have constructed the space of Boolean functions, we are ready to introduce a probability measure on $\{-1, 1\}^n$. We define $\mathbb{P}_{1/2}$ as the measure under which every bit is equal to 1 with probability $1/2$ and equal to -1 with probability $1/2$, independently of each other. We then denote the expectation with respect to this measure by $\mathbb{E}_{1/2}$. We can interpret the inner product we defined earlier as an expectation with respect to $\mathbb{P}_{1/2}$, since this corresponds to the uniform measure on $\{-1, 1\}^n$:

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x) = \mathbb{E}_{1/2}[fg].$$

The expectation of f with respect to $\mathbb{P}_{1/2}$ can be expressed in terms of its Fourier coefficients:

$$\mathbb{E}_{1/2}[f] = \hat{f}(\varnothing),$$

since $\mathbb{E}_{1/2}[\chi_S] = 0$ for $S \neq \varnothing$. We can find a similar expression for the variance of f by using Parseval's identity:

$$\mathrm{Var}_{1/2}(f) = \mathbb{E}_{1/2}[f^2] - \mathbb{E}_{1/2}[f]^2 = \|f\|^2 - \hat{f}(\varnothing)^2 = \sum_{S\neq\varnothing} \hat{f}(S)^2. \quad (2.3)$$

The most essential quantity for percolation is the influence of a variable i on the function f. We first define this for $\{-1,1\}$-valued functions.

Definition 2.1. *For $f\colon \{-1,1\}^n \to \{-1,1\}$, the influence of i on f is given by*

$$\mathrm{Inf}_i(f) = \mathbb{P}_{1/2}\big(f(x) \neq f(x^{\oplus i})\big),$$

where $x^{\oplus i}$ is the configuration x with bit i flipped:

$$x^{\oplus i} := (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n).$$

We say that i is pivotal for f, whenever changing the value of x_i changes the value of f. The influence of i on f is therefore the probability that i is pivotal for f. It can also be written using the discrete derivative operator:

$$D_i f(x) := \frac{f(x^{(i\mapsto 1)}) - f(x^{(i\mapsto -1)})}{2},$$

where $x^{(i\mapsto\pm1)}$ is given by $(x_1, \ldots, x_{i-1}, \pm1, x_{i+1}, \ldots, x_n)$. This operator is also defined for real-valued functions. In the case of $\{-1,1\}$-valued functions however, we have

$$D_i f(x) = \begin{cases} \pm1 & \text{if } x_i \text{ is pivotal for } f, \\ 0 & \text{otherwise,} \end{cases} \quad (2.4)$$

so that in this case,

$$\mathrm{Inf}_i(f) = \mathbb{E}_{1/2}\big[(D_i f)^2\big] = \|D_i f\|^2.$$

We use this as a definition for real-valued functions.

Definition 2.2. *For $f\colon\{-1,1\}^n \to \mathbb{R}$, the influence of i on f is given by*

$$\mathrm{Inf}_i(f) = \mathbb{E}_{1/2}\big[(D_i f)^2\big] = \|D_i f\|^2.$$

The influence of i on f can also be expressed using the Fourier coefficients of f.

Proposition 2.3. *For $f\colon\{-1,1\}^n \to \mathbb{R}$, we have*

$$D_i f(x) = \sum_{\substack{S\subseteq[n]\\ i\in S}} \hat{f}(S) x^{S\setminus\{i\}}, \qquad \mathrm{Inf}_i(f) = \sum_{\substack{S\subseteq[n]\\ i\in S}} \hat{f}(S)^2.$$

Proof. Since the discrete derivative operator is linear, we have

$$D_i f(x) = D_i \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x) = \sum_{S\subseteq[n]} \hat{f}(S) D_i\chi_S(x).$$

The result follows using Parseval's Identity, and considering

$$D_i\chi_S(x) = \begin{cases} x^{S\setminus\{i\}} & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

□

The total influence of f is defined as the sum over the individual influences.

Definition 2.4. *The total influence of* $f\colon \{-1,1\}^n \to \mathbb{R}$ *is defined as*

$$I(f) = \sum_{i=1}^{n} \mathrm{Inf}_i(f).$$

We are now able to prove a first concentration inequality for Boolean functions, which is known as the Poincaré inequality.

Theorem 2.5 (Poincaré inequality)**.** *Let* $f\colon \{-1,1\}^n \to \mathbb{R}$. *Then,*

$$\mathrm{Var}_{1/2}(f) \leq I(f). \tag{2.6}$$

Proof. Starting on the right hand side, and using Proposition 2.3, we find

$$I(f) = \sum_{i=1}^{n} \sum_{\substack{S \subseteq [n] \\ i \in S}} \hat{f}(S)^2 = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2 \geq \sum_{S \neq \varnothing} \hat{f}(S)^2 = \mathrm{Var}_{1/2}(f),$$

by (2.3). □

From the proof, it also follows that equality holds if and only if $\hat{f}(S) = 0$ for all sets S with $|S| \geq 2$. This is the case where f is a linear combination of the x_i's and a constant. The Poincaré inequality is not the strongest inequality, but it does set the tone for things to come. Namely, concentration inequalities where we bound the variance of f by, loosely speaking, its derivative. Sharp phase transitions are typically proven using strong differential inequalities, and these might in turn be proven using concentration inequalities. The Poincaré inequality is not strong enough for this purpose, which leads us to finding an improvement over this inequality.

2.2 THE OSSS INEQUALITY

The OSSS inequality is an improvement of the Poincaré inequality by discounting the influence of variables that are, in some sense, not necessary in determining the value of f. To make this notion precise, we introduce decision trees. Informally speaking, a decision tree is an algorithm that subsequently queries the bits until it has enough information to determine the value of f. It starts by revealing the value of x_i for some fixed variable x_i. Depending on this value, it then chooses the next variable to reveal, and so on, until it has gathered enough information. This occurs when, given the revealed information, the value of f is the same, regardless of the values of the unrevealed variables. In order to formally define a decision tree, we consider a directed tree $(V \cup U, E)$ with root $r \in V$, and define for $v \in V \cup U \backslash \{r\}$, the set $\rho(v)$ to be the ancestors of v: the set of vertices on the path from the root to v, including the root, and excluding v. We set $\rho(r) = \varnothing$.

Definition 2.6. *A decision tree T on $\{-1,1\}^n$ consists of a finite directed rooted tree $(V \cup U, E)$, and maps $\phi : V \to [n]$, $\psi : E \to \{-1,1\}$. Every vertex $v \in V$ has two children, while every $u \in U$ is a leaf. For any $v \in V$, the map ϕ satisfies $\phi(w) \neq \phi(v)$ for all $w \in \rho(v)$.*

We can use a decision tree to sequentially reveal the configuration x. This works by first revealing the value of $x_{\phi(r)}$. We then move down the tree along the edge (r, v) with $\psi((r, v)) = x_{\phi(r)}$. If this vertex is a leaf we stop the process. Otherwise we reveal $x_{\phi(v)}$ and repeat the process by moving down the edge (v, w) with $\psi((v, w)) = x_{\phi(v)}$. This process then continues until we end up in a leaf. We say that a decision tree determines the value of a Boolean function $f : \{-1,1\}^n \to \mathbb{R}$, when the revealed values determine the value of f, that is, for all $u \in U$ and all $x, y \in \{-1,1\}^n$

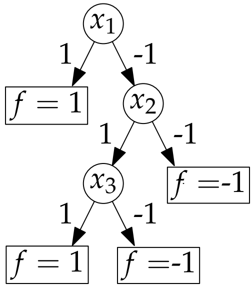


Figure 2.1: A decision tree determining the value of $f(x) = x_1 \vee (x_2 \wedge x_3)$.

with $x_{\phi(w)} = y_{\phi(w)}$ for all $w \in \rho(u)$, it holds, that $f(x) = f(y)$. We sometimes say that T computes f instead. To illustrate the above definition, we consider the example of $f\colon \{\text{-}1, 1\}^3 \to \{\text{-}1, 1\}$ given by $f(x) = x_1 \vee (x_2 \wedge x_3)$. A decision tree that determines the value of this function is shown in Figure 2.1. Note that this decision tree is not unique.

Since we consider a probability measure on $\{\text{-}1, 1\}^n$, the path that is taken in decision tree is also random. We say that a variable is revealed, if it lies on this path. We call the probability that a variable is revealed the revealment of that variable.

Definition 2.7. *Let T be a decision tree on $\{\text{-}1, 1\}^n$. The event that $i \in [n]$ is revealed, is the set of configurations $x \in \{\text{-}1, 1\}^n$ such that there exists $v \in V$ with $\phi(v) = i$, and $x_{\phi(w)} = \psi(e)$ for all edges $e = (w, w')$ with starting- and endpoints in $\rho(v) \cup \{v\}$. The revealment of i is given by*

$$\mathrm{Rev}_i(T) := \mathbb{P}_{1/2}(i \textit{ is revealed}).$$

We can now state the OSSS inequality in the simplest setting.

Theorem 2.8 (OSSS inequality for Ber(½)-random variables)**.** *Let $f\colon \{-1,1\}^n \to \mathbb{R}$, and let T be a decision tree that determines the value of f. Then,*

$$\mathbb{E}_{1/2}\big[|f(x) - f(y)|\big] \le \sum_{i=1}^{n} \mathbb{E}_{1/2}\big[|D_i f|\big]\mathrm{Rev}_i(T), \tag{2.7}$$

where x and y are independent and have law $\mathbb{P}_{1/2}$.

For $\{-1,1\}$-valued functions, we recover the more familiar form of the OSSS inequality.

Corollary 2.9. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$, and let T be a decision tree that determines the value of f. Then,*

$$\mathrm{Var}_{1/2}(f) \le \sum_{i=1}^{n} \mathrm{Inf}_i(f)\mathrm{Rev}_i(T). \tag{2.8}$$

This corollary follows directly from Theorem 2.8, since in the case of $\{-1,1\}$-valued functions, we have

$$\mathrm{Inf}_i(f) = \mathbb{E}_{1/2}\big[(D_i f)^2\big] = \mathbb{E}_{1/2}\big[|D_i f|\big],$$

and

$$\begin{aligned}\mathrm{Var}_{1/2}(f) = \mathbb{E}_{1/2}[f^2] - \mathbb{E}_{1/2}[f]^2 &= 1 - \big(\mathbb{P}_{1/2}(f=1) - \mathbb{P}_{1/2}(f=-1)\big)^2 \\ &= 4\mathbb{P}_{1/2}(f=1)\mathbb{P}_{1/2}(f=-1) \\ &= 2\mathbb{P}_{1/2}(f(x) \ne f(y)) \\ &= \mathbb{E}_{1/2}\big[|f(x) - f(y)|\big].\end{aligned}$$

Since $\mathrm{Rev}_i(T) \le 1$, Corollary 2.9 is indeed an improvement over the Poincaré inequality. To make the inequality strong for a given function f, the aim is to find a decision tree computing f for which the revealments are small. There exist functions f however, for which the OSSS inequality does not strictly improve on the Poincaré inequality, considering that the Poincaré inequality is sharp in the case where the Fourier coefficients of f are supported on sets of size 0 or 1. In these cases, it also follows that any decision tree determining the value of f always reveals the variables with positive influence.

The OSSS inequality was proven by O'Donnell, Saks, Schramm and Servedio in 2005 [48]. We will first look at their original proof, since it is in some sense the most straightforward. In a later section, we will also look at an inductive proof of the OSSS inequalty by O'Donnel [47].

Proof. Let $f\colon\{-1,1\}^n \to \mathbb{R}$, and let T be a decision tree determining the value of f. Let $x, y \in \{\text{-}1,1\}^n$ be independent and have law $\mathbb{P}_{1/2}$. The proof strategy is to take the configuration x, and change the values of its variables one by one to the values of y. In this way, we bound $\mathbb{E}_{1/2}\big[|f(x)-f(y)|\big]$ by a sum over the variables using the triangle inequality. We let $i_1,\dots,i_t$ be the sequence in $[n]$ that corresponds to the path in the decision tree on input x. Note that t depends on x as well. For $1 \le s \le n$, we let $x_s y$ be the configuration that agrees with x for the variables in $i_{s+1},\dots,i_t$, and with y for all other variables:

$$(x_s y)_i = \begin{cases} x_i & \text{if } i = i_k,\ s < k \le t, \\ y_i & \text{otherwise.} \end{cases}$$

In particular, we have $x_t y = y$, and, since T determines the value of f, $f(x_0 y) = f(y)$. Using the triangle inequality, we find

$$\begin{aligned}\mathbb{E}_{1/2}\big[|f(x)-f(y)|\big] &= \mathbb{E}_{1/2}\big[|f(x_0 y)-f(x_t y)|\big] \\ &\le \mathbb{E}_{1/2}\Big[\sum_{s=1}^{t}\big|f(x_{s-1}y)-f(x_s y)\big|\Big].\end{aligned}$$

Setting $i_s = 0$ for $s > t$, we see that $s \le t$ if and only if there exists $1 \le i \le n$ such that $i_s = i$. This gives

$$\mathbb{E}_{1/2}\big[|f(x) - f(y)|\big] \le \sum_{s=1}^{n}\sum_{i=1}^{n} \mathbb{E}_{1/2}\Big[\big|f(x_{s-1}y) - f(x_s y)\big| \mathbb{1}_{\{i_s = i\}}\Big].$$

For $s \ge 1$, we define the filtration $\mathcal{F}_s := \sigma(x_{i_1}, \ldots, x_{i_{s\wedge t}})$, and we set $\mathcal{F}_0 := \sigma(\varnothing)$, the trivial σ-algebra. Note that i_1 is $\mathcal{F}_0$-measurable, since it is deterministic. Moreover, i_t is $\mathcal{F}_{t-1}$-measurable for all $t \ge 1$, since the decision tree chooses the next variable to reveal as a function of the previously revealed vertices. Furthermore, the collection of random variables $\{x_i : i \ne i_1, \ldots, i_{s\wedge t}\}$ is independent of $\mathcal{F}_s$, and so is y, so that conditionally on $\mathcal{F}_s$ these variables retain their original distribution. We therefore find

$$\begin{aligned}
&\mathbb{E}_{1/2}\Big[\big|f(x_{s-1}y) - f(x_s y)\big| \mathbb{1}_{\{i_s = i\}} \,\Big|\, \mathcal{F}_{s-1}\Big] \\
&\quad = \mathbb{1}_{\{i_t = i\}} \mathbb{E}_{1/2}\big[|f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) - f(x)|\big] \\
&\quad = \mathbb{1}_{\{i_t = i\}} \mathbb{E}_{1/2}\big[1/4|2D_i f| + 1/4|2D_i f|\big] \\
&\quad = \mathbb{1}_{\{i_t = i\}} \mathbb{E}_{1/2}\big[|D_i f|\big].
\end{aligned} \tag{2.9}$$

Hence,

$$\begin{aligned}
&\mathbb{E}_{1/2}\Big[\big|f(x_{s-1}y) - f(x_s y)\big| \mathbb{1}_{\{i_s = i\}}\Big] \\
&\quad = \mathbb{E}_{1/2}\bigg[\mathbb{E}_{1/2}\Big[\big|f(x_{s-1}y) - f(x_s y)\big| \mathbb{1}_{\{i_s = i\}} \,\Big|\, \mathcal{F}_{s-1}\Big]\bigg] \\
&\quad = \mathbb{E}_{1/2}\big[\mathbb{1}_{\{i_t = i\}}\big] \mathbb{E}_{1/2}\big[|D_i f|\big].
\end{aligned} \tag{2.10}$$

The result now follows, since

$$\sum_{t=1}^{n} \mathbb{E}_{1/2}\big[\mathbb{1}_{\{i_t = i\}}\big] = \mathbb{P}_{1/2}(\exists t \in [n] \text{ such that } i_t = i) = \mathrm{Rev}_i(T).$$

$\square$

2.3 FINITE PRODUCT SPACES

We will now generalize the results we have seen so far to general finite product spaces. For a single variable, we consider the probability space $(\Omega, \mathcal{F}, \pi)$, for some finite set Ω with $|\Omega| \geq 2$, σ-algebra $\mathcal{F}$ given by the power set of Ω, and a probability measure π. We assume that π has full support. For $n \in \mathbb{N}$, we then consider the product space $(\Omega^n, \mathcal{F}^n, \mathbb{P})$, where the σ-algebra $\mathcal{F}^n$ is the n-fold product of $\mathcal{F}$ (which is also the power set of Ω^n), and the product probability measure defined by

$$\mathbb{P}(\{x\}) = \prod_{i=1}^{n} \pi(\{x_i\}).$$

A Boolean function in this setting is a function $f : \Omega^n \to \mathbb{R}$. Similar to the inner product that we defined in section 2.1, we define for Boolean functions f and g the inner product

$$\langle f, g \rangle = \mathbb{E}[fg],$$

where $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$. We denote the inner product space of Boolean functions on Ω^n by $L^2(\Omega^n)$. Since Ω^n is finite, any Boolean function is integrable.

The first issue we run into is that we do not a have natural set of basis functions for $L^2(\Omega^n)$. A possible basis is the set of indicator functions $\{\mathbb{1}_x : x \in \Omega^n\}$. These functions are orthogonal and span $L^2(\Omega^n)$, since for all functions f we can write

$$f(y) = \sum_{x \in \Omega^n} f(x) \mathbb{1}_x(y).$$

A nice property of this basis is that it is a product basis, that is, all basis function can be written as

$$\mathbb{1}_x(y) = \prod_{i=1}^{n} \mathbb{1}_{x_i}(y_i),$$

where the set $\{\mathbb{1}_{x_i} : x_i \in \Omega\}$ is a basis for the space of Boolean function on Ω. We could even make the indicator basis for $L^2(\Omega^n)$ orthonormal by applying an appropriate rescaling. However, the problem with this basis is that it lacks the constant function 1. If this function is included in an orthonormal basis, we have $\mathbb{E}[\phi_i] = 0$ for all other basis functions ϕ_i. This in turn implies $\mathbb{E}[f] = \langle f, 1\rangle$ and $\mathrm{Var}(f) = \sum_i \hat{f}(i)^2 - \langle f, 1\rangle^2$, where $\hat{f}(i) = \langle f, \phi_i\rangle$ are the Fourier coefficients of f. We therefore aim to find a basis that includes the constant function. As a first step, we construct a basis for $L^2(\Omega)$. Note that an orthonormal basis including 1 always exists for $L^2(\Omega)$, since we can extend $\{1\}$ by adding linearly independent functions until the set spans $L^2(\Omega)$, and then apply the Gram-Schmidt procedure to orthonormalize this set. Once we have a basis for $L^2(\Omega)$, we find a basis for $L^2(\Omega^n)$ by taking products of these basis functions. Let $m = |\Omega|$. This is the dimension of $L^2(\Omega)$, since the indicator basis consists of m functions. Let $\phi_0, \ldots, \phi_{m-1}$ be an orthonormal basis for $L^2(\Omega)$ with $\phi_0 = 1$. We define the product basis for $L^2(\Omega^n)$ by the set of functions given by

$$\phi_\alpha(x) = \prod_{i=1}^{n} \phi_{\alpha_i}(x_i), \qquad \alpha \in I_m^n := \{0, \ldots, m-1\}^n. \tag{2.11}$$

Note that $\phi_0 = 1$. We check that the above set indeed forms an orthonormal basis for $L^2(\Omega^n)$. For $\alpha, \beta \in I_m^n$, we compute

$$\begin{aligned}\langle \phi_\alpha, \phi_\beta\rangle = \mathbb{E}[\phi_\alpha \phi_\beta] &= \mathbb{E}\Big[\prod_{i=1}^{n} \phi_{\alpha_i}\phi_{\beta_i}\Big] \\ &= \prod_{i=1}^{n} \mathbb{E}[\phi_{\alpha_i}\phi_{\beta_i}] = \prod_{i=1}^{n} \mathbb{1}_{\{\alpha_i = \beta_i\}} = \mathbb{1}_{\{\alpha = \beta\}},\end{aligned}$$

since $\mathbb{P}$ is a product measure, and since the basis of $L^2(\Omega)$ is orthonormal. Now that we have a basis for $L^2(\Omega^n)$, every function f can expressed in terms of the basis functions:

$$f = \sum_{\alpha \in I_m^n} \hat{f}(\alpha)\phi_\alpha = \sum_{\alpha \in I_m^n} \langle f, \phi_\alpha \rangle \phi_\alpha.$$

This is the Fourier expansion of f, and the $\hat{f}(\alpha)$ are called the Fourier coefficients of f on α. The inner product can be expressed in terms of these coefficients:

$$\begin{aligned} \langle f, g \rangle &= \Big\langle \sum_{\alpha \in I_m^n} \hat{f}(\alpha)\phi_\alpha, \sum_{\beta \in I_m^n} \hat{g}(\beta)\phi_\beta \Big\rangle \\ &= \sum_{\alpha \in I_m^n} \sum_{\beta \in I_m^n} \hat{f}(\alpha)\hat{g}(\beta)\langle \phi_\alpha, \phi_\beta \rangle = \sum_{\alpha \in I_m^n} \hat{f}(\alpha)\hat{g}(\alpha). \end{aligned}$$

Since we have taken $\phi_0 = 1$, it now follows from the above representation of the inner product, that

$$\mathbb{E}[f] = \langle f, \phi_0 \rangle = \hat{f}(0), \qquad \mathbb{E}[f^2] = \langle f, f \rangle = \sum_{\alpha \in I_m^n} \hat{f}(\alpha)^2. \tag{2.12}$$

Furthermore, the variances and covariances satisfy

$$\mathrm{Var}(f) = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{\alpha \neq 0} \hat{f}(\alpha)^2, \tag{2.13}$$

$$\mathrm{Cov}(f, g) = \mathbb{E}[fg] - \mathbb{E}[f]\mathbb{E}[g] = \sum_{\alpha \neq 0} \hat{f}(\alpha)\hat{g}(\alpha). \tag{2.14}$$

We now turn our attention to the influence of variables. In the case of $\Omega = \{-1, 1\}$ we defined this by flipping bit i from -1 to 1, or vice versa. Now that Ω can consist of more than two elements, we cannot use this definition. Instead, we make us of the distribution of the single variable π. We do this by means of the ith expectation operator.

Definition 2.10. *The ith expectation operator on $L^2(\Omega^n)$ is given by*

$$E_i f(x) = \int_\Omega f(x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n)\, \pi(\mathrm{d}s).$$

In particular, $E_i f(x)$ does not depend on x_i. If we apply the ith expectation operator to a basis function ϕ_α, we find

$$\begin{aligned} E_i \phi_\alpha(x) = \int_\Omega \phi_{\alpha_i}(s) \prod_{j \neq i}^{n} \phi_{\alpha_j}(x_j)\, \pi(\mathrm{d}s) &= \langle \phi_{\alpha_i}, 1 \rangle \prod_{j \neq i}^{n} \phi_{\alpha_j}(x_j) \\ &= \begin{cases} \phi_\alpha & \text{if } \alpha_i = 0, \\ 0 & \text{if } \alpha_i \neq 0. \end{cases} \end{aligned}$$

The fact that E_i is linear implies the following proposition.

Proposition 2.11. *Let $f \in L^2(\Omega^n)$. The ith expectation operator satisfies*

$$E_i f = \sum_{\substack{\alpha \in I_m^n \\ \alpha_i = 0}} \hat{f}(\alpha) \phi_\alpha.$$

We now define the ith Laplacian operator.

Definition 2.12. *The ith Laplacian operator on $L^2(\Omega^n)$ is given by*

$$L_i f := f - E_i f.$$

This gives a decomposition of f into a part that depends on x_i, and a part that is independent of x_i:

$$f(x) = L_i f(x) + E_i f(x) = \sum_{\substack{\alpha \in I_m^n \\ \alpha_i \neq 0}} \hat{f}(\alpha) \phi_\alpha + \sum_{\substack{\alpha \in I_m^n \\ \alpha_i = 0}} \hat{f}(\alpha) \phi_\alpha. \tag{2.15}$$

In the case of $\Omega = \{\text{-}1, 1\}$, the basis functions are indexed by $S \subseteq [n]$, but the basis is still of the from of Definition 2.11, since

χ_S is a product of the variables indexed by S, and $\chi_\varnothing = 1$. The case $\alpha_i \neq 0$ corresponds to $i \in S$. Proposition 2.3 says that in the case $\{-1, 1\}^n$, the influences are given by

$$\mathrm{Inf}_i(f) = \sum_{\substack{S \subseteq [n] \\ i \in S}} \hat{f}(S)^2.$$

We therefore find in this case, that $\mathrm{Inf}_i(f) = \langle f, L_i f \rangle$. We use this as the definition for the influence of i in the case of general product spaces.

Definition 2.13. *For $f \in L^2(\Omega^n)$, the influence of i on f is defined by*

$$\mathrm{Inf}_i(f) := \langle f, L_i f \rangle = \sum_{\substack{\alpha \in I_m^n \\ \alpha_i \neq 0}} \hat{f}(\alpha)^2.$$

The total influence of f is given by

$$I(f) = \sum_{i=1}^{n} \mathrm{Inf}_i(f).$$

We can characterize the influence of i independently of the basis as follows.

Proposition 2.14. *Let $f \in L^2(\Omega^n)$. The influence of i on f satisfies*

$$\mathrm{Inf}_i(f) = \mathbb{E}\big[E_i f^2 - (E_i f)^2\big].$$

In particular, if f is $\{0, 1\}$-valued:

$$\mathrm{Inf}_i(f) = \tfrac{1}{2}\mathbb{P}\big(f(x) \neq f(x^{i \mapsto \pi})\big),$$

where $x^{i \mapsto \pi} = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)$, where x_i' is independent of x and has law π.

Proof. Starting on the right hand side, and using Proposition 2.11, we find

$$\begin{aligned}\mathbb{E}\big[E_i f^2 - (E_i f)^2\big] &= \mathbb{E}\Big[\sum_{\alpha:\alpha_i=0} \widehat{f^2}(\alpha)\phi_\alpha - \Big(\sum_{\alpha:\alpha_i=0} \hat{f}(\alpha)\phi_\alpha\Big)^2\Big] \\ &= \widehat{f^2}(0) - \sum_{\alpha:\alpha_i=0} \hat{f}(\alpha)^2,\end{aligned}$$

since $\mathbb{E}[\phi_\alpha] = \mathbb{1}_{\{\alpha=0\}}$, and since the basis is orthonormal. The Fourier coefficient of f^2 on 0 is given by

$$\widehat{f^2}(0) = \langle f^2, 1\rangle = \mathbb{E}[f^2] = \sum_{\alpha\in I_m^n} \hat{f}(\alpha)^2,$$

by equation (2.12). It follows, that

$$\mathbb{E}\big[E_i f^2 - (E_i f)^2\big] = \sum_{\alpha:\alpha_i\neq 0} \hat{f}(\alpha)^2 = \mathrm{Inf}_i(f).$$

We now consider the case where f is $\{0,1\}$-valued. We have

$$\begin{aligned}&\mathbb{P}\big(f(x) \neq f(x^{i\mapsto\pi})\big) = \mathbb{E}\Big[\big(f(x) - f(x^{i\mapsto\pi})\big)^2\Big] \\ &\quad = \int_{\Omega^{n-1}}\int_{\Omega}\int_{\Omega} \big(f(x) - f(x^{i\mapsto\pi})\big)^2\, \pi(\mathrm{d}x_i')\pi(\mathrm{d}x_i)\pi(\mathrm{d}x) \\ &\quad = \int_{\Omega^{n-1}}\int_{\Omega} \Big(f(x)^2 - 2f(x)E_i f(x) + E_i f^2(x)\Big)\, \pi(\mathrm{d}x_i)\pi(\mathrm{d}x) \\ &\quad = 2\mathbb{E}\big[E_i f^2 - (E_i f)^2\big]. \qquad (2.16)\end{aligned}$$

□

We end this section by showing that the Poincaré inequality also holds for general product spaces.

Theorem 2.15. *Let* $f\colon\Omega^n \to \mathbb{R}$. *Then,*

$$\mathrm{Var}(f) \leq I(f). \qquad \text{(Poincaré inequality)}$$

Proof. Using the definition of the influence, we find

$$I(f) = \sum_{i=1}^{n}\langle f, L_i\rangle = \sum_{i=1}^{n}\sum_{\substack{\alpha\in I_m^n\\ \alpha_i\neq 0}}\hat{f}(\alpha)^2 = \sum_{\alpha\in I_m^n}\hat{f}(\alpha)^2|\{i : \alpha_i \neq 0\}|$$
$$\geq \sum_{\alpha\neq 0}\hat{f}(\alpha)^2 = \mathrm{Var}(f),$$

by equation (2.13). □

2.4 THE OSSS INEQUALITY FOR FINITE PRODUCT SPACES

We now return to the OSSS inequality, and see how we can generalize it to general product spaces. In the previous section we have introduced the definition of the influence of a variable in a finite product space. Hence, in order to state the OSSS inequality in this setting, it remains to define the notion of a decision tree. The informal description remains unchanged: an algorithm that sequentially reveals the variables until it has gathered enough information to determine the value of f. Formally, we use Definition 2.6, but require that every vertex in the decision tree that is not a leaf has $|\Omega|$ children instead of 2, and that the labels of the edges are given by $\psi : E \to \{1, \dots, |\Omega|\}$. The revealment of a variable caries over directly from Definition 2.7 as well, taking $x \in \Omega^n$ instead of $\{-1, 1\}^n$. We can now state the OSSS inequality in this setting. We give the covariance version of the inequality, a more familiar form is recovered by taking $f = g$.

Theorem 2.16 (OSSS inequality for finite product spaces)**.** *Let $f, g : \Omega^n \to \{0, 1\}$, and let T be a decision tree that determines the value of f. Then,*

$$\mathrm{Cov}(f, g) \leq 2\sum_{i=1}^{n}\mathrm{Inf}_i(g)\mathrm{Rev}_i(T). \tag{2.17}$$

We have chosen to state the version for $\{0,1\}$-valued functions, since it is the most useful version for percolation models. The same statement without the factor 2 on the right-hand side can be proven for $\{-1,1\}$-valued functions with the same proof. The above version of the OSSS inequality can be proven in the same way as 2.8. In fact, the OSSS inequality is proven in [48] for the same general finite product space as we have introduced in the previous section. We choose to present a different proof here that is due to O'Donnell [47]. This proof differs substantially from the original proof, and relies more heavily on the Fourier analysis of Boolean functions.

Proof. Let $f, g : \{-1,1\}^n \to \{0,1\}$ be two Boolean functions, and let T be a decision tree determining the value of f. We define the depth of a decision tree T with vertex set V to be $\delta(T) := \max_{v \in V} |\pi(v)|$. It is therefore the maximum number of vertices that are revealed by the tree. We prove (2.17) by induction on the depth of T. If $\delta(T) = 0$, then the decision tree does not reveal any vertices, and it follows, that f is a constant function. The inequality then follows, since $\mathrm{Cov}(f,g) = 0$, and the right hand side is non-negative.

Now suppose the $\delta(T) = k \in \mathbb{N}$, and suppose that (2.17) holds for all $f', g' \in L^2(\Omega^n)$ and all decision trees of depth at most $k-1$ computing f'. Let r be the root of T, and let $i_1 = \phi(r)$. Thus, i_1 is the first variable revealed by T. In particular, it is revealed with probability 1. For $\omega \in \Omega$, we define $f_\omega, g_\omega : \Omega^n \to \{0,1\}$, given by

$$
\begin{aligned}
f_\omega(x) &= f(x^{i_1 \mapsto \omega}), \\
g_\omega(x) &= g(x^{i_1 \mapsto \omega}),
\end{aligned}
$$

where $x^{i_1 \mapsto \omega} = (x_1, \ldots, x_{i_1-1}, \omega, x_{i_1+1}, \ldots, x_n)$. Using the Fourier formulas for E_i and L_i of (2.15), we find for all $i \in [n]$,

$$\mathrm{Cov}(f,g) = \langle f, g\rangle - \mathbb{E}[f]\mathbb{E}[g] = \langle E_i f, E_i g\rangle + \langle L_i f, L_i g\rangle - \mathbb{E}[f]\mathbb{E}[g].$$

We can further write

$$\begin{aligned} \langle E_i f, E_i g\rangle - \mathbb{E}[f]\mathbb{E}[g] &= \mathbb{E}_{\omega,\omega'\sim\pi}\big[\mathbb{E}[f_\omega g_{\omega'}] - \mathbb{E}[f_\omega]\mathbb{E}[g_{\omega'}]\big] \\ &= \mathbb{E}_{\omega,\omega'\sim\pi}\big[\mathrm{Cov}(f_\omega, g_{\omega'})\big], \end{aligned}$$

where the outer expectation is taken with respect to the measure where ω and ω' are independent, and have law π. In particular we have for $i = i_1$,

$$\mathrm{Cov}(f,g) = \mathbb{E}_{\omega,\omega'\sim\pi}\big[\mathrm{Cov}(f_\omega, g_{\omega'})\big] + \langle L_{i_1} f, L_{i_1} g\rangle. \qquad (2.18)$$

Suppose $x_{i_1} = \omega$, and let T_ω be the decision tree determining the value of f_ω obtained by considering the subtree of T rooted at the child v of r with $\psi((r,v)) = \omega$. This child exists, since $\delta(T) \geq 1$. Furthermore, since we consider the subtree, $\delta(T_\omega) \leq k-1$. We use the induction hypothesis to find

$$\mathrm{Cov}(f,g) \leq \mathbb{E}_{\omega,\omega'\sim\pi}\Big[2\sum_{i\neq i_1}\mathrm{Inf}_i(g_{\omega'})\mathrm{Rev}_i(T_\omega)\Big] + \langle L_{i_1} f, L_{i_1} g\rangle.$$

Using the Fourier formulas for E_{i_1} and L_{i_1}, we have $\langle E_{i_1} f, L_{i_1} g\rangle = 0$, so that

$$\langle L_{i_1} f, L_{i_1} g\rangle = \langle f, L_{i_1} g\rangle.$$

Since $|f| \leq 1$, we find

$$\langle f, L_{i_1} g\rangle \leq \mathbb{E}\big[|f L_{i_1} g|\big] \leq \mathbb{E}\big[|L_{i_1} g|\big].$$

Since g is $\{0,1\}$-valued, we can bound the right-hand side as follows:

$$\begin{aligned}
&\mathbb{E}\big[|L_{i_1}g|\big] \\
&\quad = \int_{\Omega^{n-1}}\int_{\Omega}\big(\mathbb{1}_{\{g=1\}}(1-E_{i_1}g)+\mathbb{1}_{\{g=0\}}E_{i_1}g\big)\,\pi(\mathrm{d}x_{i_1})\pi(\mathrm{d}x) \\
&\quad = \int_{\Omega^{n-1}}\int_{\Omega}\big(\mathbb{1}_{\{g=1\}}+E_{i_1}g-2\mathbb{1}_{\{g=1\}}E_{i_1}g\big)\,\pi(\mathrm{d}x_{i_1})\pi(\mathrm{d}x) \\
&\quad = \int_{\Omega^{n-1}} 2E_{i_1}g-2(E_{i_1}g)^2\,\pi(\mathrm{d}x) \\
&\quad = 2\mathbb{E}[E_ig^2-(E_ig)^2] = 2\mathrm{Inf}_{i_1}(g),
\end{aligned}$$

by Proposition 2.14. Using the fact that $\mathrm{Rev}_{i_1}(T)=1$, we now find

$$\begin{aligned}
\mathrm{Cov}(f,g) &\le \mathbb{E}_{\omega,\omega'\sim\pi}\Big[2\sum_{i\neq i_1}\mathrm{Inf}_i(g_{|\omega'})\mathrm{Rev}_i(T_\omega)\Big]+2\mathrm{Inf}_{i_1}(g) \\
&= 2\sum_{i\neq i_1}\mathbb{E}_{\omega,\omega'\sim\pi}\Big[\mathrm{Inf}_i(g_{|\omega'})\mathrm{Rev}_i(T_\omega)\Big]+2\mathrm{Inf}_{i_1}(g) \\
&= 2\sum_{i\neq i_1}\mathrm{Inf}_i(g)\mathrm{Rev}_i(T)+2\mathrm{Inf}_{i_1}(g) \\
&= 2\sum_{i=1}^{n}\mathrm{Inf}_i(g)\mathrm{Rev}_i(T).
\end{aligned}$$

This conclusion the inductive proof of Theorem 2.16. □

2.5 INFINITE SPACES

So far we have considered finite product spaces, in the sense that $|\Omega|<\infty$, and in the sense that we only considered a finite number of variables. This suffices for most applications in percolation models, since we typically consider consider the Boolean function $\mathbb{1}_{\{0\longleftrightarrow\partial\Lambda_n\}}$, which only depends on the state of the vertices or edges inside the box. Nevertheless, applying the Boolean function framework to the contact process requires to weaken the assumption that Ω is finite, whereas for the orthant model we need to consider infinitely many variables. We will start by examining this assumption.

2.5.1 *Infinite products*

Let $(\Omega, \mathcal{F}, \pi)$ be a finite probability space. We consider the product space $(\Omega^{\mathbb{N}}, \mathcal{F}^{\mathbb{N}}, \mathbb{P})$, where the σ-algebra is generated by the cylindrical events:

$$\mathcal{F}^{\mathbb{N}} = \sigma\big(\{\{\omega_i \in A\} : i \in \mathbb{N}, A \in \mathcal{F}\}\big).$$

The product probability measure is then uniquely defined by setting

$$\mathbb{P}(\omega_i \in A_i \text{ for all } i \in I) = \prod_{i \in I} \pi(A_i),$$

for all finite $I \subseteq \mathbb{N}$, and all $A_i \in \mathcal{F}$. We denote the expectation with respect to $\mathbb{P}$ by $\mathbb{E}$.

We refrain from generalizing the Fourier framework to this setting. Instead, we introduce only the objects we need for the OSSS inequality. Nevertheless, these definitions are direct analogues to the results in Section 2.3, so that the Fourier theory is still lingering in the background. We define $L^p(\Omega^{\mathbb{N}})$ to be the set of measurable functions from $\Omega^{\mathbb{N}}$ to $\mathbb{R}$ with finite L^p-norm, which is given by

$$\|f\|_p = \mathbb{E}[f^p]^{1/p}.$$

We define the ith expectation operator on $L^2(\Omega^{\mathbb{N}})$ as

$$E_i f(x) = \int_\Omega f(x_1, \dots, x_{i-1}, s, x_{i+1}, \dots)\, \pi(\mathrm{d}s), \quad i \in \mathbb{N}.$$

Inspired by Proposition 2.14, we define for $f \in L^2(\Omega^{\mathbb{N}})$ the influence of i on f as

$$\mathrm{Inf}_i(f) = \mathbb{E}\big[E_i f^2 - (E_i f)^2\big]. \tag{2.19}$$

Jensen's inequalty implies $\mathrm{Inf}_i(f) \geq 0$. Furthermore, since $f \in L^2(\Omega^{\mathbb{N}})$, we have

$$\mathrm{Inf}_i(f) \leq \mathbb{E}\big[E_i f^2\big] = \mathbb{E}[f^2] < \infty.$$

The computation in (2.16) carries over to the infinite case, so that if f is $\{0,1\}$-valued, we find

$$\mathrm{Inf}_i(f) = {}^1\!/\!{}_2\,\mathbb{P}\big(f(x) \neq f(x^{i\mapsto\pi})\big), \tag{2.20}$$

where $x^{i\mapsto\pi} = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots)$, and x_i' is independent of x and has law π.

The definition of a decision tree that determines the value of f largely carries over from Definition 2.6, except the tree can be infinite. In other words, we do not require the algorithm to terminate. Recall that for a vertex v in the decision tree, $\rho(v)$ denotes the set of vertices on the path from the root to v, including the root and excluding v.

Definition 2.17. *A decision tree T on $\Omega^{\mathbb{N}}$ consists of a directed rooted tree $(V \cup U, E)$, and maps $\phi\colon V \to \mathbb{N}$, $\psi\colon E \to \Omega$. Every $v \in V$ has $|\Omega|$ children, while every $u \in U$ is a leaf. For any $v \in V$, the map ϕ satisfies $\phi(w) \neq \phi(v)$ for all $w \in \rho(v)$.*

We say that a decision tree on $\Omega^{\mathbb{N}}$ determines the value of a Boolean function $f\colon \Omega^{\mathbb{N}} \to \mathbb{R}$ if the following two conditions are satisfied: for all leaves $u \in U$, the values of $x_{\phi(w)}$ for $w \in \rho(v)$ determine the value of f, that is, if $x, y \in \Omega^{\mathbb{N}}$ with $x_{\phi(w)} = y_{\phi(w)}$ for all $w \in \rho(u)$, then $f(x) = f(y)$; secondly, for all infinite paths

ρ_∞ starting at the root, the values of $x_{\phi(w)}$ for $w \in \rho_\infty$ determine the value of f, that is, if $x, y \in \Omega^{\mathbb{N}}$ with $x_{\phi(w)} = y_{\phi(w)}$ for all $w \in \rho_\infty$, then $f(x) = f(y)$. Even if the decision tree does not terminate, the condition for infinite paths implies that a decision tree must acquire information that is relevant to the value of f, and that this information determines the value of f in an asymptotic sense. An example that satisfies this condition is the important case where f is $\{0,1\}$-valued, and the decision tree terminates on the set $\{f = 1\}$. If $x \in \Omega^{\mathbb{N}}$ corresponds to an infinite path in the decision tree, and thus the decision does not terminate, then it follows that $f = 0$, so that the condition on infinite paths is indeed satisfied.

The revealment of a variable is defined as in Definition 2.7, taking $x \in \Omega^{\mathbb{N}}$ instead of $\{-1,1\}^n$. We can now state the OSSS inequality for the case of countable product spaces.

Theorem 2.18 (OSSS inequality for infinite product spaces). *Let $f \in L^1(\Omega^{\mathbb{N}})$, and let T be a decision tree that determines the value of f. Then,*

$$\mathbb{E}\big[|f(x) - \mathbb{E}[f(x)]|\big] \leq \sum_{i=1}^{\infty} \mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big]\mathrm{Rev}_i(T). \quad (2.21)$$

In particular, if f is $\{0,1\}$-valued,

$$\mathrm{Var}(f) \leq \sum_{i=1}^{\infty} \mathrm{Inf}_i(f)\mathrm{Rev}_i(T). \quad (2.22)$$

In the original paper of O'Donnell, Saks, Schramm and Servedio the inequality is only given for the finite case. In the paper of Duminil-Copin, Raoufi and Tassion [15] Inequality 2.22 has been stated without proof. Hence, the following proof of Theorem 2.18 has not appeared in the literature before.

Proof. We will modify the proof of Theorem 2.8. Let $x, y \in \Omega^{\mathbb{N}}$ be independent and have law $\mathbb{P}$. We let $i_1, \ldots, i_t$ be the sequence in $\mathbb{N}$ that corresponds to the path in the decision tree on input x, where $t \in \mathbb{N} \cup \{\infty\}$. Note that $i_k \neq i_l$, for all $1 \leq k < l \leq t$, by the definition of a decision tree on $\Omega^{\mathbb{N}}$. For $n \geq 1$ we define

$$\mathcal{F}_n := \sigma(x_{i_1}, \ldots, x_{i_{n \wedge t}}), \quad \mathcal{F}_\infty = \sigma(x_{i_1}, x_{i_2}, \ldots, x_{i_t}).$$

Then $f(y)$ is independent of $\mathcal{F}_n$ for all $n \geq 1$, and $f(x)$ is $\mathcal{F}_\infty$-measurable, since T determines the value of f. We therefore find

$$\begin{aligned} \mathbb{E}\big[|f(x) - \mathbb{E}[f(x)]|\big] &= \mathbb{E}\big[|\mathbb{E}[f(x)|\mathcal{F}_\infty] - \mathbb{E}[f(y)]|\big] \\ &= \lim_{n\to\infty} \mathbb{E}\big[|\mathbb{E}[f(x)|\mathcal{F}_n] - \mathbb{E}[f(y)]|\big], \end{aligned} \tag{2.23}$$

since

$$\mathbb{E}[f(x)|\mathcal{F}_n] \to \mathbb{E}[f(x)|\mathcal{F}_\infty], \quad \text{as } n \to \infty, \tag{2.24}$$

almost surely and in L^1, see Theorem 5.5.7 of Durrett [19].

We let $x_s y$ be the configuration that agrees with x for the variables in $i_{s+1}, \ldots, i_t$, and with y for all other variables:

$$(x_s y)_i = \begin{cases} x_i & \text{if } i = i_k,\ s < k \leq t, \\ y_i & \text{otherwise.} \end{cases}$$

In particular, $f(x_0 y) = f(x)$, since T is a decision tree. Instead of applying the triangle inequality t times as in the proof of

Theorem 2.8, we now apply it $t \wedge n$ times, and later let $n \to \infty$. We find

$$\begin{aligned}\mathbb{E}\Big[\big|\mathbb{E}[f(x)|\mathcal{F}_n] - \mathbb{E}[f(y)]\big|\Big] &= \mathbb{E}\Big[\big|\mathbb{E}[f(x_0 y)|\mathcal{F}_n] - \mathbb{E}[f(y)]\big|\Big] \\ &\leq \sum_{s=1}^{t\wedge n} \mathbb{E}\Big[\big|\mathbb{E}[f(x_{s-1}y)|\mathcal{F}_n] - \mathbb{E}[f(x_s y)|\mathcal{F}_n]\big|\Big] \\ &\quad + \mathbb{E}\Big[\big|\mathbb{E}[f(x_{t\wedge n}y)|\mathcal{F}_n] - \mathbb{E}[f(y)]\big|\Big]. \end{aligned} \tag{2.25}$$

The configuration $x_{t\wedge n}y$ is independent of $\mathcal{F}_n$, and has law $\mathbb{P}$. Hence,

$$\mathbb{E}\Big[\big|\mathbb{E}[f(x_{t\wedge n}y)|\mathcal{F}_n] - \mathbb{E}[f(y)]\big|\Big] = 0.$$

We can handle the first term of (2.25) by using Jensens's inequality to bound

$$\begin{aligned}\mathbb{E}\Big[\big|\mathbb{E}[f(x_{s-1}y)|\mathcal{F}_n] - \mathbb{E}[f(x_s y)|\mathcal{F}_n]\big|\Big] & \\ &\leq \mathbb{E}\Big[\mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)| \,\big|\, \mathcal{F}_n\big]\Big] \\ &= \mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)|\big], \end{aligned} \tag{2.26}$$

by the law of total expectation. The proof now continues as the proof of Theorem 2.8:

$$\begin{aligned}\sum_{s=1}^{t\wedge n} & \mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)|\big] \\ &\leq \sum_{s=1}^{\infty}\sum_{i=1}^{\infty} \mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)| \mathbb{1}_{\{i_s = i\}}\big] \\ &= \sum_{s=1}^{\infty}\sum_{i=1}^{\infty} \mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big] \mathbb{E}\big[\mathbb{1}_{\{i_s = i\}}\big] \\ &= \sum_{i=1}^{\infty} \mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big] \mathrm{Rev}_i(T). \end{aligned} \tag{2.27}$$

where the interchange of sum and expectation is justified by Tonelli's theorem. The result follows by combining (2.25), (2.26) and (2.27).

We now suppose that f is $\{0,1\}$-valued, so that $f \in L^2(\Omega^{\mathbb{N}})$, and compute

$$\begin{aligned} 2\mathrm{Var}(f) = 2\mathbb{E}[f^2] - 2\mathbb{E}[f]^2 &= 2\mathbb{P}(f=1)\big(1-\mathbb{P}(f=1)\big) \\ &= \mathbb{E}\big[|f(x) - \mathbb{P}(f=1)|\big] \\ &= \mathbb{E}\big[|f(x) - f(y)|\big]. \end{aligned}$$

Using (2.20), we see that the factor 2 is compensated by the influence of i:

$$\mathrm{Inf}_i(f) = 1/2\,\mathbb{P}\big(f(x) \neq f(x^{i\mapsto\pi})\big) = 1/2\,\mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big].$$

□

2.5.2 *Infinite spaces*

We now consider the case where $|\Omega| = \infty$. Let $(\Omega, \mathcal{F}, \pi)$ be a probability space. Let $n \in \mathbb{N}$. For simplicity we take a n-fold product of these spaces, although we can also consider a countable product using the methods introduced earlier in this section. We consider the probability space $(\Omega^n, \mathcal{F}^n, \mathbb{P})$, where the σ-algebra is given by

$$\mathcal{F}^n = \sigma\big(\{\{\omega_i \in A\} : i \in [n], A \in \mathcal{F}\}\big),$$

and the probability measure is defined by

$$\mathbb{P}(\omega_1 \in A_1, \ldots, \omega_n \in A_n) = \prod_{i=1}^{n} \pi(A_i),$$

with $A_i \in \mathcal{F}$ for all $i \in [n]$. We again denote the expectation with respect to $\mathbb{P}$ by $\mathbb{E}$, and we define the space $L^p(\Omega^n)$ in the usual way. If $f \in L^2(\Omega^n)$, we define the inluence of i on f as in (2.19):

$$\mathrm{Inf}_i(f) = \mathbb{E}\big[E_i f^2 - (E_i f)^2\big].$$

The OSSS inequality holds in this setting, and can be proven in the same way as Theorem 2.18. However the definition of a decision tree has to be adapted, because in the case of uncountable Ω, we would have an uncountable tree. This not strictly problematic for our purposes, but it would be an uncommon notion. We therefore choose for a more abstract definition of a decision tree. Let

$$I_s = \{(i_1, \ldots, i_s) \in [n]^s \;:\; i_k \neq i_l \text{ for all } k \neq l\}.$$

Definition 2.19. *Let $i_1 \in [n]$. A decision tree T is a collection of measurable functions $(\phi_s)_{s=1}^n$, where $\phi_1 \equiv i_1 \in [n]$, and*

$$\phi_s : I_{s-1} \times \Omega^{s-1} \to [n], \qquad s = 2, \ldots, n,$$

such that

$$\phi_s(i_1, \ldots, i_{s-1}; x_{i_1}, \ldots, x_{i_{s-1}}) \neq i_1, \ldots, i_{s-1}, \qquad (i_1, \ldots, i_{s-1}) \in I_{s-1},$$

for all $(x_{i_1}, \ldots, x_{i_{s-1}}) \in \Omega^{s-1}$.

The sequence of variables that are revealed by T is given by $(i_1, \ldots, i_n)$, where

$$i_s := \phi_{s-1}(i_1, \ldots, i_{s-1}, x_{i_1}, \ldots, x_{i_{s-1}}).$$

The main difference to the original definition of decision trees is that the tree now does not terminates once it has determined the value of f. In order to encode this, we introduce a stopping with respect to the filtration $\mathcal{F}_s = \sigma(x_{i_1}, \ldots, x_{i_s})$. For a measurable $f : \Omega^n \to \mathbb{R}$ and a decision tree T we define

$$\tau := \tau(T, f) := \inf \Big\{ s \in [n] \; : f(x) = f(y) \\ \forall x, y \in \Omega^n : x_{i_k} = y_{i_k}, \, k = 1, \ldots, s \Big\}.$$

In this way, τ is the first time at which T has determined the value of f. The revealment of a variable i is then given by

$$\mathrm{Rev}_i(T) = \mathbb{P}(i = i_k \text{ for some } 1 \le k \le \tau).$$

We can now state the OSSS inequality for infinite spaces.

Theorem 2.20 (OSSS inequality for infinite spaces). *Let $f \in L^1(\Omega^n)$, let T be a decision tree, and let τ be the corresponding stopping time. Then*

$$\mathbb{E}\big[|f(x) - f(y)|\big] \le \sum_{i=1}^{n} \mathbb{E}\big[|f(x) - f(x^{i \mapsto \pi})|\big] \mathrm{Rev}_i(T), \tag{2.28}$$

where $x, y \in \Omega^n$ are independent and have law $\mathbb{P}$. In particular, if f is $\{0,1\}$-valued,

$$\mathrm{Var}(f) \le \sum_{i=1}^{n} \mathrm{Inf}_i(f) \mathrm{Rev}_i(T). \tag{2.29}$$

Proof. We can use the same proof as for Theorem 2.8, and we will omit most of it. The only change we make is to apply the triangle inequality τ times instead of t times. We then find

$$\mathbb{E}\big[|f(x) - f(y)|\big] \le \sum_{s=1}^{n} \sum_{i=1}^{n} \mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)| \mathbb{1}_{\{i_s = i\}} \mathbb{1}_{\{s \le \tau\}}\big].$$

Similar to (2.10), we obtain

$$\begin{aligned}
&\mathbb{E}\big[|f(x_{s-1}y) - f(x_s y)|\mathbb{1}_{\{i_s=i\}}\mathbb{1}_{\{s\leq\tau\}}\big] \\
&\quad = \mathbb{E}[\mathbb{1}_{\{i_s=i\}}\mathbb{1}_{\{s\leq\tau\}}]\mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big] \\
&\quad = \mathbb{E}\big[|f(x) - f(x^{i\mapsto\pi})|\big]\mathrm{Rev}_i(T).
\end{aligned}$$

where the factorization is valid, since τ is an $\mathcal{F}_s$-stopping time. The result for $\{0,1\}$-valued functions follows as in the proof of Theorem 2.18. □

The application we have in mind for this setting is the contact process, which has an underlying point process, so that we require Ω to be uncountably infinite. The above construction is sufficient for this purpose. We do mention however that a version of the OSSS inequality has been proven by Last, Peccati and Yogeshwaran that is native to Poisson point processes [40].

2.6 MONOTONIC MEASURES

So far we have only considered product measures. We will now see how we can loosen this assumption. Many percolation models do not have a product structure. The contact process and the corrupted compass model are two examples where the status of the vertices or edges are not independent, although for these models we can still find an underlying product measure. For yet other models, we can not even find an underlying product structure. An example of this is the Ising model, or more generally the random-cluster model. We will see that for measures satisfy a particular condition that is weaker than being a product measure, we can still prove the OSSS inequality.

In this section we take $\Omega^n = \{0,1\}^n$, equipped with the σ-algebra $\mathcal{F}$ being the power set, and a probability measure $\mathbb{P}$. For

$x, y \in \Omega$, we say that $x \leq y$ whenever $x_i \leq y_i$ for all $i \in [n]$. We call an event $A \in \mathcal{F}$ increasing, when

$$x \leq y \quad \Longrightarrow \quad \mathbb{1}_A(x) \leq \mathbb{1}_a(y), \quad x, y \in \Omega^n. \tag{2.30}$$

Similarly, a function $f\colon \Omega^n \to \mathbb{R}$ is called increasing, whenever

$$x \leq y \quad \Longrightarrow \quad f(x) \leq f(y), \quad x, y \in \Omega^n.$$

Definition 2.21. *A probability measure $\mathbb{P}$ is called positively-associated, whenever*

$$\mathbb{P}(A \cap B) \geq P(A)P(B) \tag{2.31}$$

for all increasing events $A, B \in \mathcal{F}$.

The inequality (2.31) is known as the FKG inequality, named after Fortuin, Kasteleyn and Ginibre, who introduced the inequality in a 1971 paper [24], along with a condition that implies the inequality, that we will introduce later in the section. Grimmett has written a detailed history on the origins of the inequality, including correspondences between him and the original authors [26].

The FKG is a very practical inequality in the analysis of percolation models, allowing us the decouple events and treat them separately, as long as the events are increasing. An even stronger property is when this inequality holds even if we condition on the state of a subset of the variables. For $I \subseteq [n]$ and $x \in \Omega^n$, let

$$\Omega_x^I := \{y \in \Omega^n \ : \ y_i = x_i \text{ for all } i \in I\}.$$

Definition 2.22. *A probability measure $\mathbb{P}$ is called strongly positively-associated, if for all $I \subseteq [n]$ and $x \in \Omega^n$ with $\mathbb{P}(\Omega_x^I) > 0$, the conditional measure $\mathbb{P}(\cdot \,|\, \Omega_x^I)$ is positively-associated, i.e., whenever*

$$\mathbb{P}\big(A \cap B \,|\, \Omega_x^I\big) \geq \mathbb{P}\big(A \,|\, \Omega_x^I\big)\mathbb{P}\big(B \,|\, \Omega_x^I\big) \tag{2.32}$$

for all increasing events $A, B \in \mathcal{F}$.

We recover positive-associativity by taking $I = \varnothing$, so that it is indeed a stronger property. In between these two properties we have the so-called downward FKG property, where we only condition on variables being 0.

Definition 2.23. *A probability measure $\mathbb{P}$ is called downward FKG if for all $I \subseteq [n]$ with $\mathbb{P}(\Omega_0^I) > 0$, the conditional measure $\mathbb{P}(\cdot \mid \Omega_0^I)$ is positively-associated.*

A closely related notion is that of the monotonicity of a measure.

Definition 2.24. *A probability measure is called monotonic, when*

$$\mathbb{P}\big(A \mid \Omega_x^I\big) \le \mathbb{P}\big(A \mid \Omega_y^I\big) \quad \textit{whenever } x \le y, \tag{2.33}$$

for all increasing events $A \in \mathcal{F}$.

It can be difficult to directly prove that a probability measure satisfies the above definitions. Luckily, there is a criterion that implies the above properties, and which is easier to check.

Definition 2.25. *A probability measure $\mathbb{P}$ satisfies the FKG lattice condition whenever*

$$\mathbb{P}\big(\{x \wedge y\}\big)\mathbb{P}\big(\{x \vee y\}\big) \ge \mathbb{P}\big(\{x\}\big)\mathbb{P}\big(\{y\}\big) \quad \textit{for all } x, y \in \Omega^n, \tag{2.34}$$

where $(x \wedge y)_i = x_i \wedge y_i$ and $(x \vee y)_i = x_i \vee y_i$, for all $i \in [n]$.

Theorem 2.26. *Let $\mathbb{P}$ be a probability measure with full support. The following statements are equivalent.*

a) The measure $\mathbb{P}$ satisfies the FKG lattice condition.

b) The measure $\mathbb{P}$ is strongly positively-associated.

c) *The measure $\mathbb{P}$ is monotonic.*

This is Theorem 2.24 of [26], and the proof can be found therein. A rather trivial example of a probability measure on $\{0,1\}^n$ that satisfies the FKG lattice condition is the product measure. If $\mathbb{P}$ is the product measure under which $\mathbb{P}(x_i = 1) = p_i$, then

$$\mathbb{P}\big(\{x \wedge y\}\big)\mathbb{P}\big(\{x \vee y\}\big) = \prod_{i=1}^{n} p_i^{x_i+y_i} = \mathbb{P}\big(\{x\}\big)\mathbb{P}\big(\{y\}\big).$$

An important and less trivial example is the random-cluster measure with parameter $q \geq 1$. This is a measure on $\{0,1\}^E$, where E is the edge set of a finite graph. For this parameter choice, this measure favours clusters of edges, which makes the measure satisfy the FKG lattice condition. This is stated in Theorem of 3.8 of [26], but was already proven by Fortuin in 1972 [23].

If a measure satisfies the FKG lattice condition, and is thus monotonic, it also satisfies the OSSS inequality, although a slight modification is necessary. In the previous sections the influence of a variable i was obtained by resampling the ith variable. Now that we no longer have a product measure, we can no longer resample only one variable. We therefore replace the influence of i on a Boolean function f by the covariance between x_i and f.

Theorem 2.27. *Let $f : \Omega^n \to \{0,1\}$ be increasing, and let T be a decision tree that determines the value of f. Suppose that $\mathbb{P}$ is a monotonic measure on Ω^n. Then,*

$$\mathrm{Var}(f) \leq \sum_{i=1}^{n} \mathrm{Cov}(f, x_i)\mathrm{Rev}_i(T). \tag{2.35}$$

This generalization of the OSSS inequality has been proven by Duminil-Copin, Raoufi and Tassion [15]. The proof follows in a broad sense the proof of Theorem 2.8, the original proof by O'Donnell et al.. However, since we no longer have a product measure, much more care has to be taken to decouple the configuration $x_s y$ from the variables we have seen before. This is done with an appropriate coupling using uniform random variables, that encode the conditional probabilities $\mathbb{E}[x_{i_s} \mid \mathcal{F}_{s-1}]$. The strong positive-associativity of the measure is then essential, since there is no control over the conditioning.

3 BERNOULLI PERCOLATION

Bernoulli percolation is the prototypical percolation model. In Bernoulli bond percolation, the edges of a graph are removed with probability $1 - p$, and thus kept with probability p, independently of each other. In Figure 1.2 this model is shown on the graph $\mathbb{Z}^2$ with nearest neighbour edges, for several values of p. In Bernoulli site percolation, the vertices are removed instead of the edges. We restrict ourselves to Bernoulli bond percolation in this chapter, although all results we will encounter are valid for Bernoulli site percolation as well.

Bernoulli percolation is one of the simplest probabilistic models that exhibits a phase transition: for small p there exist only finite clusters, but when we increase p, at some point an infinite cluster will appear. If we consider a regular tree as the underlying graph, we see that Bernoulli percolation on this tree is just a branching process with binomially distributed offspring. For this process, we know that the process dies out if and only if the expected number of children is at most 1. This is precisely the percolation phase transition if we view it from the point of view of Bernoulli percolation. Moreover, if the expected number of children in the branching process is less than 1, the process dies out exponentially quickly; we say that the phase transition is sharp. In this chapter we will prove that this not only holds on trees, but for general transitive graphs using Boolean function theory.

Bernoulli percolation was first introduced in a mathematical rigorous way by Broadbent and Hammersley in 1957 [11]. Many questions regarding the model have since been answered, and many more remain. We will restrict ourselves to introducing the basic quantities and definitions, in particular those required to prove the sharp phase transition that the model exhibits. For a broader exposition of the subject, we refer to *Percolation* by Grimmett [25].

3.1 INTRODUCTION

Let $G = (V, E)$ be an infinite, locally finite, connected, vertex-transitive graph. A locally finite graph is a graph for which every vertex has finitely many neighbours. A vertex-transitive graph looks the same from every vertex: for all $v, w \in V$ there exists a graph automorphism $\phi : V \to V$ such that $\phi(v) = \phi(w)$. In particular, every vertex of a vertex-transitive graph has the same degree. For simplicity we assume that G is vertex-transitive, although the results that follow are also valid for quasi-transitive graphs. We fix an arbitrary vertex of V as the origin and denote it by 0. For $x, y \in V$, we define $\mathrm{d}(x, y)$ to be the graph distance between x and y, that is, the length of a shortest path between x and y in G. We define the ball and the sphere of radius n around $x \in V$ as

$$\Lambda_n^x := \{y \in V \,:\, \mathrm{d}(x,y) \le n\}, \quad \partial\Lambda_n^x := \{y \in V \,:\, \mathrm{d}(x,y) = n\}.$$

For $x = 0$ we drop part of the notation: $\Lambda_n = \Lambda_n^0$ and $\partial\Lambda_n = \partial\Lambda_n^0$.

Let $0 \le p \le 1$. We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}_p)$, where

$$\Omega = \{0,1\}^E, \quad \mathcal{F} = \sigma\big(\{\{\omega_e = 1\} \,:\, e \in E\}\big),$$

and the measure is uniquely defined by setting

$$\mathbb{P}_p(\omega_e = 1, \omega_f = 0, \ \forall e \in I, \forall f \in J) = p^{|I|}(1-p)^{|J|},$$

for all finite $I, J \subseteq E$. We say that an edge e is open whenever $\omega_e = 1$, and closed otherwise. We are interested in the connectivity properties of the graph obtained by keeping only the open edges. For $x, y \in V$, we say that $x \longleftrightarrow y$ whenever there exists a path from x to y using only open edges. Similarly, for $A \subseteq V$, we say that $x \longleftrightarrow A$ whenever there exists $y \in A$ with $x \longleftrightarrow y$. We say that $x \longleftrightarrow \infty$, whenever $x \longleftrightarrow \partial\Lambda_n^x$ for all $n \in \mathbb{N}$. We define

$$\theta_n(p) := \mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n).$$

The percolation function is defined by

$$\theta(p) := \lim_{n\to\infty} \theta_n(p) = \mathbb{P}_p(0 \longleftrightarrow \infty).$$

The critical point is given by

$$p_c := \sup\{p \ : \ \mathbb{P}_p(0 \longleftrightarrow \infty) = 0\}.$$

This is the point at which the phase transition occurs. If d denotes the degree of an arbitrary vertex in the graph, we can prove that $p_c \geq 1/(d-1)$, by comparing the exploration process that explores the cluster of 0 with a branching process. First of all this shows that p_c is bounded away from zero. Secondly, we see that for some graphs $p_c = 1$. For example, the graph $G = (\mathbb{Z}, E)$, where E is the set of pairs of neighbouring integers, has degree $d = 2$, so that $p_c = 1$. On the other hand, for a large class of graphs, one can show that $\theta(p) > 0$ for some $p < 1$. This can be done with Peierls' argument, which entails counting the number of blocking surfaces that separate the origin from infinity. In particular this argument can be used to prove that $p_c < 1$ for the graph $\mathbb{Z}^d$ with nearest neighbour edges and $d \geq 2$.

The exact value of p_c is often intractable, but we will mention the celebrated result that $p_c = 1/2$ for $\mathbb{Z}^2$ with nearest neighbour edges. This was proven by Kesten in 1980 [38], building on work of Harris [28]. Without going into detail, the symmetry properties of the graph $\mathbb{Z}^2$ allow us to prove this result.

The function $p \mapsto \theta_n(p)$ is increasing in p, and hence, so is $\theta(p)$. Furthermore, $\theta_n(p)$ is a polynomial in p and is therefore differentiable. This property is not maintained when taking the limit $n \to \infty$, however, since $\lim_{n\to\infty} \theta_n(p)$ is a decreasing limit of continuous functions, and since $\theta_n(p)$ is increasing in p, we have that $\theta_n(p)$ is right-continuous. For $p < p_c$, $\theta(p) = 0$, so that $\theta(p)$ is continuous on $[0, p_c)$. It can also be shown that $\theta(p)$ is left-continuous in the supercritical phase $(p_c, 1]$. All things considered, we know that $\theta(p)$ is continuous at $p \neq p_c$, which leaves the question if $\theta(p)$ is also continuous at p_c. Since $\theta(p) = 0$ for $p < p_c$ and $\theta(p)$ is right-continuous, this is equivalent to $\mathbb{P}_{p_c}(0 \longleftrightarrow \infty) = 0$. This is an open question in general, and arguably one of the largest open problems in probability theory. For $\mathbb{Z}^2$ however, we know that $\theta(p_c) = 0$ [28], again using the symmetry properties of this graph. Furthermore, for $\mathbb{Z}^d$, with $d \geq 11$, we can also prove $\theta(p_c) = 0$. In this case, the large dimension allows for a sufficient amount of control over the dependencies between open paths [22]. This technique is know as the lace expansion. We refer to [29] for an extensive background on high-dimensional percolation.

We will focus on the sharpness of the phase transition on general transitive graphs. This means that for $p < p_c$ the clusters are exponentially small.

Theorem 3.1. *Consider Bernoulli bond percolation on a transitive graph G with parameter p. If $p < p_c$, then there exists a constant $c > 0$ such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn).$$

This theorem was first proven by Menshikov [45] and independently by Aizenman and Barsky [1] in 1986 and 1987 respectively. Both proofs are rather lengthy, and heavily rely on the specifics of the measure $\mathbb{P}_p$. It is not true that these proofs only work for product measures, since Aizenman, Barsky and Fernandez have also managed to apply this proof strategy to Ising-type models [2], but carrying over this strategy to other models with dependencies has been difficult. More recently, Duminil-Copin and Tassion have given a shorter and very elegant proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model, which is worth reading [17, 18]. However, this proof is also rather specific to these models, since it relies on decoupling the configuration inside a set $S \subset V$ from the configuration on S^c by conditioning on the configuration on the boundary of S. The proof using the OSSS inequality has proven to be more broadly applicable, which is the reason why we state this proof here. Bernoulli percolation will be one of the simplest settings where we can apply this proof strategy, and it will therefore serve as a stepping stone to more complicated models.

What all these proofs have in common, is that they all feature a differential inequality from which the sharpness of the phase transition follows. In our case we will take the derivative of $\theta_n(p)$ to p. If this derivative is large, we see a large change in the behaviour of the model for a small increase of p. Asymptotically, this should give us the sharpness of the phase transition as

$n \to \infty$. In order to implement this strategy, we first need to get a handle on the derivative of $\theta_n(p)$ to p. This will be done by means of Russo's formula. In order to state it, we introduce two definitions that we have already seen in the context of Boolean functions. We say that an event $A \in \mathcal{F}$ is increasing, whenever $\mathbb{1}_A(\omega) \le \mathbb{1}_A(\omega')$ for all $\omega \le \omega'$, where the latter inequality is interpreted pointwise. We call an edge $e \in E$ pivotal for an event A, whenever A occurs depending on the status of e:

$$\{e \text{ is pivotal for } A\} := \left\{\omega \in \Omega : \mathbb{1}_A(\omega^{e\mapsto 0}) \neq \mathbb{1}_A(\omega^{e\mapsto 1})\right\}, \quad e \in E,$$

where $\omega^{e\mapsto 0}$ is the configuration obtained taking ω and setting $\omega_e = 0$, and similarly for $\omega^{e\mapsto 1}$. Note that this event is independent of ω_e.

Proposition 3.2 (Russo's formula)**.** *Let $A \in \mathcal{F}$ be an increasing event depending on the state of finitely many edges, and let $0 < p < 1$. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(A) = \sum_{e \in E} \mathbb{P}_p(e \textit{ is pivotal for } A).$$

Proof. We first note that, since A depends on the state of only finitely many edges, $\mathbb{P}_p(A)$ is a polynomial in p, and is therefore differentiable. Let $e_1, \dots, e_n$ denote these edges. We set $\mathbf{p} = (p_1, \dots, p_n)$, with $0 < p_i < 1$ for all $i = 1, \dots, n$. We first consider the product probability measure $\mathbb{P}_{\mathbf{p}}$ under which $\omega_{e_i} = 1$ with probability p_i. Later we recover $\mathbb{P}_p$ by setting $p_i = p$ for all i. In order to make sense of the derivative to p_i, we need to couple the percolation configuration for different values of p_i. This is done by introducing a uniformly distributed random variable $U_i \in [0,1]$ that is independent of the state of the other edges. The coupling is obtained by setting $\omega_{e_i} = 1$ whenever $U_i \le p_i$.

We abuse notation and also write $\mathbb{P}_{\mathbf{p}}$ for the probability measure where the coupling is defined. Let ω_{p_i} be the configuration where $\omega_{e_i} = 1$ whenever $U_i \le p_i$, and all other edges e_j are open with probability p_j. For $i = 1, \ldots, n$, we compute

$$\frac{\partial}{\partial p_i}\mathbb{P}_{\mathbf{p}}(A) = \lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}_{\mathbf{p}}(\omega_{p_i} \notin A,\, \omega_{p_i+h} \in A)$$

The configurations ω_{p_i} and ω_{p_i+h} can only differ at edge e_i. It follows that e is pivotal for A. Furthermore since A is increasing, we see that $(\omega_{p_i})_{e_i} = 0$ and $(\omega_{p_i+h})_{e_i} = 1$. By the independence of ω_e and the pivotality of e we obtain

$$\begin{aligned}\frac{\partial}{\partial p_i}\mathbb{P}_{\mathbf{p}}(A) &= \lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}_{\mathbf{p}}(e \text{ is pivotal for } A)\mathbb{P}_{\mathbf{p}}(p_i \le U_i \le p_i + h) \\ &= \mathbb{P}_{\mathbf{p}}(e_i \text{ is pivotal for } A). \end{aligned} \tag{3.1}$$

We conclude by observing

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(A) = \sum_{i=1}^{n}\frac{\partial}{\partial p_i}\mathbb{P}_{\mathbf{p}}(A)\Big|_{\substack{p_j=p \\ j=1,\ldots,n}} = \sum_{e\in E}\mathbb{P}_p(e \text{ is pivotal for } A).$$

$\square$

In the case where A depends on infinitely many edges, we do not know if $\mathbb{P}_p(A)$ is differentiable with respect to p. Indeed, if $A = \{0 \longleftrightarrow \infty\}$, even continuity of $\mathbb{P}_p(A)$ is an important open problem. Nevertheless, in this case we can still give a weaker form of Russo's formula in terms of the lower-right Dini derivative, which is also useful.

Proposition 3.3. *Let $A \in \mathcal{F}$ be an increasing event, and let $0 \leq p < 1$. Then,*

$$\begin{aligned} D_+\mathbb{P}_p(A) &:= \liminf_{h\downarrow 0} \frac{1}{h}\big(\mathbb{P}_{p+h}(A) - \mathbb{P}_p(A)\big) \\ &\geq \sum_{e\in E} \mathbb{P}_p(e \textit{ is pivotal for } A). \end{aligned}$$

Proof. This result is obtained using a variation of the previous proof. Let $e_1, e_2, \dots$ be an enumeration of the edges on which A depends. Let $n \in \mathbb{N}$, and let $\mathbf{p} = (p_1, \dots, p_n)$, with $0 \leq p_i < 1$ for all $i = 1, \dots, n$. Denote by $\mathbb{P}_\mathbf{p}$ the product measure under which $\omega_{e_i} = 1$ with probability p_i for $i = 1, \dots, n$, and $\omega_{e_i} = 1$ with probability p for $i > n$. Let e_i denote the ith unit vector. Then similarly to (3.1), we have

$$\liminf_{h\downarrow 0} \frac{1}{h}\big(\mathbb{P}_{\mathbf{p}+h\mathrm{e}_i}(A) - \mathbb{P}_\mathbf{p}(A)\big) = \mathbb{P}_\mathbf{p}(e_i \text{ is pivotal for } A).$$

Since A is increasing, it follows for all $n \in \mathbb{N}$, that

$$\begin{aligned} \liminf_{h\downarrow 0} \frac{1}{h}\big(\mathbb{P}_{p+h}(A) - \mathbb{P}_p(A)\big) &\geq \sum_{i=1}^{n} \liminf_{h\downarrow 0} \frac{1}{h}\big(\mathbb{P}_{\mathbf{p}+h\mathrm{e}_i}(A) - \mathbb{P}_\mathbf{p}(A)\big) \\ &= \sum_{i=1}^{n} \mathbb{P}_\mathbf{p}(e_i \text{ is pivotal for } A). \end{aligned}$$

The result follows by letting $n \to \infty$. □

3.2 PROOF OF THE SHARP PHASE TRANSITION

We now present the proof by Duminil-Copin, Raoufi and Tassion [15] in the case of Bernoulli percolation. We note that this proof is also valid for the random-cluster model by using the OSSS inequality for monotonic measures, Theorem 2.27, and by replacing Russo's formula with an appropriate derivative formula for this model.

For $n \in \mathbb{N}$, let $S_n := S_n(p) := \sum_{k=1}^{n} \theta_k(p)$. We prove the following differential inequality from which Theorem 3.1 will follow.

Proposition 3.4. *Consider Bernoulli percolation with paramater* $0 < p < 1$. *Then,*

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) \geq \frac{1}{4p(1-p)} \frac{n}{S_n} \theta_n(p)\big(1-\theta_n(p)\big),$$

for all $n \in \mathbb{N}$.

Before proving this differential inequality we give some introductory considerations on how to obtain a differential inequality for θ_n such as the above inequality. We let $A_n := \{0 \longleftrightarrow \partial\Lambda_n\}$, and $f_n = \mathbb{1}_{A_n}$. Then, A_n depends only on the edges inside Λ_n. Denote this set of edges by E_n. For the purpose of proving the differential inequality, we can restrict ourselves to the probability space on $\{0,1\}^{E_n}$. Russo's formula gives

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) = \frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(A_n) = \sum_{e \in E_n} \mathbb{P}_p(e \text{ is pivotal for } A_n).$$

Since ω_e is independent of the pivotality of e, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) &= \sum_{e \in E} \mathbb{P}_p(e \text{ is pivotal for } A_n) \\ &= \sum_{e \in E} \frac{1}{2p(1-p)} \mathbb{P}_p(f_n(\omega) \neq f_n(\omega^{e \mapsto \pi})),\end{aligned}$$

where $\omega^{e\mapsto\pi}$ is the configuration obtained from ω by resampling ω_e independently with law $\mathrm{Ber}(p)$. Using Proposition 2.14, we find

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) = \frac{1}{p(1-p)} \sum_{e \in E} \mathrm{Inf}_e(f_n). \tag{3.2}$$

If we use the Poincaré inequality, Theorem 2.5, we obtain the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) \geq \frac{1}{p(1-p)}\theta_n(p)\big(1-\theta_n(p)\big).$$

However, this inequality is not strong enough to imply the sharpness of the phase transition. Using the stronger OSSS inequality will allow us to gain the factor n/S_n, which in turn will be enough to prove Theorem 3.1.

3.2.1 *Bound on the Revealment*

We now give the proof of Proposition 3.4. Since f_n only depends on the edges E_n, we can apply Theorem 2.16 for finite product spaces. We then also need a decision tree that determines the value of f_n. If T is such a decision tree, Theorem 2.16 implies

$$\theta_n(p)\big(1-\theta_n(p)\big) \leq \sum_{e\in E_n} \mathrm{Inf}_e(f_n)\mathrm{Rev}_e(T). \tag{3.3}$$

The aim is to find a decision tree such that we can find a strong bound on the revealment, that is uniform in the edges. In this way we can pull the revealment out of the sum, and recover $\theta_n'(p)$ by (3.2). If we only use one decision tree, the first edge that is revealed has $\mathrm{Rev}_e(T) = 1$, so that we cannot find a meaningful uniform bound on the revealment. We therefore introduce several decision trees, $T_1, \ldots, T_n$, and average over these trees. For $1 \leq k \leq n$, we let T_k be the decision tree that determines the value of f_n by exploring the cluster of $\partial\Lambda_k$. If 0 is connected to $\partial\Lambda_n$, this connection must go through $\partial\Lambda_k$, so that T_k indeed determines the value of f_n. We now give a more formal description of this exploration process, which in turn can be made to fit Definition 2.6. For $x,y \in V$ and $A \subset E$, we say that $x \overset{A}{\longleftrightarrow} y$ whenever there exists a path from x to y using only open edges in A. Similarly, for $B \subseteq V$, we say that $x \overset{A}{\longleftrightarrow} B$ whenever there exists $y \in B$ with $x \overset{A}{\longleftrightarrow} y$. We denote by $\mathcal{A}$ the set of active edges; that is, edges that have not been revealed yet, but are connected to $\partial\Lambda_k$ using revealed edges. We denote the set of revealed edges by $\mathcal{R}$. We fix an arbitrary ordering of the edges. The exploration process of T_k is then given by the pseudocode of Algorithm 1. A possible realization of this process is shown in Figure 3.1.

$\mathcal{A} := \{\{x,y\} \in E_n \,:\, x \in \partial\Lambda_k \text{ or } y \in \partial\Lambda_k\}$;
$\mathcal{R} := \varnothing$;
while $\mathcal{A} \neq \varnothing$ **do**
  Take minimal $e \in \mathcal{A}$;
  Reveal ω_e;
  $\mathcal{R} := \mathcal{R} \cup \{e\}$;
  $\mathcal{A} := \mathcal{A} \setminus \{e\}$;
  $\mathcal{A} := \mathcal{A} \cup \{\{x,y\} \in E_n \setminus \mathcal{R} \,:\, x \overset{\mathcal{R}}{\longleftrightarrow} \partial\Lambda_k \text{ or } y \overset{\mathcal{R}}{\longleftrightarrow} \partial\Lambda_k\}$;
  if $0 \overset{\mathcal{R}}{\longleftrightarrow} \partial\Lambda_n$ **then return** 1;
end
return 0;

Algorithm 1: The exploration algorithm T_k.

We now proceed to bound the revealment of the edges. If an edge e is revealed by T_k, then one of its endpoints is connected to $\partial\Lambda_k$. Hence, for $e = \{x,y\}$,

$$\begin{aligned}\mathrm{Rev}_e(T_k) &\leq \mathbb{P}_p(x \longleftrightarrow \partial\Lambda_k) + \mathbb{P}_p(y \longleftrightarrow \partial\Lambda_k)\\ &\leq \mathbb{P}_p\big(x \longleftrightarrow \partial\Lambda^x_{d(x,\partial\Lambda_k)}\big) + \mathbb{P}_p\big(y \longleftrightarrow \partial\Lambda^y_{d(y,\partial\Lambda_k)}\big),\end{aligned} \tag{3.4}$$

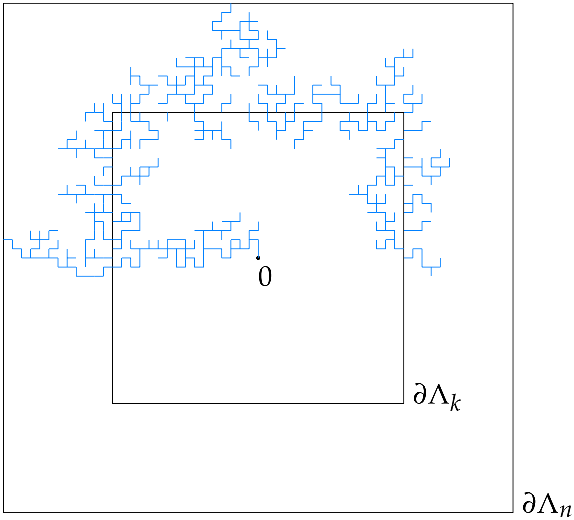


Figure 3.1: The decision tree T_k exploring the cluster of $\partial\Lambda_k$.

where $\mathrm{d}(x, \partial\Lambda_k)$ and $\mathrm{d}(y, \partial\Lambda_k)$ are the distances from x to $\partial\Lambda_k$ and y to $\partial\Lambda_k$ respectively. We now sum over k, so that we essentially average over $k = 1 \ldots, n$. We find

$$\begin{aligned}
\sum_{k=1}^{n} \mathrm{Rev}_e(T_k) &\leq \sum_{k=1}^{n} \mathbb{P}_p\big(x \longleftrightarrow \partial\Lambda^x_{d(x,\partial\Lambda_k)}\big) + \sum_{k=1}^{n} \mathbb{P}_p\big(y \longleftrightarrow \partial\Lambda^y_{d(y,\partial\Lambda_k)}\big) \\
&\leq 2\sum_{k=1}^{n} \mathbb{P}_p(x \longleftrightarrow \partial\Lambda^x_k) + 2\sum_{k=1}^{n} \mathbb{P}_p(y \longleftrightarrow \partial\Lambda^y_k) \\
&= 4\sum_{k=1}^{n} \mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_k) = 4S_n, \qquad (3.5)
\end{aligned}$$

by translation invariance. Summing (3.3) with $T = T_k$ over k gives

$$
\begin{aligned}
n\theta_n(p)\big(1-\theta_n(p)\big) &\le \sum_{k=1}^{n}\sum_{e\in E_n} \mathrm{Inf}_e(f_n)\mathrm{Rev}_e(T_k) \\
&\le 4S_n \sum_{e\in E_n} \mathrm{Inf}_e(f_n) = 4p(1-p)S_n \frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p),
\end{aligned} \tag{3.6}
$$

by (3.2). We thus arrive at the differential inequality of Proposition 3.4. □

3.2.2 *Analysis of the differential inequality*

We now derive Theorem 3.1 from the differential inequality. Let $p < p_c$. For simplicity we bound

$$\frac{1}{4p(1-p)} \ge 1,$$

as this will only result in slightly weaker constants. Let $p < p_1 < p_2 < p_c$. We further bound for all $p' \le p_2$,

$$1-\theta_n(p') \ge 1-\theta_1(p') \ge 1-\theta_1(p_2) = (1-p_2)^{2d} =: C.$$

From (4.3), it follows that

$$\frac{\mathrm{d}}{\mathrm{d}p}\log\theta_n(p) \ge C\frac{n}{S_n},$$

for all $n \in \mathbb{N}$. Integrating the above inequality from p_1 to p_2 gives

$$-\log\theta_n(p_1) \ge \log\theta_n(p_2) - \log\theta_n(p_1) \ge C(p_2-p_1)\frac{n}{S_n(p_2)},$$

so that

$$\theta_n(p_1) \leq \exp\left(-C(p_2 - p_1)\frac{n}{S_n(p_2)}\right). \tag{3.7}$$

If $S_n(p_2)$ is bounded in n, i.e., if $\sum_{k=0}^{\infty} \theta_k(p_2)$ converges, the desired exponential decay would follow from the above inequality. In fact, it suffices if $S_n(p_2) \leq n^{1-\alpha}$ for $0 < \alpha < 1$ and n large enough: from (3.7) it then follows that

$$\theta_n(p_1) \leq \exp\left(-C(p_2 - p_1)n^{\alpha}\right),$$

for n large enough, so that $\sum_{k=0}^{\infty} \theta_k(p_1)$ converges. We can then bootstrap this result by using the inequality (3.7) again to find the desired exponential decay. This motivates the definition of the following critical point, which we will show to be equal to p_c:

$$\tilde{p}_c := \sup\left\{p \,:\, \limsup_{n\to\infty} \frac{\log S_n}{\log n} < 1\right\}.$$

If $p_2 < \tilde{p}_c$, there exists $0 < \alpha < 1$ such that $S_n(p_2) \leq n^{1-\alpha}$ for n large enough. It then follows that we have stretched exponential decay at p_1 and exponential decay at p.

It remains to show that $p_c = \tilde{p}_c$. Let $\tilde{p}_c < p_1 < p$, and set $C := C(p) := (1-p)^{2d}$. Using (4.3), we find

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}\sum_{k=1}^{n}\frac{\theta_k(p)}{k} \geq C\sum_{k=1}^{n}\frac{\theta_k(p)}{S_k} \geq C\sum_{k=1}^{n}\int_{S_k}^{S_{k+1}}\frac{1}{t}\,\mathrm{d}t \\
= C\sum_{k=1}^{n}(\log S_{k+1} - \log S_k) \\
= C(\log S_{n+1} - \log S_1) \geq C\log S_{n+1}
\end{aligned}$$

We define $T_n(p) := \frac{1}{\log n} \sum_{k=1}^{n} \frac{\theta_k(p)}{k}$, and find

$$\frac{\mathrm{d}}{\mathrm{d}p} T_n(p) \geq C \frac{\log S_{n+1}}{\log n}.$$

Integrating the above inequality from p_1 to p gives

$$T_n(p) - T_n(p_1) \geq C \frac{\log S_{n+1}(p_1)}{\log n}(p - p_1).$$

Note that for all p, $\frac{1}{\log n} T_n(p) \to \theta(p)$ for $n \to \infty$, so that

$$\begin{aligned} \theta(p) \geq \theta(p) - \theta(p_1) &= \limsup_{n \to \infty} (T_n(p) - T_n(p_1)) \\ &\geq C \limsup_{n\to\infty} \frac{\log S_{n+1}(p_1)}{\log n}(p - p_1) \\ &\geq C \cdot (p - p_1) > 0, \end{aligned} \tag{3.8}$$

since $p_1 > \tilde{p}_c$. Because $p > \tilde{p}_c$ is arbitrary, it follows that $\tilde{p}_c = p_c$. This concludes the proof of Theorem 3.1. However, we proved slightly more than this theorem. Namely, the above computation also gave us an interesting result on the behaviour of the model in the supercritical parameter range $p > p_c$. If we let $p_1 \to p_c$ in 3.8, we find the following result.

Proposition 3.5. *Consider Bernoulli bond percolation on a transitive graph G with parameter p. There exists a constant $c > 0$ such that for all $p > p_c$,*

$$\theta(p) = \mathbb{P}_p(0 \longleftrightarrow \infty) \geq c(p - p_c). \tag{3.9}$$

The above bound is known as a mean-field lower bound, since is is satisfied with equality in high dimensions. To be specific, for Bernoulli percolation on $\mathbb{Z}^d$ with $d \geq 11$, we know the value of the critical exponent

$$\beta := \limsup_{p \downarrow p_c} \frac{\log \theta(p)}{\log(p - p_c)} = 1.$$

See [29] for details. For general graphs, Proposition 3.5 implies the bound $\beta \leq 1$.

3.3 THE HUTCHCROFT PROOF

There is a different way of utilizing the OSSS inequality to prove the sharp phase transition for Bernoulli percolation. This is the proof by Hutchcroft [37]. It uses a different decision tree as in the previous section, and is to some extent inspired by the original proof for the sharpness of the phase transition by Aizenman and Barsky [1]. This proof is also valid for the random-cluster model, but we will again present the proof for the case of Bernoulli percolation. Another appeal of this proof is that it gives two inequalities between certain critical exponents.

Let $\mathcal{C}$ be the cluster of 0:

$$\mathcal{C} = \{x \in V \;:\; 0 \longleftrightarrow x\}.$$

We will prove the following result regarding the sharpness of the phase transition.

Theorem 3.6. *Consider Bernoulli bond percolation on a transitive graph G with parameter p. If $p < p_c$, then there exists a constant $c > 0$ such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}_p(|\mathcal{C}| \geq n) \leq \exp(-cn).$$

This statement is stronger than Theorem 3.1, since $\mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n) \le \mathbb{P}_p(|\mathcal{C}| \ge n)$. However, if $|\partial\Lambda_n|$ grows slower than exponentially in n, then both statements are equivalent. We will prove the following differential inequality, from which Theorem 3.6 will follow.

Proposition 3.7. *Consider Bernoulli percolation with parameter* $0 < p < 1$. *Then for all* $\lambda > 0$, *and all* $n \in \mathbb{N}$,

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(|\mathcal{C}| \ge n) \ge \frac{1}{8}\frac{1}{p(1-p)}\left(\frac{1-e^{-\lambda}}{\mathbb{E}_p[1-\exp(-\lambda|\mathcal{C}|/n)]} - 1\right)\mathbb{P}_p(|\mathcal{C}| \ge n).$$

Proof. In order to prove this inequality, we define the Boolean function $f_n := \mathbb{1}\{|\mathcal{C}| \ge n\}$. Note that this function only depends on the edges in Λ_n, which we again denote by E_n. We will colour a vertex in Λ_n green with probability $q > 0$, independent of the the other vertices, and independent of the percolation configuration. Let $\mathcal{G}$ denote the set of green vertices. To encode this setting, we consider the probability space $(\Omega \times \Xi, \mathcal{F}, \mathbb{P}_{p,q})$, where $\Omega = \{0,1\}^{E_n}$, $\Xi = \{0,1\}^{\Lambda_n}$, the σ-algebra $\mathcal{F}$ is given by the power set, and $\mathbb{P}_{p,q}$ is given by

$$\mathbb{P}_{p,q}(\omega_e = 1, \omega_f = 0,\ \forall e \in I, \forall f \in I^c, |\mathcal{G}| = k) = p^{|I|}(1-p)^{|E_n|-|I|}q^k(1-q)^{|\Lambda_n|-k},$$

for all $I \subseteq E_n$, and $k = 0, 1, \ldots, |\Lambda_n|$. In this way, we retain the product structure, although not all variables have the same law. Therefore, Theorem 2.16 is not directly applicable, but its proof can be made to accompany this situation without any problems. Let $\mathbb{E}_{p,q}$ denote the expectation with respect to $\mathbb{P}_{p,q}$. We define a second Boolean function $g_n := \mathbb{1}\{0 \xleftrightarrow{\Lambda_n} \mathcal{G}\}$. It then follows from Theorem 2.16, that for a decision tree T determing the value of g_n, we have

$$\mathrm{Cov}(f_n, g_n) \leq 2 \sum_{e \in E_n} \mathrm{Inf}_e(f_n)\mathrm{Rev}_e(T) + 2 \sum_{v \in \Lambda_n} \mathrm{Inf}_v(f_n)\mathrm{Rev}_v(T). \tag{3.10}$$

We first note that the colour of v, ξ_v, is independent of f_n, so that $\mathrm{Inf}_v(f_n) = 0$ for all $v \in \Lambda_n$. Hence, the second term in the above inequality vanishes. Furthermore, the decision tree can reveal all ξ_v, without deteriorating the bound.

3.3.1 *Bound on the Revealment*

We now describe the decision tree T. It starts by revealing ξ_v for all $v \in \Lambda_n$. We then know which vertices in Λ_n are green. The decision tree subsequently explores the clusters of these vertices inside Λ_n. In particular, we then know if $\mathcal{C}_n$ contains a green vertex. The pseudocode of T is given in Algorithm 2, and a possible realization is shown in Figure 3.2.

By exploring from the green vertices, we can find a good uniform bound on the revealment of an edge. For $e = \{x, y\}$, we the obtain

$$\mathrm{Rev}_e(T) \leq \mathbb{P}_{p,q}\big(x \xleftrightarrow{\Lambda_n} \mathcal{G}\big) + \mathbb{P}_{p,q}\big(y \xleftrightarrow{\Lambda_n} \mathcal{G}\big).$$

We set $q := 1 - e^{-\lambda/n}$, with $\lambda > 0$. Since the colour of the vertices is independent of the percolation configuration, we find

$$\mathrm{Rev}_e(T) \leq \mathbb{E}_p\big[1 - \exp(-\lambda|\mathcal{C}_n^x|/n)\big] + \mathbb{E}_p\big[1 - \exp(-\lambda|\mathcal{C}_n^y|/n)\big],$$

where

$$\mathcal{C}_n^v := \{w \in \Lambda_n \,:\, v \xleftrightarrow{\Lambda_n} w\},$$

for $v \in \Lambda_n$ **do**
    Reveal ξ_v;
end
$\mathcal{G} := \{x \in \Lambda_n : \xi_v = 1\}$;
$\mathcal{A} := \{\{x, y\} \in E_n : x \in \mathcal{G} \text{ or } y \in \mathcal{G}\}$;
$\mathcal{R} := \varnothing$;
while $\mathcal{A} \neq \varnothing$ **do**
    Take minimal $e \in \mathcal{A}$;
    Reveal ω_e;
    $\mathcal{R} := \mathcal{R} \cup \{e\}$;
    $\mathcal{A} := \mathcal{A} \backslash \{e\}$;
    $\mathcal{A} := \mathcal{A} \cup \{\{x, y\} \in E_n \backslash \mathcal{R} : x \overset{\mathcal{R}}{\longleftrightarrow} \mathcal{G} \text{ or } y \overset{\mathcal{R}}{\longleftrightarrow} \mathcal{G}\}$;
    if $0 \overset{\mathcal{R}}{\longleftrightarrow} \mathcal{G}$ **then return** 1;
end
return 0;

Algorithm 2: The exploration algorithm T.

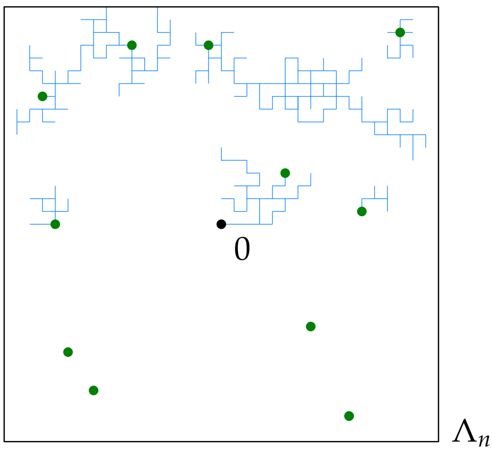


Figure 3.2: The decision tree T exploring the clusters of green vertices.

for $v \in \Lambda_n$. Disregarding the restriction of the connections being inside Λ_n, and using translation invariance, we obtain

$$\mathrm{Rev}_e(T) \leq 2\mathbb{E}_p\big[1 - \exp(-\lambda|\mathcal{C}|/n)\big]. \tag{3.11}$$

We now turn our attention to the covariance between f_n and g_n. We compute

$$\begin{aligned}
\mathrm{Cov}(f_n, g_n) &= \mathbb{E}_{p,q}[f_n g_n] - \mathbb{E}_{p,q}[f_n]\mathbb{E}_{p,q}[g_n] \\
&= \mathbb{P}_p(|\mathcal{C}| \geq n)\big(\mathbb{P}_{p,q}(0 \xleftrightarrow{\Lambda_n} \mathcal{G} \,|\, |\mathcal{C}| \geq n) - \mathbb{P}_{p,q}(0 \xleftrightarrow{\Lambda_n} \mathcal{G})\big) \\
&= \mathbb{P}_p(|\mathcal{C}| \geq n)\big(\mathbb{E}_{p,q}\big[1 - \exp(-\lambda|\mathcal{C}_n^0|/n) \,\big|\, |\mathcal{C}| \geq n\big] \\
&\qquad\qquad - \mathbb{E}_{p,q}(1 - \exp(-\lambda|\mathcal{C}_n^0|/n)]\big).
\end{aligned}$$

If $|\mathcal{C}| \geq n$, then also $|\mathcal{C}_n^0| \geq n$. Hence,

$$\mathrm{Cov}(f_n, g_n) \geq \mathbb{P}_p(|\mathcal{C}| \geq n)\big((1 - e^{-\lambda}) - \mathbb{E}_{p,q}[1 - \exp(-\lambda|\mathcal{C}|/n)]\big).$$

Combining this with (3.10) and (3.11) gives

$$\sum_{e \in E_n} \mathrm{Inf}_e(f_n) \geq \frac{1}{4}\left(\frac{1 - e^{-\lambda}}{\mathbb{E}_p[1 - \exp(-\lambda|\mathcal{C}|/n)]} - 1\right)\mathbb{P}_p(|\mathcal{C}| \geq n).$$

To prove the differential inequality it thus remains to compute the derivative of $\mathbb{P}_p(|\mathcal{C}| \geq n)$ with respect to p. We use Russo's formula, Propopsition 3.2, to find

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(|\mathcal{C}| \geq n) &= \sum_{e \in E_n} \mathbb{P}_p(e \text{ is pivotal for } |\mathcal{C}| \geq n) \\
&= \sum_{e \in E} \frac{1}{2p(1-p)}\mathbb{P}_p(f_n(\omega) \neq f_n(\omega^{e \mapsto \pi})) \\
&= \sum_{e \in E} \frac{1}{2p(1-p)}\mathrm{Inf}_e(f_n),
\end{aligned}$$

by Proposition 2.14, and where $\omega^{e \mapsto \pi}$ is the configuration obtained from ω by resampling ω_e independently with law $\mathrm{Ber}(p)$. This completes the proof of Proposition 3.7. □

3.3.2 *Analysis of the differential inequality*

We will now Theorem 3.6. The analysis of the differential inequality of Proposition 3.7 is quite similar to the analysis of the inequality in Section 3.2.2. Let $\psi_n(p) := \mathbb{P}_p(|\mathcal{C}| \geq n)$, and $S_n := S_n(p) = \sum_{k=1}^n \psi_k$. Since $1 - e^{-x} \leq x \wedge 1$, we can bound

$$\mathbb{E}_p[1 - \exp(-\lambda|\mathcal{C}|/n)] \leq \frac{\lambda}{n}\mathbb{E}\big[|\mathcal{C}| \wedge \frac{n}{\lambda}\big] \leq \frac{\lambda}{n} S_{\lceil n/\lambda \rceil}.$$

We further bound $\frac{1}{4}\frac{1}{p(1-p)} \geq 1$, and obtain for all $n \in \mathbb{N}$,

$$\frac{\mathrm{d}}{\mathrm{d}p}\psi_n(p) \geq \frac{1}{2}\left(\frac{\frac{n}{\lambda}(1-e^{-\lambda})}{S_{\lceil n/\lambda \rceil}} - 1\right)\psi_n(p). \tag{3.12}$$

We again define an auxiliary critical point:

$$\tilde{p}_c := \sup\left\{p \,:\, \limsup_{n\to\infty} \frac{\log S_n(p)}{\log n} < 1\right\}.$$

Let $p < p_1 < p_2 < \tilde{p}_c$. Then there exists a constant $\alpha > 0$ such that $S_n(p_2) \leq n^{1-\alpha}$, for n large enough. Taking $\lambda = 1$, it then follows from (3.12), that for n large,

$$\frac{\mathrm{d}}{\mathrm{d}p}\log\psi_n(p_2) \geq \frac{1}{2}n^{\alpha}(1 - 1/e) - \frac{1}{2}.$$

Since $S_n(p') \leq S_n(p_2)$ for all $p' \leq p_2$, we can integrate the above inequality from p_1 to p_2, to obtain

$$\begin{aligned} -\log\psi_n(p_1) &\geq \log\psi_n(p_2) - \log\psi_n(p_1) \\ &\geq \Big(\frac{1}{2}n^{\alpha}(1-1/e) - \frac{1}{2}\Big)(p_2 - p_1), \end{aligned}$$

so that

$$\psi_n(p_1) \le \sqrt{e}\exp\Big(-\frac{1}{2}(1-1/e)(p_2-p_1)n^\alpha\Big).$$

In particular, $\sum_{k=1}^\infty \psi_k(p_1)$ converges. Applying (3.12) again with $\lambda \downarrow 0$, gives for all $n \in \mathbb{N}$,

$$\frac{\mathrm{d}}{\mathrm{d}p}\psi_n(p) \ge \frac{1}{2}\Big(\frac{n}{\sum_{k=1}^\infty \psi_k(p_1)} - 1\Big)\psi_n(p),$$

so that

$$\mathbb{P}_p(|\mathcal{C}| \ge n) = \psi_n(p) \le \sqrt{e}\exp\Big(\frac{p_1 - p}{\sum_{k=1}^\infty \psi_k(p_1)}n\Big).$$

This proves the desired exponential decay below $\tilde{p}_c$. If desired, the constant $\sqrt{e}$ can be removed by changing the constant in the exponential accordingly.

It thus remains to show that $\tilde{p}_c = p_c$. This part is also similar to the calculation in Section 3.2.2, apart of the -1 term, which is somewhat of a nuisance. Let $\tilde{p}_c < p_1 < p$. We define $T_n(p) := \frac{1}{\log n}\sum_{k=1}^n \frac{\psi_k(p)}{k}$. We set $\lambda = 1$ in (3.12), and compute

$$\frac{\mathrm{d}}{\mathrm{d}p}T_n(p) \ge \frac{1}{2\log n}\sum_{k=1}^n\Big((1-1/e)\frac{\psi_k(p)}{S_k} - \frac{\psi_k(p)}{k}\Big) \tag{3.13}$$

For the first term we set $C = (1-1/e)/2$, and find

$$\begin{aligned}\frac{1-1/e}{2\log n}\sum_{k=1}^n\frac{\psi_k(p)}{S_k} &\ge \frac{C}{\log n}\sum_{k=1}^n\int_{S_k}^{S_{k+1}}\frac{1}{t}\,\mathrm{d}t\\ &= \frac{C}{\log n}\sum_{k=1}^n(\log S_{k+1} - \log S_k)\\ &= C\frac{\log S_{n+1} - \log S_1}{n} \ge C\frac{\log S_{n+1}}{\log n},\end{aligned}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}p}T_n(p) \geq C\frac{\log S_{n+1}}{\log n} - \frac{1}{2\log n}\sum_{k=1}^{n}\frac{\psi_k(p)}{k}.$$

Integrating the above inequality from p_1 to p, gives

$$T_n(p) - T_n(p_1) \geq (p-p_1)\Big(C\frac{\log S_{n+1}(p_1)}{\log n} - \frac{1}{2\log n}\sum_{k=1}^{n}\frac{\psi_k(p_1)}{k}\Big)$$

For $n \to \infty$, we have $T_n(p) \to \mathbb{P}_p(|\mathcal{C}| = \infty)$, so that by the definition of $\tilde{p}_c$, we have

$$\begin{aligned}\mathbb{P}_p(|\mathcal{C}| = \infty) &\geq \mathbb{P}_p(|\mathcal{C}| = \infty) - \mathbb{P}_{p_1}(|\mathcal{C}| = \infty)\\ &\geq (p-p_1)\Big(C - \frac{\mathbb{P}_p(|\mathcal{C}| = \infty)}{2}\Big).\end{aligned}$$

We conclude

$$\mathbb{P}_p(|\mathcal{C}| = \infty) \geq \frac{C(p-p_1)}{1+(p-p_1)/2} > 0.$$

This shows $\tilde{p}_c = p_c$, and completes the proof of Theorem 3.6. □

We remark that by taking $p_1 \downarrow p_c$, we have a different proof for Proposition 3.5, since $|\mathcal{C}| = \infty$ if and only if $0 \longleftrightarrow \infty$. Furthermore, Proposition 3.7 implies two other inequalities for certain critical exponents, namely

$$\gamma \leq \delta - 1, \qquad \Delta \leq \gamma + 1,$$

if the exponents exists, where

$$\begin{aligned}\mathbb{E}_p[|\mathcal{C}|^k] &\approx (p_c - p)^{-(k-1)\Delta-\gamma}, \quad &&\text{as } p \uparrow p_c, \quad k \in \mathbb{N}\\ \mathbb{P}_{p_c}(|\mathcal{C}| \geq n) &\approx n^{-1/\delta}, &&\text{as } n \to \infty.\end{aligned} \tag{3.14}$$

See [37] for the proof of these inequalities.

4 THE CORRUPTED COMPASS MODEL

Let $G = (V, E)$ be an infinite, connected, locally finite, vertex-transitive graph. We consider the corrupted compass model on G, which is informally defined as follows. Each vertex $v \in V$ is corrupted with probability p, independently of each other. For each corrupted vertex, we declare each neighbouring edge to be open. On the other hand, for an uncorrupted vertex, we choose one neighbouring edge to be open uniformly at random. A possible configuration of this model on the triangular lattice is shown in Figure 4.1.

The corrupted compass model was introduced by Hirsch, Holmes and Kleptsyn [30] in the context of reinforcement models for neural networks. They show that in a class of reinforcement models the reinforced edges almost surely do not form an infinite cluster if the reinforcement is strong enough. They show this by making a coupling between the reinforcement model and the corrupted compass model, and subsequently showing that in the latter model there exists only finite clusters almost surely for p small enough.

The corrupted compass model is not only relevant to reinforcement models, as the model was also used in the context of alignment percolation by Beaton, Grimmett and Holmes [3]. In the one-choice alignment percolation model on $\mathbb{Z}^d$ introduced by these authors, a Bernoulli site percolation configuration with parameter p is taken. Subsequently, for each occupied vertex, one of the $2d$ directions is chosen uniformly at random the entire

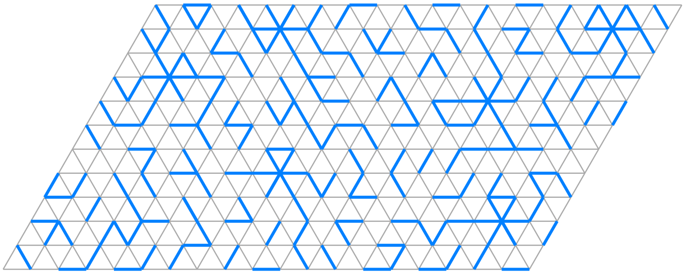


Figure 4.1: The corrupted compass model on the triangular lattice.

line segment in this direction until the next occupied vertex is declared blue. The authors then ask the question whether there exists an infinite blue cluster. The main problem in the analysis of this model is the lack of monotonicity in p. Nevertheless, the authors show that for p large enough there exists no infinite blue clusters almost surely. They show this by dominating the alignment percolation model by a corrupted compass model with parameter $1 - p$. Since the corrupted compass model does not have infinite clusters for $1 - p$ small enough, the one-choice alignment percolation model does not have any infinite clusters for p large enough.

This chapter is an adaption of a paper that has appeared in *Indagationes Mathematicae* [4].

4.1 FRAMEWORK AND MAIN RESULT

We will show that this model exhibits a sharp phase transition as well. Since the state of the edges depend on each other, the proofs of Aizenman and Barsky, and of Menshikov cannot be applied in this model. The $\phi_p(S)$ proof of Duminil-Copin and

Tassion faces the same issue. Therefore, we have to rely on the strategy using the OSSS inequality.

For $v \in V$, let $\mathcal{N}(v)$ denote the set of edges that include v, and let $d = |\mathcal{N}(v)|$ (which is independent of v). We fix an arbitrary vertex $0 \in V$ to be the origin. For $v, w \in V$, let $\mathrm{d}(v,w)$ denote the graph distance between v and w in G. For $n \in \mathbb{N}$, we define the balls

$$\Lambda_n^v := \{w \in V : \mathrm{d}(v,w) \le n\}, \qquad \partial\Lambda_n^v := \{w \in V : \mathrm{d}(v,w) = n\}.$$

For $v = 0$, we drop part of the notation: $\Lambda_n = \Lambda_n^0$ and $\partial\Lambda_n = \partial\Lambda_n^0$. For a bond configuration η and $v, w \in V$, we say that $v \longleftrightarrow w$, if there is a path of open edges starting in v and ending in w. Similarly, for $A \subset V$ we say that $v \longleftrightarrow A$, whenever there exists $w \in A$, such that $v \longleftrightarrow w$. We say that $0 \longleftrightarrow \infty$, if for all $n \in \mathbb{N}$, we have $0 \longleftrightarrow \partial\Lambda_n$. We define the critical value for percolation as

$$p_c := \sup\{p : \mathbb{P}_p(0 \longleftrightarrow \infty) = 0\}.$$

Hirsch, Holmes and Kleptsyn [30] have shown that, for p small enough, all clusters are finite almost surely. From this it follows that $p_c > 0$. On the other hand the corrupted compass model dominates the Bernoulli site percolation model that only uses the corrupted compasses. Therefore, we have $p_c \le p_c^{\text{site}}(G)$, where $p_c^{\text{site}}(G)$ is the critical threshold for Bernoulli site percolation on G. Depending on the graph G, this threshold is nontrivial, so that also $p_c < 1$.

The corrupted compass model is defined by the probability space $(\Omega, \mathcal{F}, \mathbb{P}_p)$, where

$$\Omega = \prod_{v \in V} [0,1] \times \mathcal{N}(v),$$

the σ-algebra $\mathcal{F}$ is generated by the cylindrical events, and $\mathbb{P}_p$ is the product measure of the uniform measures on $[0,1] \times \mathcal{N}(v)$. For $\omega \in \Omega$, we denote $U_v := \omega_{v,1}$, i.e., the uniform random variable on $[0,1]$ associated to v, and $A_v := \omega_{v,2}$, the uniformly chosen edge in $\mathcal{N}(v)$. We define $X_v := (U_v, A_v)$. Let $\mathcal{K}$ denote the set of corrupted vertices, i.e.,

$$\mathcal{K} := \{v \in V \,:\, U_v < p\}.$$

We can obtain the bond configuration η as follows. Let $\eta : \Omega \to \{0,1\}^E$ be given by

$$\eta_e(\omega) := \mathbb{1}\left\{e \in \bigcup_{v \in \mathcal{K}} \mathcal{N}(v) \cup \bigcup_{v \in \mathcal{K}^c} \{A_v\}\right\}.$$

We say that an edge e is open whenever $\eta_e = 1$, and closed otherwise.

The sharpness of the phase transition of this model is formulated as follows.

Theorem 4.1. *Consider the corrupted compass model with parameter p. For all $p < p_c$, there exists a constant $c > 0$, such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn).$$

Similar to the case of Bernoulli percolation, we also obtain a lower bound in the supercritical regime.

Proposition 4.2. *There exists a constant $c > 0$, such that for all $p > p_c$,*

$$\mathbb{P}_p(0 \longleftrightarrow \infty) \geq c(p - p_c).$$

4.2 PROOF OF THE SHARP PHASE TRANSITION

We will apply the OSSS inequality to the Boolean function $f_n := \mathbb{1}\{0 \longleftrightarrow \partial\Lambda_n\}$ for fixed $n \in \mathbb{N}$. This function only depends on the variables in $\{X_v : v \in \Lambda_n\}$. Therefore, we define the truncated space

$$\Omega_n = \bigotimes_{v \in \Lambda_n} [0,1] \times \mathcal{N}(v),$$

so that we can directly apply the OSSS inequality for finite probability spaces, Theorem 2.16. We write $\theta_n(p) := \mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n)$. For $1 \leq k \leq n$, let T_k be the decision tree that explores the cluster of $\partial\Lambda_k$. This decision tree determines f, since a path from 0 to $\partial\Lambda_n$ must go through $\partial\Lambda_k$. To precisely describe the decision tree T_k, we need a subalgorithm $\mathsf{Determine}(v)$, for $v \in \Lambda_n$. When $\mathsf{Determine}(v)$ is called, X_v is revealed, as well as X_w for all neighbours w of v. This determines the state of the edges in $\mathcal{N}(v)$. The decision tree T_k then start by setting the active set of vertices, $\mathcal{A}$, equal to $\partial\Lambda_k$. The algorithm then goes trough the vertices in the active set, according to some predetermined ordering of the vertices. If a vertex v is taken from the active set, $\mathsf{Determine}(v)$ is called, and the active set is updated by removing v, and adding vertices that are now connected to $\partial\Lambda_v$ by the revealed edges. This process continues until a connection $0 \longleftrightarrow \partial\Lambda_n$ is found, or until the active set is empty. In either case, it is then determined whether there is a connection from 0 to $\partial\Lambda_n$. For $x \in \Lambda_n$, and $A \subseteq \Lambda_n$, we say that $x \overset{\mathcal{R}}{\longleftrightarrow} A$, whenever there is a path from x to some $y \in A$, using only open edges in $\{\mathcal{N}(v) : v \in \mathcal{R}\}$.

The pseudocode of T_k is given in Algorithm 3. A possible realization of the exploration process carried out by T_k for the model on the triangular lattice is shown in Figure 4.2.

Function `Determine`(v)**:**
    for $w \in \Lambda_1^v$ **do**
        Reveal X_w;
        $\mathcal{R} := \mathcal{R} \cup \{w\}$;
        $\mathcal{A} := \mathcal{A} \backslash \{w\}$;
    end

$\mathcal{A} := \partial\Lambda_k$;
$\mathcal{R} := \varnothing$;
while $\mathcal{A} \neq \varnothing$ **do**
    Take minimal $v \in \mathcal{A}$;
    `Determine`(v);
    $\mathcal{A} := \mathcal{A} \cup \{w \in \Lambda_n \backslash \mathcal{R} : w \overset{\mathcal{R}}{\longleftrightarrow} \partial\Lambda_k\}$;
    if $0 \overset{\mathcal{R}}{\longleftrightarrow} \partial\Lambda_n$ **then return** 1;
end
return 0;

Algorithm 3: The exploration algorithm T_k.

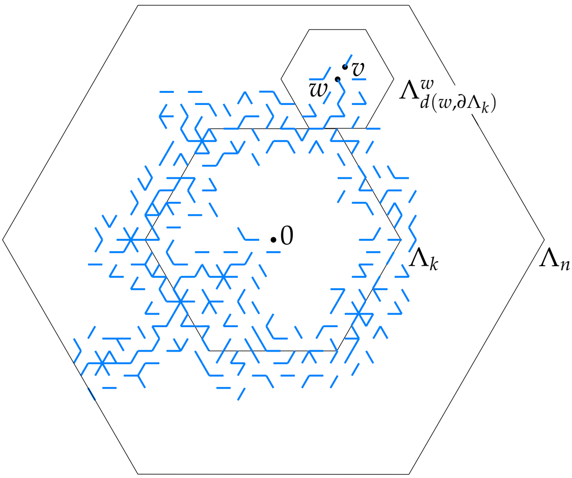


Figure 4.2: The decision tree T_k exploring the cluster of $\partial\Lambda_k$. When X_v is revealed, there must be a neighbour w of v that is connected to $\partial\Lambda_k$.

Applying the OSSS inequality to f and T_k, and summing over k gives

$$n\theta_n(p)(1-\theta_n(p)) \le \sum_{v\in\Lambda_n}\sum_{k=1}^{n} \mathrm{Rev}_v(T_k)\mathrm{Inf}_v(f). \tag{4.1}$$

4.2.1 *Bound on the Revealment*

By summing over k, we essentially average over all spheres $\partial\Lambda_k$ with radius up to n, so that the average revealment is small. This is in spirit the same as taking $1 \le k \le n$ uniformly at random. We note that if X_v is revealed by T_k, it follows that $\Lambda_1^v \longleftrightarrow \partial\Lambda_k$. We obtain

$$\begin{aligned}\sum_{k=1}^{n} \mathrm{Rev}_v(T_k) &\le \sum_{k=1}^{n} \mathbb{P}_p(\Lambda_1^v \longleftrightarrow \partial\Lambda_k)\\ &\le \sum_{k=1}^{n}\sum_{w\in\Lambda_1^v} \mathbb{P}_p(w \longleftrightarrow \partial\Lambda_k)\\ &\le \sum_{k=1}^{n}\sum_{w\in\Lambda_1^v} \mathbb{P}_p\big(w \longleftrightarrow \partial\Lambda^w_{d(w,\partial\Lambda_k)}\big).\end{aligned}$$

Using translation invariance, we have

$$\sum_{k=1}^{n} \mathbb{P}_p\big(w \longleftrightarrow \partial\Lambda^w_{d(w,\partial\Lambda_k)}\big) \le 2\sum_{k=1}^{n} \mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_k).$$

If we define $S_n = S_n(p) := \sum_{k=1}^{n} \theta_k(p)$, it follows that

$$\sum_{k=1}^{n} \mathrm{Rev}_v(T_k) \le 2dS_n.$$

4.2.2 *Bound on the Influence*

For $\omega \in \Omega$, we say that $v \in V$ is a pivotal corrupted compass for an event A, whenever $\mathbb{1}_A(\omega) \neq \mathbb{1}_A(\hat{\omega}_v)$, where $\hat{\omega}_v$ is obtained from ω by corrupting v if v is uncorrupted in ω, or by uncorrupting v if v is corrupted in ω. Russo's formula gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) &= \frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(0 \longleftrightarrow \partial\Lambda_n)\\ &= \sum_{v\in\Lambda_n} \mathbb{P}_p(v \text{ piv. corr. compass for } 0 \longleftrightarrow \partial\Lambda_n).\end{aligned}$$

This can be proven similarly as Theorem 3.2. The aim is to relate the above quantity to the total influence, so that we obtain a differential inequality. By Proposition 2.14, we have

$$\sum_{v\in V} \mathrm{Inf}_v(f) = \frac{1}{2}\sum_{v\in V} \mathbb{P}_p(f(\omega) \neq f(\tilde{\omega}_v)),$$

where $\tilde{\omega}_v$ is obtained from ω by resampling X_v independently. If $f(\omega) = 0$, and $f(\tilde{\omega}_v) = 1$, it follows that v is not corrupted in ω. Therefore, corrupting v will put f to 1, because this will open at least as much edges as the resampling of X_v. Thus, v is a pivotal corrupted compass. The same argumentation holds when $f(\omega) = 1$ and $f(\tilde{\omega}_v) = 0$. We obtain

$$\sum_{v\in V} \mathrm{Inf}_v(f_n) \leq \frac{1}{2}\sum_{v\in V} \mathbb{P}_p(v \text{ piv. corr. compass for } 0 \longleftrightarrow \partial\Lambda_n).$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) \geq 2\sum_{v\in V} \mathrm{Inf}_v(f_n).$$

Combining the OSSS inequality, Theorem 2.16, and the bounds on the revealment and the influence gives

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) \geq \frac{n}{2dS_n}\theta_n(p)(1-\theta_n(p)). \tag{4.2}$$

4.2.3 *Analysis of the differential inequality*

To finish the proof, we distinguish between the cases $p_c = 1$ and $p_c < 1$. First we assume that $p_c = 1$. Let $p_0 < p_c$. We have $\theta_n(p_0) \to 0$ as $n \to \infty$. Let N be such that $\theta_n(p_0) \leq \frac{1}{2}$ for all $n > N$. Then for all $p \leq p_0$ and for all $n > N$ we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\theta_n(p) &\geq \frac{n}{2dS_n}\theta_n(p)(1-\theta_n(p)) \\ &\geq \frac{n}{2dS_n}\theta_n(p)(1-\theta_n(p_0)) \geq \frac{1}{4d}\frac{n}{S_n}\theta_n(p).\end{aligned}$$

From this inequality we can obtain the sharpness of the phase transition, which we will show in the next section. First we will find the same differential inequality, but with a different constant for the case $p_c < 1$. We can assume that $d \geq 3$, since the only infinite, connected, transitive graph with $d = 2$ is $\mathbb{Z}$ with nearest neighbour edges, for which $p_c = 1$. Let $p_c < \delta < 1$. For $n \geq 2$ and $p \leq \delta$, we bound

$$1-\theta_n(p) \geq 1-\theta_2(\delta) = \mathbb{P}_\delta(0 \not\leftrightarrow \partial\Lambda_2).$$

We can construct a configuration in which $0 \not\leftrightarrow \partial\Lambda_2$, and which has positive probability, as follows. Let v be the vertex that the compass of 0 points to, i.e., $A_0 = \{0, v\}$. We require that the compass of v points back to 0, which happens with probability $1/d$. Furthermore we want 0, v, and all other neighbours of 0 and v to be uncorrupted, which costs at most $(1-\delta)^{2d}$. Finally, we want that the compasses of the other neighbours of 0 and v do not point towards 0 or v, which happens with probability $(\frac{d-2}{d})^{2d-2}$. All together we find

$$\mathbb{P}_\delta(0 \not\leftrightarrow \partial\Lambda_2) \geq (1-\delta)^{2d}\frac{1}{d}\left(\frac{d-2}{d}\right)^{2d-2} =: C_0 > 0,$$

so that for all $n \geq 2$ and all $p \leq \delta$, we have

$$\frac{\mathrm{d}}{\mathrm{d}p}\theta_n'(p) \geq C_1\frac{n}{S_n}\theta_n(p), \tag{4.3}$$

where $C_1 := C_0/2d$. Since $C_1 \leq 1/4d$, the above inequality holds for the case where $p_c = 1$ as well, for $n > N$. The remainder of the proof of Theorem 4.1 and Proposition 4.2 involves analyzing the differential inequality as in Section 3.2.2, and we omit this part of the proof.

5 THE CONTACT PROCESS

The contact process is a stochastic process that can be used to model the spread of an infection on a network. We use a graph to model the network, and each vertex of the graph has a state associated with it that is either 0 or 1. The vertices in state 1 are thought of as infected, while the vertices in state 0 are thought of as healthy. The dynamics of the system are as follows: the state of a vertex changes from 1 to 0 with rate 1, independently of the rest of the system. On the other hand, if a vertex is in state 0, its state changes to 1 with rate λ times the number of neighbours that are infected, where $\lambda \geq 0$ is the parameter of the model. A common choice for the graph is the hypercubic lattice $\mathbb{Z}^d$, with $d \geq 1$, which is also the graph we will work with.

The contact process undergoes two phase transitions as λ increases. The first one occurs at the critical point λ_c: above this point the infection survives indefinitely with positive probability, if at least one vertex is infected at the start. The second phase transition occurs at $\lambda_p \geq \lambda_c$, this is the point above which there exists an infinite cluster of infected vertices with positive probability in the limit as the time $t \to \infty$, again in the case where some vertex is infected initially. For both transitions we can talk about a sharp phase transition. For the transition at λ_c, this means that the infection only survives for an exponentially small time if $\lambda < \lambda_c$. This is a classical result of Bezuidenhout and Grimmett [10], but we will give a new proof using the OSSS inequality in this chapter. Afterwards, we will focus on the transition at λ_p, and prove several results regarding the sharpness of this phase transition.

5.1 PRELIMINARIES

We start by introducing the objects and results pertaining to the contact process that we need to prove the sharpness of the phase transitions, starting with the definition of the model. For a wider exposition of the contact process we refer to *Stochastic Interacting Systems* by Liggett [42].

We consider the graph $\mathbb{Z}^d$ with nearest neighbour edges. For $n \in \mathbb{N}$, and $x \in \mathbb{Z}^d$, we define the boxes

$$\Lambda_n^x := \{y \in \mathbb{Z}^d : \|x-y\|_\infty \le n\}, \quad \partial\Lambda_n^x := \{y \in \mathbb{Z}^d : \|x-y\|_\infty = n\}.$$

For $x = 0$ we suppress part of the notation: $\Lambda_n := \Lambda_n^0$, and $\partial\Lambda_n := \partial\Lambda_n^0$. We denote the space of configurations by $S = \{0,1\}^{\mathbb{Z}^d}$. This set is equipped with the product topology τ, that is, the topology generated by the sets

$$\big\{\{\eta \in S : \eta_v = a\} : v \in \mathbb{Z}^d, a \in \{0,1\}\big\}.$$

This topology is metrized by the metric ρ on S, given by

$$\rho(\eta,\xi) := \sum_{v\in\mathbb{Z}^d} \alpha(v)|\eta_v - \xi_v|,$$

for a summable $\alpha : \mathbb{Z}^d \to (0,\infty)$. The space S is a compact set with respect to this topology, since, informally speaking, sets in the topology only fix the state of finitely many vertices. We further denote the set of continuous functions from S to $\mathbb{R}$ by $C(S)$. We also associate the σ-algebra $\mathcal{S}$ with S, generated by

$$\mathcal{S} := \sigma\big(\big\{\{\eta_v = 1\} : v \in \mathbb{Z}^d\big\}\big),$$

similar to the σ-algebra we defined for Bernoulli percolation. For a configuration $\eta \in S$, we write $\eta^{\oplus v}$ for the configuration with the status of v flipped:

$$\eta^{\oplus v}_w = \begin{cases} 1 - \eta_v & \text{if } v = w, \\ \eta_w & \text{if } v \neq w. \end{cases}$$

For $\lambda \geq 0$, we define the flip rates $c : \mathbb{Z}^d \times S \to [0, \infty)$, by

$$c(v, \eta) = \begin{cases} \lambda \sum_{w \sim v} \eta_w & \text{if } \eta_v = 0, \\ 1 & \text{if } \eta_v = 1, \end{cases}$$

where $w \sim v$ denotes that w is a neighbour of v. The above sum is therefore the number of infected neighbours of v. Note that the flip rates are uniformly bounded by $2d\lambda + 1$. To define the contact process by means of a generator, we define for all $\lambda \geq 0$ the operator $\mathcal{L}_\lambda : D \to C(S)$, given by

$$\mathcal{L}_\lambda f(\eta) = \sum_{v \in \mathbb{Z}^d} c(v, \eta)\big(f(\eta^{\oplus v}) - f(\eta)\big),$$

on the domain

$$D := \Big\{ f \in C(S) \ : \ \sum_{v \in \mathbb{Z}^d} \sup_{\eta \in S} |f(\eta^{\oplus v}) - f(\eta)| < \infty \Big\}.$$

The closure of $\mathcal{L}_\lambda$ is a probability generator that generates the contact process.

Definition 5.1. *The contact process with parameter $\lambda \geq 0$ is the Feller process with generator $\overline{\mathcal{L}}_\lambda$ on the domain $\mathcal{D}(\overline{\mathcal{L}})$.*

The fact that $\overline{\mathcal{L}}_\lambda$ is a probability generator follows from the fact that the flip rates are uniformly bounded. See Theorem 4.3

of Liggett [44] for details. We do not have an explicit form of the domain $\mathcal{D}(\overline{\mathcal{L}}_\lambda)$, but we know that $D \subset \mathcal{D}(\overline{\mathcal{L}}_\lambda)$ is a core for $\overline{\mathcal{L}}_\lambda$.

Feller processes are defined on the space of càdlàg functions:

$$X = \big\{\sigma\colon [0,\infty) \to S \,:\, \sigma \text{ is right-continuous and has left limits}\big\}.$$

This space is equipped with a $\sigma-$algebra $\mathcal{A}$, which is generated by the one-dimensional projections, as well as the right-continuous filtration $(\mathcal{A}_t)_{t\geq 0}$ such that σ is adapted to this filtration. Furthermore, there exists a unique family of probability measures $(\mathbb{P}^\eta)_{\eta\in S}$ corresponding to the contact process that we have defined by means of the generator, one measure for each starting configuration, satisfying $\mathbb{P}^\eta\big(\sigma_v(0) = \eta_v,\, v \in \mathbb{Z}^d\big) = 1$.

5.2 THE GRAPHICAL REPRESENTATION

It is more convenient to work with an alternative construction of the process than to work with the generator directly. In this section we introduce the graphical representation for the contact process. This a construction of the process which is essential in the analysis of the process. For every vertex $v \in \mathbb{Z}^d$, we create a time axis $[0,\infty)$, so that we obtain the space-time $\mathbb{Z}^d \times [0,\infty)$. We then create a marked Poisson process on every axis $\{v\} \times [0,\infty)$: with rate 1 there appears a point marked by a cross, which corresponds to the healing of v, i.e., a change of σ_v from 1 to 0. Furthermore, for every neighbour w of v, we have a point marked with an arrow that points to w, which corresponds to the infection of w by v. We denote the set of marks by

$$\mathcal{M} := \{\times\} \cup \mathcal{I}, \qquad \mathcal{I} := \{\pm \mathsf{e}_1, \pm \mathsf{e}_2, \ldots, \pm \mathsf{e}_d\},$$

where e_i is the ith unit vector, so that the mark $\pm\mathsf{e}_i$ corresponds to an infection of $v \pm \mathsf{e}_i$ by v. The marked Poisson point process has rate $2d\lambda + 1$, however for our purposes it is convenient to rescale time so that the combined process has rate 1.

We now consider a Poisson point process on $\mathbb{Z}^d \times [0, \infty)$ with intensity 1. In order to obtain a coupling between contact processes with different values of λ, we associate to every Poisson point x two random variables U_x and ρ_x, which are independent of each other and of the point process. Furthermore, U_x is uniformly distributed between 0 and 1, and $\rho_x \in \mathcal{I}$ uniformly at random. From these random variables, we obtain the mark m_x of x as follows:

$$m_x := \begin{cases} \times & \text{if } U_x \leq \frac{1}{2d\lambda+1}, \\ \rho_x & \text{otherwise.} \end{cases} \tag{5.1}$$

Let ω be the configuration of the Poisson point process along with the marks (U_x, ρ_x) for all Poisson points x. We denote the probability space associated to ω by $(\Omega, \mathcal{F}, \mathbb{P}_\lambda)$. If we want to emphasize the initial configuration, we write $\mathbb{P}_\lambda^\eta$. We further denote by $\mathbb{P}$ the coupling of the processes for different values of λ.

Let $s < t$. An active space-time path from (v, s) to (w, t) is a path in $\mathbb{Z}^d \times [0, \infty)$ starting in (v, s) and ending in (w, t) that is allowed to move upward in time without hitting points marked with a cross and is allowed to move to a neighbouring vertex along arrows that point to that particular vertex. We denote the event that such a path exists by $(v, s) \longrightarrow (w, t)$. Moreover, for $\Lambda \subseteq \mathbb{Z}^d$, we define $(v, s) \xrightarrow{\Lambda} (w, t)$ to be the event that a space-time path exists from (v,s) to (w,t) using only vertices in Λ. In general we define for $A, B, \Lambda \subseteq \mathbb{Z}^d$, and $T_1, T_2 \subseteq [0,\infty)$,

$$\big(A, T_1\big) \xrightarrow{\Lambda} (B, T_2) \iff$$
$$\exists v \in A,\ w \in B,\ s \in T_1,\ t \in T_2 \text{ s.t. } (v,s) \xrightarrow{\Lambda} (w,t).$$

We will now define a stochastic process $\sigma : [0,\infty) \to S$. We fix an initial configuration $\eta \in S$, and set $\sigma_v(0) := \eta_v$. For $t > 0$, let $\sigma_v(t)$ be the status of v at time t, that is,

$$\sigma_v(t) = 1 \iff \exists w \in \mathbb{Z}^d \text{ s.t. } \sigma_w(0) = 1,\ (w,0) \longrightarrow (v,t).$$

A realization of the graphical representation for the process on $\mathbb{Z}$ is shown in Figure 5.1, along with the values of $\sigma_v(t)$, for several vertices $v \in \mathbb{Z}$. In this realization all vertices are set to be infected initially.

The first useful consequence of the graphical representation of the contact process is the self-duality property of the process.

Proposition 5.2 (Self-duality). *Let $A, B \subseteq \mathbb{Z}^d$. The contact process is self-dual in the following sense:*

$$\mathbb{P}_\lambda\big((A,0) \longrightarrow (B,t)\big) = \mathbb{P}_\lambda\big((B,0) \longrightarrow (A,t)\big), \tag{5.2}$$

for all $t \geq 0$ and $\lambda \geq 0$.

Proof. The statement follows from considering the graphical representation with time reversed, and reversing the direction of the arrows: we call a path from (w,t) to (v,s), $t \geq s$ a reversed active space-time path when it starts in (w,t) and ends in (v,s), and that is allowed to move backward in time without hitting points marked with a cross and is allowed to move from $u \in \mathbb{Z}^d$ to a neighbouring vertex $u \pm \mathrm{e}_i$ at time t', whenever there is a

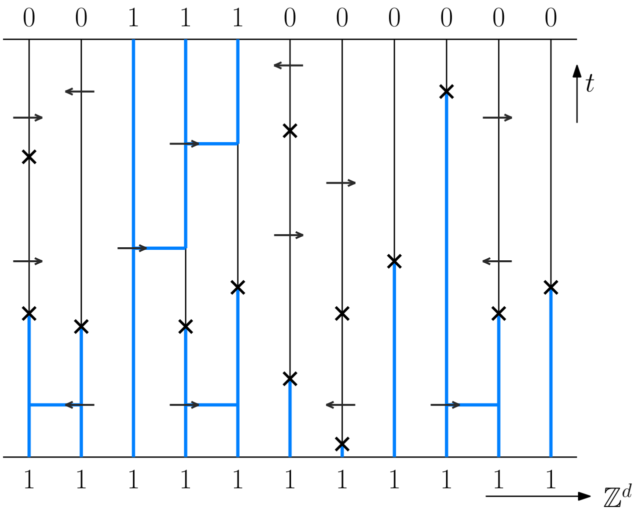


Figure 5.1: The graphical representation of the contact process

point $x \in \omega \cap \{u \pm e_i\} \times \{t'\}$ with $m_x = \mp e_i$. Note that the existence of such paths have the same law as normal active space-time paths. This then implies the duality relation. We refer to Theorem 1.7 of Chapter 4 of [43] for a more formal proof. □

A second useful result is a bound on the rate of growth of the infection.

Proposition 5.3. *For all* $\lambda > 0$ *and all* $c > 0$*, there exists* $a > 0$*, such that for all* $t > 0$*, and* $s \leq t$*,*

$$\sum_{\substack{v \in \mathbb{Z}^d \\ \|v\|_\infty \geq at}} \mathbb{P}_\lambda\big((0,0) \longrightarrow (v,s)\big) \leq a \exp(-ct). \tag{5.3}$$

Moreover, for all $c > 0$*, there exists* $a > 0$ *and* $C > 0$*, such that for all* $t > 0$*,*

$$\mathbb{P}_\lambda\big((0,0) \longrightarrow (\partial\Lambda_{\lceil at \rceil}, [0,t])\big) \leq C \exp(-ct). \tag{5.4}$$

Proof. We refer to [42], Proposition 1.21 and Lemma 1.22, for the full proof of (5.3), and only give a short sketch here. The growth of the infection can be dominated by suppressing all healings. This corresponds to ignoring all crosses in the graphical representation. The resulting process is a branching random walk, which is well-understood. In particular we have exponential bounds on the probability that this process grows linearly. For (5.4) we assume without loss of generality that $t \geq 0$ is integer-valued. We note that $(0,0) \longrightarrow \big(\partial\Lambda_{\lceil at\rceil}, [0,t]\big)$ is independent of the configuration of the Poisson points on $\partial\Lambda_{\lceil at\rceil} \times [0,t]$. Therefore, we find for $1 \leq s \leq t$,

$$\begin{aligned}
&\mathbb{P}_\lambda\Big((0,0) \longrightarrow \big(\partial\Lambda_{\lceil at\rceil}, [s-1,s]\big)\Big)\\
&\quad\leq \frac{1}{\exp(-1/(2d\lambda+1))} \sum_{v\in\partial\Lambda_{\lceil at\rceil}} \mathbb{P}_\lambda\Big((0,0) \longrightarrow (v,[s-1,s]),\\
&\qquad\qquad \big|\{x \in \omega \cap \{v\} \times [s-1,s] \,:\, m_x = \times\}\big| = 0\Big)\\
&\quad\leq e \sum_{v\in\partial\Lambda_{\lceil at\rceil}} \mathbb{P}_\lambda\big((0,0) \longrightarrow (v,s)\big)\\
&\quad\leq ae\exp(-ct),
\end{aligned}$$

by (5.3). Summing k from 1 to t gives

$$\begin{aligned}
\mathbb{P}_\lambda\big((0,0) \longrightarrow &(\partial\Lambda_{\lceil at\rceil}, [0,t])\big)\\
&\leq \sum_{k=1}^{t} \mathbb{P}_\lambda\Big((0,0) \longrightarrow \big(\partial\Lambda_{\lceil at\rceil}, [s-1,s]\big)\Big)\\
&\leq aet\exp(-ct)\\
&\leq C\exp(-ct/2),
\end{aligned}$$

for some $C > 0$, and all $t \geq 0$. Since $c > 0$ can be chosen arbitrarily, the result follows. $\square$

5.2.1 *Russo's Formula*

We can prove a version of Russo's formula that is valid on the graphical representation, similar to Russo's formula for Bernoulli percolation, Proposition 3.2. This formula will be essential to obtain the desired differential inequalities. In order to state the formula, we first introduce the concept of an increasing event. Informally speaking, an event is increasing if it still occurs after the addition of infection arrows or the removal of healings.

Definition 5.4. *An event $A \in \mathcal{F}$ is called increasing, whenever $\mathbb{1}_A(\omega) \le \mathbb{1}_A(\omega')$ for all $\omega, \omega' \in \Omega$ such that*

$$\{x \in \omega \,:\, m_x = \rho_x\} \subseteq \{x \in \omega' \,:\, m_x = \rho_x\}, \text{ and}$$
$$\{x \in \omega \,:\, m_x = \times\} \supseteq \{x \in \omega' \,:\, m_x = \times\}.$$

If an event A is increasing, then $\mathbb{1}_A(\omega) \le \mathbb{1}_A(\tilde{\omega})$, where $\tilde{\omega}$ is obtained from ω by changing the mark of a point $x \in \omega$ with $m_x = \times$ to $\pm \mathrm{e}_i$ for any $1 \le i \le d$. We further define the notion of pivotality in this context:

$$\mathrm{Piv} := \left\{x \in \omega \,:\, \mathbb{1}_A(\omega) \neq \mathbb{1}_A(\omega^{\oplus x})\right\},$$

where $\omega^{\oplus x}$ is the configuration obtained from ω by changing the mark of x from $\times$ to ρ_x or vice versa, and leaving all other points unchanged.

Proposition 5.5 (Russo's formula for the contact process)**.** *Let $A \in \mathcal{F}$ be an increasing event that depends only on the space-time $\Lambda_n \times [0,t]$, for some $n \in \mathbb{N}$ and $t \ge 0$. Then, $\mathbb{P}_\lambda(A)$ is differentiable with respect to λ for any $\lambda > 0$, and*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{P}_\lambda(A) = \frac{2d}{(2d\lambda+1)^2}\mathbb{E}_\lambda|\mathrm{Piv}|. \tag{5.5}$$

Proof. Let A be an increasing event that depends on the space-time $\Lambda_n \times [0,t]$, for $n \in \mathbb{N}$ and $t \geq 0$. We will condition on the location of the points in ω. Let $\mathcal{G}$ be the σ-Algebra generated by the locations of the points in ω. The marks m_x are thus independent of $\mathcal{G}$. For $\delta > 0$, we can write

$$\mathbb{P}_{\lambda+\delta}(A) - \mathbb{P}_\lambda(A) = \mathbb{E}\Big[\mathbb{E}_{\lambda+\delta}[\mathbb{1}_A|\mathcal{G}] - \mathbb{E}_\lambda[\mathbb{1}_A|\mathcal{G}]\Big],$$

since increasing λ only changes the marks of the points, not the location of the points. For $x \in \omega$, denote by $\mathbb{E}^x_{\lambda+\delta}[\cdot|\mathcal{G}]$ the conditional expectation with respect to the measure that samples the mark of x with parameter $\lambda+\delta$, and all other marks with parameter λ. Then, using the fact that A is increasing, we obtain

$$\begin{aligned}
&\mathbb{E}^x_{\lambda+\delta}[A \mid \mathcal{G}] - \mathbb{E}^x_\lambda[A \mid \mathcal{G}] \\
&\qquad = \mathbb{E}^x_\lambda\Big[\mathbb{1}\Big\{\frac{1}{2d(\lambda+\delta)+1} \leq U_x \leq \frac{1}{2d\lambda+1},\ x \in \mathrm{Piv}\Big\}\ \Big|\ \mathcal{G}\Big] \\
&\qquad = \frac{2d\delta}{(2d\lambda+1)(2d(\lambda+\delta)+1)}\mathbb{E}^x_\lambda[\mathbb{1}\{x \in \mathrm{Piv}\} \mid \mathcal{G}],
\end{aligned}$$

since the mark of x is independent of the event that x is pivotal, and of $\mathcal{G}$. It now follows that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[\mathbb{1}_A \mid \mathcal{G}] &= \lim_{\delta\downarrow 0}\sum_{x\in\omega}\frac{\mathbb{E}^x_{\lambda+\delta}[\mathbb{1}_A \mid \mathcal{G}] - \mathbb{E}^x_\lambda[\mathbb{1}_A \mid \eta]}{\delta} \\
&= \sum_{x\in\eta} C(\lambda)\mathbb{E}_\lambda[\mathbb{1}\{x \in \mathrm{Piv}\} \mid \mathcal{G}] = C(\lambda)\mathbb{E}_\lambda\big[|\mathrm{Piv}|\ \big|\ \mathcal{G}\big],
\end{aligned}$$

where $C(\lambda) = 2d/(2d\lambda+1)^2$. Therefore, we find

$$\begin{aligned}
\mathbb{P}_{\lambda+\delta}(A) - \mathbb{P}_\lambda(A) &= \mathbb{E}\Big[\int_0^\delta C(\lambda+s)\mathbb{E}_{\lambda+s}\big[|\mathrm{Piv}|\ \big|\ \mathcal{G}\big]\mathrm{d}s\Big] \\
&= \int_0^\delta C(\lambda+s)\mathbb{E}\Big[\mathbb{E}_{\lambda+s}\big[|\mathrm{Piv}|\ \big|\ \mathcal{G}\big]\Big]\mathrm{d}s \qquad (5.6) \\
&= \int_0^\delta C(\lambda+s)\mathbb{E}_{\lambda+s}|\mathrm{Piv}|\mathrm{d}s, \qquad (5.7)
\end{aligned}$$

by Fubini's theorem. Since A only depends on $\omega \cap \Lambda_n \times [0,t]$, we have the following domination

$$|\mathrm{Piv}| \le \big|\omega \cap \Lambda_n \times [0,t]\big|,$$

which is integrable. It follows that $\mathbb{E}_\lambda|\mathrm{Piv}|$ is continuous in λ. Therefore, if we divide both sides of (5.7) by δ and take the limit $\delta \downarrow 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{P}_\lambda(A) = C(\lambda)\mathbb{E}_\lambda|\mathrm{Piv}|,$$

where we used the continuity of $\mathbb{E}_\lambda|\mathrm{Piv}|$. □

5.2.2 *BK Inequality*

In this section we state a BK inequality for the graphical representation of the contact process. This inequality bounds the probability of the disjoint occurrence of two events by the product of the probabilities of both events. We say that $A, B \in \mathcal{F}$ occur disjointly, denoted $A \circ B$, when there are two disjoint subsets $\mathbb{Z}^d \times [0,\infty)$ such that the configuration on each of these subsets assures the occurrence of one of the events. More precisely,

$$A \circ B := \Big\{\omega \in \Omega \,:\, \exists K, L \subseteq \mathbb{Z}^d \times [0,\infty),\, K \cap L = \varnothing,\ \text{s.t.}\\ [\omega_K] \subseteq A, [\omega_L] \subseteq B\Big\},$$

where $[\omega_K]$ and $[\omega_L]$ are the cylinder events generated by K and L:

$$[\omega_M] := \{\omega' \in \Omega \,:\, \omega \cap M = \omega' \cap M\}, \qquad M \subseteq \mathbb{Z}^d \times [0,\infty).$$

Theorem 5.6 (BK Inequality for the contact process). *Let $A, B \in \mathcal{F}$ be increasing events that depends only on the space-time $\Lambda_n \times [0,t]$, for some $n \in \mathbb{N}$ and $t \ge 0$. Then, for any $\lambda > 0$,*

$$\mathbb{P}_\lambda(A \circ B) \le \mathbb{P}_\lambda(A)\mathbb{P}_\lambda(B). \tag{5.8}$$

This inequality is names after Van den Berg and Kesten, who proved the result for Bernoulli percolation [8]. In this case the assumption that the events are increasing can be removed, which is the celebrated result of Reimer [50]. In the case of the graphical representation of the contact process the result is proven by Bezuidenhout and Grimmett in Section 2 of [10]. They make a minor technical assumption regarding the stability of the events with respect to a discretization of the space. The same result for more general marked Poisson processes and without the technical assumption is given by Van den Berg. [6].

5.2.3 *The graphical representation yields the contact process*

In this section we show that the process obtained by the graphical representation is a Feller process, and furthermore that it is the contact process, which we defined by means of a probability generator.

Proposition 5.7. *The process $(\sigma_t)_{t\geq 0}$ obtained by the graphical representation is a Feller process.*

Proof. The properties that make up the definition of a Feller process, as given in Definition 3.1 of [44], follow directly from the properties of the Poisson process, with the exception of the Feller property. In particular the Markov property follows from the domain Markov property of the Poisson process. For the Feller property, we let $t \geq 0$, $f \in C(S)$, and $(\eta_n)_{n\in\mathbb{N}} \subseteq S$, such that $\eta_n \to \eta \in S$ with respect to the metric ρ. In particular, $\eta_n \to \eta$ pointwise, as $n \to \infty$.

The graphical representation allows for a natural coupling between the processes with different initial configurations. We underline the dependence on the initial configuration $\xi \in S$ by defining

$$\sigma_v^{\xi}(t) = 1 \iff \exists w \in \mathbb{Z}^d \text{ s.t. } \xi_w(0) = 1,\ (w,0) \to (v,t).$$

If we let $\tilde{\mathbb{P}}_\lambda$ be the coupling between different initial configurations, we can use Jensen's inequality to find

$$\begin{aligned}\lim_{n\to\infty} \left|\mathbb{E}_\lambda^{\eta_n}\left[f(\sigma(t))\right] - \mathbb{E}_\lambda^{\eta}\left[f(\sigma(t))\right]\right| \\ &\leq \lim_{n\to\infty} \tilde{\mathbb{E}}_\lambda\left[\left|f(\sigma^{\eta_n}(t)) - f(\sigma^{\eta}(t))\right|\right] \\ &= \tilde{\mathbb{E}}_\lambda\left[\lim_{n\to\infty}\left|f(\sigma^{\eta_n}(t)) - f(\sigma^{\eta}(t))\right|\right],\end{aligned}$$

by the dominated convergence theorem, since f is bounded. We will now show that $\sigma^{\eta_n}(t)$ converges pointwise to $\sigma^{\eta}(t)$ almost surely, und thus also almost surely with respect to ρ. Let $v \in \mathbb{Z}^d$. Using Proposition 5.2 and Proposition 5.3, we see that there exists a random $N \in \mathbb{N}$, such that

$$(w,0) \not\to (v,t) \qquad \text{for all } w \in \mathbb{Z}^d \backslash \Lambda_N^v.$$

Using this observation, it follows that $\sigma^{\eta_n}(t)$ converges pointwise to $\sigma^{\eta}(t)$, almost surely, since $\eta_n \to \eta$ pointwise. We conclude

$$\begin{aligned}\lim_{n\to\infty} \left|\mathbb{E}_\lambda^{\eta_n}\left[f(\sigma(t))\right] - \mathbb{E}_\lambda^{\eta}\left[f(\sigma(t))\right]\right| \\ &\leq \tilde{\mathbb{E}}_\lambda\left[\lim_{n\to\infty}\left|f(\sigma^{\eta_n}(t)) - f(\sigma^{\eta}(t))\right|\right] = 0,\end{aligned}$$

since f is continuous. □

Proposition 5.8. *The Feller process $(\sigma_{(2d\lambda+1)t})_{t\geq 0}$ is the contact process with parameter λ.*

Proof. Let $\mathcal{L}' : D' \to C(S)$ be the generator of $(\sigma_t)_{t\geq 0}$, which is given by

$$\mathcal{L}'f(\eta) = \lim_{t\downarrow 0} \frac{\mathbb{E}_\lambda^\eta[f(\sigma_t) - f(\eta)]}{t},$$

where the domain D' is the set of functions in $C(S)$ for which the above limit exists with respect to the supremum norm. Using the chain rule, we see that $(\sigma_{2d(\lambda+1)t})_{t\geq 0}$ has generator $(2d\lambda+1)\mathcal{L}'$. It suffices to show that this generator coincides with $\mathcal{L}_\lambda$ on D and that $D \subseteq D'$. It then immediately follows that $\mathcal{D}(\overline{\mathcal{L}}) = D'$, since D is a core of $\overline{\mathcal{L}}_\lambda$ and since the domain of a probability generator cannot be extended. Let $\varepsilon > 0$ and $f \in D$. There exists $n \in \mathbb{N}$ such that

$$\sum_{v\in\Lambda_n^c} \sup_{\eta\in S} |f(\eta^{\oplus v}) - f(\eta))| < \varepsilon.$$

If we can show that

$$\lim_{t\downarrow 0} \left\| (2d\lambda+1)\frac{1}{t}\mathbb{E}^\eta[f(\sigma_t) - f(\eta)] - \mathcal{L}_\lambda f(\eta) \right\|_\infty = 0,$$

then $f \in D'$ and $\mathcal{L}_\lambda f = (2d\lambda+1)\mathcal{L}'f$. The configuration σ_t can be obtained from η by subsequently changing the status of the vertices at which the configurations disagree. We do this such that the state of the first vertex we change is inside Λ_n. Let $A_n(t)$ be the number of Poisson points in $\Lambda_{n+1} \times [0,t)$. Then, $A_n(t) = 1$ with probability $e^{-|\Lambda_{n+1}|t}|\Lambda_{n+1}|t$, and $A_n(t) \geq 2$ with probability at most Ct^2, for a constant $C > 0$. Since f is a continuous function on a compact set, we can find $M > 0$ such that $|f| \leq M$. If $A_n(t) \geq 2$, we bound $|f(\sigma_t) - f(\eta)| \leq 2M$. If on the other hand $A_n(t) = 1$, the configurations σ_t and η differ in at most one vertex of Λ_n. We obtain

$$\begin{aligned}
&\lim_{t\downarrow 0}\Big\|(2d\lambda+1)\frac{1}{t}\mathbb{E}^\eta[f(\sigma_t)-f(\eta)]-\mathcal{L}_\lambda f(\eta)\Big\|_\infty \\
&\quad\le \lim_{t\downarrow 0}\Big\|(2d\lambda+1)\frac{1}{t}\mathbb{E}^\eta_\lambda\Big[\mathbb{1}_{\{A_n(t)=1\}}\sum_{v\in\Lambda_n}\mathbb{1}_{\{(\sigma_t)_v\neq\eta_v\}} \\
&\qquad\qquad\cdot\big(f(\eta^{\oplus v})-f(\eta)\big)\Big]-\mathcal{L}_\lambda f(\eta)\Big\|_\infty \\
&\qquad+\Big\|(2d\lambda+1)\frac{1}{t}\mathbb{E}^\eta_\lambda\Big[\mathbb{1}_{\{A_n(t)=1\}}\sum_{w\in\Lambda_n^c}\mathbb{1}_{\{(\sigma_t)_w\neq\eta_w\}} \\
&\qquad\qquad\cdot\sup_{\xi\in S}\big(f(\xi^{\oplus w})-f(\xi)\big)\Big]\Big\|_\infty+2MCt.
\end{aligned}$$

By considering all possible locations and marks of the Poisson point, we see that the status of $v\in\Lambda_n$ changes with conditional probability $c(v,\eta)/(2d\lambda+1)|\Lambda_{n+1}|$, given that $A_n(t)=1$. Similarly, the configurations σ_t and η differ at a vertex $w\in\Lambda_n^c$ with probability $c(w,\eta)t+O(t^2)$. We find

$$\begin{aligned}
&\lim_{t\downarrow 0}\Big\|(2d\lambda+1)\frac{1}{t}\mathbb{E}^\eta[f(\sigma_t)-f(\eta)]-\mathcal{L}_\lambda f(\eta)\Big\|_\infty \\
&\quad\le \lim_{t\downarrow 0}\Big\|\frac{1}{t}\mathbb{P}_\lambda(A_n(t)=1)\sum_{v\in\Lambda_n}\frac{c(v,\eta)}{|\Lambda_{n+1}|}\big(f(\eta^{\oplus v})-f(\eta)\big) \\
&\qquad\qquad-\mathcal{L}_\lambda f(\eta)\Big\|_\infty+(2d\lambda+1)\varepsilon \\
&\quad=\Big\|\sum_{v\in\Lambda_n^c}c(v,\eta)\big(f(\eta^{\oplus v})-f(\eta)\big)\Big\|_\infty+(2d\lambda+1)\varepsilon \\
&\quad\le(4d\lambda+4)\varepsilon.
\end{aligned}$$

Taking $\varepsilon\to 0$ finishes the proof. □

5.3 SHARP PHASE TRANSITION AT λ_c

Let $\eta^0 \in S$ be the configuration for which only 0 is infected: $\eta^0_v = \mathbb{1}\{v = 0\}$. We define the critical point above which the infection survives indefinitely with positive probability:

$$\lambda_c := \lambda_c\big(\mathbb{Z}^d\big) := \sup\big\{\lambda \,:\, \mathbb{P}^{\eta^0}(\sigma(t) \not\equiv 0,\ \forall t \geq 0) = 0\big\}. \tag{5.9}$$

The first thing to note about this critical point is that $\lambda_c \geq 1/2d$. This bound follows by dominating the process by a branching random walk, which arises if we suppress all healings. This branching random walk can in turn be compared with a branching process by disregarding all spatial information, and recording only the number of particles. The branching process is well-known to die out if the expectation of its offspring distribution, $2d\lambda$ in our case, is less than 1. On the other hand we can bound

$$\lambda_c\big(\mathbb{Z}^d\big) \leq \lambda_c(\mathbb{Z}), \quad d \geq 1,$$

by a natural domination that arises from the graphical representation. For the process on $\mathbb{Z}$ it is known that the infection survives if $\lambda \geq 2$, which was proven by Holley and Liggett [31]. In particular it follows that $\lambda_c(\mathbb{Z}) \leq 2$. On the whole, we see that λ_c is non-trivial.

We record one more result regarding the phase transition at λ_c before we move on to the sharpness of this transition. The celebrated result of Bezuidenhout and Grimmett is that the contact process dies out for $\lambda = \lambda_c$ [9]. This is in particular remarkable, since the corresponding question remains open for Bernoulli percolation, which is in many ways a simpler model. A minor consequence of this result is that $\lambda_c(\mathbb{Z}) < 2$, which is however relevant for our purposes in Section 5.5.

We now focus on the sharpness of the phase transition at λ_c.

Theorem 5.9. *Consider the contact process on $\mathbb{Z}^d$ with parameter $\lambda < \lambda_c$. Then there exists a constant $c > 0$, such that for all $t \geq 0$,*

$$\mathbb{P}^{\eta^0}(\sigma(t) \not\equiv 0) \leq \exp(-ct). \tag{5.10}$$

This result is due to Bezuidenhout and Grimmett [10], who exploit the dynamical renormalization scheme of Grimmett and Marstrand [27]. A simpler proof was given by Swart using harmonic functions [51]. We will give a new proof using the OSSS inequality here.

Proof. Let $t > 0$. Instead of considering the process on $\mathbb{Z}^d$ directly, it is more convenient to first consider the process on the torus of size $N \gg t$. Let $\mathbb{T}_N = (\Lambda_N, E_N^{\mathbb{T}})$ be the graph on the vertex set Λ_N with nearest neighbour edges, complemented by the edges that join the boundaries together. Note that this is a transitive graph. Let $0 < \varepsilon \ll t$ be such that $t \in \varepsilon\mathbb{Z}$. We partition the time axis of every vertex in intervals of size ε. For $v \in \Lambda_n$ and $s \in \varepsilon\mathbb{Z}$ with $0 \leq s < t$, Let $(\Omega_{v,s}, \mathcal{F}_{v,s}, \mathbb{P}_{\lambda;v,s})$ be the probability space for the marked Poisson process on $\{v\} \times [s, s+\varepsilon)$. Then $\Omega = \prod_{v,s} \Omega_{v,s}$, and $\mathbb{P}_\lambda$ is the associated product measure. We can apply the OSSS inequality, Theorem 2.20, to this product space.

Let $f_t := f_t(N) := \mathbb{1}\{(0,0) \xrightarrow{\mathbb{T}_N} (\Lambda_N, t)\}$, and

$$\theta_t^N(\lambda) := \mathbb{P}_\lambda^{\eta^0}((0,0) \xrightarrow{\mathbb{T}_N} (\Lambda_N, t)).$$

For $z \in \varepsilon\mathbb{Z}$, $\varepsilon \leq z < t$, we introduce a decision tree T_z that determines the value of f_t. Informally speaking, this decision tree explores the graphical representation starting from time z. It explores both forward and backward in time. If there is a space-time connection from $(0,0)$ to some vertex at time t, then this connection must go through time z, so that this decision tree determines the value of f_t.

We now describe the decision tree in more detail. We introduce the active sets $\mathcal{A}, \mathcal{B} \subseteq \{(v,s) : v \in \Lambda_N, s \in \varepsilon\mathbb{Z}, 0 \leq s < t\}$, where $\mathcal{A}$ is the set of variables from which we will explore forward in time, and $\mathcal{B}$ is the set of variables from which we will explore backward in time. At the start of the algorithm we set

$$\mathcal{A} := \mathcal{A}_0 := \{(v,z) : v \in \Lambda_N\}, \qquad \mathcal{B} := \mathcal{B}_0 := \{(v,z-\varepsilon) : v \in \Lambda_N\}.$$

The forward exploration process and the backward exploration process are two separate processes that do not interact with each other, since the forward exploration takes place in the space-time after time z, and the backward exploration before time z. We choose to subsequently do one step for each process in order to have a single algorithm. One step in the forward process is to reveal a variable from $\mathcal{A}$, and add any variables to $\mathcal{A}$ that have not been revealed yet and can now be reached using the revealed variables.

The backward exploration is slightly more complicated, since it is unknown from which variables we can reach the explored space-time cluster. During one step of the backward exploration process we reveal one variable $(v,s) \in \mathcal{B}$, and add the unrevealed variables (w,s') to $\mathcal{B}$ that can potentially be connected to the explored space-time cluster, depending on the configuration $\omega_{w,s'}$. The pseudocode of the decision tree is given in Algorithm 4. A realization of this exploration algorithm is shown in Figure 5.2.

$\mathcal{A} := \{(v,z) : v \in \Lambda_N\}$;
$\mathcal{B} := \{(v,z-\varepsilon) : v \in \Lambda_N\}$;
$\mathcal{R} := \varnothing$;
while $\mathcal{A} \cup \mathcal{B} \neq \varnothing$ **do**
    Take minimal $(v,s) \in \mathcal{A}$;
    Reveal $\omega_{v,s}$;
    $\mathcal{R} := \mathcal{R} \cup \{(v,s)\}$;
    $\mathcal{A} := \mathcal{A} \backslash \{(v,s)\}$;
    $\mathcal{A} := \mathcal{A} \cup \left\{(w,s) \in (\mathbb{Z}^d \times \varepsilon\mathbb{Z}) \backslash \mathcal{R} : z \leq s < t,\ (\Lambda_n, z) \xrightarrow{\mathcal{R}} (w,s)\right\}$;
    Take minimal $(v,s) \in \mathcal{B}$;
    Reveal $\omega_{v,s}$;
    $\mathcal{B} := \mathcal{B} \backslash \{(v,s)\}$;
    $\mathcal{B} := \mathcal{B} \cup \left\{(w,s') \in (\Lambda_N \times \varepsilon\mathbb{Z}) \backslash \mathcal{R} : 0 \leq s' \leq z-\varepsilon,\ \exists x \sim w, s'' \in [s', s'+\varepsilon), (x,s'') \xrightarrow{\mathcal{R}} (\Lambda_N, z)\right\}$;
    $\mathcal{B} := \mathcal{B} \cup \left\{(w,s') \in (\Lambda_N \times \varepsilon\mathbb{Z}) \backslash \mathcal{R} : 0 \leq s' \leq z-\varepsilon,\ (w, s'+\varepsilon) \xrightarrow{\mathcal{R}} (\Lambda_N, z)\right\}$;
    if $(0,0) \xrightarrow{\mathcal{R}} (\Lambda_N, t)$ **then return** 1;
end
return 0;

Algorithm 4: The exploration algorithm T_z.

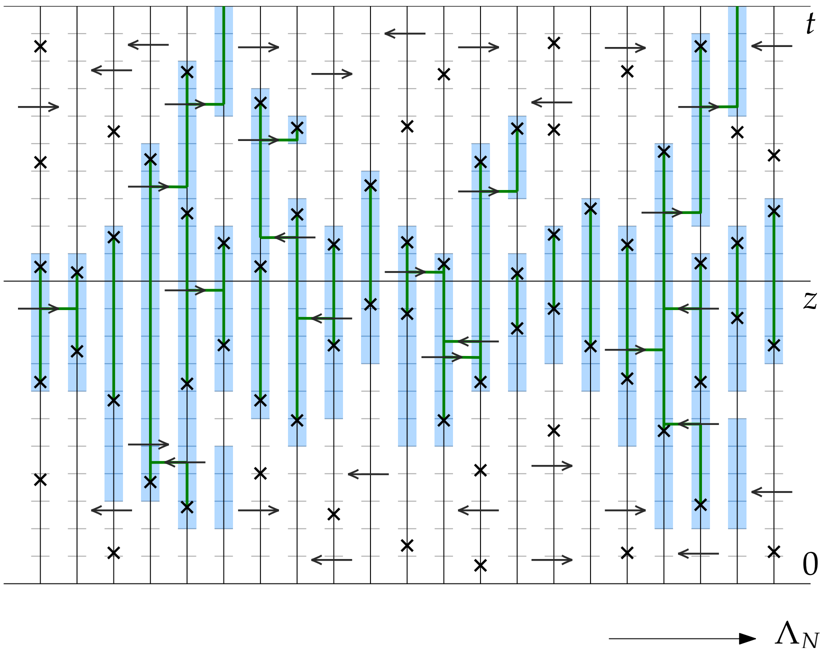


Figure 5.2: The algorithm T_z exploring the forward and backward space-time clusters of (Λ_N, z). The revealed variables are shaded blue, while the forward and backward space-time clusters of (Λ_N, z) are shown in green.

5.3.1 *Bound on the revealment*

We will now bound the revealment of a variable $\omega_{v,s}$ by the decision tree T_z. We first assume that $z \le s < t$, so that the variable is revealed in the forward exploration process. We find

$$\mathrm{Rev}_{v,s}(T_z) \le \mathbb{P}_\lambda\big((\Lambda_n, z) \longrightarrow (v,s)\big) = \mathbb{P}_\lambda\big((v,0) \longrightarrow (\Lambda_N, s-z)\big),$$

by the duality property of the contact process. We now use translation invariance, which follows the fact that $\mathbb{T}_N$ is transitive, to find

$$\mathrm{Rev}_{v,s}(T_z) \le \mathbb{P}_\lambda\big((0,0) \longrightarrow (\Lambda_N, s-z)\big) = \theta^N_{s-z}(\lambda).$$

Now suppose $0 \le s \le z - \varepsilon$, so that the variable is revealed in the backward exploration cluster. Then either $(v, s+\varepsilon) \longrightarrow (\Lambda_N, z)$, or there exists a vertex $w \sim v$, such that $(w, s') \longrightarrow (\Lambda_N, z)$, for some $s' \in [s, s+\varepsilon)$. We can therefore bound

$$\begin{aligned}\mathrm{Rev}_{v,s}(T_z) &\le \mathbb{P}_\lambda\big((v, s+\varepsilon) \longrightarrow (\Lambda_N, z)\big) \\ &\qquad + \sum_{w\sim v} \mathbb{P}_\lambda\big(\exists s' \in [s, s+\varepsilon) : (w, s') \longrightarrow (\Lambda_N, z)\big) \\ &\le \mathbb{P}_\lambda\big((v, s+\varepsilon) \longrightarrow (\Lambda_N, z)\big) \\ &\qquad + \sum_{w\sim v} \mathbb{P}_\lambda\big((w, s+\varepsilon) \longrightarrow (\Lambda_N, z)\big) + 2d\lambda\varepsilon,\end{aligned}$$

since if $(w, s') \longrightarrow (\Lambda_N, z)$ for some $s' \in [s, s+\varepsilon)$, but $(w, s+\varepsilon) \not\longrightarrow (\Lambda_N, z)$, there must be an infection arrow in $v \times [s, s+\varepsilon)$. Using the duality property and translation invariance, we find

$$\begin{aligned}\mathrm{Rev}_{v,s}(T_z) &\le (2d+1)\mathbb{P}_\lambda\big((0,0) \longrightarrow (\Lambda_N, z-s-\varepsilon)\big) + 2d\lambda\varepsilon \\ &= (2d+1)\theta^N_{z-s-\varepsilon}(\lambda) + 2d\lambda\varepsilon.\end{aligned}$$

Summing over z gives

$$\begin{aligned}
\sum_{\substack{z\in\varepsilon\mathbb{Z}\\ \varepsilon\le z<t}} \mathrm{Rev}_{v,s}(T_z) &\le \sum_{z<s} \mathrm{Rev}_{v,s}(T_z) + \sum_{s\le z<t} \mathrm{Rev}_{v,s}(T_z)\\
&\le \sum_{0\le z<t} \theta_z^N(\lambda) + (2d+1)\sum_{0\le z<t} \theta_z^N(\lambda) + 2d\lambda t\\
&\le (2d+2)\sum_{\substack{z\in\varepsilon\mathbb{Z}\\ 0\le z<t}} \theta_z^N(\lambda) + 2d\lambda t. \qquad (5.11)
\end{aligned}$$

5.3.2 *Analysis of the differential inequality*

We now apply the OSSS inequality, Theorem 2.20, and sum over z to find

$$\lfloor t/\varepsilon\rfloor \theta_t^N(\lambda)\big(1-\theta_t^N(\lambda)\big) \le \sum_{v\in\Lambda_N} \sum_{\substack{s\in\varepsilon\mathbb{Z}\\ 0\le s<t}} \sum_{\substack{z\in\varepsilon\mathbb{Z}\\ \varepsilon\le z<t}} \mathrm{Rev}_{v,s}(T_z)\mathrm{Inf}_{v,s}(f_t). \quad (5.12)$$

Applying (5.11) gives

$$\begin{aligned}
&\frac{\lfloor t/\varepsilon\rfloor}{(2d+2)\sum_{\substack{z\in\varepsilon\mathbb{Z}\\ 0\le z<t}} \theta_z^N(\lambda) + 2d\lambda t}\theta_t^N(\lambda)\big(1-\theta_t^N(\lambda)\big)\\
&\qquad\qquad \le \sum_{v\in\Lambda_N} \sum_{\substack{s\in\varepsilon\mathbb{Z}\\ 0\le s\le t-\varepsilon}} \mathrm{Inf}_{v,s}(f_t). \qquad (5.13)
\end{aligned}$$

In order to obtain a differential inequality, we bound the sum of the influences by the expected number of pivotal Poisson points and relate this to the derivative to λ by Russo's formula. Let $A := (0,0) \xrightarrow{\mathbb{T}_N} (\Lambda_N, t)$. Note that this is an increasing event. We define the pivotal set of Poisson points for A and a configuration ω:

$$\mathrm{Piv} := \{x \in [\omega] \, : \, \mathbb{1}_A(\omega) \neq \mathbb{1}_A(\omega^{\oplus x})\},$$

where $\omega^{\oplus x}$ is the configuration obtained from ω by changing the mark of x from $\times$ to ρ_x or vice versa, and leaving all other points unchanged. When we resample $\omega_{v,s}$, the value of $\mathbb{1}_A$ can only change when there is at least one point in $\omega_{v,s} \cup \tilde{\omega}_{v,s}$. Furthermore, the probability that there are two or more points in $\omega_{v,s} \cup \tilde{\omega}_{v,s}$ is of order $O(\varepsilon^2)$. Therefore, since ω and $\tilde{\omega}$ are interchangeable, we can bound

$$\begin{aligned}\mathrm{Inf}_{v,s}(A) \le \tfrac{1}{2} 2\mathbb{P}_\lambda\big(\mathbb{1}_A(\omega) \neq \mathbb{1}_A(\tilde{\omega}), |[\omega_{v,s}]| = 1, |[\tilde{\omega}_{v,s}]| = 0\big) \\ + O\left(\varepsilon^2\right).\end{aligned}$$

Under the above event, $[\omega_{v,s}]$ must contain a pivotal point. Therefore,

$$\begin{aligned}\mathrm{Inf}_{v,s}(A) &\le \mathbb{P}_\lambda\big(\mathrm{Piv} \cap [\omega_{v,s}] \neq \varnothing\big) + O\left(\varepsilon^2\right) \\ &\le \mathbb{E}_\lambda \big|\mathrm{Piv} \cap [\omega_{v,s}]\big| + O\left(\varepsilon^2\right).\end{aligned}$$

Summing over all v and s gives

$$\sum_{v \in \Lambda_N} \sum_{\substack{s \in \varepsilon\mathbb{Z} \\ 0 \le s \le t-\varepsilon}} \mathrm{Inf}_{v,s}(A) \le \mathbb{E}_\lambda |\mathrm{Piv}| + O(\varepsilon).$$

Combining this inequality with (5.13), and letting $\varepsilon \to 0$ gives

$$\mathbb{E}_\lambda |\mathrm{Piv}| \ge \lim_{\varepsilon \downarrow 0} \frac{\lfloor t/\varepsilon \rfloor}{(2d+2)\sum_{\substack{z \in \varepsilon\mathbb{Z} \\ 0 \le z < t}} \theta_z^N(\lambda) + 2d\lambda t} \theta_t^N(\lambda)\big(1 - \theta_t^N(\lambda)\big).$$

Using Russo's formula, Proposition 5.5, and bounding $\big(1 - \theta_t^N(\lambda)\big) \ge 1/(2d\lambda + 1)$, gives

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_t^N(\lambda) \ge \frac{1}{(2d+2)(2d\lambda+1)} \frac{t}{S_t^N} \theta_t^N(\lambda), \tag{5.14}$$

for all $t > 0$, where

$$S_t^N(\lambda) := \int_0^t \theta_z^N(\lambda)\,\mathrm{d}z.$$

From here the analysis is similar to the analysis in Section 3.2.2, however we still have to let $N \to \infty$ at the right moment, so that we have a result on $\mathbb{Z}^d$ instead of on the torus. Let $\theta_t(\lambda) := \mathbb{P}_\lambda\big((0,0) \longrightarrow (\mathbb{Z}^d, t)\big)$, and $S_t := S_t(\lambda) := \int_0^t \theta_z(\lambda)\,\mathrm{d}z$. Note that $\theta_t^N(\lambda) \to \theta_t(\lambda)$ as $N \to \infty$. We define

$$\tilde{\lambda}_c := \sup\Big\{\lambda \,:\, \limsup_{t\to\infty} \frac{\log S_t}{\log t} < 1\Big\}.$$

We will show that this alternative critical point is equal to λ_c, and prove Theorem 5.9 in the process. Let $\lambda_1 < \lambda_2 < \tilde{\lambda}_c$. Then there exists $\alpha > 0$ and $t_0 > 0$ such that $S_t(\lambda_2) \le t^{1-\alpha}$ for all $t \ge t_0$. Let $t \ge t_0$ and let $a > 0$ be such that for $N \ge at$, we have

$$\mathbb{P}_{\lambda_2}\big((0,0) \longrightarrow (\Lambda_N^c, t)\big) \le a\exp(-t).$$

The constant $a > 0$ exists by Proposition 5.3. We then find

$$\begin{aligned} S_t^N(\lambda_2) = \int_0^t \theta_z^N(\lambda_2)\,\mathrm{d}z &\le \int_0^t \theta_z(\lambda_2)\,\mathrm{d}z + (t+1)\exp(-t) \\ &\le t^{1-\alpha} + (t+1)\exp(-t). \end{aligned}$$

We write $C(\lambda) := 1/(2d+2)(2d\lambda+1)$, and use (5.14) to obtain

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\log\theta_t^N(\lambda) \ge C(\lambda_2)\frac{t}{t^{1-\alpha} + (t+1)\exp(-t)} \ge \frac{C(\lambda_2)}{2}t^\alpha,$$

for all $\lambda \le \lambda_2$, and all $t > t_1$, for some $t_1 \ge t_0$. Integrating the above inequality from λ_1 to λ_2 gives

$$-\log\theta_t^N(\lambda_1) \ge \log\theta_t^N(\lambda_2) - \log\theta_t^N(\lambda_1) \ge \frac{C(\lambda)}{2}t^\alpha(\lambda_2 - \lambda_1).$$

Letting $N \to \infty$ gives

$$\theta_t(\lambda_1) = \lim_{N\to\infty} \theta_t^N(\lambda_1) \leq \exp\Big(-\frac{C(\lambda_2)}{2}(\lambda_2 - \lambda_1)t^\alpha \Big).$$

This proves stretched exponential decay below $\tilde{\lambda}_c$. From this fact we can conclude that $\lim_{t\to\infty} S_t(\lambda_1) < \infty$. We can then bootstrap the argument to find proper exponential decay for all $\lambda_0 < \lambda_1$:

$$\theta_t(\lambda_0) \leq \exp\Big(-\frac{C(\lambda_1)}{2}(\lambda_1 - \lambda_0)t \Big),$$

for all $t \geq t_1$. It follows that $\tilde{\lambda}_c \leq \lambda_c$.

Now suppose $\tilde{\lambda}_c < \lambda_1 < \lambda_2$. Similar as in Section 3.2.2, we define for $t > 1$

$$T_t^N(\lambda) := \frac{1}{\log t}\int_0^t \frac{\theta_z^N(\lambda)}{z}\,\mathrm{d}z, \quad T_t(\lambda) := \frac{1}{\log t}\int_0^t \frac{\theta_z(\lambda)}{z}\,\mathrm{d}z.$$

We use 5.14 to find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}T_t^N = \frac{1}{\log t}\int_0^t \frac{\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_z^N(\lambda)}{z}\,\mathrm{d}z &\geq \frac{C(\lambda)}{\log t}\int_0^t \frac{\theta_z^N(\lambda)}{S_z^N(\lambda)}\,\mathrm{d}z \\ &= \frac{C(\lambda)}{\log t}\int_0^t \frac{\mathrm{d}}{\mathrm{d}z}\log S_z^N(\lambda)\,\mathrm{d}z \\ &= C(\lambda)\frac{\log S_t^N(\lambda)}{\log t}.\end{aligned}$$

Hence,

$$T_t^N(\lambda_2) - T_t^N(\lambda_1) \geq C(\lambda_2)\frac{\log S_t^N(\lambda_2)}{\log t}(\lambda_2 - \lambda_1).$$

Letting $N \to \infty$, gives

$$T_t(\lambda_2) - T_t(\lambda_1) \geq C(\lambda_2)\frac{\log S_t(\lambda_2)}{\log t}(\lambda_2 - \lambda_1).$$

Defining $\theta(\lambda) := \lim_{t\to\infty} \theta_t(\lambda)$, we can let $t \to \infty$ and use the definition of $\tilde{\lambda}_c$ to conclude

$$\begin{aligned}\theta(\lambda_2) \geq \theta(\lambda_2) - \theta(\lambda_1) &= \lim_{t\to\infty} T_t(\lambda_2) - T_t(\lambda) \\ &\geq C(\lambda_2)(\lambda_2 - \lambda_1). \end{aligned} \tag{5.15}$$

This proves $\tilde{\lambda}_c = \lambda_c$, and completes the proof of Theorem 5.9. □

Remark 5.10. *We could have avoided the use of the torus, by applying the OSSS inequality directly on the full space $\mathbb{Z}^d$. This would have required a version of the OSSS inequality that is a combination of Theorem 2.20 and 2.27: defined for Boolean functions on infinite products of infinite probability spaces. We have chosen not to include this most general version of the OSSS inequality for simplicity. Furthermore, the application of Russo's formula would have been more delicate, since the event in question would depend on the entire space-time.*

The above analysis of the diffential inequality gave us more than the exponential decay in the subcritical phase. Similar to Proposition 3.5, we obtain a lower-bound in the supercritical regime. Inequality (5.15) implies the following corollary, since $C(\lambda)$ is bounded away from 0 around λ_c.

Proposition 5.11. *There exists a constant $c > 0$, such that for all $\lambda > \lambda_c$,*

$$\mathbb{P}^{\eta^0}(\sigma(t) \not\equiv 0, \forall t > 0) \geq c\big((\lambda - \lambda_c) \wedge 1\big). \tag{5.16}$$

This lower bound also appeared as a by-product of the original proof for the sharp phase transition at λ_c by Bezuidenhout and Grimmett [10].

5.3.3 *Exponential bound in the supercritical regime*

The exponential decay of the survival time of the infection in the subcritical regime is complemented by an exponential decay result in the supercritical regime. If $\lambda > \lambda_c$, there is a positive probability that the infection survives indefinitely. However, if the infection dies out, it must do so exponentially quickly. To state this result, we introduce the stopping time

$$\tau := \inf\{t \geq 0 \,:\, \sigma(t) \equiv 0\},$$

so that $\tau = \infty$ if the infection survives indefinitely.

Theorem 5.12. *Let $\lambda > \lambda_c$. There exists a constant $c > 0$ such that for all $t \geq 0$,*

$$\mathbb{P}^{\eta^0}(t \leq \tau < \infty) \leq \exp(-ct). \tag{5.17}$$

This result can be found as Theorem 2.30 in [44]. The proof uses ideas of Durrett [20], building on the comparison between the contact process with oriented percolation by Bezuidenhout and Grimmett [9]. The OSSS inequality does not seem to lend itself to proving this result: if we take $f_t := \mathbb{1}\{t \leq \tau < \infty\}$ as the Boolean function, it seems hard to find a decision tree such that the revealment of a variable by this decision tree can be bounded by an event of the form $\{s \leq \tau < \infty\}$, for some $s > 0$ depending on the variable in question.

5.4 SHARP PERCOLATION PHASE TRANSITION FOR μ_t

In this section we will consider a different phase transition than in the previous section. Instead of looking at the spread of infection through time, we take a fixed time $t \geq 0$, and investigate the configuration at this fixed time. In particular we ask whether there exists an infinite cluster of infected vertices at that time. We first introduce the relevant measure on S, which is obtained from the process by projecting on time t. We consider the process started with all vertices infected: $\bar{\eta} \equiv 1$. For $t \geq 0$ we define the measure on $(S, \mathcal{S})$ by

$$\mu_t(A) = \mathbb{P}_\lambda^{\bar{\eta}}(\sigma(t) \in A),$$

for $A \in \mathcal{S}$. Note that μ_s dominates μ_t for all $s \leq t$, in the sense that

$$\mu_t(A) \leq \mu_s(A), \quad s \leq t, \tag{5.18}$$

for all increasing events $A \in \mathcal{S}$. (We call $A \in \mathcal{S}$ increasing whenever $\mathbb{1}_A(\eta) \leq \mathbb{1}_A(\eta')$ for all $\eta, \eta' \in S$ with $\eta \leq \eta'$.) We write $\mu \leq \nu$ if ν dominates μ. Naturally, the measure μ_t exhibits dependencies between the states of different vertices, which makes the analysis much more involved compared to the case of Bernoulli percolation, for which there are no dependencies under $\mathbb{P}_p$. Nevertheless, the question we investigate is the same as for Bernoulli percolation: is there a non-trivial percolation phase transition, and is this transition sharp? In order to define this phase transition, we recall the events from Chapter 3 on Bernoulli percolation: For $v, w \in V$, we say that $v \longleftrightarrow w$ whenever $\eta_w = 1$ and there exists a path from v to w for which every intermediate vertex u satisfies $\eta_u = 1$. In particular, we do not require v to be infected. This is a slightly uncommon definition of a connection event in site percolation, but is technically convenient for our purposes. For $A \subseteq V$, we say that $v \longleftrightarrow A$ whenever there exists $w \in A$ with $v \longleftrightarrow w$. We say that $v \longleftrightarrow \infty$, whenever $v \longleftrightarrow \partial\Lambda_n^v$ for all $n \in \mathbb{N}$. We define the critical point for percolation as

$$\lambda_p(t) := \sup\left\{\lambda \,:\, \mu_t(0 \longleftrightarrow \infty) = 0\right\}.$$

From the domination of μ_t, (5.18), it follows that $\lambda_p(t)$ is increasing in t. There is no simple argument to show that $\lambda_p(t)$ is non-trivial. In fact, if $t < -\log p_c(\mathbb{Z}^d)$, then $\lambda_p(t) = 0$, since by taking $\lambda = 0$ we are left with only recoveries, which results in a Bernoulli percolation process. However, we can prove that for large t, $\lambda_p(t)$ is non-trivial.

Proposition 5.13. *The critical thresholds for the contact process satisfy*

$$\lim_{t\to\infty} \lambda_p(t) \geq \lambda_c.$$

In particular, there exists $t_0 > 0$ such that $\lambda_p(t) > 0$, for all $t > t_0$.

Proof. Suppose $\lambda < \lambda_c$. We will show that $\mu_t(|\mathcal{C}| \geq n) \to 0$ as $n \to \infty$, for t large enough, where $\mathcal{C} = \mathcal{C}(t)$ is the cluster of 0 at time t. We partition $\mathbb{Z}^d$ into boxes of size $L \in \mathbb{N}$. We identify a box with its minimal corner:

$$B_x : xL + \{0, \dots, L-1\}^d, \quad x \in \mathbb{Z}^d.$$

In particular $B_x \cap B_y = \varnothing$, for $x \neq y$. We say that a box B_x is good whenever $0 \longleftrightarrow B_x$. Suppose $|\mathcal{C}| \geq n$. Then there are at least $\lceil n/L^d \rceil$ good boxes. Moreover, B_0 is good, and the indices of the boxes form a connected subset of $\mathbb{Z}^d$ that contains 0. Such a subset of $\mathbb{Z}^d$ is called a lattice animal. There exists a constant $A \in \mathbb{N}$ independent of n such that the number of lattice animals of size n is at most A^n. See Lemma 9.3 of Penrose [49] for details. Let $\mathcal{D} = \mathcal{D}(\mathcal{C})$ be the lattice animal of size equal to $\lceil n/L^d \rceil$ with $\mathcal{D} \subseteq \mathcal{C}$ obtained from $\mathcal{C}$ in a deterministic way. (E.g. we can choose the good boxes with lexicographical minimal indices that form a connected subset of $\mathbb{Z}^d$.) We find

$$\mu_t(|\mathcal{C}| \geq n) = \sum_{\substack{D \text{ lattice animal} \\ |D| = \lceil n/L^d \rceil}} \mu_t(\mathcal{C} \geq n, \mathcal{D} = D).$$

We can define truncated versions of $\sigma_v(t)$ as follows. For $L \in \mathbb{N}$ we define

$$\sigma_v^L(t) := \mathbb{1}\{(\Lambda_L^v, 0) \longrightarrow (v,t) \cup (\partial\Lambda_L^v, [0,t]) \longrightarrow (v,t)\}, \quad v \in \mathbb{Z}^d.$$

Then $\sigma_v^L(t) \geq \sigma_v(t)$. Furthermore, $\sigma_v^L(t)$ only depends on the space-time $(\Lambda_L^v, [0,t])$, so that $\sigma_v^L(t)$ and $\sigma_w^L(t)$ are independent whenever $\|v - w\|_\infty > 2L$. We now take $D' \subseteq D$ with $|D'| \geq \lceil n/L^d \rceil / 3^d$ deterministically, such that $\|x - y\|_\infty \geq 3$ for all $x, y \in D'$. It then follows that $\sigma_v^L(t)$ and $\sigma_w^L(t)$ are independent for all $v \in B_x$ and $w \in B_y$, for all $x, y \in D'$ with $x \neq y$. We say that a box B_x is L-infected if there exists $v \in B_x$ satisfying $\sigma_v^L(t) = 1$. The event that B_x is L-infected is independent of the event that B_y is L-infected for all $x, y \in D'$, $x \neq y$. Hence,

$$\begin{aligned}
&\mu_t(|\mathcal{C}| \geq n) \\
&\quad = \sum_{\substack{D \text{ lattice animal} \\ |D| = \lceil n/L^d \rceil}} \mathbb{P}_\lambda(\mathcal{C}(t) \geq n, \mathcal{D} = D, B_x \text{ is } L\text{-infected } \forall x \in D') \\
&\quad \leq \sum_{\substack{D \text{ lattice animal} \\ |D| = \lceil n/L^d \rceil}} \mathbb{P}_\lambda(\exists v \in B_x : \sigma_v^L(t) = 1 \; \forall x \in D') \\
&\quad \leq A^{\lceil n/L^d \rceil} \mathbb{P}_\lambda(\exists v \in B_0 : \sigma_v^L(t) = 1)^{\lceil n/L^d \rceil / 3^d} \\
&\quad \leq \Big(\big(A^{3^d} \mathbb{P}_\lambda(\exists v \in B_0 : \sigma_v^L(t) = 1) - A^{3^d} \mu_t(\exists v \in B_0 : \eta_v = 1) \\
&\qquad + A^{3^d} \mu_t(\exists v \in B_0 : \eta_v = 1) \big)^{1/3^d} \Big)^{\lceil n/L^d \rceil}.
\end{aligned}$$

We now take $L := t \in \mathbb{N}$ large enough such that

$$\begin{aligned}
A^{3^d} &\big(\mathbb{P}_\lambda(\exists v \in B_0 : \sigma_v^L(t) = 1) - \mu_t(\exists v \in B_0 : \eta_v = 1) \big) \\
&= A^{3^d} \big(\mathbb{P}_\lambda(\exists v \in B_0 : (\partial\Lambda_L^v, [0,t]) \longrightarrow (v,t)) \big) \\
&\leq C A^{3^d} t^d \exp(-t) < \frac{1}{4},
\end{aligned}$$

where we used the union bound and Proposition 5.3. We can simultaneously take t large enough such that

$$A^{3d}\mu_t(\exists v \in B_0 : \eta_v = 1) \le A^{3d}t^d\mu_t(\eta_0 = 1) < \frac{1}{4},$$

which is possible by Theorem 5.9 and Proposition 5.2. It follows that

$$\mu_t(|\mathcal{C}| \ge n) \le \exp\Big(-\frac{\log(2)}{(3t)^d}n\Big) \to 0, \quad n \to \infty.$$

□

In Section 5.5 we give an argument for $\lim_{t\to\infty}\lambda_p(t) < \infty$, so that $\lambda_p(t)$ is indeed non-trivial for large t. We will now show that the corresponding phase transition is sharp.

Theorem 5.14. *Let $t \ge 0$, and suppose $\lambda < \lambda_p(t)$. There exists a constant $c := c(t) > 0$, such that for all $n \in \mathbb{N}$,*

$$\mu_t(0 \longleftrightarrow \partial\Lambda_n) \le \exp(-cn). \tag{5.19}$$

Proof. This proof is in some sense a combination of the proof for Bernoulli percolation, Section 3.2, and the proof for the sharp phase transition at λ_c. Similar to the latter proof, we first consider the process on a finite domain. Let $t \ge 0$, and $N, n \in \mathbb{N}$. We then consider the space-time $\Lambda_{N+n} \times [0, t)$. We take $\varepsilon > 0$, such that $t \in \varepsilon\mathbb{Z}$, and partition this space by taking $\Omega_{v,s} = \{v\} \times [s, s+\varepsilon)$, for $s \in \varepsilon\mathbb{Z}$ with $0 \le s < t$, and $v \in \Lambda_{N+n}$. The probability space then satisfies $\Omega = \prod_{v,s}\Omega_{v,s}$. We define the Boolean function $f_n := \mathbb{1}\big\{0 \overset{\Lambda_{N+n}}{\longleftrightarrow} \partial\Lambda_n\big\}$. We further define $\theta_n^N(\lambda) := \mu_t\big(0 \overset{\Lambda_{N+n}}{\longleftrightarrow} \partial\Lambda_n\big)$.

Unsurprisingly, we define a decision tree that determines the value of f_n. It is in this algorithm that the union of both proofs is most apparent. As in the proof for Bernoulli percolation, we introduce several decision trees T_k, for $1 \leq k \leq n$. The tree T_k then explores the connections from $\partial\Lambda_k$ at time t. To determine whether a vertex is infected at this time, we need to explore space-time to see if there is a space-time connection $(\Lambda_{N+n}, 0) \longrightarrow (v, t)$. For this exploration process we define a subalgorithm $\mathsf{Explore}(v)$ that explores the backward space-time cluster of v, the same as the backward exploration process of Algorithm 4, however we do not stop this process once we have a connection $(\Lambda_{N+n}, 0) \longrightarrow (v, t)$. Instead, we explore the entire backward space-time cluster of v. After this process is completed, the state of v is known, and we either add it to the set of known infected vertices $\mathcal{C}$, or the set of known healthy vertices $\mathcal{T}$. The main algorithm then resembles Algorithm 1. It explores the space cluster of $\partial\Lambda_k$ by taking vertices from the set of active set $\mathcal{A}$ and determining their state. If at some point the algorithm sees $0 \stackrel{\mathcal{C}}{\longleftrightarrow} \partial\Lambda_n$, it returns 1. Otherwise, if the algorithm runs out of vertices to explore, it returns 0. The pseudocode of T_k is given in Algorithm 5.

5.4.1 *Bound on the Revealment*

Recall that the state of w is irrelevant for the event $w \longleftrightarrow \partial\Lambda_k$. Therefore, for every vertex $w \in \mathcal{A}$, we have $w \longleftrightarrow \partial\Lambda_k$. When a variable (v, s) is revealed, this occurs during the exploration of the backward space-time cluster of a vertex $w \in \mathcal{A}$, when $(\Lambda_v^1, s) \longrightarrow (w, t)$. Since $w \in \mathcal{A}$, we know $w \longleftrightarrow \partial\Lambda_k$. Moreover, the variable (v, s) was not revealed during the exploration processes of the vertices that make up this path. It follows that

Function `Explore`(v)**:**

$\mathcal{B} := \{(v, t-\varepsilon)\}$;
while $\mathcal{B} \neq \emptyset$ **do**
  Take minimal $(v,s) \in \mathcal{B}$;
  Reveal $\omega_{v,s}$;
  $\mathcal{R} := \mathcal{R} \cup \{(v,s)\}$;
  $\mathcal{B} := \mathcal{B} \backslash \{(v,s)\}$;
  $\mathcal{B} := \mathcal{B} \cup \big\{(w,s') \in (\Lambda_{N+n} \times \varepsilon\mathbb{Z}) \backslash \mathcal{R} :$
   $0 \le s' < t,\ \exists x \sim w, s'' \in [s', s'+\varepsilon), (x,s'') \xrightarrow{\mathcal{R}} (v,t)\big\}$;
  $\mathcal{B} := \mathcal{B} \cup \big\{(w,s') \in (\Lambda_{N+n} \times \varepsilon\mathbb{Z}) \backslash \mathcal{R} :$
   $0 \le s' < t,\ (w, s'+\varepsilon) \xrightarrow{\mathcal{R}} (v,t)\big\}$;
end

$\mathcal{A} := \partial\Lambda_k$;
$\mathcal{C} := \emptyset$;
$\mathcal{R} := \emptyset$;
$\mathcal{T} := \emptyset$;
while $\mathcal{A} \neq \emptyset$ **do**
  Take minimal $v \in \mathcal{A}$;
  `Explore` (v);
  $\mathcal{A} = \mathcal{A} \backslash \{v\}$;
  if $\sigma_v(t) = 1$ **then** $\mathcal{C} = \mathcal{C} \cup \{v\}$;
  if $\sigma_v(t) = 0$ **then** $\mathcal{T} = \mathcal{T} \cup \{v\}$;
  $\mathcal{A} := \mathcal{A} \cup \big\{w \in \Lambda_{N+n} \backslash (\mathcal{C} \cup \mathcal{T}) : w \xleftrightarrow{\mathcal{C}} \partial\Lambda_k\big\}$;
  if $0 \xleftrightarrow{\mathcal{C}} \partial\Lambda_n$ **then return** 1;
end
return 0;

Algorithm 5: The exploration algorithm T_k.

$(\Lambda_v^1, s) \longrightarrow (w,t)$ and $w \in \mathcal{A}$, $w \longleftrightarrow \partial\Lambda_k$ occur disjointly. Using the union bound and the BK inequality, Theorem 5.6, we obtain

$$\begin{aligned}\mathrm{Rev}_{v,s}(T_k) &\leq \sum_{w\in\Lambda_{N+n}} \mathbb{P}_\lambda\Big(w \longleftrightarrow \partial\Lambda_k \circ (\Lambda_v^1, s) \xrightarrow{\Lambda_{N+n}} (w,t)\Big)\\ &\leq \sum_{w\in\Lambda_{N+n}} \mathbb{P}_\lambda\Big(w \overset{\Lambda_{N+n}}{\longleftrightarrow} \partial\Lambda_k\Big)\mathbb{P}_\lambda\Big((\Lambda_v^1, s) \xrightarrow{\Lambda_{N+n}} (w,t)\Big).\end{aligned}$$

Summing over k gives

$$\begin{aligned}\sum_{k=1}^n \mathrm{Rev}_{v,s}(T_k) &\leq \sum_{w\in\Lambda_{N+n}} \sum_{k=1}^n \mathbb{P}_\lambda\Big(w \overset{\Lambda_{N+n}}{\longleftrightarrow} \partial\Lambda_k\Big)\mathbb{P}_\lambda\Big((\Lambda_v^1, s) \xrightarrow{\Lambda_{N+n}} (w,t)\Big)\\ &\leq 2S_n^N \sum_{w\in\Lambda_{N+n}} \mathbb{P}_\lambda\Big((\Lambda_v^1, s) \xrightarrow{\Lambda_{N+n}} (w,t)\Big),\end{aligned}$$

similar to 3.5, where $S_n^N := S_n^N(\lambda) := \sum_{k=0}^n \theta_k^N(\lambda)$. We now use the bound on the growth of the infection, Proposition 5.3, with $c = 1$, to find a constant $a > 0$ such that

$$\begin{aligned}&\sum_{w\in\Lambda_{N+n}} \mathbb{P}_\lambda\Big((\Lambda_v^1, s) \xrightarrow{\Lambda_{N+n}} (w,t)\Big)\\ &\qquad\leq 2d \sum_{w\in\Lambda_{N+n}} \mathbb{P}_\lambda\Big((v,0) \xrightarrow{\Lambda_{N+n}} (w,t-s)\Big)\\ &\qquad\leq 2d(a(t-s))^d + 2d \sum_{\substack{w\in\Lambda_{N+n}:\\ \|v-w\|\geq a(t-s)}} \mathbb{P}_\lambda\Big((v,0) \xrightarrow{\Lambda_{N+n}} (w,t)\Big)\\ &\qquad\leq 2d(at)^d + 2da\exp(-(t-s))\\ &\qquad\leq 2d(at)^d + 2da.\end{aligned}$$

Setting $C(t) := 4d(at)^d + 4da$, gives $\sum_{k=1}^n \mathrm{Rev}_{v,s}(T_k) \leq C(t)S_n^N(\lambda)$.

5.4.2 *Analysis of the differential inequality*

From here the proof resembles the proof for the sharpness of the phase transition at λ_c to a large extent. We set $A := 0 \overset{\Lambda_{N+a}}{\longleftrightarrow} \partial\Lambda_n$, which is an increasing event. By using Russo's formula, Proposition 5.5, as in Section 5.3.2 Similar as we obtain

$$\sum_{v\in\Lambda_{N+n}} \sum_{\substack{s\in\varepsilon\mathbb{Z}\\ 0\le s<t}} \mathrm{Inf}_{v,s}(f_n) \le \mathbb{E}|\mathrm{Piv}| + O(\varepsilon) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_n^N(\lambda) + O(\varepsilon),$$

where Piv is the set of Poisson points pivotal for A. Using the OSSS inequality, Theorem 2.20, the bound on the revealment and letting $\varepsilon \to 0$ gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_n^N(\lambda) &\ge \sum_{k=1}^{n} \sum_{v\in\Lambda_{N+n}} \sum_{\substack{s\in\varepsilon\mathbb{Z}\\ 0\le s<t}} \mathrm{Inf}_{v,s}(f_n)\mathrm{Rev}_{v,s}(T_k)\\ &\ge \frac{1}{C(t)}\frac{n}{S_n^N}\theta_n^N(\lambda)\big(1-\theta_n^N(\lambda)\big). \qquad (5.20)\end{aligned}$$

We can bound

$$1-\theta_n^N(\lambda) \ge \frac{1}{(2d\lambda+1)^{2d}},$$

since if the last Poisson point on each of the time axis $\{v\}\times[0,t)$ is a healing for all $v \sim 0$, 0 is not connected to $\partial\Lambda_n$. We thus obtain

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_n^N(\lambda) \ge \frac{1}{C(t)(2d\lambda+1)^{2d}}\frac{n}{S_n^N}\theta_n^N(\lambda),$$

from which we can obtain the sharpness of the phase transition similarly to Section 5.3.2. During this analysis we also let $N \to \infty$, noting that $\theta_n^N(\lambda)$ is increasing in N. We define

$$\theta_n(\lambda) = \lim_{N\to\infty} \theta_n^N(\lambda) = \mu_t(0 \longleftrightarrow \partial\Lambda_n), \quad S_n := \lim_{N\to\infty} S_n^N.$$

We then consider the alternative critical point

$$\tilde{\lambda}_p(t) := \sup\Big\{\lambda \,:\, \limsup_{n\to\infty} \frac{\log S_n}{n} < 1\Big\}.$$

If $\lambda_0 < \lambda_1 < \lambda_2 < \tilde{\lambda}_p$, we obtain with the same bootstrap argument as in Section 5.3.2

$$\theta_n(\lambda_0) \le \exp\Big(-\frac{1}{C(t)(2d\lambda_2+1)^{2d}}n\Big),$$

so that $\tilde{\lambda}_p(t) \le \lambda_p(t)$. If on the other $\tilde{\lambda}_p(t) < \lambda_1 < \lambda_2$, we define

$$T_n^N(\lambda) := \frac{1}{\log n}\sum_{k=1}^n \frac{\theta_k^N(\lambda)}{k},$$

$$T_n(\lambda) := \lim_{N\to\infty} T_n^N(\lambda) = \frac{1}{\log n}\sum_{k=1}^n \frac{\theta_k(\lambda)}{k}.$$

We then obtain as in Section 5.3.2,

$$T_n^N(\lambda_2) - T_n^N(\lambda_1) \ge \frac{1}{C(t)(2d\lambda_2+1)^{2d}} \frac{\log S_n^N(\lambda_2)}{\log n}(\lambda_2-\lambda_1),$$

so that by letting $N \to \infty$ and subsequently $n \to \infty$, we find

$$\begin{aligned}\theta(\lambda_2) \ge \theta(\lambda_2) - \theta(\lambda_1) &:= \lim_{n\to\infty}\big(T_n(\lambda_2) - T_n(\lambda_1)\big) \\ &\ge \frac{1}{C(t)(2d\lambda_2+1)^{2d}}(\lambda_2-\lambda_1).\end{aligned}$$

It follows that $\tilde{\lambda}_p(t) = \lambda_p(t)$, which finishes the proof. □

From the analysis of the differential inequality we obtained a lower bound on the probability that 0 is in an infinite cluster in the supercritical regime.

Proposition 5.15. *There exists constant $C, c > 0$ such that for all $t \ge 0$ and all $\lambda > \lambda_p(t)$*

$$\mu_t(0 \longleftrightarrow \infty) \ge \frac{1}{ct^d + C}\big((\lambda - \lambda_p(t)) \wedge 1\big). \tag{5.21}$$

5.5 THE UPPER INVARIANT MEASURE

In the previous section we have introduced the time t measure μ_t. In this section we will investigate what happens for $t \to \infty$. As a first consideration we can use the graphical representation and the duality of the process to find

$$\begin{aligned}\lim_{t\to\infty} \mu(\sigma_v = 1) &= \lim_{t\to\infty} \mathbb{P}_\lambda\big((0,0) \longrightarrow (\mathbb{Z}^d, t)\big)\\ &= \mathbb{P}^{\eta^0}(\sigma(t) \not\equiv 0, \ \forall t \geq 0),\end{aligned}$$

where η^0 is the configuration where only 0 is infected. This last probability is positive if and only if $\lambda > \lambda_c$. This hints that μ_t converges to a non-trivial measure if $\lambda > \lambda_c$. We will now make this notion precise. Let $\mathcal{P}(S)$ denote the set of probability measures on $(S, \mathcal{S})$.

Proposition 5.16. *There exists $\bar{\nu}_\lambda \in \mathcal{P}(S)$, such that*

$$\mu_t \to \bar{\nu}_\lambda, \quad \textit{as } t \to \infty,$$

with respect to the topology of weak convergence. I.e.,

$$\int_S f \,\mathrm{d}\mu_t \to \int_S f \,\mathrm{d}\bar{\nu}_\lambda, \quad \textit{as } t \to \infty,$$

for all bounded $f \in C(S)$.

Proof. The fact that S is compact with respect to the chosen topology ρ implies that $\mathcal{P}(S)$ is compact. (See Klenke [39, Corollary 13.30].) It thus follows that $(\mu_t)_{t\geq 0}$ has a convergent subsequence. We denote the limit of this subsequence by ν_1. Then the monotonicity of μ_t in t implies

$$\int_S f \,\mathrm{d}\nu_1 \leq \int_S f \,\mathrm{d}\mu_t, \quad t \geq 0$$

for all increasing $f \in C(S)$. This follows from (5.18) and a standard measure-theoretic approximation of f by simple functions. ($f \in C(S)$ is increasing if $f(\eta) \le f(\eta')$ for all $\eta, \eta' \in S$ with $\eta \le \eta'$.) Suppose $(\mu_t)_{t\ge 0}$ has another convergent subsequence converging to $\nu_2 \in \mathcal{P}$. Then

$$\int_S f \,\mathrm{d}\nu_1 = \int_S f \,\mathrm{d}\nu_2, \tag{5.22}$$

for all increasing functions f. Functions of the form $f = \mathbb{1}_A$, where $A \in \mathcal{T}$, with

$$\mathcal{T} := \Big\{ \{\eta \in S \,:\, \eta(v_1) = \cdots = \eta(v_n) = 1\} \,:\, n \in \mathbb{N}, v_1, \ldots, v_n \in \mathbb{Z}^d \Big\},$$

are continuous and increasing. It then follows from (5.22), that $\nu_1(A) = \nu_2(A)$, and since $\mathcal{T}$ is an intersection stable generator of $\mathcal{S}$, that $\nu_1 = \nu_2$. We conclude that all converging subsequences of $(\mu_t)_{t\ge 0}$ converge to the same limit, which implies that μ_t itself converges, as $t \to \infty$, since $\mathcal{P}(S)$ is compact. □

The limiting measure $\bar{\nu}_\lambda \in \mathcal{P}(S)$ is called the upper invariant measure. This refers to the fact that we have started the process with all vertices infected, and to the fact that this measure is invariant under the probability semigroup T_t of the contact process. The latter is a direct consequence of the fact that $\bar{\nu}_\lambda$ is a limiting measure for $t \to \infty$.

We can obtain a second invariant measure by starting the process with all vertices healthy, and letting $t \to \infty$. However, this lower invariant measure is trivial: $\underline{\nu}_\lambda = \delta_0$, the measure with all mass on the configuration with all vertices healthy, since there are no spontaneous infections in the model. Starting with all vertices infected or all vertices healthy are the most extreme starting configurations, which also result in the extreme invariant measures. To formalize this statement, let $\mu \in \mathcal{P}(s)$, and let $\nu := \lim_{t\to\infty} \mu T_t$ be the weak limiting measure when we start the process in a configuration with law μ and let $t \to \infty$. Then we can directly see from the graphical representation, that

$$\int_S f \, \mathrm{d}\underline{\nu}_\lambda \le \int_S f \, \mathrm{d}\nu \le \int_S f \, \mathrm{d}\bar{\nu}_\lambda,$$

for all increasing functions $f \in C(S)$. This is further complemented by the complete convergence theorem.

Theorem 5.17 (Complete convergence theorem)**.** *Let $\mu \in \mathcal{P}$. Then there exists $0 \le \alpha \le 1$ such that*

$$\mu T_t \to \alpha \underline{\nu}_\lambda + (1-\alpha)\bar{\nu}_\lambda, \quad \textit{weakly as } t \to \infty.$$

This result is due to Bezuidenhout and Grimmett [9], who proved this result with the same technique used to prove that the infection dies out at λ_c. It shows that $\bar{\nu}_\lambda$ is the measure that is naturally of interest when analyzing the contact process, and the rest of the chapter is dedicated to the analysis of the properties of this measure.

Now that we have established the existence of the limiting measure, we are interested in the rate of convergence. Control over this rate of convergence is useful in analyzing $\bar{\nu}_\lambda$, since the measure μ_t is often more tractable. The next result shows that there is an exponential rate of convergence in the case $f = \mathbb{1}\{\sigma_0 = 1\}$.

Theorem 5.18. *Consider the contact process on $\mathbb{Z}^d$ with parameter $\lambda \neq \lambda_c$. Then there exists a constant $c > 0$, such that for all $t \ge 0$,*

$$\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1) \le \exp(-ct). \tag{5.23}$$

Proof. Suppose first that $\lambda < \lambda_c$. Then

$$\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1) \le \mu_t(\eta_0 = 1) = \mathbb{P}_\lambda^{\eta^0}(\sigma(t) \not\equiv 0),$$

by the duality relation for the contact process, Proposition 5.2, where η^0 is the configuration where only 0 is infected. We conclude from Theorem 5.9, that

$$\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1) \le \mathbb{P}_\lambda^{\eta^0}(\sigma(t) \not\equiv 0) \le \exp(-ct),$$

for some constant $c > 0$, independent of t.

Now suppose $\lambda > \lambda_c$. We can again use the duality of the process to find

$$\begin{aligned} &\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1) \\ &\quad = \mathbb{P}_\lambda\big((0,0) \longrightarrow (\mathbb{Z}^d, t),\ \exists s > t \text{ s.t. } (0,0) \nrightarrow (\mathbb{Z}^d, s)\big). \end{aligned}$$

In other words, the infection survives up to time t, but dies out eventually. We can bound this probability using Theorem 5.12, and find

$$\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1) \le \exp(-ct),$$

for some constant $c > 0$, independent of t. □

We can harvest the previous theorem to find a bound for events that depend on the state of multiple vertices.

Corollary 5.19. *Consider the contact process on $\mathbb{Z}^d$ with parameter $\lambda \neq \lambda_c$. Then there exists a constant $c > 0$, such that for all $t \ge 0$, and all events $A \in \mathcal{S}$,*

$$\big|\mu_t(A) - \bar{\nu}_\lambda(A)\big| \le |\Lambda_A| \exp(-ct), \tag{5.24}$$

where $\Lambda_A \subseteq \mathbb{Z}^d$ is any set of vertices such that A is measurable with respect to the status of these vertices.

The above bound is essentially a bound on the total variation distance between μ_t and $\bar{\nu}_\lambda$, except that the bound deteriorates as $|\Lambda_A|$ grows. The corollary follows directly from Theorem 5.18 and the union bound. Theorem 5.18 can also be used to show that correlations with repsect to the upper invariant measure between the state of two vertices decays exponentially as the distance between the vertices increases.

Corollary 5.20. *Suppose $\lambda \geq 0$. There exits a constant $c > 0$ such that for all $v \in \mathbb{Z}^d$*

$$\mathrm{Cov}_{\bar{\nu}_\lambda}(\eta_0, \eta_v) \leq \exp\big(-c\|v\|_\infty\big).$$

Proof. The result is trivial if $\lambda \leq \lambda_c$, since in this case $\bar{\nu}_\lambda(\eta_0 = 0) = 1$. We thus henceforth assume $\lambda > \lambda_c$. Let $v \in \mathbb{Z}^d$. For $t \geq 0$ and $a > 0$ we define the events

$$\begin{aligned} A_0 &:= \Big\{\big(\partial\Lambda_t^0, [0, t/a]\big) \longrightarrow (0, t/a)\Big\}, \\ A_v &:= \Big\{\big(\partial\Lambda_t^v, [0, t/a]\big) \longrightarrow (v, t/a)\Big\}. \end{aligned}$$

We fix $a > 0$ obtained from Proposition 5.3, such that

$$\mathbb{P}_\lambda(A_0) = \mathbb{P}_\lambda(A_v) \leq aC\exp(-t),$$

for a constant $C > 0$, and all $t > 0$. We now take $t = \|v\|_\infty/3$ and use Theorem 5.18 to show that for some $c_0 > 0$,

$$\begin{aligned} \mathrm{Cov}_{\bar{\nu}_\lambda}(\eta_0, \eta_v) &= \mathbb{E}_{\bar{\nu}_\lambda}[\eta_0\eta_v] - \mathbb{E}_{\bar{\nu}_\lambda}[\eta_0]^2 \\ &\leq \mathbb{P}_\lambda\big(\sigma_0(t) = \sigma_v(t) = 1, A_0^c, A_v^c\big) + 2aC\exp(-t). \end{aligned}$$

We note that $A_0^c \cap \{\sigma_0(t) = 1\}$ is independent of $A_v^c \cap \{\sigma_v(t) = 1\}$, since the events depend on disjoint regions in space-time. Hence, using translation invariance,

$$\begin{aligned}\mathrm{Cov}_{\bar{\nu}_\lambda}(\eta_0, \eta_v) &\leq \mu_t(\eta_0 = 1)^2 - \bar{\nu}_\lambda(\eta_0 = 1)^2 + 2aC\exp(-t) \\ &\leq 2\big(\mu_t(\eta_0 = 1) - \bar{\nu}_\lambda(\eta_0 = 1)\big) + 2aC\exp(-t) \\ &\leq 2\exp\Big(-\frac{c}{3}\|v\|_\infty\Big) + 2aC\exp\Big(-\frac{1}{3}\|v\|_\infty\Big),\end{aligned}$$

by Theorem 5.18, for all $v \in \mathbb{Z}^d$. The result follows by adjusting the constant in the exponent to accommodate vertices v close to 0. □

A major source of difficulty in analyzing the upper invariant measure is the lack of monotonicity of this measure. Here we mean monotonicity in the sense that the measure is strongly positively-associated, see Definition 2.24 and Theorem 2.26. We now give a proof that $\bar{\nu}_\lambda$ is not monotonic for the process on $\mathbb{Z}$, and a specific choice of the parameter. This shows that $\bar{\nu}_\lambda$ is not monotonic in general, and that we cannot rely on the OSSS inequality for monotonic measures, Theorem 2.27, when analyzing the properties of the measure. It is conceivable that similar counterexamples to the monotonicity can be constructed for $\mathbb{Z}^d$ and other values of λ.

Proposition 5.21. *Consider the contact process on $\mathbb{Z}$. The upper invariant measure $\bar{\nu}_\lambda$ is not monotonic for $\lambda_c < \lambda < 2$.*

Proof. We follow Liggett [41]. We suppose $\bar{\nu}_\lambda$ is monotonic, and work toward a contradiction. For $n \in \mathbb{N}$, we define the conditional probability

$$F(n) := \bar{\nu}_\lambda\big(\eta_v = 0, 1 \leq v \leq n-1 \,\big|\, \eta_0 = 1\big),$$

which exists, since $\lambda > \lambda_c$. Using the assumed monotonicty of $\bar{\nu}_\lambda$, we find for $1 \le k \le n$,

$$\begin{aligned}
&\bar{\nu}_\lambda(\eta_k = 1, \eta_v = 0, 1 \le v \le n, v \ne k) \\
&\qquad = \bar{\nu}_\lambda\big(\eta_v = 0, 1 \le v \le n, v \ne k \,\big|\, \eta_k = 1\big)\bar{\nu}_\lambda(\eta_k = 1) \\
&\qquad \ge \bar{\nu}_\lambda\big(\eta_v = 0, 1 \le v < k \,\big|\, \eta_k = 1\big)\bar{\nu}_\lambda(\eta_k = 1) \\
&\qquad\quad \cdot \bar{\nu}_\lambda\big(\eta_v = 0, k < v \le n \,\big|\, \eta_k = 1\big),
\end{aligned}$$

since $\{\eta_v = 0, 1 \le v < k\}$ and $\{\eta_v = 0, k < v \le n\}$ are both decreasing events. Using the symmetry and translation invariance of the process, we obtain

$$\bar{\nu}_\lambda(\eta_k = 1, \eta_v = 0, 1 \le v \le n, v \ne k) \ge F(k)F(n-k+1)\bar{\nu}_\lambda(\eta_0 = 1).$$

We compute

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t}\bar{\nu}_\lambda(\eta_v = 0, 1 \le v \le n)\Big|_{t=0} \\
&= -\lambda\bar{\nu}_\lambda(\eta_0 = 1, \eta_v = 0, 1 \le v \le n) \\
&\quad - \lambda\bar{\nu}_\lambda(\eta_{n+1} = 1, \eta_v = 0, 1 \le v \le n) \\
&\quad + \sum_{k=1}^{n} \bar{\nu}_\lambda(\eta_k = 1, \eta_v = 0, 1 \le v \le n, v \ne k).
\end{aligned}$$

It follows that

$$\sum_{k=1}^{n} F(k)F(n-k+1)\bar{\nu}_\lambda(\eta_0 = 1) \le 2\lambda F(n+1)\bar{\nu}_\lambda(\eta_0 = 1).$$

Summing this inequality over n and defining $M := \sum_{n=1}^{\infty} F(n)$ gives

$$M^2 \le 2\lambda(M-1),$$

since $F(1) = 1$. Assuming $M < \infty$, the above inequality only has solutions in M if $\lambda \geq 2$. Which contradicts the assumption $\lambda_c < \lambda < 2$.

It remains to show that $M < \infty$. Using the FKG inequality we obtain

$$M \leq \frac{1}{\bar{\nu}_\lambda(\eta_0 = 1)} \sum_{l=0}^{\infty} \bar{\nu}_\lambda(\eta_1 = \cdots = \eta_l = 0).$$

We take $t = \sqrt{l}$, so that by Corollary 5.19 there exists $c \geq 0$ such that

$$\bar{\nu}_\lambda(\eta_1 = \cdots = \eta_l = 0) \leq \mu_t(\eta_1 = \cdots = \eta_l = 0) + l\exp(-ct). \quad (5.25)$$

For $k = 1, \dots, \lfloor\sqrt{t}\rfloor$, we define $v_k := k\lfloor\sqrt{l}\rfloor$, so that $|v_k - v_{k+1}| \geq \lfloor\sqrt{l}\rfloor$, for all $k = 1, \dots, \lfloor\sqrt{t}\rfloor - 1$. We define the event

$$A_k := \left\{\left(\partial\Lambda^{v_k}_{\lfloor\sqrt{l}\rfloor/2}, [0,t]\right) \longrightarrow (v_k, t)\right\},$$

Then, using the union bound,

$$\begin{aligned}\mu_t(\eta_1 = \cdots = \eta_l = 0) &\leq \mu_t\left(\bigcap_{k=1}^{\lfloor\sqrt{l}\rfloor} A_k^c \cap \{\eta_{v_k} = 0\}\right) + \sum_{k=1}^{\lfloor\sqrt{l}\rfloor} \mu_t(A_k) \\ &\leq \prod_{k=1}^{\lfloor\sqrt{l}\rfloor} \mu_t(\eta_{v_k} = 0) + \sum_{k=1}^{\lfloor\sqrt{l}\rfloor} \mu_t(A_k),\end{aligned}$$

since the events $A_k^c \cap \{\eta_{v_k} = 0\}$ are independent for all k, since we have choosen the vertices v_k far enough apart. We use Theorem 5.18 to bound the first term, and the bound on the growth of the infection, Proposition 5.3 for the second term to obtain

$$\begin{aligned}\mu_t(\eta_1 = \cdots = \eta_l = 0) \leq{}& \left(\bar{\nu}_\lambda(\eta_{v_k} = 0) + \exp(-c\lfloor\sqrt{l}\rfloor)\right)^{\lfloor\sqrt{l}\rfloor} \\ &+ \lfloor\sqrt{l}\rfloor\exp(-\lfloor\sqrt{l}\rfloor),\end{aligned}$$

for some constant $c > 0$. Since $\lambda > \lambda_c$, we have $\bar{\nu}_\lambda(\eta_{v_k} = 0) < 1$. Using (5.25), it follows that there exists a constant $c_0 > 0$, such that for l large enough,

$$\begin{aligned}\bar{\nu}_\lambda(\eta_1 = \cdots = \eta_l = 0) &\leq \mu_t(\eta_1 = \cdots = \eta_l = 0) + l\exp(-c\lfloor\sqrt{l}\rfloor)\\ &\leq \exp(-c_0\lfloor\sqrt{l}\rfloor) + l\exp(-c\lfloor\sqrt{l}\rfloor).\end{aligned}$$

We conclude that $\sum_{l=0}^{\infty}\bar{\nu}_\lambda(\eta_1 = \cdots = \eta_l = 0)$ converges, so that $M < \infty$. $\square$

5.6 SHARP PERCOLATION PHASE TRANSITION FOR $\bar{\nu}_\lambda$

We are interested in the percolation properties of the upper invariant measure. We define the critical threshold for percolation as

$$\lambda_p := \lambda_p(\mathbb{Z}^d) := \sup\left\{\lambda : \bar{\nu}_\lambda(0 \longleftrightarrow \infty) = 0\right\}. \tag{5.26}$$

As a first consideration we see that this critical value satisfies $\lambda_c \leq \lambda_p$, so that in particular $\lambda_p > 0$. It is more involved to show that $\lambda_p < \infty$. In other words, it is more difficult to show that the cluster of 0 is infinite with positive probability for λ large enough. In fact, $\lambda(\mathbb{Z}) = \infty$, since $\bar{\nu}_\lambda(\sigma_0 = 0) > 0$ for all $\lambda > 0$. Nevertheless, it is true that $\lambda_p < \infty$ for $d \geq 2$, and this follows from the domination result, dominating a Bernoulli percolation measure with an arbitrary density by $\bar{\nu}_\lambda$ for some large λ.

Theorem 5.22. *Let $\mathbb{P}_p$ denote the Bernoulli percolation measure on $\mathbb{Z}^d$, $d \geq 1$ with $0 < p < 1$. There exists $\lambda > 0$ such that the upper invariant measure for the contact process on $\mathbb{Z}^d$ dominates $\mathbb{P}_p$:*

$$\mathbb{P}_p \leq \bar{\nu}_\lambda.$$

By taking $p > p_c(\mathbb{Z}^d)$ in the above theorem, we obtain the existence of $\lambda > 0$, such that

$$\bar{\nu}_\lambda(0 \longleftrightarrow \infty) \geq \mathbb{P}_p(0 \longleftrightarrow \infty) > 0,$$

so that $\lambda_p \leq \lambda < \infty$. Theorem 5.22 has been proven by Liggett and Steif [43]. It suffices to prove this theorem for $d = 1$: for $d \geq 2$ we can suppress all infection arrows of the graphical representation with direction other than $\pm e_1$. This results in family of independent one dimensional contact processes for which we can use the result for $d = 1$. The proof for $d = 1$ is then heavily one dimensional in nature. Furthermore, the proof is rather quantitative: $\bar{\nu}_\lambda$ dominates $\mathbb{P}_p$ for a given $0 < p < 1$ if λ is such that $\lambda - 2 \geq p\lambda$. This then implies $\lambda_p \leq 6.25$ for all $d \geq 2$, using a bound on the critical value of Bernoulli site percolation on $\mathbb{Z}^2$.

We are naturally interested in the question if the phase transition at λ_p is sharp. So far this question has only been answered for the two dimensional process.

Theorem 5.23. *Consider the contact process on $\mathbb{Z}^2$. Suppose $\lambda < \lambda_p$. Then there exists a constant $c > 0$, such that for all $n \in \mathbb{N}$,*

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn). \tag{5.27}$$

This theorem has been proven by Van den Berg [7]. It uses a generization of an influence bound due to Talagrand [52] applied to box-crossing events. This makes the proof very two dimensional in nature, and difficult to be generalized to higher dimensions. In the remainder of this section we reflect on ways to proof the generalization of Theorem 5.23 to higher dimensions.

Proposition 5.24. *Consider the contact process on $\mathbb{Z}^d$, $d \geq 2$ with parameter $\lambda < \lambda_p$. Assume $\lambda_p(t) \to \lambda_p$ as $t \to \infty$. Then there exists a constant $c > 0$, such that for all $n \in \mathbb{N}$,*

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn). \tag{5.28}$$

Proof. The proof is immediate from Theorem 5.14, since $\lambda < \lambda_p$ combined with $\lambda_p(t) \to \lambda_p$ implies $\lambda < \lambda_p(t)$ for some $t \geq 0$ large enough, and since $\bar{\nu}_\lambda \leq \mu_t$. □

The convergence of $\lambda_p(t) \to \lambda_p$ as $t \to \infty$ is also a necessary condition for the sharp phase transition at λ_p. The following proposition states that for $\lambda \in (\lim_{t\to\infty} \lambda_p(t), \lambda_p)$, $\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n)$ decays polylogarithmically in n. Thus if $\lim_{t\to\infty} \lambda_p(t) < \lambda_p$, we have a nonempty subcritical parameter range for which $\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n)$ does not decay exponentially quickly.

Proposition 5.25. *Consider the contact process on $\mathbb{Z}^d$, $d \geq 2$. Suppose $\lambda > \lim_{t\to\infty} \lambda_p(t)$. Then there exists a constant $c_0 > 0$, such that*

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \geq \frac{c_0}{(\log n)^d}. \tag{5.29}$$

In particular, if $\lambda < \lambda_p$ implies

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) < \frac{c_0}{(\log n)^d},$$

for all $n \in \mathbb{N}$, then there exists a constant $c > 0$, such that for all $n \in \mathbb{N}$,

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn). \tag{5.30}$$

Proof. Suppose $\lambda > \lim_{t\to\infty} \lambda_p(t)$. We use proposition 5.15 and Corollary 5.19 to bound

$$\begin{aligned}
&\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \\
&\quad \geq \mu_t(0 \longleftrightarrow \partial\Lambda_n) - (2n+1)^d \exp(-c_1 t) \\
&\quad \geq \frac{1}{c_2 t^d + C}(\lambda - \lambda_p(t)) - \exp\big(d\log(2n+1) - c_1 t\big) \\
&\quad \geq \frac{1}{c_2 t^d + C}\big(\lambda - \lim_{t\to\infty} \lambda_p(t)\big) - \exp\big(d\log(2n+1) - c_1 t\big),
\end{aligned}$$

for constants $c_1, c_2 > 0$, all $n \in \mathbb{N}$, and all $t > 0$ such that $\lambda > \lambda_p(t)$. Taking $t := t(n) = 2d/c_1 \log(2n+1)$ implies (5.29). The second statement of the proposition follows directly from Theorem 5.14, since if the implication holds, then $\lambda < \lambda_p = \lim_{t\to\infty} \lambda_p(t)$, so that we can take t large enough such that $\lambda < \lambda_p(t)$, from which it follows that

$$\bar{\nu}_\lambda(0 \longleftrightarrow \partial\Lambda_n) \leq \mu_t(0 \longleftrightarrow \partial\Lambda_n) \leq \exp(-cn).$$

□

The above proof strategy cannot be used directly to prove that $\lambda_p(t) \to \lambda_p$ as $t \to \infty$, since $1/(c_2 t^d + C) \to 0$ as $t \to \infty$. Conversely, if we can prove Proposition 5.15 with a constant that is independent of t, the convergence of the critical points follows. The factor $1/(c_2 t^d + C)$ comes from the bound of on the realement of variable $\omega_{v,s}$ in the graphical representation by the decision tree T_k determining the value of $\mathbb{1}\{0 \longleftrightarrow \partial\Lambda_n\}$. To determine the state of vertex at time t, we revealed the backward space-time cluster of v, which size is of order t^d if $\sigma_v(t) = 1$. Thus, we can not expect to bound the revealment of the variables uniformly in t.

The convergence of critical points such as the convergence of $\lambda_p(t) \to \lambda_p$ might be proven by comparing the partial derivatives with respect to t and with respect to λ of $\theta_n(t,\lambda) := \mu_t(0 \longleftrightarrow \partial\Lambda_n)$. This strategy is not unlike the approach to proving strict inequalities between critical as described in Section 3.2 of [25]. A Russo's formula for the derivative with respect to t of $\theta_n(t,\lambda)$ would state

$$\begin{aligned}
-&\frac{\partial}{\partial t}\theta_n(t,\lambda) \\
&= \sum_{v\in\Lambda_n} \mathbb{P}_\lambda\big(v \text{ pivotal for } 0 \longleftrightarrow \partial\Lambda_n,\ \sigma_v(-t) \neq \lim_{s\to-\infty} \sigma_v(-s)\big),
\end{aligned}$$

where we made use of the graphical representation of the contact process defined on $\mathbb{Z}^d \times (-\infty, 0]$, which is typically more convenient when analyzing the upper invariant measure. (In this case $\sigma_v(0) = 1$ if and only if $(\mathbb{Z}^d, -t) \longrightarrow (v, 0)$, for all $t \geq 0$, so that $\sigma(0)$ has law $\bar{\nu}_\lambda$.) The difficulty then lies in decoupling the two events in the above probability. If this decoupling is possible, we can use Theorem 5.18 to obtain

$$\begin{aligned}
-\frac{\partial}{\partial t}\theta_n(t,\lambda) &= \sum_{v\in\Lambda_n} \mathbb{P}_\lambda(v \text{ pivotal for } 0 \longleftrightarrow \partial\Lambda_n) \\
&\qquad\qquad \cdot \mathbb{P}_\lambda(\sigma_v(-t) \neq \lim_{s\to-\infty} \sigma_v(s)) \\
&\leq \exp(-ct) \sum_{v\in\Lambda_n} \mathbb{P}_\lambda(v \text{ pivotal for } 0 \longleftrightarrow \partial\Lambda_n) \\
&= \frac{1}{C}\exp(-ct)\frac{\partial}{\partial\lambda}\theta_n(t,\lambda),
\end{aligned}$$

for some constant C depending on λ, which is the expected surface area in space-time where an infection pointing towards v can be placed to ensure $\sigma_v(t) = 1$. It can be shown that this constant is bounded away from 0. Once the above inequality between the partial derivatives is established for all $t \geq t_0$, and all λ in a neighbourhood around λ_p, we then take $\lim_{t\to\infty} \lambda_p(t) < \lambda_1 < \lambda_2 < \lambda_p$, assuming $\lim_{t\to\infty} \lambda_p(t) < \lambda_p$, and aim to find a contradiction. We take $t_1 \geq t_0$ such that $c(\lambda_2 - \lambda_1)e^{ct_1} \geq 1/C$, and define

$$\begin{aligned}\lambda\colon [t_1, \infty) &\to (0, \infty),\\ \lambda(t) &= \lambda_1 + (\lambda_2 - \lambda_1)\big(1 - e^{-c(t-t_1)}\big),\end{aligned}$$

so that $\lambda(t_1) = \lambda_1$, and $\lambda(t) \to \lambda_2$ as $t \to \infty$. We can then compute the derivative of $\theta_n(t) := \theta_n\big(t, \lambda(t)\big)$, and obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\theta_n(t) &= c(\lambda_2 - \lambda_1)e^{-c(t-t_1)}\frac{\partial}{\partial\lambda}\theta_n(t,\lambda)\Big|_{\lambda=\lambda(t)} + \frac{\partial}{\partial t}\theta_n(t,\lambda)\Big|_{\lambda=\lambda(t)}\\ &\geq \Big(c(\lambda_2 - \lambda_1)e^{-c(t-t_1)} - \frac{1}{C}\exp(-ct)\Big)\frac{\partial}{\partial\lambda}\theta_n(t,\lambda)\Big|_{\lambda=\lambda(t)}\\ &\geq 0,\end{aligned}$$

by the choice of t_1. It then follows that

$$\theta_n(t_1, \lambda_1) = \theta_n(t_1) \leq \lim_{t\to\infty} \theta_n(t) = \bar{\nu}_{\lambda_2}(0 \longleftrightarrow \partial\Lambda_n).$$

Letting n tend to infinity, and recalling $\lim_{t\to\infty} \lambda_p(t) < \lambda_1 < \lambda_2 < \lambda_p$, gives the contradiction:

$$0 < \lim_{n\to\infty} \theta_n(t_1, \lambda_1) \leq \bar{\nu}_{\lambda_2}(0 \longleftrightarrow \infty) = 0.$$

Note that it is essential that t_1 is chosen independently of n.

The above ansatz to prove the sharp phase transition for $\bar{\nu}_\lambda$ uses the time t measure μ_t to approximate the invariant measure. A different way to approximate the process is to consider the contact process with spontaneous infections: the dynamics of the original process are altered such that a vertex in state 0 becomes infected with rate $\alpha > 0$, independent of the rest of the process. For this process the sharpness of the phase transition can be shown. That is, there exists $0 \leq \lambda_p(\alpha) \leq \lambda_p$ such that for $\lambda < \lambda_p(\alpha)$ clusters are exponentially small with respect to the upper invariant measure of this process, and for $\lambda > \lambda_p(\alpha)$ there exists an infinite cluster almost surely with respect to this measure. Similar to Proposition 5.13, it can then be shown that $\lambda_p(\alpha) > 0$ for small enough $\alpha > 0$. The sharpness of this phase transition can be shown by using the OSSS inequality in the same way as in the proof for the sharpness of the phase transition for μ_t. The spontaneous infections ensure that not too many variables are revealed by the decision tree, since once an spontaneous infection event is found by the exploration subalgorithm of Algorithm 5, we know the state of the vertex whose space-time we explored. However, this approach suffers from the same limitations as the approximation by μ_t: the relevant constants depend on α, and deteriorate as $\alpha \downarrow 0$. Another sufficient condition for the sharpness of the phase transition for $\bar{\nu}_\lambda$ is now $\lim_{\alpha\downarrow 0} \lambda_p(\alpha) = \lambda_p$. This convergence might be shown in a similar fashion as the suggested approach for the convergence of $\lambda_p(t) \to \lambda_p$. However, a similar problem arises here: namely we need to decouple the event that v is pivotal for $0 \longleftrightarrow \partial\Lambda_n$ with the size of the backward space-time cluster of v, given that $\sigma_v(t) = 0$. If that is possible, we see that the expected backward space-time cluster size of v given that $\sigma_v(0)$ is exponentially small, by Theorem 5.12. We can take $\alpha > 0$ small enough, similar to taking t_1 large enough in the case of $\lambda_p(t)$.

6 THE ORTHANT MODEL

In this chapter we look at the directed percolation model known as the orthant model. It takes place on the directed graph $\mathbb{Z}^d$, $d \geq 2$, with nearest neighbour edges. The model is informally described as follows. Let $e_1, \ldots, e_d$ be the standard unit basis vectors of $\mathbb{R}^d$. We set $\mathcal{E}_+ := \{e_1, \ldots, e_d\}$, and $\mathcal{E}_- := \{-e_1, \ldots, -e_d\}$, as well as $\mathcal{E} = \mathcal{E}_+ \cup \mathcal{E}_-$. A vertex $v \in \mathbb{Z}^d$ is connected to the vertices $v + e$ for all $e \in \mathcal{E}_+$ with a directed edge with probability p, independently of the other vertices. Otherwise, so with probability $1 - p$, the vertex v is connected to the vertices $v + e$ for all $e \in \mathcal{E}_-$. The model is shown in Figure 6.1 for $d = 2$.

The random directed graph obtained in this way has the property that every vertex is in an infinite cluster, since, for example, either the edge in the direction e_1 or in the direction $-e_2$ is always available. Therefore, there is no classical percolation phase transition in this model, where for small p there are only finite clusters, and for large p an infinite cluster exists. Instead, for small p the clusters will have a tendency to move in the directions of $\mathcal{E}_-$, and for large p a tendency in the directions of $\mathcal{E}_+$. We will make this notion precise, and see that this phase transition is sharp in an appropriate sense.

The orthant model was introduced by Holmes and Salisbury [32, 33]. It arose from the context of random walks in random environment. In this branch of probability theory most work has been done in the elliptic setting. This means that at every step, all directions are available to the random walk, although some

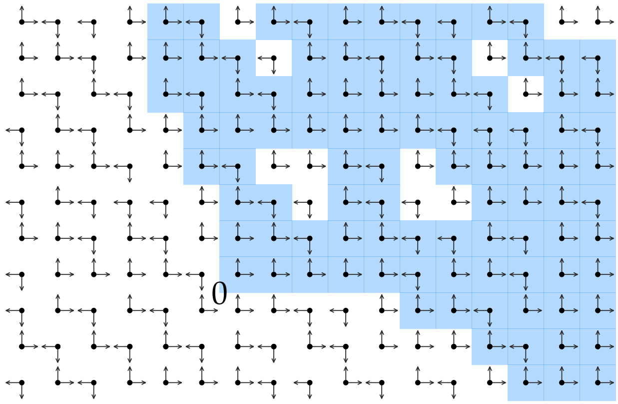


Figure 6.1: The orthant model on $\mathbb{Z}^2$. The cluster of the origin is shaded blue.

directions are more likely to be chosen than others, depending on the environment. In a non-elliptic random environment this is no longer the case: the random walk cannot take every direction in some locations in the graph. The random environment obtained by the orthant model is an example of a non-elliptic environment. Hence, the main motivation of studying this model is to understand how the random walk behaves in this environment. By proving the sharpness of the phase transition in this model, we will be able to conclude detailed results for the random walk.

This chapter has appeared in *Mathematical Physics, Analysis and Geometry* [5].

6.1 FRAMEWORK AND MAIN RESULT

In order to define the phase transition in this model, we introduce a cone oriented in the direction $\mathbf{1} := e_1 + \cdots + e_d$. For $0 \le \eta \le 1$, we define the convex cone

$$\mathcal{K}_\eta = \left\{ x \in \mathbb{R}^d \, : \, x \cdot \mathbf{1} \ge \eta \|x\|_1 \right\}.$$

Note that $\mathcal{K}_0$ is a half-space, and that $\mathcal{K}_1$ is the positive orthant.

For $v, w \in \mathbb{Z}^d$, we say that $v \longrightarrow w$, whenever there is a directed path from v to w. Note that this is not a symmetrical event, since we are working with a directed graph. Furthermore, for $A \subset \mathbb{Z}^d$, we say that $v \longrightarrow A$, whenever there exists $w \in A$ such that $v \longrightarrow w$. For $v \in \mathbb{Z}^d$, let $\mathcal{C}_v$ denote the forward cluster of v, i.e.,

$$\mathcal{C}_v := \{ w \in \mathbb{Z}^d \, : \, v \longrightarrow w \}.$$

We can define the critical point above which $\mathcal{C}_0$ is contained in a translated cone with parameter η:

$$\tilde{p}_c(\eta) := \inf \left\{ p \, : \, \mathbb{P}_p(\mathcal{C}_0 \subset -n\mathbf{1} + \mathcal{K}_\eta \text{ for some } n \in \mathbb{N}) = 1 \right\}.$$

Note that this critical point is increasing in η, and that $\tilde{p}_c(1) = 1$, so that $\tilde{p}_c(\eta)$ is of interest for $\eta \in [0,1)$. In fact, Holmes and Salisbury [34] have proven that $p_c^{\mathrm{OSP}} \le \tilde{p}_c(\eta) < 1$ for all $0 < \eta < 1$, where p_c^{OSP} is the critical parameter for oriented site percolation on the triangular lattice. Furthermore, from considerations later in this section, it follows that $\tilde{p}_c(0) > 0$. Therefore, $\tilde{p}_c(\eta)$ is nontrivial for $0 \le \eta < 1$. The phase transition associated with $\tilde{p}_c(\eta)$ is sharp in the following sense.

Theorem 6.1. *Consider the orthant model on $\mathbb{Z}^d$ with parameter p. Let $0 \le \eta < 1$ and suppose $p > \tilde{p}_c(\eta)$. Then, there exists a constant $c_p > 0$, such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}_p\big(0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c\big) \le \exp(-c_p n).$$

The above result for $d = 2$ was proven by Holmes and Salisbury [32] by making a connection with oriented site percolation on the triangular lattice. As a consequence of the above sharp threshold result, we can prove a shape theorem for $\mathcal{C}_0$ above $\tilde{p}_c := \lim_{\eta \downarrow 0} \tilde{p}_c(\eta)$. This critical point can also be written as

$$\tilde{p}_c = \inf\big\{p \,:\, \exists \eta > 0 \text{ s.t. } \mathbb{P}_p(\mathcal{C}_0 \subset -n\mathbf{1} + \mathcal{K}_\eta \text{ for some } n \in \mathbb{N}) = 1\big\}.$$

A shape theorem for the orthant model was first proven by Holmes and Salisbury [35] for large p. Using Theorem 6.1, we can extend this result to all $p > \tilde{p}_c$. In order to state the shape theorem, we introduce for $u \in \mathbb{Z}^d$

$$\beta_n(u) := \inf\{k \in \mathbb{Z} \,:\, k\mathbf{1} + nu \in \mathcal{C}_0\}.$$

Furthermore, let $\Lambda_r := \{v \in \mathbb{Z}^d \,:\, \|v\|_\infty \le r\}$ be the closed ball around 0 with radius r with respect to the L^∞-norm. Borrowing the notation from [35], the shape theorem for the orthant model can be stated as follows.

Corollary 6.2 (Shape theorem). *Let $p > \tilde{p}_c$. The following hold for the orthant model on $\mathbb{Z}^d$ with parameter p.*

1. *For $u \in \mathbb{Z}^d$, there is a deterministic $\gamma(u) \in \mathbb{R}$ such that $\frac{\beta_n(u)}{n} \to \gamma(u)$, as $n \to \infty$, $\mathbb{P}_p$-almost surely.*
2. *This limit satisfies $\gamma(u + w) \le \gamma(u) + \gamma(w)$, $\gamma(ru) = r\gamma(u)$, $\gamma(u + r\mathbf{1}) = \gamma(u) - r$, for $u, w \in \mathbb{Z}^d$, and $r \in \mathbb{N}$. Furthermore, γ is symmetric under permutation of coordinates, $\gamma(u) \ge 0$ if $u \cdot \mathbf{1} \le 0$, and $\gamma(u) \le 0$ if u lies in the positive orthant.*
3. *The limit γ extends to a Lipschitz map $\mathbb{R}^d \to \mathbb{R}$ with these same properties, but for $r \in [0, \infty)$ and $u, w \in \mathbb{R}^d$.*

4. *The set* $C := \{z \in \mathbb{R}^d : \gamma(z) \le 0\}$ *is a closed convex cone, which is symmetric under permutations of the coordinates, contains the positive orthant, and is contained in the half-space* $\mathcal{K}_0 = \{z : z \cdot \mathbf{1} \ge 0\}$.

5. *Let* $\mathcal{C}_0^* := \mathcal{C}_0 + e_1\mathbb{N}_0$*, i.e., "*$\mathcal{C}_0$ *with its holes filled in". It holds that* $\frac{1}{n}\mathcal{C}_0^* \to C$*, in the sense that for every* $\varepsilon > 0$ *and every* $r > 0$*, the following holds* $\mathbb{P}_p$*-a.s. for sufficiently large (random) n:*

$$\left(\Lambda_r \cap \frac{1}{n}\mathcal{C}_0^*\right) \subset \Lambda_\varepsilon + C, \quad \text{and} \quad (\Lambda_r \cap C) \subset \Lambda_\varepsilon + \frac{1}{n}\mathcal{C}_0^*.$$

To prove this theorem for all $p > \tilde{p}_c$, we modify the proof in [35] by using Theorem 6.1 in the places where they require p to be large. Another consequence of Theorem 6.1 is the ballisticity of the random walk on $\mathcal{C}_0$.

Corollary 6.3 (Ballisticity of the Random Walk)**.** *Consider the orthant model on* $\mathbb{Z}^d$ *with parameter* $p > \tilde{p}_c$*. Let* X_n *be a simple random walk on* $\mathcal{C}_0$ *and let* P *be the annealed law of this random walk (i.e., averaged over* $\mathcal{C}_0$*). Then there exists* $v > 0$ *such that* $\frac{1}{n}X_n \to v\mathbf{1}$ P*-a.s. as* $n \to \infty$*, and*

$$\left(\frac{X_{\lfloor nt \rfloor} - v\mathbf{1}nt}{\sqrt{n}}\right)_{t \ge 0} \Rightarrow (B_t)_{t\ge 0}, \quad \text{as } n \to \infty,$$

weakly under P*, in the space of càdlàg functions endowed with the Skorohod topology, where* $(B_t)_{t\ge0}$ *is a d-dimensional Brownian motion with nonsingular covariance matrix* Σ*.*

This is Theorem 1.4 combined with Corollary 1.9 of [34] by Holmes and Salisbury applied to the orthant model. Their theorem is stated for more general models, and requires two conditions, one of which they show to hold for the orthant model with any value of p. The other condition is the existence of $\eta > 0$ and $c > 0$ such that $\mathbb{P}_p\big(0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c\big) \le \exp(-cn^\beta)$, for some $\beta > 0$. By taking $\beta = 1$ and assuming $p > \tilde{p}_c$, it follows from Theorem 6.1 that this condition holds for the orthant model with parameter p. Corollary 6.3 is therefore an immediate consequence of combining Theorem 6.1 with Theorem 1.4 of [34].

Despite the above results, the theoretical picture of the orthant model is still incomplete. We will use the remainder of this section to formulate two open questions for the model. The shape theorem and the ballisticity of the random walk have now been shown to hold for $p > \tilde{p}_c := \lim_{\eta\downarrow 0} \tilde{p}_c(\eta)$. A natural extension would be to prove these results for $p > \tilde{p}_c(0)$. This would follow from the continuity of $\tilde{p}_c(\eta)$.

Open Problem 6.4. *Consider the orthant model on $\mathbb{Z}^d$. The function $\eta \mapsto \tilde{p}_c(\eta)$ is continuous on $[0,1]$.*

A critical value other than $\tilde{p}_c$ can be defined for the orthant model. In order to state this definition, we introduce for $v \in \mathbb{Z}^d$,

$$L_v := \inf\{k \in \mathbb{Z} : v + ke_1 \in \mathcal{C}_0\}.$$

The critical value p_c is defined as

$$p_c := \sup\{p : L_0 = -\infty \text{ a.s.}\}.$$

Holmes and Salisbury [36] have shown that this critical value is nontrivial, i.e., $0 < p_c < 1$. From the definitions of the critical values, it is clear that $p_c \le \tilde{p}_c(0) \le \tilde{p}_c$. However, it is as of yet unclear that above p_c there exists a cone with parameter $\eta > 0$ that contains the forward cluster of 0.

Open Problem 6.5. *Consider the orthant model on $\mathbb{Z}^d$. It holds that*

$$p_c = \tilde{p}_c.$$

In order to prove this, perhaps it is most natural to first show that $p_c = \tilde{p}_c(0)$, and subsequently show the continuity of $\tilde{p}_c(\eta)$.

The sharp threshold result of Theorem 6.1 will be proven in Section 6.2, while some preliminaries required for this proof are introduced in Section 6.2.1. The proof for Corollary 6.2 is given in Section 6.3.

6.2 PROOF OF THE SHARP PHASE TRANSITION

6.2.1 *Preliminaries*

We can couple the model for different values of p by considering a coupling similar to the standard coupling in Bernoulli percolation: we consider a family of i.i.d. random variables $(U_v)_{v \in \mathbb{Z}^d}$, and connect v to $v+e$ for all $e \in \mathcal{E}_+$ whenever $U_v < p$, and to $v+e$ for all $e \in \mathcal{E}_-$ if $U_v \geq p$. One difficulty in analysing the orthant model is the lack of monotonicity in p, i.e., a path from v to w might be lost if we increase p. To deal with this issue, we introduce the half-orthant model. In this model a vertex v is always connected to $v+e$ for all $e \in \mathcal{E}_+$, whereas v is connected to $v+e$, for all $e \in \mathcal{E}_-$, with probability $1-p$. This model is monotone in p, in the sense that $\mathbb{1}\{v \longrightarrow w\}$ is monotonically decreasing in p, under the coupling where v is connected to $v+e$, for all $e \in \mathcal{E}_-$, whenever $U_v > p$. Let $\mathcal{C}_v^*$ denote the forward cluster of v in the half-orthant model. The half-orthant model dominates the orthant model, in the sense that $\mathcal{C}_v \subseteq \mathcal{C}_v^*$, almost surely under a suitable coupling between the two models. For $v \in \mathbb{Z}^d$, we further define

$$L_v^* := \inf\{k \in \mathbb{Z} \,:\, v + ke_1 \in \mathcal{C}_0^*\}.$$

From the domination it follows that $L_v \geq L_v^*$. However, it turns out that equality holds: $L_v^* = L_v$ for all $v \in \mathbb{Z}^d$ [36, Thm. 1.9]. So, loosely speaking, if we only care about the leftmost boundary of $\mathcal{C}_0$, it does not matter if we consider the orthant model or the half-orthant model. This allows us to prove statements for the orthant model by making use of the monotonicity of the half orthant model. In light of this, we remark that the above definition of $\mathcal{C}_0^*$ coincides with the definition stated in Corollary 6.2. Furthermore, we note that $L_v < \infty$ for all $v \in \mathbb{Z}^d$, since $L_v^* < \infty$, but it might be the case that $L_v = -\infty$ for some $v \in \mathbb{Z}^d$. In fact, Holmes and Salisbury proved that if L_v is finite for some $v \in \mathbb{Z}^d$, then it is finite for all $v \in \mathbb{Z}^d$ [36, Lemma 2.2]. For $p < p_c$ it follows that $L_v^* = -\infty$ for all $v \in \mathbb{Z}^d$, and in this case, $\mathcal{C}_0^* = \mathbb{Z}^d$. On the other hand, if $p > p_c$, L_v is finite for all $v \in \mathbb{Z}^d$ using the monotonicity of the half-orthant model.

To prove Theorem 6.1, it therefore suffices to work with the half-orthant model. We start by giving a formal description of this model. For $p \in [0,1]$, we consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}_p)$, where

$$\Omega = \{0,1\}^{\mathbb{Z}^d},$$

the σ-algebra $\mathcal{F}$ is generated by the cylindrical events, and $\mathbb{P}_p$ is the product measure on Ω such that $\mathbb{P}_p(\omega_v = 1) = p$ for all $v \in \mathbb{Z}^d$. From $\omega \in \Omega$ we obtain the edge configuration $\xi \subseteq \{(v, v+e) : v \in \mathbb{Z}^d, e \in \mathcal{E}\}$ by adding the edge $(v, v+e)$ to the graph for all $e \in \mathcal{E}_+$, and for all $e \in \mathcal{E}_-$ whenever $\omega_v = 0$. Let $f\colon \Omega \to \{0,1\}$ be a Boolean function, and let T be a decision tree that determines the value of f. We apply the OSSS inequality for infinite product spaces, Theorem 2.18, to obtain,

$$\mathrm{Var}_p(f) \leq \sum_{v \in \mathbb{Z}^d} \mathrm{Inf}_v(f)\mathrm{Rev}_v(T).$$

For $v, w \in \mathbb{Z}^d$, we say $v \sim w$ when v is a neighbour of w, i.e., whenever $w = v + e$ for some $e \in \mathcal{E}$. Furthermore, we say that $v \rightsquigarrow w$, whenever $(v, w) \in \xi$. For $A \subset \mathbb{Z}^d$, we say that $v \xrightarrow{A} w$, whenever there is a path from v to w using only edges in ξ with starting points in A. Note that w does not have to be an element A for this event to hold. For $A = \mathbb{Z}^d$ we use the shorthand notation $\{v \longrightarrow w\} := \left\{v \xrightarrow{\mathbb{Z}^d} w\right\}$. Furthermore, the event $v \xrightarrow{A} v$ trivially holds for all $v \in \mathbb{Z}^d$, and all $A \subset \mathbb{Z}^d$. The Boolean function we are interested in is

$$f_n := \mathbb{1}\{0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c\},$$

for $\eta \geq 0$ and $n \in \mathbb{N}$.

6.2.2 *Exploration algorithm*

We now introduce decision trees that determine the value of f_n. A vital point in the proof is that we can uniformly bound the revealment of the vertices. If we only use one decision tree with a deterministic starting point, then the starting vertex will have revealment 1, so that we cannot find a nontrivial uniform bound on the revealment. Therefore, we will introduce the decision trees T_k, for $1 \leq k \leq n$, which all start at different vertices. In this way, we can average over k and have a meaningful uniform bound on the revealment. The basic idea of the decision tree T_k is that it explores the cluster of the boundary of $-k\mathbf{1} + \mathcal{K}_\eta$. If $0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c$, this path must go through the boundary of the cone $-k\mathbf{1} + \mathcal{K}_\eta$, so that T_k determines f_n. Furthermore, T_k terminates in a finite number of steps when $f_n = 1$.

We will now describe the exploration algorithm of T_k more precisely. We define the boundary and the outer boundary of the cone as

$$\partial(-k\mathbf{1}+\mathcal{K}_\eta) := \Big\{v \in (-k\mathbf{1}+\mathcal{K}_\eta)\cap\mathbb{Z}^d \; : \\ \exists w \in (-k\mathbf{1}+\mathcal{K}_\eta)^c \cap \mathbb{Z}^d,\, v \sim w\Big\},$$

$$\partial^+(-k\mathbf{1}+\mathcal{K}_\eta) := \Big\{v \in (-k\mathbf{1}+\mathcal{K}_\eta)^c\cap\mathbb{Z}^d \; : \\ \exists w \in (-k\mathbf{1}+\mathcal{K}_\eta) \cap \mathbb{Z}^d,\, v \sim w\Big\}.$$

The decision tree T_k consists of two phases. In the first phase, T_k explores the backward cluster of $\partial(-k\mathbf{1}+\mathcal{K}_\eta)$ inside the cone, that is, it explores the set $\{v \in -k\mathbf{1}+\mathcal{K}_\eta \;:\; v \longrightarrow \partial(-k\mathbf{1}+\mathcal{K}_\eta)\}$. When this is finished, the set of vertices

$$\Big\{v \in \partial^+(-k\mathbf{1}+\mathcal{K}_\eta) \;:\; 0 \xrightarrow{-k\mathbf{1}+\mathcal{K}_\eta} v\Big\}$$

has been determined. In the second phase, the algorithm explores the forward clusters of these vertices. If for one of these vertices we find that $v \longrightarrow (-n\mathbf{1}+\mathcal{K}_\eta)^c$, then we also have $0 \longrightarrow (-n\mathbf{1}+\mathcal{K}_\eta)^c$. A schematic visualisation of the algorithm is shown in Figure 6.2.

There is however one technical issue: since f_n depends on the state of infinitely many vertices, it is possible that the algorithm gets stuck exploring inside $-k\mathbf{1}+\mathcal{K}_\eta$, and never gets to explore the forward clusters outside $-k\mathbf{1}+\mathcal{K}_\eta$. In order to deal with this, the decision tree operates in rounds, denoted by $i \in \mathbb{N}$. Recall that Λ_r is the ball of radius r around 0 with respect to L^∞-norm. In round i we only explore inside Λ_i, so it is not possible to get stuck in any particular phase. Note that if $0 \longrightarrow (-n\mathbf{1}+\mathcal{K}_\eta)^c$, there exists $i \in \mathbb{N}$ such that $0 \xrightarrow{\Lambda_i} (-n\mathbf{1}+\mathcal{K}_\eta)^c$.

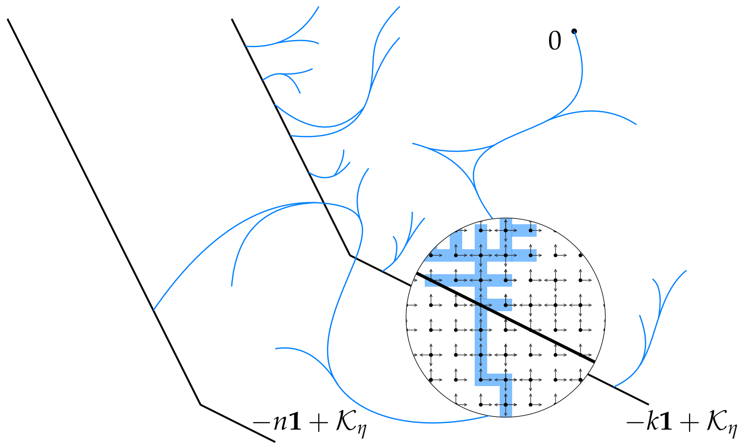


Figure 6.2: The algorithm T_k exploring the cluster of $\partial(-k\mathbf{1} + \mathcal{K}_\eta)$ to find a path from 0 to $(-n\mathbf{1} + \mathcal{K}_\eta)^c$. The blue vertices are revealed.

We denote by $\mathcal{R}$ the set of revealed vertices. Furthermore, we denote by $\mathcal{A}$ the set of active vertices for the first phase and by $\mathcal{B}$ the set of active vertices for the second phase. We start the algorithm by setting $\mathcal{A} := \mathcal{A}_0 := \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$, and $\mathcal{B} := \varnothing$. The pseudocode of T_k is given in Algorithm 6. We have to be careful when updating $\mathcal{A}$ in the first phase: by revealing v it is possible that we create a new path $x \xrightarrow{\mathcal{R}} \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$ for some $x \neq v$. Therefore, it is not sufficient to only consider $w \sim v$ for the update of $\mathcal{A}$. Instead, we add w to $\mathcal{A}$ if and only if $w \notin \mathcal{R} \cap \mathcal{B}$, and if there exists $x \in \mathcal{R}$ such that $x \sim w$, and $x \xrightarrow{\mathcal{R}} \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$.

$i := n$;
$\mathcal{A} := \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \cap \Lambda_n$;
$\mathcal{B} := \varnothing$;
$\mathcal{R} := \varnothing$;
while $0 \overset{\mathcal{R}}{\not\longrightarrow} (-n\mathbf{1} + \mathcal{K}_\eta)^c$ **do**
  while $\mathcal{A} \cap \Lambda_i \neq \varnothing$ **do**
    Take lexicographical minimal $v \in \mathcal{A} \cap \Lambda_i$;
    Reveal ω_v;
    $\mathcal{R} := \mathcal{R} \cup \{v\}$;
    $\mathcal{A} := \mathcal{A} \backslash \{v\}$;
    $\mathcal{A} := \mathcal{A} \cup \{w \in (-k\mathbf{1} + \mathcal{K}_\eta) \cap \mathbb{Z}^d \, : \, w \notin \mathcal{R} \cup \mathcal{B}, \exists x \in \mathcal{R}, x \sim w, \text{ s.t. } x \xrightarrow{\mathcal{R}} \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)\}$;
    $\mathcal{B} := \mathcal{B} \cup \{w \in \mathbb{Z}^d \, : \, w \notin \mathcal{R}, \exists x \in \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \text{ s.t. } 0 \xrightarrow{\mathcal{R}} x, x \xrightarrow{\mathcal{R}} w\}$;
    $\mathcal{A} := \mathcal{A} \backslash \{w \in \mathbb{Z}^d \, : \, w \notin \mathcal{R}, \exists x \in \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \text{ s.t. } 0 \xrightarrow{\mathcal{R}} x, x \xrightarrow{\mathcal{R}} w\}$;
    if $0 \xrightarrow{\mathcal{R}} (-n\mathbf{1} + \mathcal{K}_\eta)^c$ **then return** 1;
  end

    while $\mathcal{B} \cap \Lambda_i \neq \varnothing$ **do**
        Take lexicographical minimal $v \in \mathcal{B} \cap \Lambda_i$;
        Reveal ω_v;
        $\mathcal{R} := \mathcal{R} \cup \{v\}$;
        $\mathcal{B} := \mathcal{B} \backslash \{v\}$;
        $\mathcal{B} := \mathcal{B} \cup \{w \in \mathbb{Z}^d \,:\, w \notin \mathcal{R}, \exists x \in \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \text{ s.t. } 0 \xrightarrow{\mathcal{R}} x, x \xrightarrow{\mathcal{R}} w\}$;
        $\mathcal{A} := \mathcal{A} \backslash \{w \in \mathbb{Z}^d \,:\, w \notin \mathcal{R}, \exists x \in \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \text{ s.t. } 0 \xrightarrow{\mathcal{R}} x, x \xrightarrow{\mathcal{R}} w\}$;
        if $0 \xrightarrow{\mathcal{R}} (-n\mathbf{1} + \mathcal{K}_\eta)^c$ **then return** 1;
    end
    $i := i + 1$;
end
return 0;

Algorithm 6: The exploration algorithm T_k.

At the start of any iteration of the inner loops of the algorithm, the following hold for the active sets $\mathcal{A}$ and $\mathcal{B}$:

$$\begin{aligned}\mathcal{A} \subseteq \mathcal{A}_0 \backslash \mathcal{R} \cup \Big\{ v \in (-k\mathbf{1} + \mathcal{K}_\eta) \cap \mathbb{Z}^d \ :& \\ v \notin \mathcal{R}, \exists w \sim v, \text{ s.t. } w \xrightarrow{\mathcal{R}} \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \Big\},& \\ \mathcal{B} = \Big\{ v \in (-n\mathbf{1} + \mathcal{K}_\eta) \cap \mathbb{Z}^d \ :& \\ v \notin \mathcal{R}, \exists w \in \partial^+(-k\mathbf{1} + \mathcal{K}_\eta) \text{ s.t. } 0 \xrightarrow{\mathcal{R}} w, w \xrightarrow{\mathcal{R}} v \Big\}.&\end{aligned} \tag{6.1}$$

6.2.3 *Bound on the revealment*

Let $\theta_n(p) := \mathbb{P}_p(f_n = 1)$. Summing the OSSS inequality over k gives

$$n\theta_n(p)(1 - \theta_n(p)) \leq \sum_{v \in \mathbb{Z}^d} \mathrm{Inf}_v \sum_{k=1}^{n} \mathrm{Rev}_v(T_k). \tag{6.2}$$

We will now bound $\sum_{k=1}^n \mathrm{Rev}_v(T_k)$ uniformly in v. Let k_v be such that $v \in \partial^+(-k_v\mathbf{1} + \mathcal{K}_\eta)$. Note that $k_{v-\mathbf{1}} = k_v + 1$. Suppose first that $k > k_v + 1$. If v is revealed by T_k in the second phase, we have $0 \longrightarrow \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$ by (6.1). On the other hand, if v is revealed by T_k in the first phase, there exists $w \sim v$ such that $w \longrightarrow \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$. Since $k > k_v + 1$, we know that $v - \mathbf{1} \in -k\mathbf{1} + \mathcal{K}_\eta$, which implies that every neighbour of v is also contained in $-k\mathbf{1} + \mathcal{K}_\eta$. Applying the union bound gives

$$\begin{aligned}&\sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\} \mathrm{Rev}_v(T_k) \\ &\quad \leq \sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\} \Big(\theta_k(p) + \sum_{w \sim v} \mathbb{P}_p(w \longrightarrow \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)) \Big).\end{aligned}$$

Let

$$d_k^w := \sup\{l \in \mathbb{Z} \,:\, w - l\mathbf{1} \in -k\mathbf{1} + \mathcal{K}_\eta\}.$$

Note that $d_k^w \geq 0$ for all $k > k_v + 1$. Furthermore $w - d_k^w \mathbf{1} \in -k\mathbf{1} + \mathcal{K}_\eta$, and

$$w - d_k^w \mathbf{1} + \mathcal{K}_\eta \subseteq -k\mathbf{1} + \mathcal{K}_\eta,$$

since $-k\mathbf{1} + \mathcal{K}_\eta$ is a convex cone. Therefore, using translation invariance, it follows that

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\} \mathrm{Rev}_v(T_k) \\
&\quad \leq \sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\}\Big(\theta_k(p) + \sum_{w \sim v} \mathbb{P}_p\big(w \longrightarrow \partial^+\big(w - d_k^w \mathbf{1} + \mathcal{K}_\eta\big)\big)\Big) \\
&\quad = \sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\}\Big(\theta_k(p) + \sum_{w \sim v} \mathbb{P}_p\big(0 \longrightarrow \partial^+\big(-d_k^w \mathbf{1} + \mathcal{K}_\eta\big)\big)\Big).
\end{aligned} \tag{6.3}$$

We have $d_{k_v+2}^w \geq 0$, and using the fact that $d_{k+1}^w = d_k^w + 1$, we know $d_{k_v+l}^w \geq l - 2$, for all $l \geq 2$. We can thus bound

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{1}\{k > k_v + 1\} \mathrm{Rev}_v(T_k) \\
&\qquad \leq \sum_{k=1}^{n} \theta_k(p) + 2d \sum_{k=0}^{n-1} \mathbb{P}_p\big(0 \longrightarrow \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)\big) \\
&\qquad = \sum_{k=1}^{n} \theta_k(p) + 2d \sum_{k=0}^{n-1} \theta_k(p).
\end{aligned} \tag{6.4}$$

Now suppose $k < k_v$, so $v \notin -k\mathbf{1} + \mathcal{K}_\eta \cup \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$. If v is revealed, it holds that $0 \xrightarrow{\mathcal{R}} v$. In particular, we have $0 \longrightarrow \partial^+(-k\mathbf{1} + \mathcal{K}_\eta)$. We find

$$\sum_{k=1}^{n} \mathbb{1}\{k < k_v\} \mathrm{Rev}_v(T_k) \leq \sum_{k=1}^{n} \theta_k(p). \tag{6.5}$$

Combining (6.4) and (6.5) gives

$$\sum_{k=1}^{n} \mathrm{Rev}_v(T_k) \leq 2 + 2\sum_{k=1}^{n} \theta_k(p) + 2d \sum_{k=0}^{n} \theta_k(p) = (2d+2) \sum_{k=0}^{n} \theta_k(p).$$

Writing $S_n := \sum_{k=0}^{n} \theta_k(p)$, gives

$$\sum_{v \in \mathbb{Z}^d} \mathrm{Inf}_v \geq \frac{1}{2d+2} \frac{n}{S_n} \theta_n(p)(1 - \theta_n(p)). \tag{6.6}$$

6.2.4 *Analysis of the differential inequality*

We are now able to complete the proof of Theorem 6.1. We can obtain a differential inequality by using Russo's formula. However, since f_n depends on infinitely many vertices, $\theta_n(p)$ is not necessarily differentiable in p. Instead we have to work with the upper-right Dini derivative:

$$D^+\theta_n(p) := \limsup_{h \downarrow 0} \frac{\theta_n(p+h) - \theta_n(p)}{h}.$$

Using the fact that $0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c$ is a decreasing event, i.e., f_n is a decreasing function of ω, Russo's formula gives

$$\begin{aligned} -D^+\theta_n(p) &\geq \sum_{v \in \mathbb{Z}^d} \mathbb{P}_p(v \text{ is pivotal for } \{f_n = 1\}) \\ &= \frac{1}{2p(1-p)} \sum_{v \in \mathbb{Z}^d} \mathrm{Inf}_v(f_n). \end{aligned}$$

This version of Russo's formula can be found in the book on Percolation by Grimmett [25]. This is the point in the proof where we use the monotonicity of the half-orthant model, as well as the coupling given at the start of Section 6.2.1. Combining the above inequality with (6.6) gives

$$-D^{+}\theta_n(p) \geq \frac{1}{2d}\frac{n}{S_n}\theta_n(p)(1-\theta_n(p)), \tag{6.7}$$

where we use $(2d+2)2p(1-p) \leq 2d$ for simplicity. The rest of the proof consists of analysing the above differential inequality. This analysis follows the line of Duminil-Copin, Raoufi and Tassion [15], but since it differs on several points, we choose to include it. We have to work with Dini derivatives instead of regular derivatives, and, more importantly, in our case we cannot give a simple lower bound on $1-\theta_n(p)$.

To analyse the differential inequality, we introduce the auxiliary critical point

$$\hat{p}_c(\eta) := \sup\left\{p \,:\, \limsup_{n\to\infty}\frac{\log S_n(p)}{\log n} = 1\right\}.$$

Note that by the monotonicity of the model, $\limsup_{n\to\infty}\frac{\log S_n(p)}{\log n} = 1$ for all $p < \hat{p}_c(\eta)$, and $\limsup_{n\to\infty}\frac{\log S_n(p)}{\log n} < 1$ for all $p > \hat{p}_c(\eta)$. We will first show that $\hat{p}_c(\eta) \leq \tilde{p}_c(\eta)$, for $\eta \geq 0$. To prove this, we assume the contrary, and let $p \in (\tilde{p}_c(\eta), \hat{p}_c(\eta))$. Since $p > \tilde{p}_c(\eta)$, we can fix $l \in \mathbb{N}$, such that for all $n > l$ it holds that $1-\theta_n(p) \geq 1/2$.

We define $T_n(p) := \frac{1}{\log n}\sum_{k=l}^{n}\frac{\theta_k(p)}{k}$. Taking the upper-right Dini derivative and using (6.7) gives

$$\begin{aligned}
-D^+T_n &\geq \frac{1}{2d}\frac{1}{\log n}\sum_{k=l}^{n}\frac{\theta_k(p)}{S_k}(1-\theta_k(p)) \\
&\geq \frac{1}{4d}\frac{1}{\log n}\sum_{k=l}^{n}\frac{\theta_k(p)}{S_k} \geq \frac{1}{4d}\frac{\log S_{n+1}-\log S_l}{\log n},
\end{aligned} \tag{6.8}$$

where in the last inequality we used

$$\frac{\theta_k(p)}{S_k} \geq \int_{S_k}^{S_{k+1}} \frac{1}{x}\,\mathrm{d}x = \log S_{k+1} - \log S_k.$$

Now let $p_1 \in (p, \hat{p}_c(\eta))$. We will integrate the differential inequality between p and p_1 and use the following result regarding Dini derivatives: the Dini derivative of a decreasing function $f\colon [a,b] \to \mathbb{R}$ satisfies

$$f(b) - f(a) \leq \int_a^b D^+f(x)\,\mathrm{d}x. \tag{6.9}$$

Applying this to $T_n(p)$ and using (6.8) gives

$$\begin{aligned}
T_n(p_1) - T_n(p) &\leq \int_p^{p_1} D^+T_n(s)\,\mathrm{d}s \\
&\leq -(p_1-p)\frac{1}{4d}\frac{\log S_{n+1}(p_1) - \log S_l(p)}{\log n}.
\end{aligned}$$

Furthermore, $T_n(p)$ converges to $\theta(p) := \lim_{n\to\infty}\theta_n(p)$ for $n \to \infty$, since for all $l < m < n$:

$$\theta_n(p)\frac{\sum_{k=l}^{n}\frac{1}{k}}{\log n} \leq T_n(p) \leq \theta_m(p)\frac{\sum_{k=m}^{n}\frac{1}{k}}{\log n} + \frac{\sum_{k=l}^{m-1}\frac{1}{k}}{\log n},$$

from which the limit follows by first taking $n \to \infty$, and then $m \to \infty$. We find

$$\theta(p_1) - \theta(p) \le -(p_1 - p)\frac{1}{4d} \limsup_{n\to\infty} \frac{\log S_{n+1}(p_1) - \log S_l(p)}{\log n}.$$

Since $p < \hat{p}_c(\eta)$, we have that $\limsup_{n\to\infty} \frac{\log S_n(p)}{\log n} = 1$ and the same holds for p_1, so that also

$$\limsup_{n\to\infty} \frac{\log S_{n+1}(p) - \log S_l(p_1)}{\log n} = 1.$$

We conclude

$$\theta(p) \ge \theta(p) - \theta(p_1) \ge \frac{p_1 - p}{4d} > 0, \tag{6.10}$$

which contradicts $p > \tilde{p}_c(\eta)$, so that we have established that $\hat{p}_c(\eta) \le \tilde{p}_c(\eta)$.

Now suppose $p > \hat{p}_c(\eta)$. Then there exists $N_1 \in \mathbb{N}$ and $\beta < 1$ such that $S_n(p) \le n^\beta$ for all $n \ge N_1$, and there exists $N_2 \in \mathbb{N}$ such that $\theta_n(p) \le \frac{1}{2}$ for all $n \ge N_2$. Combining this with (6.7) and using the chain rule for Dini derivatives gives

$$D^+ \log \theta_n(p) \le -\frac{1}{2d} n^{1-\beta}(1 - \theta_n(p)) \le -\frac{1}{4d} n^{1-\beta},$$

for all $n > N := N_1 \vee N_2$. Let $p_1 := (\tilde{p}_c(\eta) + p)/2$. Integrating the above inequality between p_1 and p and using (6.9) gives

$$\log \theta_n(p) \le \log \theta_n(p) - \log \theta_n(p_1) \le -\frac{1}{4d}(p - p_1) n^{1-\beta}.$$

It follows that

$$\theta_n(p) \le \exp\left(-\frac{1}{8d}(p - \tilde{p}_c(\eta)) n^{1-\beta}\right).$$

It remains to improve the above stretched exponential decay to proper exponential decay. From the stretched exponential decay it follows that $S(p) := \lim_{n\to\infty} S_n(p) < \infty$. Combining this fact with (6.7), and using that $\theta_n(p) \leq \frac{1}{2}$ for $n > N$, since $p > \hat{p}_c(\eta)$, gives

$$D^+ \log \theta_n(p) \leq -\frac{1}{4dS(p)} n.$$

From here the proof is similar as for the stretched exponential decay, and we conclude

$$\theta_n(p) \leq \exp\left(-\frac{1}{8dS(p)}(p - \tilde{p}_c(\eta))n\right).$$

It follows that $\tilde{p}_c(\eta) = \hat{p}_c(\eta)$, and that Theorem 6.1 holds with

$$\begin{aligned} c_p &:= \frac{1}{8dS(p)}(p - \tilde{p}_c(\eta)) \\ &\quad \wedge \sup\left\{C > 0 \,:\, \theta_n(p) \leq \exp(-Cn) \text{ for all } n \leq N\right\} \\ &> 0. \end{aligned}$$

□

Remark 6.6. *A mean-field lower bound such as Proposition 3.5 can often be obtained from the analysis of a differential inequality such as the above one, which would be* $\theta(p) \geq c(\tilde{p}_c(\eta) - p)$ *in our case, for* $p < \tilde{p}_c(\eta)$*, and some constant* $c > 0$*, independent of* p*. This does not directly follow from the above analysis, since we have assumed* $p \in (\tilde{p}_c(\eta), \hat{p}_c(\eta))$ *in the first part of the analysis. If instead we take* $p < \hat{p}_c(\eta)$ *such that* $\theta(p) < 1/2$*, then we can still bound* $1 - \theta_n(p) \geq 1/2$*, for* n *large enough, and obtain the mean-field lower bound at (6.10). However, this bound is of little interest for the orthant model, since if* $p < p_c$*,* $C^*(0) = \mathbb{Z}^d$ *almost surely, so that* $\theta(p) = 1$*. On the* *other hand, if* $p \in (\tilde{p}_c, \tilde{p}_c(\eta))$ *for some* $\eta > 0$*, Corollary 6.2 implies that* $\theta(p) = 1$ *as well. This leaves the interval* $(p_c, \tilde{p}_c)$ *to be considered, but we conjecture this interval to be empty.*

6.3 PROOF OF THE SHAPE THEOREM

To prove Corollary 6.2, we modify the proof of Holmes and Salisbury [35] in the places where they require p to be large. Their proof is structured in seven lemmas, two of which require a large p. The first of these is Lemma 1 of [35]. This lemma asserts the existence of $\theta > 1$, such that for every $\eta \in [0, 1)$, there exists $p_0 = p_0(\eta, d) < 1$, such that for $p > p_0$, there exists $c_1 > 0$ such that $\mathbb{P}_p(0 \longrightarrow (-n\mathbf{1} + \mathcal{K}_\eta)^c) \leq c_1\theta^{-nd}$, for all $n \in \mathbb{N}$. In the remainder of their proof, this lemma is only used for the case $\eta = 0$. Therefore, we can replace this lemma by Theorem 6.1, and require $p > \tilde{p}_c$, instead of $p > p_0$.

The second lemma in the proof of Holmes and Salisbury which require large p is Lemma 5 of [35]. We will prove this lemma for $p > \tilde{p}_c$, instead of for large p, using Theorem 6.1. To state this lemma, we let $u \in \mathbb{Z}^d \backslash \mathbb{Z}\mathbf{1}$, and fix $v \in \mathbb{R}^d$ such that $u \cdot v > 0$ and $v \cdot \mathbf{1} = 0$. We define the slab

$$\Lambda_{u,v}(m, n) := \{z \in \mathbb{Z}^d \; : \; mu \cdot v \leq z \cdot v < nu \cdot v\}.$$

We are interested in the following three events. Let $A'_n(M)$ be the event there exists a path starting in o and ending in a point $k\mathbf{1} + nu$ with $k < n\gamma(u)$ that hits $\Lambda_{u,v}(-\infty, -M)$. Let $A''_n(M)$ be the event there exists a path starting in o and ending in a point $k\mathbf{1} + nu$ with $k < n\gamma(u)$ that hits $\Lambda_{u,v}(M + n, \infty)$. Lastly, let $\hat{A}_n$ be the event that there is a path starting in o and ending in some point $k\mathbf{1}$, with $k < 0$, and reaches $\Lambda_{u,v}(n, \infty)$. We will prove the following lemma regarding these events:

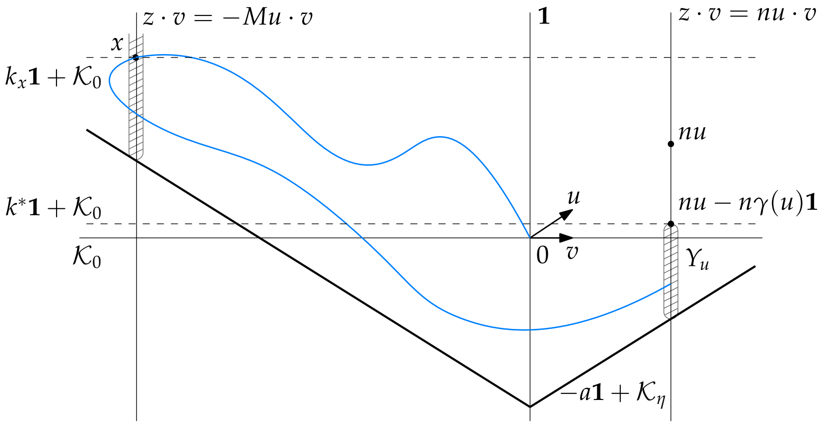


Figure 6.3: When the event $A_n'(M)$ occurs, there exists a path from 0 to Y_u going through the shaded region

Lemma 6.7. *Let $p > \tilde{p}_c$. There exists $c > 0$, such that*

$$\mathbb{P}_p(A_n'(\lfloor cn \rfloor) \text{ i.o.}) = \mathbb{P}_p(A_n''(\lfloor cn \rfloor) \text{ i.o.}) = \mathbb{P}_p(\hat{A}_n(\lfloor cn \rfloor) \text{ i.o.}) = 0.$$

We will prove the above lemma for the event $A_n'(\lfloor cn \rfloor)$, the other two events can be proven similarly. The event $A_n'(\lfloor cn \rfloor)$ is shown in Figure 6.3. Let $p > \tilde{p}_c$. By the definition of this critical point there exists $\eta > 0$ such that $p > \tilde{p}_c(\eta)$. We fix such an η. Let $c > 0$, and let $M := M(n) := \lfloor cn \rfloor$. We will choose the precise value of c later on. Let $a > 0$ and suppose $\mathcal{C}_0^* \subseteq -a\mathbf{1} + \mathcal{K}_\eta$. If $A_n'(M)$ occurs, there exists $x \in \mathbb{Z}^d$ satisfying

$$\begin{cases} (x + a\mathbf{1}) \cdot \mathbf{1} \geq \eta \|x + a\mathbf{1}\|_1, \\ x \cdot v = -Mu \cdot v, \end{cases} \implies \begin{cases} x \cdot \mathbf{1} \geq -2da + \eta \|x\|_1, \\ x \cdot v = -Mu \cdot v, \end{cases} \tag{6.11}$$

such that $0 \longrightarrow x$, and $x \longrightarrow y$, with $y = k\mathbf{1} + nu$ for some $k < n\gamma(u)$. Since the L^1-norm is equivalent to the L^2-norm, and since the L^2-norm is invariant under an orthonormal basis change, it follows from the above equation that $\|x\|_1 \geq c_0 M$, for some constant $c_0 = c_0(u,v) > 0$. Combining this with the above inequality gives

$$x \cdot \mathbf{1} \geq -2da + c_0 \eta M.$$

Define $k_x := x \cdot \mathbf{1}/d$, so that $x \in k_x \mathbf{1} + \mathcal{K}_0$. Then the above inequality implies

$$k_x \geq -2a + \frac{c_0 \eta M}{d}.$$

We define the set

$$Y_u := \{y \in \mathbb{Z}^d \ : \ y = k\mathbf{1} + nu, \ \text{with } k < n\gamma(u)\}.$$

We use Theorem 6.1 and the union bound to obtain

$$\begin{aligned}
\mathbb{P}_p(A_n'(M)) &\leq \exp(-c_p a) + \sum_{\substack{k_x \geq \\ -2a + c_0\eta M/d}}^{\infty} \mathbb{P}_p(\exists x \in \Lambda_{u,v}(-\infty, -M) \ : \\
&\qquad\qquad\qquad\qquad x \cdot \mathbf{1} = dk_x,\, 0 \longrightarrow x,\, x \longrightarrow Y_u) \\
&\leq \exp(-c_p a) + \sum_{\substack{k_x \geq \\ -2a + c_0\eta M/d}}^{\infty} \sum_{\substack{x \in \Lambda_{u,v}(-\infty,-M), \\ x \cdot \mathbf{1} = dk_x}} \mathbb{P}_p(x \longrightarrow Y_u).
\end{aligned}$$

We define $k^* := k^*(n) := n(\gamma(u) + u \cdot \mathbf{1}/d)$. With this choice, it follows that $y \in (k^*\mathbf{1} + \mathcal{K}_0)^c$ for all $y \in Y_u$, and all $n \in \mathbb{N}$. We can now use translation invariance to bound

$$\begin{aligned}
\mathbb{P}_p(x \longrightarrow Y_u) &\leq \mathbb{P}_p\big(x \longrightarrow (k^*\mathbf{1} + \mathcal{K}_0)^c\big) \\
&= \mathbb{P}_p\big(x - k_x\mathbf{1} \longrightarrow (-(k_x - k^*)\mathbf{1} + \mathcal{K}_0)^c\big).
\end{aligned}$$

We now fix

$$c := \Big(\frac{d\gamma(u) + u \cdot \mathbf{1}}{c_0 \eta} + 1\Big) \vee 1.$$

Let $k' \geq 0$ such that $k_x = -2a + \frac{c_0 \eta M}{d} + k'$. It holds, that

$$\begin{aligned} k_x - k^* &\geq -2a + \frac{c_0\eta}{d}\Big\lfloor n\Big(\frac{d\gamma(u) + u \cdot \mathbf{1}}{c_0\eta} + 1\Big)\Big\rfloor + k' - n\Big(\gamma(u) + \frac{u \cdot \mathbf{1}}{d}\Big) \\ &\geq -2a + \frac{c_0\eta}{d}(n-1) + k' =: f(n,k'). \end{aligned}$$

It follows, that

$$\begin{aligned} \mathbb{P}_p(A'_n(M)) &\leq \exp(-c_p a) \\ &\quad + \sum_{k'=0}^{\infty} \sum_{\substack{x \in \Lambda_{u,v}(-\infty,-M), \\ x \cdot \mathbf{1} = -2da + c_0\eta M + dk'}} \mathbb{P}_p\Big(x - k_x \cdot \mathbf{1} \longrightarrow (-f(n,k')\mathbf{1} + \mathcal{K}_0)^c\Big). \end{aligned}$$

Combining $x \cdot \mathbf{1} = -2da + c_0\eta M + dk'$ with (6.11), shows that

$$\|x\|_1 \leq c_0 M + \frac{dk'}{\eta}.$$

Using another union bound, translation invariance, and Theorem 6.1, we find

$$\begin{aligned} &\mathbb{P}_p(A'_n(M)) \\ &\quad \leq \exp(-c_p a) + \sum_{k'=0}^{\infty} \Big|\Big\{x \in \mathbb{Z}^d \;:\; \|x\|_1 \leq c_0 M + \frac{dk'}{\eta}\Big\}\Big| \\ &\qquad\qquad \cdot \mathbb{P}_p(0 \longrightarrow (-f(n,k')\mathbf{1} + \mathcal{K}_0)^c) \\ &\quad \leq \exp(-c_p a) + \sum_{k'=0}^{\infty} \Big(2c_0 c n + 2\frac{dk'}{\eta}\Big)^d \exp(-c_p f(n,k')). \qquad (6.12) \end{aligned}$$

We now take

$$a := a(n) := \frac{c_0\eta}{4}n,$$

so that

$$f(n,k') = -2a + \frac{c_0\eta}{d}(n-1) + k' = c_0\eta\Big(\frac{n}{2} - 1\Big) + k'.$$

A careful examination of (6.12) shows that the sum over k' converges for all $n \in \mathbb{N}$, and that the result is summable with respect to n, so that by the Borel-Cantelli lemma $\mathbb{P}_p(A'_n(\lfloor cn \rfloor) \text{ i.o.}) = 0$. The same result can be proven similarly for the events $A''_n(\lfloor cn \rfloor)$ and $\hat{A}_n(\lfloor cn \rfloor)$, and we omit the proof. $\square$

BIBLIOGRAPHY

[1] Michael Aizenman and David J. Barsky. "Sharpness of the phase transition in percolation models." In: *Communications in Mathematical Physics* 108.3 (1987), pp. 489–526.

[2] Michael Aizenman, David J. Barsky, and Roberto Fernández. "The phase transition in a general class of Ising-type models is sharp." In: *Journal of Statistical Physics* 47.3 (1987), pp. 343–374.

[3] Nicholas R. Beaton, Geoffrey R. Grimmett, and Mark Holmes. "Alignment percolation." In: *arXiv:1908.07203* (2019).

[4] Thomas Beekenkamp. "Sharpness of the phase transition for the corrupted compass model on transitive graphs." In: *Indagationes Mathematicae. New Series* 32.3 (2021), pp. 736–744.

[5] Thomas Beekenkamp. "Sharpness of the phase transition for the orthant model." In: *Mathematical Physics, Analysis and Geometry* 24 (2021).

[6] Jacob van den Berg. "A note on disjoint-occurrence inequalities for marked Poisson point processes." In: *Journal of Applied Probability* 33.2 (1996), pp. 420–426.

[7] Jacob van den Berg. "Sharpness of the percolation transition in the two-dimensional contact process." In: *The Annals of Applied Probability* 21.1 (2011), pp. 374–395.

[8] Jacob van den Berg and Harry Kesten. "Inequalities with applications to percolation and reliability." In: *Journal of Applied Probability* 22.3 (1985), pp. 556–569.

[9] Carol Bezuidenhout and Geoffrey Grimmett. "The critical contact process dies out." In: *The Annals of Probability* (1990), pp. 1462–1482.

[10] Carol Bezuidenhout and Geoffrey Grimmett. "Exponential Decay for Subcritical Contact and Percolation Processes." In: *The Annals of Probability* 19 (1991), pp. 984–1009.

[11] Simon R. Broadbent and John M. Hammersley. "Percolation processes. I. Crystals and mazes." In: *Mathematical proceedings of the Cambridge philosophical society* 53 (1957), pp. 629–641.

[12] David Dereudre and Pierre Houdebert. "Sharp phase transition for the continuum Widom-Rowlinson model." In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 57.1 (2021), pp. 387–407.

[13] Hugo Duminil-Copin, Subhajit Goswami, Pierre-François Rodriguez, and Franco Severo. "Equality of critical parameters for percolation of Gaussian free field level-sets." In: *arXiv preprint arXiv:2002.07735* (2020).

[14] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. "Exponential decay of connection probabilities for subcritical Voronoi percolation in $\mathbb{R}^d$." In: *Probability Theory and Related Fields* (2017), pp. 1–12.

[15] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. "Sharp phase transition for the random-cluster and Potts models via decision trees." In: *Annals of Mathematics* 189.1 (2019), pp. 75–99.

[16] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. "Subcritical phase of d-dimensional Poisson–Boolean percolation and its vacant set." In: *Annales Henri Lebesgue* 3 (2020), pp. 677–700.

[17] Hugo Duminil-Copin and Vincent Tassion. "A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model." In: *Communications in Mathematical Physics* 343.2 (2016), pp. 725–745.

[18] Hugo Duminil-Copin and Vincent Tassion. "A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$." In: *L'Enseignement Mathématique* 62.1 (2016), pp. 199–206.

[19] Richard Durrett. *Probability. Theory and examples*. Fourth edition. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 2010.

[20] Rick Durrett. "The Contact Process, 1974-1989." In: *Lectures in Applied Mathematics*. Vol. 27. AMS, 1991, pp. 1–18.

[21] Robert Fitzner. *Simulator for random spatial structures*. URL: `http://fitzner.nl/simulator/index.html`.

[22] Robert Fitzner and Remco van der Hofstad. "Mean-field behavior for nearest-neighbor percolation in $d > 10$." In: *Electronic Journal of Probability* 22 (2017), pp. 1–65.

[23] Cornelius M. Fortuin. "On The Random-cluster Model III: The Simple Random-cluster Model." In: *Physica* 59 (1972), pp. 545–570.

[24] Cornelius M. Fortuin, Piet W. Kasteleyn, and Jean Ginibre. "Correlation inequalities on some partially ordered sets." In: *Communications in Mathematical Physics* 22.2 (1971), pp. 89–103.

[25] Geoffrey Grimmett. *Percolation*. Second Edition. Grundlehren der Mathematischen Wissenschaften 321. Springer, 1999.

[26] Geoffrey Grimmett. *The Random-Cluster Model*. Grundlehren der Mathematischen Wissenschaften 333. Springer, 2006.

[27] Geoffry. R. Grimmett and John M. Marstrand. "The supercritical phase of percolation is well behaved." In: *Proceedings of the Royal Society of London, Series A* 430.1879 (1990), pp. 439–457.

[28] Theodore E. Harris. "A lower bound for the critical probability in a certain percolation process." In: *Mathematical Proceedings of the Cambridge Philosophical Society* 56 (1960), pp. 13–20.

[29] Markus Heydenreich and Remco van der Hofstad. *Progress in high-dimensional percolation and random graphs*. CRM Short Courses. Springer, 2017.

[30] Christian Hirsch, Mark Holmes, and Victor Kleptsyn. "Absence of WARM percolation in the very strong reinforcement regime." In: *The Annals of Applied Probability* 31.1 (2021), pp. 199–217.

[31] Richard Holley and Thomas M. Liggett. "The Survival of Contact Processes." In: *The Annals of Probability* 6.2 (1978).

[32] Mark Holmes and Thomas S. Salisbury. "Degenerate random environments." In: *Random Structures & Algorithms* 45.1 (2014), pp. 111–137.

[33] Mark Holmes and Thomas S. Salisbury. "Random walks in degenerate random environments." In: *Canadian Journal of Mathematics* 66.5 (2014), pp. 1050–1077.

[34] Mark Holmes and Thomas S. Salisbury. "Conditions for ballisticity and invariance principle for random walk in non-elliptic random environment." In: *Electronic Journal of Probability* 22 (2017), pp. 1–18.

[35] Mark Holmes and Thomas S. Salisbury. "A shape theorem for the Orthant model." In: *The Annals of Probability* 49.3 (2021), pp. 1237–1256.

[36] Mark Holmes and Thomas S. Salisbury. "Phase transitions for degenerate random environments." In: *ALEA* 18 (2021), pp. 707–725.

[37] Tom Hutchcroft. "New critical exponent inequalities for percolation and the random cluster model." In: *Probability and Mathematical Physics* 1.1 (2020), pp. 147–165.

[38] Harry Kesten. "The critical probability of bond percolation on the square lattice equals $1/2$." In: *Communications in Mathematical Physics* 74.1 (1980), pp. 41–59.

[39] Achim Klenke. *Wahrscheinlichkeitstheorie*. Zweite, korrigierte Auflage. Springer, 2008.

[40] Günter Last, Giovanni Peccati, and D. Yogeshwaran. "Phase transitions and noise sensitivity on the Poisson space via stopping sets and decision trees." In: *arXiv preprint arXiv:2101.07180* (2021).

[41] Thomas M. Liggett. "Survival and coexistence in interacting particle systems." In: *Probability and phase transition*. Kluwer Academic Publishers, 1994, pp. 209–226.

[42] Thomas M. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Grundlehren der Mathematischen Wissenschaften 324. Springer, 1999.

[43] Thomas M. Liggett. *Interacting particle systems*. Reprint of the 1985 edition. Classics in mathematics. Springer, 2005.

[44] Thomas M. Liggett. *Continuous time Markov processes. An introduction*. Graduate studies in mathematics 113. American Mathematical Society, 2010.

[45] Mikhail V. Menshikov. "Coincidence of critical points in percolation problems." In: *Doklady Akademii Nauk SSSR* 288.6 (1986), pp. 1308–1311.

[46] Stephen Muirhead and Hugo Vanneuville. "The sharp phase transition for level set percolation of smooth planar Gaussian fields." In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56.2 (2020), pp. 1358–1390.

[47] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[48] Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco A Servedio. "Every decision tree has an influential variable." In: *46th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2005, pp. 31–39.

[49] Mathew Penrose. *Random geometric graphs*. Oxford Studies in Probability 5. Oxford university press, 2003.

[50] David Reimer. "Proof Of The Van Den Berg-Kesten Conjecture." In: *Combinatorics, Probability and Computing* 9.1 (2000), pp. 27–32.

[51] Jan M. Swart. "A simple proof of exponential decay of subcritical contact processes." In: *Probability Theory and Related Fields* 170.1 (2018), pp. 1–9.

[52] Michel Talagrand. "On Russo's approximate zero-one law." In: *The Annals of Probability* 22.3 (1994), pp. 1576–1587.

EIDESSTATTLICHE VERSICHERUNG

(Siehe Promotionsordnung vom 12.07.11, §8, Abs. 2, Pkt. 5.)

Hiermite erkläre ich am Eidesstatt, dass die Disertation von mir selbständig, ohne unerlaubte Beihilfe angefertigt ist.

München, 11. August 2021

Thomas Beekenkamp