Efficient Sampling and Counting Algorithms for the Potts Model on $\mathbb{Z}^d$ at All Temperatures*

Christian Borgs
University of California Berkeley
Berkeley, California, USA
borgs@berkeley.edu

Jennifer Chayes
University of California Berkeley
Berkeley, California, USA
jchayes@berkeley.edu

Tyler Helmuth
University of Bristol
Bristol, UK
jhelmt@gmail.com

Will Perkins
University of Illinois at Chicago
Chicago, Illinois, USA
math@willperkins.org

Prasad Tetali
Georgia Institute of Technology
Atlanta, Georgia, USA
tetali@math.gatech.edu

ABSTRACT

For $d \geq 2$ and all $q \geq q_0(d)$ we give an efficient algorithm to approximately sample from the q-state ferromagnetic Potts and random cluster models on the torus $(\mathbb{Z}/n\mathbb{Z})^d$ for any inverse temperature $\beta \geq 0$. This stands in contrast to Markov chain mixing time results: the Glauber dynamics mix slowly at and below the critical temperature, and the Swendsen–Wang dynamics mix slowly at the critical temperature. We also provide an efficient algorithm (an FPRAS) for approximating the partition functions of these models.

Our algorithms are based on representing the random cluster model as a contour model using Pirogov–Sinai theory, and then computing an accurate approximation of the logarithm of the partition function by inductively truncating the resulting cluster expansion. One important innovation of our approach is an algorithmic treatment of unstable ground states; this is essential for our algorithms to apply to all inverse temperatures β. By treating unstable ground states our work gives a general template for converting probabilistic applications of Pirogov–Sinai theory to efficient algorithms.

CCS CONCEPTS

• **Theory of computation → Design and analysis of algorithms**; **Generating random combinatorial structures**.

KEYWORDS

approximate counting, sampling, Potts model, cluster expansion

ACM Reference Format:
Christian Borgs, Jennifer Chayes, Tyler Helmuth, Will Perkins, and Prasad Tetali. 2020. Efficient Sampling and Counting Algorithms for the Potts Model on $\mathbb{Z}^d$ at All Temperatures. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3357713.3384271

*TH supported by EPSRC EP/P003656/1, WP supported in part by NSF DMS- 1847451, PT supported in part by NSF DMS-1811935.

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384271

1 INTRODUCTION

The Potts model is a probability distribution on assignments of q colors to the vertices of a finite graph G. Let

$$H_G(\sigma) := \sum_{(i,j)\in E(G)} \delta_{\sigma_i \neq \sigma_j}, \qquad \sigma \in [q]^{V(G)} := \{1, 2, \ldots, q\}^{V(G)},$$

be the the number of bichromatic edges of G under the coloring σ. Then the *q-state ferromagnetic Potts model* at inverse temperature $\beta \geq 0$ is the probability distribution μ_G^{Potts} on $[q]^{V(G)}$ defined by

$$\mu_G^{\text{Potts}}(\sigma) := \frac{e^{-\beta H_G(\sigma)}}{Z_G^{\text{Potts}}(\beta)}, \qquad Z_G^{\text{Potts}}(\beta) := \sum_{\sigma \in [q]^{V(G)}} e^{-\beta H_G(\sigma)}.$$

The normalizing constant $Z_G^{\text{Potts}}(\beta)$ is the Potts model partition function. Since $\beta \geq 0$, monochromatic edges are preferred and the model is ferromagnetic.

From a computational point of view, Z_G^{Potts} and μ_G^{Potts} define families of functions and probability measures indexed by finite graphs G, and there are two main computational tasks associated to these objects. The first is the *approximate counting* problem: for a partition function Z_G and error tolerance $\epsilon > 0$, compute a number $\hat{Z}$ so that $e^{-\epsilon}\hat{Z} \leq Z_G \leq e^{\epsilon}\hat{Z}$. We say that such a $\hat{Z}$ is an *ϵ-relative approximation* to Z_G. The second is the *approximate sampling* problem: for a probability measure μ_G and error tolerance $\epsilon > 0$, output a random configuration $\hat{\sigma}$ with distribution $\hat{\mu}$ so that $\|\hat{\mu} - \mu_G\|_{TV} < \epsilon$. We say $\hat{\sigma}$ is an *ϵ-approximate sample* from μ_G.

A *fully polynomial-time approximation scheme* (FPTAS) is an algorithm that given G and $\epsilon > 0$ returns an ϵ-relative approximation to Z_G and runs in time polynomial in $|V(G)|$ and $1/\epsilon$. If the algorithm uses randomness it is a *fully polynomial-time randomized approximation scheme* (FPRAS). A randomized algorithm that given G and ϵoutputs an ϵ-approximate sample from μ_G and runs in time polynomial in both $|V(G)|$ and ϵ^{-1} is an *efficient sampling scheme.*

In this paper we give an FPRAS and an efficient sampling scheme for the q-state Potts model on the discrete torus $\mathbb{T}_n^d = (\mathbb{Z}/n\mathbb{Z})^d$ for *all* inverse temperatures $\beta \geq 0$, provided q is large enough as a function of d.

Theorem 1.1. *For all $d \geq 2$ there exists $q_0 = q_0(d)$ so that for $q \geq q_0$ and all inverse temperatures $\beta \geq 0$ there is an FPRAS and efficient*

sampling scheme for the q-state Potts model at inverse temperature β on the torus $\mathbb{T}_n^d$.

If ϵ is not too small, meaning $\epsilon \geq \exp(-O(n^{d-1}))$, our approximate counting algorithm is deterministic. We comment on this further below Theorem 1.2 in Section 1.4, but before stating our more general results we briefly discuss the aspects of the Potts model relevant to this paper. For a more comprehensive discussion see, e.g., [15]. In the remainder of this paper we will focus on the counting aspect of Theorem 1.1, deferring a discussion of the sampling aspects to [7].

The Potts model is known to exhibit a *phase transition* on the infinite lattice $\mathbb{Z}^d$ when $d \geq 2$. Roughly speaking, this means that there is a critical value $0 < \beta_c < \infty$ such that the model is sensitive to boundary conditions when $\beta > \beta_c$, and insensitive when $\beta < \beta_c$. To make this more precise, consider the Potts model on a box in $\mathbb{Z}^d$. Sensitivity can be formalized as follows. *Monochrome boundary conditions* mean that all spins at the boundary of the box have a single fixed color, say, red. Given monochromatic red boundary conditions, what is the probability a spin at the center of the box is red? If this probability is q^{-1} as the box size tends to infinity, the boundary conditions are asymptotically irrelevant: it is as if the spin is assigned a color uniformly at random from the q possible colors. However, if this probability exceeds q^{-1}, the boundary conditions have a macroscopic influence, and there is a sensitivity to the boundary conditions. These notions can be formalized in terms of *Gibbs measures* [17]; for our purposes what is important is that there is a unique Gibbs measure when boundary conditions do not play a role, and there are multiple Gibbs measure when there is a sensitivity to boundary conditions.

When q is sufficiently large the phase diagram for the q-state Potts model has been completely understood for some time [30, 32]. There is a critical temperature $\beta_c = \beta_c(d, q)$ satisfying

$$\beta_c = \frac{\log q}{d} + O(q^{-1/d}) \tag{1}$$

such that for $\beta < \beta_c$ there is a unique infinite-volume Gibbs measure, while if $\beta > \beta_c$ there are q extremal translation-invariant Gibbs measures. Each of these low-temperature measures favors one of the q colors. At the transition point $\beta = \beta_c$ there are $q + 1$ extremal translation-invariant Gibbs measures; q of these measures favor one of the q colors, and the additional measure is the 'disordered' measure from $\beta < \beta_c$. The precise meaning of these statements will not be needed in what follows; we merely wish to emphasize that there is a complete probabilistic understanding of the model. We note that the phenomenology of the model is q-dependent [15]. The preceding results require q large as they use q^{-1} as a small parameter in proofs.

The existence of multiple measures in the low-temperature phase is reflected in the dynamical aspects of the model on finite graphs like $\mathbb{T}_n^d$. This is exemplified by the behavior of a simple Markov chain with local updates called the Potts model Glauber dynamics. These dynamics mix rapidly on $\mathbb{T}_n^d$ at sufficiently high temperatures, but mix slowly, in time $\exp(\Theta(n^{d-1}))$, when $\beta \geq \beta_c$ [8, 10]. Even the global-move Swensden–Wang dynamics take time $\exp(\Theta(n^{d-1}))$ to mix when $\beta = \beta_c$ [8].

The results just discussed were primarily obtained by making use of a sophisticated form of *Pirogov–Sinai theory*. Pirogov–Sinai theory is an important tool from statistical physics which involves representing a lattice spin model in terms of *contours*. Roughly, contours are geometric objects that separate spins into regions that are aligned with different ground states. This is quite similar to what is done in the famous Peierls argument for proving the existence of a phase transition for the Ising model. It was recently shown that in some special circumstances Pirogov–Sinai theory can be used to develop efficient algorithms for approximate counting and sampling [26]. The major restriction of this method is that it only applied to models in which all ground states are *stable*, i.e., all ground states have the same truncated free energy. Most applications of Pirogov–Sinai theory, including the results concerning the Potts model described in the previous paragraphs, involve working with both stable and unstable ground states, and the main achievement of this paper is to show how to develop efficient algorithms when unstable ground states play a significant role. We discuss our methods in more detail in the subsections that follow.

1.1 Methods and Related Results

Prior to this paper, efficient algorithmic results for the Potts model on $\mathbb{Z}^d$ for $d \geq 2$ were restricted to either $\beta < \beta_c$ (see [6] and references therein), $\beta \gg \beta_c$ [4, 26], or the special cases of $q = 2$ [28] or the planar case of $\mathbb{Z}^2$ [5, 21, 22, 42]. The results in the planar case make use of planar duality to map results from $\beta < \beta_c$ to $\beta > \beta_c$.

More broadly, meaning beyond $\mathbb{Z}^d$ and beyond the Potts model, algorithms for low-temperature models have only recently been developed, and have been based primarily on cluster expansion methods [4, 11–13, 27, 34]. Our algorithms make use of cluster expansion methods as well, albeit with significant refinements that we describe below. All of these low temperature algorithms belong to the same circle of ideas as Barvinok's interpolation method [3] and subsequent improvements due to Patel and Regts [36].

Barvinok's method relies on the existence of a region in the complex plane that is free of zeros for the partition functions that one wants to approximate. The zero-free hypothesis presents a significant difficulty for using the method to study the Potts model partition functions $Z^{\text{Potts}}_{\mathbb{T}_n^d}$ at criticality. Namely, there is no open set in the complex plane centered at β_c that is zero-free uniformly in the side length n of the torus, precisely because there is a phase transition at β_c, which implies complex zeroes of the partition functions must approach β_c as $n \to \infty$. This connection between complex zeroes of partition functions and phase transitions is the Lee–Yang theory of phase transitions [33]. To deal with this, we develop algorithms based directly on the cluster expansion, and avoid the use of complex zero-free regions and Barvinok's interpolation method that were used in [26].

To describe the most important advance of the present paper we must first recall the basic idea of Pirogov–Sinai theory. For a discrete statistical mechanics model, *ground states* are locally optimal configurations, i.e., configurations whose energy is only increased by making local changes. For example, the all-plus and all-minus configurations are ground states for the Ising model, and the monochromatic configurations are ground states for the Potts model. Informally, a ground state is *stable* if it is globally

optimal in the sense that a significant fraction of the partition function is accounted for by this ground state and small deviations from it. For the Ising model with a positive magnetic field the all-plus configuration is stable, while the all-minus configuration is unstable. In [26] it was shown how to create efficient algorithms using Pirogov–Sinai theory at very low temperatures in models in which *all* ground states are stable. The restriction to very low temperatures plays an important role, as β^{-1} plays the role of a small parameter that ensures certain power series expansions converge.

To control the Potts model at all temperatures, we make use of a sophisticated form of Pirogov–Sinai theory for the Potts model that was initiated in [8]. This theory begins by considering the *random cluster model* representation of the Potts model using the Edwards–Sokal coupling (see Section 1.4 below). The random cluster model is a family of probability distributions on edge sets of graph parameterized by q and β. The random cluster model has two ground states: the ordered ground state consisting of all edges occupied and the disordered ground state consisting of no occupied edges. Significantly, these two ground states are *not* symmetric, unlike the q ground states of the Potts model.

The Pirogov–Sinai representation of the random cluster model is a representation of the partition function in terms of contours separating ordered and disordered regions. This rewriting of the partition function allows one to use $q^{-1} \ll 1$ as the small parameter that controls convergence of the relevant power series. In particular, it is possible to obtain convergence when $\beta \geq \beta_h$ for an explicit β_h *smaller* than β_c. There is a price to pay, however: unlike the spin representation of the Potts model with symmetric (and hence stable) ground states, this formulation has *unstable* ground states when $\beta \neq \beta_c$. To make use of this representation, an essential task is therefore the development of algorithms that work with unstable ground states. We outline how this is done in more detail in Section 1.3 below.

We remark that while the present paper directly discusses only the Potts and random cluster models, our approach to the treatment of unstable ground states can be carried out for other models. Thus the present paper provides a blueprint for developing efficient algorithms *whenever* Pirogov–Sinai theory can be applied.

One further remark concerning related literature is in order. Our algorithms for $\beta < \beta_c$ are also based on the cluster expansion and Pirogov–Sinai theory. It is believed that the Glauber dynamics mix rapidly on the torus for all $\beta < \beta_c$, which would yield a much faster sampling algorithm than the one we have given here. It may be possible to combine results and proof techniques from [2, 16, 35] to prove this. However, we were not able to do this and are not aware of any existing statement in the literature which would directly imply rapid mixing in the whole range $\beta < \beta_c$. We leave this as an open problem. Further open problems can be found in the conclusion of this paper, Section 6.

1.2 Discussion of Broader Context

Recall the results of Weitz and Sly [39, 43] and subsequent extensions [18, 40] that show that the phase transition point of the hard-core model on the infinite Δ-regular tree coincides with the computational threshold for efficient sampling algorithms for the class of graphs with maximum degree Δ. Recall also the connection mentioned above between non-uniqueness of Gibbs measure and slow mixing of certain Markov chains. Finally, recall that in the setting of search and optimization problems on random graphs, there is a beautiful series of predictions that link certain phase transitions and the accompanying structural changes to the solution space to computational hardness (e.g. [1, 14, 20, 31]). All of these results raise a fascinating, if imprecise, question: what is the relation between phase transitions in statistical mechanics and phase transitions in computational complexity?

Closely related to the results of Sly and Weitz is the question of #BIS, the complexity class defined by the task of approximately counting independent sets on *bipartite* graphs [23]. It is conjectured that (approximate) #BIS lies between P and NP-hard; that is, there is no polynomial-time algorithm yet the problem is not NP-hard. Resolving this question is a central open problem in the field of approximate counting. For some intuition regarding the significance of the bipartite restriction, recall that while finding a maximum size independent set in a general graph is an NP-hard problem, finding a maximum size independent set in a *bipartite* graph is in P.

The connection of these results and conjectures to the subject of the present paper is the fact that approximating the partition function of the ferromagnetic Potts model is #BIS-hard [19, 23]. One motivation for this article was the desire to rule out a connection between the phase transition for the Potts model on $\mathbb{Z}^d$ and any (conjectural) computational barriers to #BIS. While our results do not allow us to draw this conclusion in total generality, they represent a significant step forward, precisely because Theorem 1.1 (and Theorem 1.2 below) applies to *all* temperatures including the critical temperature. Extensions of our results to more general boundary conditions, which we expect to be possible, will be further evidence against any connection.

Beyond this, we also expect that it will also be possible to generalize our methods beyond $\mathbb{Z}^d$. In this way we view our results as a step towards clarifying the connection between physical and computational phase transitions. In this respect, it is important to note that our results rely crucially on the fact that the q-state Potts model and random cluster model with $q \gg 1$ undergo entropy-driven first order phase transitions, a type of transition that probabilists and mathematical physicists understand deeply and have sophisticated tools for analyzing. More precisely, the methods that we use in this paper at $\beta = \beta_c$ rely on the fact that the random cluster model with $q \gg 1$ has a finite correlation length at β_c, i.e., correlations decay exponentially fast in a large box under either the ordered or disordered boundary conditions. We expect that developing algorithms for models with second-order phase transitions, which do not have finite correlation length at criticality, will require making use of a similarly deep understanding of the underlying physical phenomena.

1.3 Proof Overview

This section contains an outline of the main steps of our proof. Since the proof itself involves a substantial amount of machinery, we focus here on giving an outline that does not enter into non-essential technicalities.

As will be clear from this outline, this paper uses the methods and framework developed in [8] and [26]. For the ease of the reader

who wishes to see the proofs of results we use from [8] we have largely stuck to the definitions presented in that paper, and have made careful note whenever we have chosen alternative definitions that facilitate our algorithms. Due to space constraints, we omit many proofs in the present article. Full details can be found in [7].

1.3.1 Section 2. We recall the notion of a polymer model along with a convenient convergence criterion for the associated cluster expansion. We then recall from [26] how this can be used for approximation algorithms, and present an important improvement upon [26] by showing how algorithms can be obtained by working directly with the cluster expansion (and using approximate instead of exact weights) rather than using Barvinok's method [4]. The essential idea is to make use of the explicit error estimates provided by a convergent cluster expansion. We recall from Section 1.1 that this avoidance of Barvinok's method is crucial for what follows.

The second part of this section makes use of the polymer model algorithm to obtain algorithms for the random cluster model at very high temperatures, meaning $\beta \le \beta_h := \frac{3\log q}{4d}$. Note that $\beta_h < \beta_c$. The precise value of β_h is not important; it was chosen to cover the range of β where the Pirogov–Sinai methods of subsequent sections do not apply. This is the first place in the paper where we exploit $q \gg 1$ as a small parameter. Similar ideas are also behind the Pirogov–Sinai methods used in the subsequent sections.

1.3.2 Section 3. We recall and develop further the Pirogov–Sinai theory tools for the Potts model that were introduced in [8]. The upshot of the section is a so-called 'contour model' formulation of the random cluster model, together with lemmas that ensure that it is possible to efficiently construct the polymers relevant for our algorithms. To help the reader, we provide a few wayposts.

The first target of this section is the derivation of the formulas (14) for Z_{dis} and Z_{ord} along with the geometric interpretations given by (12) and (13). Estimates established in later sections will show that the sum of the partition functions Z_{dis} and Z_{ord} gives a good approximation of the random cluster partition function Z.

Having introduced contours as geometric objects, it is then necessary to show that we can efficiently construct the contours used in our algorithms. This occupies the majority of Sections 3.5 and 3.6. The reader who wishes to first see how to obtain the contour model representation used for our algorithms should read up to Lemma 3.4 and then skip to Section 3.7.

1.3.3 Section 4. This section concerns estimates for the contour model representation derived in Section 3 when $q \gg 1$. The initial parts of this section recall inputs that we need from [8]. Using these inputs we then prove some new results that are needed for our algorithms. The crucial new estimate is Lemma 4.6. Roughly speaking, this estimate controls how quickly unstable contours 'flip' to becoming stable contours. The fact that this flipping occurs quickly is an essential property for our algorithms to be efficient and the key new ingredient for dealing with unstable ground states.

This section focuses on the most interesting case of $\beta \ge \beta_c$. The case $\beta_h < \beta < \beta_c$, which is very similar to $\beta > \beta_c$ and also uses estimates from [8], is discussed in [7].

1.3.4 Section 5. This section presents our approximate counting algorithms. The central difficulty in using the polymer model algorithms of Section 2 for the contour models that result from Pirogov–Sinai theory is that the weights of contours models involve ratios of partition functions. In [26] it was shown how to accurately approximate these weights in an inductive manner when all ground states are stable. The importance of this last assumption arises because the partition functions that appear in contour weights have boundary conditions; the possible boundary conditions are the ground states. Pirogov–Sinai theory based algorithms give good approximations for partition functions with boundary conditions from *stable* ground states, and this was an important part of the inductive approach in [26].

When $\beta = \beta_c$ all ground states are stable for our contour model, and implementing the ideas of [26] is fairly straightforward. However, when $\beta > \beta_c$ not all ground states are stable for our contour model, and this necessitates significant new ideas. As mentioned above, the main idea is to use the fact that unstable contours 'flip' quickly, i.e., deep inside an unstable contour it is likely to look like the interior of a stable contour. We implement this idea by showing that it is possible to explicitly enumerate the most likely ways in which an unstable boundary conditions reverts to being a stable boundary condition. This is formalized in Lemma 5.3; given this lemma the remainder of the analysis is similar to the case $\beta = \beta_c$.

Similar refinements for the treatment of unstable ground states are used for the development of sampling algorithms, see [7].

1.4 Random Cluster Model

As remarked above, an essential component of our algorithms and analysis is using the random cluster representation of the Potts model. This also means that our results apply more generally than to the Potts model on the torus: they also give efficient approximation algorithms for the more general random cluster model on both the torus and on a broad class of subsets of $\mathbb{Z}^d$. To make this more precise, recall that given a finite graph $G = (V(G), E(G))$ the random cluster model is a probability distribution on edge sets of G given by

$$\mu_G^{\text{RC}}(A) := \frac{p^{|A|}(1-p)^{|E(G)|-|A|}q^{c(G_A)}}{Z_G^{\text{RC}}(p,q)}, \qquad A \subseteq E(G), \quad (2)$$

where $c(G_A)$ is the number of connected components of the graph $G_A = (V(G), A)$ and

$$Z_G^{\text{RC}}(p,q) := \sum_{A \subseteq E(G)} p^{|A|}(1-p)^{|E(G)|-|A|}q^{c(G_A)}$$

is the random cluster model partition function.

The Potts model and the random cluster model can be put onto the same probability space via the Edwards–Sokal coupling. This result, see, e.g., [7, Appendix A], gives the relation, for $\beta \ge 0$ and integer $q \ge 2$,

$$Z_G^{\text{Potts}}(\beta) = e^{\beta|E(G)|}Z_G^{\text{RC}}(1-e^{-\beta}, q).$$

With the parameterization $p = 1 - e^{-\beta}$ the random cluster model on $\mathbb{Z}^d$, $d \ge 2$, also has a critical inverse temperature $\beta_c = \beta_c(q,d)$ that satisfies (1) and that coincides with the Potts critical inverse temperature for integer q. For $\beta < \beta_c$ the random cluster model

has a unique infinite volume measure (the *disordered* measure), while for $\beta > \beta_c$ the *ordered* measure is the unique infinite volume measure. For $\beta = \beta_c$ the two measures coexist for large q.

Our counting and sampling algorithms extend to the random cluster model on finite subgraphs of $\mathbb{Z}^d$ with two different types of boundary conditions. To make this precise requires a few definitions. Let Λ be a finite set of vertices of $\mathbb{Z}^d$ and let G_Λ be the subgraph induced by Λ. We say G_Λ is *simply connected* if G_Λ is connected and the subgraph induced by $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$ is connected. The random cluster model with *free boundary conditions* on G_Λ is just the random cluster model on the induced subgraph G_Λ as defined by (2). The random cluster model with *wired boundary conditions* on G_Λ is the random cluster model on the (multi-)graph G'_Λ obtained from G_Λ by identifying all of the vertices on the boundary of Λ to be one vertex; see [15, Section 1.2.2] for a formal definition. We refer to the Gibbs measures and partition functions with free and wired boundary conditions as $\mu^f_\Lambda, \mu^w_\Lambda, Z^f_\Lambda, Z^w_\Lambda$. Explicitly,

$$Z^f_\Lambda := \sum_{A \subset E(G_\Lambda)} p^{|A|}(1-p)^{|E(G_\Lambda)|-|A|} q^{c(G_A)}, \qquad \text{and}$$
$$Z^w_\Lambda := \sum_{A \subset E(G'_\Lambda)} p^{|A|}(1-p)^{|E(G'_\Lambda)|-|A|} q^{c(G'_A)},$$

where $c(G_A)$ is the number of connected components of the graph (Λ, A) and $c(G'_A)$ is the number of components of the graph (Λ', A) in which we identify all vertices on the boundary of Λ.

Theorem 1.2. *For $d \geq 2$ there exists $q_0 = q_0(d)$ so that for $q \geq q_0$ the following is true.*

For $\beta \geq \beta_c$ there is an FPTAS and efficient sampling scheme for the random cluster model on all finite, simply connected induced subgraphs of $\mathbb{Z}^d$ with wired boundary conditions.

For $\beta \leq \beta_c$ there is an FPTAS and efficient sampling scheme for the random cluster model on all finite, simply connected induced subgraphs of $\mathbb{Z}^d$ with free boundary conditions.

Theorem 1.2 yields an FPTAS, while Theorem 1.1 gave an FPRAS for the torus. The reason for this is that our Pirogov–Sinai based methods become more difficult to implement on the torus if the error parameter ϵ is smaller than $\exp(-O(n^{d-1}))$. The algorithm for Theorem 1.1 circumvents this by making use of the Glauber dynamics for this range of ϵ. This is possible because, despite being slow mixing, the Glauber dynamics are fast enough when given time $O(\epsilon^{-1})$ for ϵ this small by [8]. By using Glauber dynamics in a similar manner we could obtain an FPRAS for the random cluster model on $\mathbb{T}^d_n$.

We note that our methods are certainly capable of handling boundary conditions other than those described above, but we leave an investigation of this for the future.

2 POLYMER MODELS, CLUSTER EXPANSIONS, AND ALGORITHMS

This section describes how two related tools from statistical physics, abstract polymer models and the cluster expansion, can be used to design efficient algorithms to approximate partition functions.

An *abstract polymer model* [24, 29] consists of a set $\mathcal{C}$ of *polymers* each equipped with a complex-valued *weight* w_γ and a non-negative *size* $\|\gamma\|$. The set $\mathcal{C}$ also comes equipped with a symmetric compatibility relation $\sim$ such that each polymer is incompatible with itself, denoted $\gamma \nsim \gamma$. Let $\mathcal{G}$ denote the collection of all sets of pairwise compatible polymers from $\mathcal{C}$, including the empty set of polymers. The polymer model partition function is defined to be

$$Z(\mathcal{C}, w) := \sum_{\Gamma \in \mathcal{G}} \prod_{\gamma \in \Gamma} w_\gamma. \tag{3}$$

In (3) w is shorthand for the collection of polymer weights.

Let Γ be a non-empty tuple of polymers. The *incompatibility graph* H_Γ of Γ has vertex set Γ and edges linking any two incompatible polymers, i.e., $\{\gamma, \gamma'\}$ is an edge if and only if $\gamma \nsim \gamma'$. A non-empty ordered tuple Γ of polymers is a *cluster* if its incompatibility graph H_Γ is connected. Let $\mathcal{G}^c$ be the set of all clusters of polymers from $\mathcal{C}$. The cluster expansion is the following formal power series for $\log Z(\mathcal{C}, w)$ in the variables w_γ:

$$\log Z(\mathcal{C}, w) = \sum_{\Gamma \in \mathcal{G}^c} \phi(H_\Gamma) \prod_{\gamma \in \Gamma} w_\gamma. \tag{4}$$

In (4) $\phi(H)$ denotes the *Ursell function* of the graph $H = (V(H), E(H))$, i.e.,

$$\phi(H) := \frac{1}{|V(H)|!} \sum_{\substack{A \subseteq E(H) \\ (V(H), A) \text{ connected}}} (-1)^{|A|}.$$

For a proof of (4) see, e.g., [17, 29]. Define $\|\Gamma\| := \sum_{\gamma \in \Gamma} \|\gamma\|$, and define the truncated cluster expansion by

$$T_m(\mathcal{C}, w) := \sum_{\substack{\Gamma \in \mathcal{G}^c \\ \|\Gamma\| < m}} \phi(H_\Gamma) \prod_{\gamma \in \Gamma} w_\gamma .$$

Henceforth we will restrict our attention to a special class of polymer models defined in terms of a graph G with maximum degree Δ on N vertices. Namely, we will assume that each polymer is a connected subgraph $\gamma = (V(\gamma), E(\gamma))$ of G. The compatibility relation is defined by disjointness in G: $\gamma \sim \gamma'$ iff $V(\gamma) \cap V(\gamma') = \emptyset$. We write $|\gamma|$ for $|V(\gamma)|$, the number of vertices in the polymer γ.

A useful criteria for convergence of the formal power series in (4) is given by the following adaptation of a theorem of Kotecký and Preiss [29].

Lemma 2.1. *Suppose that polymers are connected, induced subgraphs of a graph G of maximum degree Δ on N vertices. Suppose further that for some $b > 0$ and all $\gamma \in \mathcal{C}$,*

$$\|\gamma\| \geq b|E(\gamma)|, \tag{5}$$
$$|w_\gamma| \leq e^{-(\frac{4+\log\Delta}{b}+3)\|\gamma\|}. \tag{6}$$

Then the cluster expansion (4) converges absolutely, and for $m \in \mathbb{N}$,

$$|T_m(\mathcal{C}, w) - \log Z(\mathcal{C}, w)| \leq N e^{-3m}.$$

Moreover, if instead all polymers are connected, induced subgraphs of G, and for some $b > 0$ and all $\gamma \in \mathcal{C}$,

$$\|\gamma\| \geq b|\gamma|, \quad \text{and} \quad |w_\gamma| \leq e^{-(\frac{3+\log\Delta}{b}+3)\|\gamma\|},$$

then the same conclusion holds.

This lemma implies that if conditions (5) and (6) hold, then $\exp(T_m(C,w))$ is an ϵ-relative approximation to $Z(C,w)$ for $m \geq \log(N/\epsilon)/3$.

Because clusters are connected objects arising from a bounded-degree graph, the truncated cluster expansion can be computed efficiently. Recall that $N = |V(G)|$.

Lemma 2.2. *Suppose the conditions of Lemma 2.1 hold. Then given a list of all polymers γ of size at most m along with the weights w_γ of these polymers, the truncated cluster expansion $T_m(C,w)$ can be computed in time $O(N\exp(O(m)))$.*

The next lemma says that, for the purposes of approximating a polymer partition function, it is sufficient to have approximate evaluations $\tilde{w}_\gamma$ of the weights w_γ.

Lemma 2.3. *Let $v\colon C \to [0,\infty)$ be a non-negative function on polymers such that $v(\gamma) \leq \|\gamma\|^2$. Suppose $0 < \epsilon < N^{-1}$, and let $m = \log(8/\epsilon)/3$. Suppose the conditions of Lemma 2.1 hold and that for all $\gamma \in C$ with $\|\gamma\| \leq m$, $\tilde{w}_\gamma$ is an $\epsilon v(\gamma)$-relative approximation to w_γ. Then $\exp(T_m(C,\tilde{w}))$ is an $N\epsilon/4$-relative approximation to $Z(C,w)$.*

Proof. Using the definition of m and applying Lemma 2.1, we have

$$|\log Z_G(C,w) - T_m(C,w)| \leq N\epsilon/8,$$

so by the triangle inequality it is enough to show that

$$|T_m(C,\tilde{w}) - T_m(C,w)| \leq N\epsilon/8. \tag{7}$$

Define r_γ by $\log \tilde{w}_\gamma = \log w_\gamma + r_\gamma$. To prove (7), note the identity

$$T_m(C,\tilde{w}) - T_m(C,w) = \sum_{\substack{\Gamma \in \mathcal{G}^c(G) \\ \|\Gamma\| < m}} \phi(H_\Gamma) \prod_{\gamma\in\Gamma} w_\gamma \cdot \left[\exp\left(\sum_{\gamma\in\Gamma} r_\gamma \right) - 1 \right].$$

Our hypotheses imply $|r_\gamma| \leq \epsilon v(\gamma)$, and hence by the triangle inequality we obtain

$$|T_m(C,\tilde{w}) - T_m(C,w)| \leq \sum_{\substack{\Gamma \in \mathcal{G}^c(G) \\ \|\Gamma\| < m}} (\exp(\sum_{\gamma\in\Gamma} \epsilon v(\gamma)) - 1) \left| \phi(H_\Gamma) \prod_{\gamma\in\Gamma} w_\gamma \right|,$$

where we have used the elementary inequality $|e^a - 1| \leq e^b - 1$ when $|a| \leq b$ to bound the term in square brackets. Since $v(\gamma) \leq \|\gamma\|^2$ this yields, after ordering the sum over clusters according to their size k,

$$\begin{aligned} |T_m(C,\tilde{w}) - T_m(C,w)| &\leq \sum_{k=1}^{m-1} (\exp(\epsilon k^2) - 1) \sum_{\substack{\Gamma \in \mathcal{G}^c(G) \\ \|\Gamma\| = k}} \left| \phi(H_\Gamma) \prod_{\gamma\in\Gamma} w_\gamma \right| \\ &\leq \sum_{k=1}^{m-1} (\exp(\epsilon k^2) - 1) N e^{-3k}. \end{aligned}$$

The last inequality follows from the (proof of the) convergence of the cluster expansion. Since $\epsilon < N^{-1}$ we can bound $e^{\epsilon k^2} - 1$ by $2\epsilon k^2$, and (7) follows since $\sum_{k\geq 1} k^2 e^{-3k} < 1/16$. □

Putting Lemmas 2.1, 2.2, and 2.3 together we see that the partition function $Z(C,w)$ can be approximated efficiently if: (1) conditions (5) and (6) hold, (2) polymers of size at most m can be enumerated efficiently, i.e., in time polynomial in N and exponential in m, and (3) the polymer weights w_γ can be approximated efficiently, i.e., in time polynomial in the size of γ.

2.1 High Temperature Expansion

The polymer model algorithm of the previous section yields efficient counting and sampling algorithms for the random cluster model when q is sufficiently large and $\beta \leq \beta_h = \frac{3\log q}{4d}$. In fact, the simpler setting of $\beta \leq \beta_h$ allows for greater generality: we will derive an algorithm that applies to the random cluster model on *any* graph G of maximum degree at most $2d$.

Theorem 2.4. *Suppose $d \geq 2$ and $q = q(d)$ is sufficiently large. Then for $\beta \leq \beta_h$ there is an FPTAS and efficient sampling scheme for the Potts model and the random cluster model with $p = 1 - e^{-\beta}$ on all graphs of maximum degree at most $2d$.*

Proof of Theorems 1.1 and 1.2 for $\beta \leq \beta_h$. Theorem 1.1 follows immediately from Theorem 2.4 since $\mathbb{T}_n^d$ is $2d$-regular.

By (1), $\beta_h < \beta_c$ when q is large enough. Thus Theorem 1.2 requires we provide approximate counting and sampling algorithms for free boundary conditions. Since induced subgraphs of $\mathbb{Z}^d$ have degree bounded by $2d$, the result follows by Theorem 2.4. □

3 CONTOUR MODEL REPRESENTATIONS

Contour models refer to a class of polymer models that arise in Pirogov–Sinai theory [38]. For a given spin configuration, contours represent geometric boundaries between regions dominated by different ground states; the precise definition for the purposes of this paper will be given below. This section describes an important contour model representation for the random cluster model on the torus $\mathbb{T}_n^d$ that is the basic combinatorial object in our algorithms. This contour representation was originally developed for obtaining optimal lower bounds on the mixing time for Glauber and Swensden–Wang dynamics [8]. In addition to recalling the construction from [8] this section also develops the additional ingredients necessary for algorithmic applications of the representation.

3.1 Continuum Embedding

The contour model representation from [8] is based on the natural embedding of the discrete torus $\mathbb{T}_n^d = (\mathbb{Z}/n\mathbb{Z})^d$ of side-length $n \in \mathbb{N}$ into the continuum torus $\boldsymbol{T}_n^d := (\mathbb{R}/n\mathbb{R})^d$. This subsection recalls the basic definitions, and explains how they can be rephrased in terms of discrete graph-theoretic notions.[1]

In what follows we abuse notation slightly and write $\mathbb{T}_n^d$ for the graph $(\mathbb{T}_n^d, E)$, where E is the edge set of the discrete torus. We will follow the convention that bold symbols, e.g., $\boldsymbol{V}$, denote subsets of $\boldsymbol{T}_n^d$, while objects denoted by non-bold symbols like V reside in $\mathbb{T}_n^d$. Thus each vertex $v \in \mathbb{T}_n^d$ is identified with a point $\boldsymbol{v} \in \boldsymbol{T}_n^d$, and we will identify each edge $e = \{u,v\} \in E$ with the unit line segment

[1] This continuum construction allows for tools from algebraic topology to be used. We have chosen to follow the continuum terminology to allow the interested reader to easily consult [8].

$\boldsymbol{e} \subset \boldsymbol{T}_n^d$ that joins $\boldsymbol{u}$ to $\boldsymbol{v}$. We will also drop $\mathbb{T}_n^d$ from the notation when possible, e.g., E for $E(\mathbb{T}_n^d)$.

Recall that $\Omega = 2^E$ is the set of configurations of the random cluster model on $\mathbb{T}_n^d$. Let $\boldsymbol{c} \subset \boldsymbol{T}_n^d$ denote a closed k-dimensional hypercube with vertices in $\mathbb{T}_n^d$ for some $k = 1, \dots, d$. We say a hypercube $\boldsymbol{c}$ is *occupied* with respect to $A \in \Omega$ if for all edges e with $\boldsymbol{e} \subset \boldsymbol{c}$, e is in A. Define

$$\boldsymbol{A} \coloneqq \left\{ \boldsymbol{x} \in \boldsymbol{T}_n^d \mid \text{there exists } \boldsymbol{c} \text{ occupied s.t. } d_\infty(\boldsymbol{x}, \boldsymbol{c}) \le \frac{1}{4} \right\},$$

where d_∞ is the ℓ_∞-distance, and the distance from a point to a set is defined in the standard way: $d_\infty(\boldsymbol{x}, \boldsymbol{c}) = \inf_{\boldsymbol{y} \in \boldsymbol{c}} d_\infty(\boldsymbol{x}, \boldsymbol{y})$. Thus $\boldsymbol{A}$ is the closed 1/4-neighborhood of the occupied hypercubes of A. The connected components of the (topological) boundary $\partial \boldsymbol{A}$ of the set $\boldsymbol{A}$ are the crucial objects in what follows. Since each connected component arises from an edge configuration in Ω, it is clear that the set of possible connected components is a finite set. As the connected components of $\partial \boldsymbol{A}$ are continuum objects, it may not be immediately apparent how to represent them in a discrete manner. We briefly describe how to do this now.

Let $\frac{1}{2}\mathbb{T}_n^d$ denote the graph $(\frac{1}{2}\mathbb{Z}/n\mathbb{Z})^d$; as a graph this is equivalent to the discrete torus $(\mathbb{Z}/(2n)\mathbb{Z})^d$. The notation $\frac{1}{2}\mathbb{T}_n^d$ is better because we will embed $\frac{1}{2}\mathbb{T}_n^d$ in $\boldsymbol{T}_n^d$ such that (i) $\boldsymbol{0}$ coincides in $\mathbb{T}_n^d$ and $\frac{1}{2}\mathbb{T}_n^d$, and (ii) the nearest neighbors of 0 in $\frac{1}{2}\mathbb{T}_n^d$ are the midpoints of the edges e containing 0 in $\mathbb{T}_n^d$.[2]

An important observation is that $\boldsymbol{A}$ can be written as a union of collections of adjacent closed d-dimensional hypercubes of side-length 1/2 centered at vertices in $\frac{1}{2}\mathbb{T}_n^d$, where two hypercubes are called *adjacent* if they share a $(d-1)$-dimensional face. Adjacency of a set of hypercubes means the set of hypercubes is connected under the binary relation of being adjacent. By construction the connected components of $\boldsymbol{A}$ correspond to the connected components of the edge configuration A.

The boundary $\partial \boldsymbol{A}$ of $\boldsymbol{A}$ is just the sum, modulo two, of the boundaries of the hypercubes whose union gives A. These boundaries are $(d-1)$-dimensional hypercubes dual to edges in $\frac{1}{2}\mathbb{T}_n^d$; here dual means that the barycenter of the $(d-1)$-dimensional hypercube is the same as barycenter of the edge in $\frac{1}{2}\mathbb{T}_n^d$. The $(d-1)$-dimensional hypercubes that arise from this duality are the vertices in $(\frac{1}{2}\mathbb{T}_n^d)^\star$, the graph dual to $\frac{1}{2}\mathbb{T}_n^d$; two vertices in $(\frac{1}{2}\mathbb{T}_n^d)^\star$ are connected by an edge if and only if the corresponding $(d-1)$-dimensional hypercubes intersect in one $(d-2)$-dimensional hypercube. The preceding discussion implies $\partial \boldsymbol{A}$ can be identified with a subgraph of $(\frac{1}{2}\mathbb{T}_n^d)^\star$.

In the sequel we will discuss components of $\partial \boldsymbol{A}$ as continuum objects; by the preceding discussion this could be reformulated in terms of subgraphs of $(\frac{1}{2}\mathbb{T}_n^d)^\star$, see [7, Appendix C].

3.2 Contours and Interfaces

An important aspect of the analysis in [8] is that it distinguishes topologically trivial and non-trivial components of $\partial \boldsymbol{A}$. To make this precise, for $i = 1, \dots, d$ we define the *ith fundamental loop* $\boldsymbol{L}_i$ to be the set $\{\boldsymbol{y} \in \boldsymbol{T}_n^d \mid \boldsymbol{y}_j = 1 \text{ for all } j \neq i\}$. The *winding vector* $N(\boldsymbol{\gamma}) \in \{0,1\}^d$ of a connected component $\boldsymbol{\gamma} \in \partial \boldsymbol{A}$ is the vector whose ith component is the number of intersections (mod 2) of $\boldsymbol{\gamma}$ with $\boldsymbol{L}_i$.

Definition 1. *Let $A \in \Omega$ be an edge configuration.*

(1) The set of contours $\Gamma(A)$ associated to A *is the set of connected components of $\partial \boldsymbol{A}$ with winding vector* 0.

(2) The interface network $\mathcal{S}(A)$ associated to A *is the set of connected components of $\partial \boldsymbol{A}$ with non-zero winding vector. Each connected component of an interface network is an* interface.

Without reference to any particular edge configuration, a subset $\boldsymbol{\gamma} \subset \boldsymbol{T}_n^d$ is a contour *if there is an $A \in \Omega$ such that $\boldsymbol{\gamma} \in \Gamma(A)$. Interfaces and interface networks are defined analogously.*

Since each fundamental loop intersects each $(d-1)$-dimensional face of a hypercube centered on $\frac{1}{2}\mathbb{T}_n^d$ exactly zero or one times, we have the following lemma, which ensures contours can be efficiently distinguished from interfaces.

Lemma 3.1. *Suppose $\boldsymbol{\gamma} \in \partial \boldsymbol{A}$ is comprised of K $(d-1)$-dimensional faces. Then the winding vector of $\boldsymbol{\gamma}$ can be computed in time $O(nK)$.*

Proof. Fix $i \in \{1, 2, \dots, d\}$. Each fundamental loop L_i has length $O(n)$, and hence the set F_i of faces that have non-trivial intersection with L_i has cardinality $|F_i| = O(n)$. Given the list of faces in $\boldsymbol{\gamma}$ we can compute the ith component of the winding vector by (i) iterating through the list of faces of $\boldsymbol{\gamma}$ and adding one each time we find a face in F_i, and (ii) taking the result modulo two. □

The connected components of $\boldsymbol{T}_n^d \setminus \partial \boldsymbol{A}$ are subsets of either $\boldsymbol{A}$ or $\boldsymbol{T}_n^d \setminus \boldsymbol{A}$. In the former case we call a component *ordered* and in the latter case *disordered*. We write $\boldsymbol{A}_{\text{ord}}$ (resp. $\boldsymbol{A}_{\text{dis}}$) for the union of the ordered (resp. disordered) components associated to A.

Definition 2. *The* labelling ℓ_A associated to A *is the map from the connected components of $\boldsymbol{T}_n^d \setminus \partial \boldsymbol{A}$ to the set $\{dis, ord\}$ that assigns ord to components in $\boldsymbol{A}_{ord}$ and dis to components in $\boldsymbol{A}_{dis}$.*

Definition 3. *Two contours $\boldsymbol{\gamma}_i$, $i = 1, 2$ are* compatible *if $d_\infty(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \ge \frac{1}{2}$. We extend this definition analogously to two interfaces, or one interface and one contour.*

Definition 4. *A* matching collection of contours Γ and interfaces $\mathcal{S}$ *is a triple $(\Gamma, \mathcal{S}, \ell)$ such that $\mathcal{S}$ is an interface network and*

(1) The contours and interfaces in $\Gamma \cup \mathcal{S}$ are pairwise compatible, and

(2) ℓ is a map from the set of connected components of $\boldsymbol{T}_n^d \setminus \cup_{\boldsymbol{\gamma} \in \Gamma \cup \mathcal{S}} \boldsymbol{\gamma}$ to the set $\{dis, ord\}$ such that for every $\boldsymbol{\gamma} \in \Gamma \cup \mathcal{S}$, distinct components adjacent to $\boldsymbol{\gamma}$ are assigned different labels.

Lemma 3.2. *The map from edge configurations $A \in \Omega$ to triples $(\Gamma, \mathcal{S}, \ell)$ of matching contours and interfaces is a bijection.*

Proof. See [8, p.15]. □

3.3 Contour and Interface Formulation of Z

By Lemma 3.2 we can rewrite the partition function in terms of matching collections of contours and interfaces by re-writing the weight $w(A)$ of a configuration A in terms of its contours and

[2] More formally, since $\mathbb{Z}^d \subset \frac{1}{2}\mathbb{Z}^d \subset \mathbb{R}^d$, we obtain a common embedding of $\frac{1}{2}\mathbb{T}_n^d$ and $\mathbb{T}_n^d$ in $\boldsymbol{T}_n^d$.

interfaces. By weight $w(A)$ we mean the numerator of (2), i.e., $w(A) = p^{|A|}(1-p)^{|E\setminus A|}q^{c(V,A)}$. To this end, define

$$e_{\text{ord}} := -d\log(1-e^{-\beta}), \quad e_{\text{dis}} := d\beta - \log q, \quad \kappa := \frac{1}{2}\log(e^{\beta}-1).$$

Further, define the *size* $\|\boldsymbol{\gamma}\|$ of a contour $\boldsymbol{\gamma}$ (resp. *size* $\|S\|$ of an interface S) by

$$\|\boldsymbol{\gamma}\| := \left|\boldsymbol{\gamma} \cap \bigcup_{e\in E} \boldsymbol{e}\right|, \quad \|S\| := \left|S \cap \bigcup_{e\in E} \boldsymbol{e}\right|. \tag{8}$$

This is the number of intersections of $\boldsymbol{\gamma}$ (resp. S) with $\bigcup_{e\in E}\boldsymbol{e}$. For a continuum set $\boldsymbol{\Lambda}$ we write $|\boldsymbol{\Lambda}|$ for $|\boldsymbol{\Lambda}\cap\mathbb{T}_n^d|$, that is, the number of vertices of $\mathbb{T}_n^d$ in $\boldsymbol{\Lambda}$ in the embedding of $\mathbb{T}_n^d$ into $\boldsymbol{T}_n^d$. This will cause no confusion as we never need to measure the volume of a continuum set.

Using these definitions, $w(A)$ can be written as

$$w(A) = q^{c(\boldsymbol{A}_{\text{ord}})}e^{-e_{\text{dis}}|\boldsymbol{A}_{\text{dis}}|}e^{-e_{\text{ord}}|\boldsymbol{A}_{\text{ord}}|}\prod_{S\in\mathcal{S}} e^{-\kappa\|S\|}\prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|}, \tag{9}$$

where $c(\boldsymbol{A}_{\text{ord}})$ is the number of connected components of $\boldsymbol{A}_{\text{ord}}$. The products run over the sets of interfaces and contours associated to the edge configuration A, respectively. We indicate the derivation of (9) in Section 3.3.1 below; see also [8, p.13-15]. Since

$$Z = Z^{\text{RC}}_{\mathbb{T}_n^d}(1-e^{-\beta},q) = \sum_{A\in\Omega} w(A),$$

it follows from (9) and Lemma 3.2 that

$$Z = \sum_{(\mathcal{S},\Gamma)} q^{c(\boldsymbol{A}_{\text{ord}})}e^{-e_{\text{dis}}|\boldsymbol{A}_{\text{dis}}|}e^{-e_{\text{ord}}|\boldsymbol{A}_{\text{ord}}|}\prod_{S\in\mathcal{S}} e^{-\kappa\|S\|}\prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|}, \tag{10}$$

where the sum runs over matching collections of contours and interfaces. This is the contour and interface network representation of the random cluster model partition function.

In what follows it will be necessary to divide the contributions to Z. To this end, let

$$\Omega_{\text{tunnel}} := \{A\in\Omega \mid \mathcal{S}(A)\neq\emptyset\}, \quad \Omega_{\text{rest}} := \Omega\setminus\Omega_{\text{tunnel}},$$

and define the corresponding partition functions

$$Z_{\text{tunnel}} := \sum_{A\in\Omega_{\text{tunnel}}} w(A), \quad Z_{\text{rest}} := \sum_{A\in\Omega_{\text{rest}}} w(A).$$

By (10) Z_{rest} can be expressed in terms of contours alone. We will see later that Z_{tunnel} is small compared to Z_{rest}, and so the task of approximating Z is essentially the task of approximating Z_{rest}.

We briefly indicate how to obtain (9). Recall that G_A denotes the graph $(V(A), A)$. Let $\|\delta A\| = |\delta_1 A| + |\delta_2 A|$, where $\delta_k A$ is the set of edges in $E\setminus A$ that contain k vertices in $V(A)$. Observe

$$c(V,A) = c(G_A) + |V\setminus V(A)|$$
$$2|A| = 2d|V(A)| - \|\delta A\|.$$

The first of these relations follows since every vertex not contained in an edge of A belongs to a singleton connected component, and the second is a counting argument. Using these relations one obtains

$$w(A) = q^{c(G_A)}e^{-e_{\text{dis}}|V\setminus V(A)|}e^{-e_{\text{ord}}|V(A)|}e^{-\kappa\|\delta A\|}. \tag{11}$$

To pass from (11) to (9) requires just a few observations. First, $c(G_A)$ equals the number of components of $\boldsymbol{A}$, which is the number of connected components of $\boldsymbol{A}_{\text{ord}}$. Second, $|V(A)| = |\boldsymbol{A}_{\text{ord}}|$, and similarly $|V\setminus V(A)| = |\boldsymbol{A}_{\text{dis}}|$. Lastly, $\|\delta A\|$ is precisely the sum of sizes of the contours and interfaces, as each contribution to $\|\delta A\|$ is given by a transverse intersection of an edge $\boldsymbol{e}$ with the boundary of $\boldsymbol{A}$.

3.4 External Contour Representations

Next we will take the first steps to construct a representation of Z_{rest} as a sum of polymer model partition functions. We begin with basic results and definitions. Fix an arbitrary point $\boldsymbol{x}_0 \in \boldsymbol{T}_n^d$ that cannot be contained in any contour, and let $\sqcup$ denote disjoint union.

Lemma 3.3 ([8, Lemma 4.3]). *For any contour $\boldsymbol{\gamma}$, $\boldsymbol{T}_n^d\setminus\boldsymbol{\gamma}$ has exactly two components.*

Definition 5. *Let $\boldsymbol{\gamma}$ be a contour, and suppose $\boldsymbol{T}_n^d\setminus\boldsymbol{\gamma} = \boldsymbol{C}\sqcup\boldsymbol{D}$. Then the* exterior *$Ext\,\boldsymbol{\gamma}$ of $\boldsymbol{\gamma}$ is $\boldsymbol{C}$ if $|\boldsymbol{C}| > |\boldsymbol{D}|$, and is $\boldsymbol{D}$ if the inequality is reversed. In the case of equality the exterior is the component containing $\boldsymbol{x}_0$. The* interior *$Int\,\boldsymbol{\gamma}$ of $\boldsymbol{\gamma}$ is the component of $\boldsymbol{T}_n^d\setminus\boldsymbol{\gamma}$ that is not $Ext\,\boldsymbol{\gamma}$.*

Note that the notion of exterior is defined relative to $\boldsymbol{T}_n^d$, though we omit this from the notation.

Remark. *This is a different definition of exterior than is used in [8]; our definition is more convenient for algorithmic purposes. Most of the results of [8] concerning the interiors/exteriors of contours apply verbatim with this change, and whenever we use these results we will remark on why they apply.*

If two contours $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ are compatible, then we write (i) $\boldsymbol{\gamma} < \boldsymbol{\gamma}'$ if $\operatorname{Int}\boldsymbol{\gamma}\subset\operatorname{Int}\boldsymbol{\gamma}'$ and (ii) $\boldsymbol{\gamma}\perp\boldsymbol{\gamma}'$ if $\operatorname{Int}\boldsymbol{\gamma}\cap\operatorname{Int}\boldsymbol{\gamma}' = \emptyset$. Given a matching collection of contours Γ, $\boldsymbol{\gamma}\in\Gamma$ is an *external contour* if there does not exist $\boldsymbol{\gamma}'\in\Gamma$ such that $\boldsymbol{\gamma}' < \boldsymbol{\gamma}$. The *exterior* of a matching collection of contours Γ is

$$\operatorname{Ext}\Gamma := \bigcap_{\boldsymbol{\gamma}\in\Gamma}\operatorname{Ext}\boldsymbol{\gamma}.$$

If Γ is matching, then $\operatorname{Ext}\Gamma$ is a connected subset of $\mathbb{T}_n^d$. This follows by noting that [8, Lemma 5.5] holds with Definition 5 of the interior and exterior, and given this, the connectedness of $\operatorname{Ext}\Gamma$ follows by the argument in [8, Lemma 5.6]. Note that since $\operatorname{Ext}\Gamma$ is contained in $\mathbb{T}_n^d\setminus\bigcup_{\boldsymbol{\gamma}\in\Gamma}\boldsymbol{\gamma}$, this implies that $\operatorname{Ext}\Gamma$ is labelled either ord or dis.

As usual in Pirogov–Sinai theory, see, e.g. [8, Section 6.2], it is useful to resum the matching compatible contours that contribute to (10) according to the external contours of the configuration. To make this precise, we require several definitions. A matching collection of contours Γ is *mutually external* if $\boldsymbol{\gamma}\perp\boldsymbol{\gamma}'$ for all $\boldsymbol{\gamma}\neq\boldsymbol{\gamma}'\in\Gamma$. For a continuum set $\boldsymbol{\Lambda}\subseteq\boldsymbol{T}_n^d$, we say a contour $\boldsymbol{\gamma}$ is *a contour in $\boldsymbol{\Lambda}$* if $d_\infty(\boldsymbol{\gamma}, \boldsymbol{T}_n^d\setminus\boldsymbol{\Lambda})\geq 1/2$. The distance to the empty set is infinite by convention.

Write $C(\boldsymbol{\Lambda})$ for the set of contours in $\boldsymbol{\Lambda}$, and $C = C(\boldsymbol{T}_n^d)$ for the set of all contours. For $\boldsymbol{\Lambda}\subseteq\boldsymbol{T}_n^d$ define $\mathcal{G}^{\text{ext}}(\boldsymbol{\Lambda})$ to be the set of matching mutually external contours in $\boldsymbol{\Lambda}$, and then define

$$Z_{\text{ord}}(\boldsymbol{\Lambda}) := \sum_{\Gamma\in\mathcal{G}^{\text{ext}}_{\text{ord}}(\boldsymbol{\Lambda})} e^{-e_{\text{ord}}|\boldsymbol{\Lambda}\cap\operatorname{Ext}\Gamma|}\prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|}Z_{\text{dis}}(\operatorname{Int}\boldsymbol{\gamma}) \tag{12}$$

$$Z_{\text{dis}}(\boldsymbol{\Lambda}) := \sum_{\Gamma\in\mathcal{G}^{\text{ext}}_{\text{dis}}(\boldsymbol{\Lambda})} e^{-e_{\text{dis}}|\boldsymbol{\Lambda}\cap\operatorname{Ext}\Gamma|}\prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|}qZ_{\text{ord}}(\operatorname{Int}\boldsymbol{\gamma}), \tag{13}$$

where the sums in (12) and (13) run over sets of matching mutually external contours in which Ext Γ is labelled ord and dis, respectively. This is the desired resummation. In the special case $\boldsymbol{\Lambda} = \boldsymbol{T}_n^d$ these partition functions represent the sums of $w(A)$ over

$$\Omega_{\text{ord}} \coloneqq \{A \in \Omega \setminus \Omega_{\text{tunnel}} \mid \text{Ext}\,\Gamma(A) \text{ is labelled ord}\},$$
$$\Omega_{\text{dis}} \coloneqq \{A \in \Omega \setminus \Omega_{\text{tunnel}} \mid \text{Ext}\,\Gamma(A) \text{ is labelled dis}\}.$$

That is, we get a decomposition $Z_{\text{rest}} = qZ_{\text{ord}} + Z_{\text{dis}}$, where

$$Z_{\text{ord}} = q^{-1} \sum_{A \in \Omega_{\text{ord}}} w(A), \quad Z_{\text{dis}} = \sum_{A \in \Omega_{\text{dis}}} w(A). \tag{14}$$

3.5 Labelled Contours

This subsection introduces labelled contours and establishes some basic properties of these objects. These properties will ensure that we can efficient enumerate labelled contours.

In Definition 2 we associated a labelling to an entire collection of matching and compatible contours and interfaces. For collections of contours it is more convenient to associate the labelling to individual contours. We do this by assigning a label to Int $\boldsymbol{\gamma}$ (resp. Ext $\boldsymbol{\gamma}$) according to the label of the region of $\mathbb{T}_n^d \setminus \cup_{\boldsymbol{\gamma} \in \Gamma} \boldsymbol{\gamma}$ adjacent to $\boldsymbol{\gamma}$ contained in Int $\boldsymbol{\gamma}$ (resp. Ext $\boldsymbol{\gamma}$).

A *compatible set of labelled contours* Γ is a set of compatible contours Γ such that the connected components of $\mathbb{T}_n^d \setminus \cup_{\boldsymbol{\gamma} \in \Gamma} \boldsymbol{\gamma}$ are assigned the same labels by the labelled contours. More precisely, for a component $\boldsymbol{B}$ of $\mathbb{T}_n^d \setminus \cup_{\boldsymbol{\gamma} \in \Gamma} \boldsymbol{\gamma}$, $\partial \boldsymbol{B}$ is a union of compatible contours $\boldsymbol{\gamma}_0, \dots, \boldsymbol{\gamma}_k$ for some $k \geq 0$, and (up to relabelling) either (i) $\boldsymbol{\gamma}_i < \boldsymbol{\gamma}_0$ for $i = 1, \dots, k$ or (ii) $\boldsymbol{\gamma}_i \perp \boldsymbol{\gamma}_j$ for $i \neq j$. The condition of compatibility of the labels in the first case is that the interior label of $\boldsymbol{\gamma}_0$ is the same as the exterior label of $\boldsymbol{\gamma}_i$ for all $i = 1, \dots k$, and in the second case is that all exterior labels agree.

By construction, the set of collections of matching and compatible contours is the same as the set of collections of compatible labelled contours. The advantage of the latter is that it enables us to define a labelled contour $\boldsymbol{\gamma}$ to be *ordered* if its exterior label is ord, and *disordered* if its exterior label is dis. We let $C_{\text{ord}}(\boldsymbol{\Lambda})$ and $C_{\text{dis}}(\boldsymbol{\Lambda})$ denote the sets of labelled contours in $\boldsymbol{\Lambda}$ with external labels ord and dis, respectively, with $C_{\text{ord}} = C_{\text{ord}}(\boldsymbol{T}_n^d)$ and $C_{\text{dis}} = C_{\text{dis}}(\boldsymbol{T}_n^d)$. The next lemma gives a way to construct a labelled contour $\boldsymbol{\gamma}$ from an edge configuration.

Lemma 3.4. *Let $\ell \in \{ord, dis\}$, let $\boldsymbol{\gamma} \in C_\ell$, and $\boldsymbol{\Lambda} = Int\,\boldsymbol{\gamma}$. Then*

- *If $\ell = dis$, let $E'(\boldsymbol{\Lambda})$ be set of edges contained in $\boldsymbol{\Lambda}$. Then $\boldsymbol{\gamma}$ is the unique component of $\partial \boldsymbol{A}$ where $A = E'(\boldsymbol{\Lambda}) \subset E$.*
- *If $\ell = ord$, let $E'(\boldsymbol{\Lambda})$ be the set of edges whose midpoints are contained in $\boldsymbol{\Lambda}$. Then $\boldsymbol{\gamma}$ is the unique component of $\partial \boldsymbol{A}$ where $A = E \setminus E'(\boldsymbol{\Lambda})$.*

Proof. These claims follows from [8, Lemma 5.1]; see the proof of [8, Lemma 5.11].[3] □

Lemma 3.4 gives a way to construct a given contour from some set of edges A. For our algorithms it will be important to be able to generate contours from a relatively small set of edges. We first explain how to do this for disordered contours.

[3]These results rely only on the geometry of hypercubes and not on the definitions of interior/exterior.

Suppose $\boldsymbol{\gamma} \in C_{\text{dis}}$ and let $\Lambda = \text{Int}\,\boldsymbol{\gamma} \cap \mathbb{T}_n^d$. Define

$$\mathcal{E}_{\boldsymbol{\gamma}} \coloneqq \{e = \{i, j\} \mid i, j \in \Lambda,\ d_\infty(\text{mid}(\boldsymbol{e}), \boldsymbol{\gamma}) \geq 3/4\},$$

where mid($\boldsymbol{e}$) denotes the midpoint of the edge $\boldsymbol{e}$; this is the vertex of $\frac{1}{2}\mathbb{T}_n^d$ on the two-step path from i to j in $\frac{1}{2}\mathbb{T}_n^d$.

Lemma 3.5. *Suppose $\boldsymbol{\gamma} \in C_{dis}$ and let $\boldsymbol{\Lambda} = Int\,\boldsymbol{\gamma}$. Suppose $F \subseteq \mathcal{E}_{\boldsymbol{\gamma}}$ and let $A = E' \setminus F$, where $E' = E'(\Lambda)$ is defined as in Lemma 3.4. Let Γ be the set of contours in $\partial \boldsymbol{A}$. Then $\boldsymbol{\gamma} \in \Gamma$, and for all $\boldsymbol{\gamma}' \in \Gamma$ with $\boldsymbol{\gamma}' \neq \boldsymbol{\gamma}$ we have $\boldsymbol{\gamma}' < \boldsymbol{\gamma}$. Moreover, all sets of matching contours consisting of γ and contours in Int γ arise from such F.*

Proof. We begin by recalling an alternate construction of $\boldsymbol{A}$ from [8]. Let $E \subset E(\mathbb{T}_n^d)$, and let $D \subset E$. Set $D^\star$ to be the set of $(d-1)$-dimensional unit hypercubes dual to the edges of D, and set

$$V_-(D) = \left\{x \in V(\mathbb{T}_n^d) \mid \{x, y\} \in D \text{ if } \{x, y\} \in E\right\}.$$

Set $\boldsymbol{D}_{\text{dis}}$ to be the union of the open 3/4-neighborhood of $V_-(D)$ and the open 1/4-neighborhood of $D^\star$. Then by [8, Lemma 5.1, (iv)], if $D = E \setminus A$, $\boldsymbol{E} \setminus \boldsymbol{A} = \boldsymbol{D}_{\text{dis}}$. I.e., $\boldsymbol{D}_{\text{dis}}$ is the disordered region associated to A (relative to the region $\boldsymbol{E}$).

To prove the lemma, we apply this construction with $E = E'(\Lambda)$ and $D = F$. The definition of $\mathcal{E}_{\boldsymbol{\gamma}}$ ensures that both the open 3/4-neighborhoods of the included vertices and the open 1/4-neighborhoods of the included dual facets are at distance at least 1/2 from $\boldsymbol{\gamma}$. This implies that $\boldsymbol{\gamma}$ is a boundary component of $\boldsymbol{E} \setminus F$, and the first claim follows as all other boundary components are adjacent to $\boldsymbol{D}_{\text{dis}}$. The second claim follows from the bijection of Lemma 3.2, which restricts to a bijection in this setting. □

Lemma 3.6. *Suppose $\boldsymbol{\gamma} \in C_{dis}$. Then there is a connected graph with edge set A such that (i) $|A| \leq 2d\|\boldsymbol{\gamma}\|$ and (ii) $\boldsymbol{\gamma}$ is the outermost contour in $\partial \boldsymbol{A}$.*

We now establish a similar way to construct an ordered contour from a small edge set. The situation is slightly different due to the differences between ordered and disordered contours in Lemma 3.4. Define, for $\boldsymbol{\gamma} \in C_{\text{ord}}$, $\Lambda = \text{Int}\,\boldsymbol{\gamma} \cap \mathbb{T}_n^d$,

$$\mathcal{E}_{\boldsymbol{\gamma}} \coloneqq \{\{i, j\} \mid i, j \in \Lambda\}.$$

Lemma 3.7. *Suppose $\boldsymbol{\gamma} \in C_{ord}$ and $F \subseteq \mathcal{E}_{\boldsymbol{\gamma}}$. Let $A = (E \setminus E'(\Lambda)) \cup F$, where $E'(\Lambda)$ is defined as in Lemma 3.4. Let Γ be the set of contours in $\partial \boldsymbol{A}$. Then $\boldsymbol{\gamma} \in \Gamma$, and for all $\boldsymbol{\gamma}' \in \Gamma$ with $\boldsymbol{\gamma}' \neq \boldsymbol{\gamma}$ we have $\boldsymbol{\gamma}' < \boldsymbol{\gamma}$. Moreover, all sets of matching contours consisting of γ and contours in Int γ arise from such F.*

Two edges $e, f \in E$ are called 1-*adjacent* if $d_\infty(\boldsymbol{e}, \boldsymbol{f}) \leq 1$. A set of edges A is 1-*connected* if for any $e, f \in A$, there is a sequence of 1-adjacent edges in A from e to f. In the next lemma, $\partial \boldsymbol{A}^{\boldsymbol{c}}$ is the boundary of the thickening of the edge set $A^c = E \setminus A$.

Lemma 3.8. *Suppose $\boldsymbol{\gamma} \in C_{ord}$. Then there is a 1-connected set of edges A of size at most $\|\boldsymbol{\gamma}\|$ such that $\boldsymbol{\gamma}$ is the outermost contour in $\partial \boldsymbol{A}^{\boldsymbol{c}}$.*

Proof. Let A be the set of all edges that intersect $\boldsymbol{\gamma}$. By the definition of $\|\cdot\|$, $|A| \leq \|\boldsymbol{\gamma}\|$. By Lemma 3.7 $\boldsymbol{\gamma}$ is the outermost contour in $\boldsymbol{A}^{\boldsymbol{c}}$, as $A^c = E'(\Lambda) \cup \mathcal{E}_{\text{dis}}(\Lambda)$. The 1-connectedness of A

follows from the connectedness of $\boldsymbol{\gamma}$ and the observation that every point of $\boldsymbol{\gamma}$ is at most d_∞ distance $1/2$ from an edge in A. □

3.6 Contour Enumeration

This section uses the results of the previous subsection to guarantee the existence of an efficient algorithm for enumerating contours. This requires a few additional lemmas.

Lemma 3.9. *For all $\boldsymbol{\gamma} \in C$, $\left|Int\,\boldsymbol{\gamma}\right| \le \|\boldsymbol{\gamma}\|^2$, and $\left|Int\,\boldsymbol{\gamma}\right| \le (n/2)\|\boldsymbol{\gamma}\|$.*

Proof. This follows by [8, Lemma 5.7], as the interior of a contour as defined by Definition 5 is always smaller than the definition of the interior of a contour in [8]. □

Lemma 3.10. *There is an algorithm that determines the vertex set $Int\,\boldsymbol{\gamma} \cap \mathbb{T}_n^d$ in time $O(\|\boldsymbol{\gamma}\|^3)$.*

Lemma 3.11. *Fix an edge $e \in E$. There is an algorithm to construct all contours $\boldsymbol{\gamma} \in C_{ord}$ that (i) can arise from a connected edge set A that contains e and (ii) have $\|\boldsymbol{\gamma}\| \le m$. The algorithm runs in time* $\exp(O(m))$.

Similarly, there is an $\exp(O(m))$-time algorithm to construct all contours $\boldsymbol{\gamma} \in C_{dis}$ that (i) can arise from an edge set A such that A^c is 1-connected and contains e and (ii) have $\|\boldsymbol{\gamma}\| \le m$.

The next definition is useful for inductive arguments.

Definition 6. *The* level $\mathcal{L}(\boldsymbol{\gamma})$ *of a contour $\boldsymbol{\gamma}$ is defined inductively as follows. If $\boldsymbol{\gamma}$ is* thin, *meaning $C(Int\,\boldsymbol{\gamma}) = \emptyset$, then $\mathcal{L}(\boldsymbol{\gamma}) = 0$. Otherwise, $\mathcal{L}(\boldsymbol{\gamma}) = 1 + \max\{\mathcal{L}(\boldsymbol{\gamma}') \mid \boldsymbol{\gamma}' < \boldsymbol{\gamma}\}$.*

Call a set $\Lambda \subseteq \boldsymbol{T}_n^d$ a *region* if $\Lambda = \boldsymbol{T}_n^d$ or if Λ is a connected component of $\boldsymbol{T}_n^d \setminus \partial \boldsymbol{A}$ for some $A \subset E$. In the former case set $\partial\Lambda = \emptyset$, and in the latter case set $\partial\Lambda$ to be the union of all connected components of $\partial \boldsymbol{A}$ incident to Λ. In particular if $\Lambda = \mathrm{Int}\,\boldsymbol{\gamma}$ for some contour $\boldsymbol{\gamma}$, then Λ is a region and $\partial\Lambda = \boldsymbol{\gamma}$. Finally, for compatible contours $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_t$, define $\|\boldsymbol{\gamma}_1 \cup \cdots \cup \boldsymbol{\gamma}_t\| = \|\boldsymbol{\gamma}_1\| + \cdots + \|\boldsymbol{\gamma}_t\|$. We conclude this subsection by stating our main algorithmic result on efficiently computing sets of contours.

Proposition 3.12. *There is an $O((|\Lambda|+\|\partial\Lambda\|)\exp(O(m)))$-time algorithm that, for all regions Λ, (i) enumerates all contours in $C_{ord}(\Lambda) \cup C_{dis}(\Lambda)$ with size at most m and (ii) sorts this list consistent with the level assignments.*

3.7 Polymer Representations for $Z_{\mathbf{ord}}$ and $Z_{\mathbf{dis}}$

To obtain polymer model representations of Z_{ord} and Z_{dis}, define $\tilde{\Omega}_{\mathrm{ord}}(\Lambda)$ and $\tilde{\Omega}_{\mathrm{dis}}(\Lambda)$ to be the sets of compatible collections of contours in Λ that are labelled ord and dis, respectively. Define

$$K_{\mathrm{ord}}(\boldsymbol{\gamma}) = e^{-\kappa\|\boldsymbol{\gamma}\|}\frac{Z_{\mathrm{dis}}(\mathrm{Int}\,\boldsymbol{\gamma})}{Z_{\mathrm{ord}}(\mathrm{Int}\,\boldsymbol{\gamma})}, \qquad K_{\mathrm{dis}}(\boldsymbol{\gamma}) = e^{-\kappa\|\boldsymbol{\gamma}\|}\frac{qZ_{\mathrm{ord}}(\mathrm{Int}\,\boldsymbol{\gamma})}{Z_{\mathrm{dis}}(\mathrm{Int}\,\boldsymbol{\gamma})}.$$

By following a well-trodden path in Pirogov–Sinai theory (see, e.g., [8, p.28] or [26, p.28]), these definitions give the following representations for Z_{ord} and Z_{dis} as partition functions of abstract polymer models:

$$Z_{\mathrm{ord}}(\Lambda) = e^{-e_{\mathrm{ord}}|\Lambda|}\sum_{\Gamma\in\tilde{\Omega}_{\mathrm{ord}}(\Lambda)}\prod_{\boldsymbol{\gamma}\in\Gamma}K_{\mathrm{ord}}(\boldsymbol{\gamma}) \tag{15}$$

$$Z_{\mathrm{dis}}(\Lambda) = e^{-e_{\mathrm{dis}}|\Lambda|}\sum_{\Gamma\in\tilde{\Omega}_{\mathrm{dis}}(\Lambda)}\prod_{\boldsymbol{\gamma}\in\Gamma}K_{\mathrm{dis}}(\boldsymbol{\gamma})\,. \tag{16}$$

where the sums run over collections of compatible labelled contours in Λ with external label ord and dis, respectively.

In fact, for $\ell \in \{\mathrm{ord}, \mathrm{dis}\}$, the above formulas represent $Z_\ell(\Lambda)$ as the partition function of a polymer model in the form discussed in Section 2, i.e., where polymers are subgraphs of a fixed graph G with bounded degree. In detail, recalling the discussion in Section 3.1, we consider contours as subgraphs of (a subgraph of) the bounded-degree graph $(\frac{1}{2}\mathbb{T}_n^d)^\star$. Thus $|\boldsymbol{\gamma}|$ is the number of vertices in a contour when represented as a subgraph. Condition (5) holds with $b = 1$ since $\|\boldsymbol{\gamma}\| \ge |\gamma|$ by (8). The more substantial hypothesis (6) will be verified in later sections for appropriate choices of the label and of β.

In the sequel we will write $|\Lambda|_{(\frac{1}{2}\mathbb{T}_n^d)^\star}$ for the size of set of vertices of $(\frac{1}{2}\mathbb{T}_n^d)^\star$ that are part of some contour $\boldsymbol{\gamma}$ in $C_\ell(\Lambda)$ for some ℓ. The next technical lemma shows it is enough to find algorithms that are polynomial time in $|\Lambda|_{(\frac{1}{2}\mathbb{T}_n^d)^\star}$.

Lemma 3.13. *For all Λ, $|\Lambda|_{(\frac{1}{2}\mathbb{T}_n^d)^\star}$ is polynomial in $|\Lambda|$.*

4 CONTOUR MODEL ESTIMATES

In this section we state several estimates related to the contour representations from the previous section.

Lemma 4.1 (Lemma 6.1 (a) [8]). *There are constants $c > 0$, $q_0 = q_0(d) < \infty$, and $n_0 < \infty$ such that if $q \ge q_0$, $n \ge n_0$, and $\beta \ge \beta_c$,*

$$\frac{Z_{tunnel}}{Z} \le \exp(-c\beta n^{d-1}).$$

In what follows c will always denote the constant from Lemma 4.1, and q_0 and n_0 will always be at least as large as the constants in the lemma. Lemma 4.1 ensures that Z_{tunnel} is neglectable when approximating Z up to relative errors $\epsilon \gg \exp(-c\beta n^{d-1})$. We will also need to know that Z_{dis} is neglectable when $\beta > \beta_c$. This requires two lemmas.

Lemma 4.2. *If $q \ge q_0$, $n \ge n_0$, and $\beta > \beta_c$ there exist $a_{dis} > 0$ and $f > 0$ so that if $\epsilon_n := 2\exp(-c\beta n)$, then*

$$Z_{ord} \ge \exp(-(f+\epsilon_n)n^d),$$

$$Z_{dis} \le \exp((-f+\epsilon_n)n^d)\max_{\Gamma\in\mathcal{G}_{dis}^{ext}} e^{-\frac{a_{dis}}{2}|Ext\,\Gamma|}\prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\frac{c}{2}\beta\|\boldsymbol{\gamma}\|},$$

Proof. With $a_{\mathrm{dis}} \ge 0$ this follows from [8, Lemma 6.3] provided $f = f_{\mathrm{ord}}$ for $\beta \ge \beta_c$, and that $f = f_{\mathrm{ord}}$ follows from [8, Lemma A.3]. What remains is to prove $a_{\mathrm{dis}} > 0$ when $\beta > \beta_c$. The results of [32] imply that there is a unique Gibbs measure for the random cluster model when $\beta > \beta_c$. If a_{dis} was 0 for some $\beta > \beta_c$, then the argument establishing [8, Lemma 6.1 (b)] implies the existence of multiple Gibbs measures, a contradiction. □

Lemma 4.3. *If $q \ge q_0$, $n \ge n_0$, and $\beta > \beta_c$, then there exists a constant $b_{dis} > 0$ so that*

$$\frac{Z_{dis}}{Z} \le 2\exp(-b_{dis}n^{d-1})\,.$$

Proof. Suppose $\Gamma \in \mathcal{G}_{\mathrm{dis}}^{\mathrm{ext}}$. Then we claim that

$$|\mathrm{Ext}\,\Gamma| + \sum_{\boldsymbol{\gamma}\in\Gamma}\|\boldsymbol{\gamma}\| \ge 2n^{d-1}. \tag{17}$$

To see this, note that

$$|\mathrm{Ext}\,\Gamma| + \sum_{\boldsymbol{\gamma}\in\Gamma} |\mathrm{Int}\,\boldsymbol{\gamma}| = n^d,$$

which combined with Lemma 3.9 implies

$$|\mathrm{Ext}\,\Gamma| + \frac{n}{2}\sum_{\boldsymbol{\gamma}\in\Gamma} \|\boldsymbol{\gamma}\| \geq n^d$$

which implies (17) when $n \geq 2$.

By Lemma 4.2, if n is large enough,

$$\frac{Z_{\mathrm{dis}}}{Z_{\mathrm{ord}}} \leq 2 \max_{\Gamma\in\mathcal{G}^{\mathrm{ext}}_{\mathrm{dis}}} e^{-\frac{a_{\mathrm{dis}}}{2}|\mathrm{Ext}\,\Gamma|} \prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\frac{c}{2}\beta\|\boldsymbol{\gamma}\|}. \tag{18}$$

Set $b_{\mathrm{dis}} := \min\{a_{\mathrm{dis}}, c\beta\} > 0$. By (17),

$$e^{-\frac{a_{\mathrm{dis}}}{2}|\mathrm{Ext}\,\Gamma|} \prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\frac{c}{2}\beta\|\boldsymbol{\gamma}\|} \leq \exp(-b_{\mathrm{dis}} n^{d-1})$$

for all $\Gamma \in \mathcal{G}^{\mathrm{ext}}_{\mathrm{dis}}$. The lemma now follows from (18). □

The next two lemmas will allow us to verify the Kotecký–Preiss condition for the contour models defined in the previous section.

Lemma 4.4 (Lemma 6.3 [8]). *If $q \geq q_0$ and $\beta = \beta_c$, then*

$$K_{ord}(\boldsymbol{\gamma}) \leq e^{-c\beta\|\boldsymbol{\gamma}\|}, \quad \text{and} \quad K_{dis}(\boldsymbol{\gamma}) \leq e^{-c\beta\|\boldsymbol{\gamma}\|},$$

for all $\boldsymbol{\gamma}$ in C_{ord} and C_{dis}, respectively.

Lemma 4.5 (Lemma 6.3 [8]). *If $q \geq q_0$ and $\beta > \beta_c$, then*

$$K_{ord}(\boldsymbol{\gamma}) \leq e^{-c\beta\|\boldsymbol{\gamma}\|}, \qquad \boldsymbol{\gamma} \in C_{ord}$$

In particular, since $\beta \geq \frac{3\log q}{d}$, then for sufficiently large q the contour weights K_{ord} (for $\beta \geq \beta_c$) and K_{dis} (for $\beta = \beta_c$) will satisfy condition (6).

Next we will show that when $\beta > \beta_c$ and the disordered ground state is unstable, that regions with disordered boundary conditions 'flip' quickly to ordered regions by way of a large contour; more precisely, the dominant contribution to $Z_{\mathrm{dis}}(\boldsymbol{\Lambda})$ from collections of contours with small external volume.

For a region $\boldsymbol{\Lambda}$ and $M > 0$ we define

$$\mathcal{H}^{\mathrm{flip}}_{\mathrm{dis}}(\boldsymbol{\Lambda}, M) := \{\Gamma \in \mathcal{G}^{\mathrm{ext}}_{\mathrm{dis}}(\boldsymbol{\Lambda}) \mid |\mathrm{Ext}\,\Gamma \cap \boldsymbol{\Lambda}| \leq M\},$$

and

$$Z^{\mathrm{flip}}_{\mathrm{dis}}(\boldsymbol{\Lambda}, M) := \sum_{\Gamma\in\mathcal{H}^{\mathrm{flip}}_{\mathrm{dis}}(\boldsymbol{\Lambda},M)} e^{-e_{\mathrm{dis}}|\mathrm{Ext}\,\Gamma\cap\boldsymbol{\Lambda}|} \prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|} q Z_{\mathrm{ord}}(\mathrm{Int}\,\boldsymbol{\gamma}).$$

Thus, c.f. (13), $Z^{\mathrm{flip}}_{\mathrm{dis}}(\boldsymbol{\Lambda}, M)$ is the contribution to $Z_{\mathrm{dis}}(\boldsymbol{\Lambda})$ from contour configurations with small exterior volume.

Lemma 4.6. *Suppose $q \geq q_0$ and $\beta > \beta_c$. Then there exists $a_{dis} > 0$ so that the following holds for all $n \geq n_0$. Suppose $\boldsymbol{\gamma} \in C_{ord}$. For any $\epsilon > 0$, if*

$$M \geq \frac{2}{a_{dis}} \log\frac{8q}{\epsilon} + \frac{2}{a_{dis}}(\kappa+3)\|\boldsymbol{\gamma}\|$$

then $Z^{flip}_{dis}(Int\,\boldsymbol{\gamma}, M)$ is an ϵ-relative approximation to $Z_{dis}(Int\,\boldsymbol{\gamma})$.

We end this section with a lemma concerning $\mathcal{H}^{\mathrm{flip}}_{\mathrm{dis}}$.

Proposition 4.7. *There is an algorithm that given $\boldsymbol{\gamma} \in C_{ord}$ and $M \in \mathbb{N}$ outputs $\mathcal{H}^{flip}_{dis}(Int\,\boldsymbol{\gamma}, M)$ in time $\|\boldsymbol{\gamma}\| e^{O(\|\boldsymbol{\gamma}\|+M)}$.*

5 APPROXIMATE COUNTING ALGORITHMS

This section describes our approximate counting algorithms for $\beta > \beta_h$. The algorithms differ depending on whether $\beta = \beta_c$, $\beta > \beta_c$, or $\beta_h < \beta < \beta_c$. Recall that $Z_\ell(\boldsymbol{\Lambda})$ was defined for all regions $\boldsymbol{\Lambda}$ in (12)–(13). The heart of this section is the following lemma.

Lemma 5.1. *For $d \geq 2$ and $q \geq q_0$ the following hold.*

1. *If $\beta = \beta_c$ there is an FPTAS to approximate $Z_{ord}(\boldsymbol{\Lambda})$ and $Z_{dis}(\boldsymbol{\Lambda})$.*
2. *If $\beta > \beta_c$ there is an FPTAS to approximate $Z_{ord}(\boldsymbol{\Lambda})$.*
3. *If $\beta_h < \beta < \beta_c$ there is an FPTAS to approximate $Z_{dis}(\boldsymbol{\Lambda})$.*

In each case the FPTAS applies to any region $\boldsymbol{\Lambda}$, with running time polynomial in $|\boldsymbol{\Lambda}|$, the number of vertices of $\mathbb{T}^d_n$ in $\boldsymbol{\Lambda}$.

Sections 5.1 and 5.2 prove the first two cases of Lemma 5.1. The case $\beta_h < \beta < \beta_c$ is very similar to $\beta > \beta_c$, and we defer the details to [7]. In Section 5.3 we show how these results, together with a result from [8], suffice to give an FPRAS for Z on the torus.

5.1 Proof of Lemma 5.1, $\beta = \beta_c$

We begin by defining a useful variant of the truncated cluster expansion for $Z_{\mathrm{ord}}(\boldsymbol{\Lambda})$ and $Z_{\mathrm{dis}}(\boldsymbol{\Lambda})$. Let K be a function from contours to positive real numbers. For $\ell \in \{\mathrm{ord}, \mathrm{dis}\}$ define

$$T_{\ell,m}(\boldsymbol{\Lambda}, K) := \sum_{\substack{\Gamma\in\mathcal{G}^c_\ell(\boldsymbol{\Lambda}) \\ \|\Gamma\|<m}} \phi(\Gamma) \prod_{\boldsymbol{\gamma}\in\Gamma} K(\boldsymbol{\gamma}).$$

so that by (15) and (16) $Z_\ell(\Lambda) = \exp(-e_\ell|\Lambda|)T_{\ell,\infty}(\boldsymbol{\Lambda}, K_\ell)$ provided the cluster expansion for the polymer models converge.

Recall that the level of a contour was defined in Definition 6, and that $|\boldsymbol{\Lambda}|_{(\frac{1}{2}\mathbb{T}^d_n)^\star}$ was defined immediately prior to Lemma 3.13.

Lemma 5.2. *Suppose $d \geq 2$, $q \geq q_0$ and $\beta = \beta_c$. Given $\boldsymbol{\Lambda}$ with $|\boldsymbol{\Lambda}|_{(\frac{1}{2}\mathbb{T}^d_n)^\star} = N$, and an error parameter $\epsilon > 0$, let $m = \log(8N^2/\epsilon)/3$. Inductively (by level) define weights $\tilde{K}_{ord}(\boldsymbol{\gamma})$ and $\tilde{K}_{dis}(\boldsymbol{\gamma})$ for all contours $\boldsymbol{\gamma}$ in $C_{ord}(\boldsymbol{\Lambda})$ and $C_{dis}(\boldsymbol{\Lambda})$ with size $\|\boldsymbol{\gamma}\| \leq m$ by:*

1. *If $\boldsymbol{\gamma}$ is thin, then set*

$$\tilde{K}_{ord}(\boldsymbol{\gamma}) = e^{-\kappa\|\boldsymbol{\gamma}\|-(e_{dis}-e_{ord})|Int\,\boldsymbol{\gamma}|},$$
$$\tilde{K}_{dis}(\boldsymbol{\gamma}) = qe^{-\kappa\|\boldsymbol{\gamma}\|-(e_{ord}-e_{dis})|Int\,\boldsymbol{\gamma}|}.$$

2. *If $\boldsymbol{\gamma}$ is not thin, then define $\tilde{K}_{ord}$ by*

$$e^{-\kappa\|\boldsymbol{\gamma}\|-(e_{dis}-e_{ord})|Int\,\boldsymbol{\gamma}|} \exp\left[T_{m,dis}(Int\,\boldsymbol{\gamma}, \tilde{K}) - T_{m,ord}(Int\,\boldsymbol{\gamma}, \tilde{K})\right],$$

and $\tilde{K}_{dis}$ by

$$qe^{-\kappa\|\boldsymbol{\gamma}\|-(e_{ord}-e_{dis})|Int\,\boldsymbol{\gamma}|} \exp\left[T_{m,ord}(Int\,\boldsymbol{\gamma}, \tilde{K}) - T_{m,dis}(Int\,\boldsymbol{\gamma}, \tilde{K})\right].$$

Then for N sufficiently large $e^{-e_\ell|\Lambda|}\exp(T_{\ell,m}(\boldsymbol{\Lambda}, \tilde{K}_\ell))$ is an ϵ-relative approximation to $Z_\ell(\boldsymbol{\Lambda})$ for $\ell \in \{ord, dis\}$.

Proof. Suppose $\ell \in \{\mathrm{dis}, \mathrm{ord}\}$. First note that the inductive definition of the weights $\tilde{K}_\ell(\boldsymbol{\gamma})$ makes sense: to compute $\tilde{K}_\ell(\boldsymbol{\gamma})$ for a contour $\boldsymbol{\gamma}$ of level $t+1$ only requires knowing $\tilde{K}_\ell(\boldsymbol{\gamma}')$ for contours $\boldsymbol{\gamma}'$ of level t and smaller.

Since $\beta = \beta_c$ and $q \geq q_0$, Lemma 4.4 tells us that

$$K_\ell(\boldsymbol{\gamma}) \leq e^{-c\beta\|\boldsymbol{\gamma}\|} \tag{19}$$

for $\ell \in \{\text{dis}, \text{ord}\}$ and for all $\boldsymbol{\gamma} \in C_\ell(\Lambda)$. If q_0 is large enough then (19) implies condition (??) holds since β_c grows like $\log q$ by (1). Thus by Section 3.7 the hypotheses of Lemma 2.1 are satisfied and the cluster expansion for $Z_\ell(\boldsymbol{\Lambda})$ converges for $\ell \in \{\text{ord}, \text{dis}\}$.

Now let $\epsilon' = \epsilon/N$, so that $m = \log(8N/\epsilon')/3$. We will apply Lemma 2.3 with $v(\boldsymbol{\gamma}) = |\text{Int}\,\boldsymbol{\gamma}|$. This is a valid choice of $v(\boldsymbol{\gamma})$ by Lemma 3.9. Lemma 2.3 says that

$$e^{-e_{\text{ord}}|\Lambda|} \exp\left(T_{\text{ord},m}(\boldsymbol{\Lambda}, \tilde{K}_{\text{ord}})\right)$$

and

$$e^{-e_{\text{dis}}|\Lambda|} \exp\left(T_{\text{dis},m}(\boldsymbol{\Lambda}, \tilde{K}_{\text{dis}})\right)$$

are ϵ-relative approximations to $Z_{\text{ord}}(\boldsymbol{\Lambda})$ and $Z_{\text{dis}}(\boldsymbol{\Lambda})$ if for all $\boldsymbol{\gamma} \in C_\ell(\boldsymbol{\Lambda})$ of size at most m, $\tilde{K}_\ell(\boldsymbol{\gamma})$ is an $\epsilon'|\text{Int}\,\boldsymbol{\gamma}|$-relative approximation to $K_\ell(\boldsymbol{\gamma})$. We will prove this by induction on the level of $\boldsymbol{\gamma}$.

For a thin contour, $\tilde{K}_\ell(\boldsymbol{\gamma}) = K_\ell(\boldsymbol{\gamma})$. Now suppose that for all contours $\boldsymbol{\gamma}$ of level at most t and size at most m, $\tilde{K}_\ell(\boldsymbol{\gamma})$ is an $\epsilon'|\text{Int}\,\boldsymbol{\gamma}|$-relative approximation of $K_\ell(\boldsymbol{\gamma})$. Consider a contour $\boldsymbol{\gamma}$ of level $t+1$ and size at most m. Then all contours $\boldsymbol{\gamma}'$ that appear in the expansions

$$T_{m,\text{dis}}(\text{Int}\,\boldsymbol{\gamma}, \tilde{K}_{\text{dis}}) \quad \text{and} \quad T_{m,\text{ord}}(\text{Int}\,\boldsymbol{\gamma}, \tilde{K}_{\text{ord}})$$

are of level at most t and size at most m, and so for each such $\boldsymbol{\gamma}'$, by the inductive hypothesis $\tilde{K}_\ell(\boldsymbol{\gamma}')$ is an $\epsilon'|\text{Int}\,\boldsymbol{\gamma}'|$-relative approximation to $K_\ell(\boldsymbol{\gamma}')$. Then by Lemma 2.3, we have that

$$e^{-(e_{\text{dis}}-e_{\text{ord}})|\text{Int}\,\boldsymbol{\gamma}|} \exp\left[T_{m,\text{dis}}(\text{Int}\,\boldsymbol{\gamma}, \tilde{K}_{\text{dis}}) - T_{m,\text{ord}}(\text{Int}\,\boldsymbol{\gamma}, \tilde{K}_{\text{ord}})\right]$$

is an $|\text{Int}\,\boldsymbol{\gamma}|\epsilon'$-relative approximation to $\frac{Z_{\text{dis}}(\text{Int}\,\boldsymbol{\gamma})}{Z_{\text{ord}}(\text{Int}\,\boldsymbol{\gamma})}$ (and likewise for dis and ord swapped). Multiplying by the prefactor $e^{-\kappa\|\boldsymbol{\gamma}\|}$ for ord and by $qe^{-\kappa\|\boldsymbol{\gamma}\|}$ for dis shows that $\tilde{K}_\ell(\boldsymbol{\gamma})$ is an $\epsilon'|\text{Int}\,\boldsymbol{\gamma}|$-relative approximation to $K_\ell(\boldsymbol{\gamma})$ as desired. □

Proof of Lemma 5.1 when $\beta = \beta_c$. Let $N = |\boldsymbol{\Lambda}|_{(\frac{1}{2}\mathbb{T}_n^d)^\star}$ and let $m = \log(8N^2/\epsilon)/3$. We need to show that the expansion $T_{\ell,m}(\boldsymbol{\Lambda}, \tilde{K}_\ell)$ and the weights $\tilde{K}_\ell(\boldsymbol{\gamma})$ for all $\boldsymbol{\gamma}$ of size at most m in $C_\ell(\boldsymbol{\Lambda})$ can be computed in time polynomial in N and $1/\epsilon$ for $\ell \in \{\text{dis}, \text{ord}\}$. We can list the sets of contours in $C_{\text{ord}}(\boldsymbol{\Lambda})$ and $C_{\text{dis}}(\boldsymbol{\Lambda})$ of size at most m, together with their labels and levels, in time $O(N\exp(O(m)))$ by Proposition 3.12. Since $m = \log(8N^2/\epsilon)/3$, $O(N\exp(O(m)))$ is polynomial in N and $1/\epsilon$. N itself is polynomial in $|\Lambda|$ by Lemma 3.13.

To prove the lemma we must compute the weights $\tilde{K}_\ell(\boldsymbol{\gamma})$ and the truncated cluster expansions $T_{m,\ell}(\text{Int}\,\boldsymbol{\gamma}, \tilde{K}_\ell)$ for each contour in the list. We do this inductively by level. For level zero contours $\tilde{K}_\ell(\boldsymbol{\gamma}) = K_\ell(\boldsymbol{\gamma})$ only depends on $\|\boldsymbol{\gamma}\|$ and $|\text{Int}\,\boldsymbol{\gamma}|$, so $\tilde{K}_\ell(\boldsymbol{\gamma})$ can be computed in time $O(\|\boldsymbol{\gamma}\|^3)$ by computing these quantities by using Lemma 3.10. We then continue inductively; each $\tilde{K}_\ell(\boldsymbol{\gamma})$ can be computed efficiently since the truncated cluster expansions can be computed in time polynomial in N and $1/\epsilon$ using Lemma 2.2. □

5.2 Proof of Lemma 5.1, $\beta > \beta_c$

When $\beta > \beta_c(q,d)$ the ordered ground state is stable, but the disordered state is unstable. For a definition of stability of ground states, see, e.g., [9]; the upshot for this paper is that we cannot use the cluster expansion to approximate $Z_{\text{dis}}(\boldsymbol{\Lambda})$ for a region $\boldsymbol{\Lambda}$.

To deal with this complication we will appeal to Lemma 4.6. In words, this lemma says that for $\beta > \beta_c$, a typical contour configuration in a region with disordered boundary conditions will have very few external vertices. We will exploit this fact to enumerate all sets of typical external contours in the region. This is possible since the number of external vertices is small. Once we have fixed a set of external contours we are back to the task of approximating partition functions with ordered boundary conditions.

We now make the preceding discussion precise. Given $K\colon C_{\text{ord}}(\boldsymbol{\Lambda}) \to [0,\infty)$, define $\Xi^M_{\text{dis}}(\boldsymbol{\Lambda}, K)$ to be

$$e^{e_{\text{dis}}|\Lambda|} \sum_{\Gamma \in \mathcal{H}^{\text{flip}}_{\text{dis}}(\boldsymbol{\Lambda}, M)} e^{-e_{\text{dis}}|\text{Ext}\,\Gamma|} \prod_{\boldsymbol{\gamma}\in\Gamma} e^{-\kappa\|\boldsymbol{\gamma}\|} q \exp\left[T_{m,\text{ord}}(\text{Int}\,\boldsymbol{\gamma}, K)\right].$$

Lemma 5.3. *Suppose $d \geq 2$, $q \geq q_0$ and $\beta > \beta_c$. Let $\boldsymbol{\Lambda}$ be a region with $|\boldsymbol{\Lambda}|_{(\frac{1}{2}\mathbb{T}_n^d)^\star} = N$, fix $\epsilon > 0$, and let $m = \log(8N^2/\epsilon)/3$. Inductively (by level) define $\tilde{K}_{ord}(\boldsymbol{\gamma})$ for $\boldsymbol{\gamma} \in C_{ord}(\boldsymbol{\Lambda})$ with size $\|\boldsymbol{\gamma}\|$ at most m by*

(1) If $\boldsymbol{\gamma}$ is thin, then

$$\tilde{K}_{ord}(\boldsymbol{\gamma}) = e^{-\kappa\|\boldsymbol{\gamma}\| - (e_{dis}-e_{ord})|Int\,\boldsymbol{\gamma}|}.$$

(2) If $\boldsymbol{\gamma}$ is not thin, define $\tilde{K}_{ord}$ by

$$e^{-\kappa\|\boldsymbol{\gamma}\| - (e_{dis}-e_{ord})|Int\,\boldsymbol{\gamma}|} \exp\left[-T_{m,ord}(Int\,\boldsymbol{\gamma}, \tilde{K})\right] \Xi^M_{dis}(Int\,\boldsymbol{\gamma}, \tilde{K}_{ord}),$$

with $M = \frac{2}{a_{dis}}\left(\log(\frac{32q}{\epsilon'}) + (\kappa+3)m\right)$.

Then for all N large enough, $e^{-e_{ord}|\Lambda|}\exp\left(T_{ord,m}(\boldsymbol{\Lambda}, \tilde{K}_{ord})\right)$ is an ϵ-relative approximation to $Z_{ord}(\boldsymbol{\Lambda})$.

5.3 Proof of Theorem 1.1

To prove Theorem 1.1 we will need the following result from [8] about the mixing time of the Glauber dynamics.

Theorem 5.4 ([8, Theorem 1.1]). *The mixing time of the Glauber dynamics for the q-state ferromagnetic Potts model satisfies*

$$\tau_{q,\beta}(\mathbb{T}_n^d) = e^{O(n^{d-1})},$$

where the $O(\cdot)$ in the exponent hides constants that depend on q, β.

We will use this result to give an approximation algorithm when the approximation parameter ϵ is extremely small. The reason we are able to combine the Glauber dynamics with our contour-based algorithm to give an FPRAS is that [8] proves *optimal* slow mixing results for the Glauber and Swendsen–Wang dynamics. That is, up to a constant in the exponent, the upper bound of the mixing time of the Glauber dynamics (or Swendsen–Wang dynamics) is the inverse of the bound on Z_{tunnel}/Z from Lemma 4.1. Thus when ϵ is too small for the contour algorithms to work, the Glauber dynamics can take over.

Proof of Theorem 1.1. Let $N = n^d$ be the number of vertices of $\mathbb{T}_n^d$. We will use a simple fact several times below: if $\epsilon \in (0,1)$, $Z, Z^* > 0$, and $Z^*/Z < \epsilon/2$, then $(Z - Z^*)$ is an ϵ-relative approximation to Z.

We first consider the case $\beta = \beta_c$. To give an FPRAS for $Z = Z_{\mathbb{T}_n^d}$ we consider two subcases. Let c be the constant from Lemma 4.1.

Suppose $\epsilon < 4e^{-c\beta n^{d-1}}$. Since $e^{O(n^{d-1})}$ is polynomial in N and $1/\epsilon$, we can use Glauber dynamics to obtain an ϵ-approximate sample in polynomial time. By using simulated annealing (e.g. [41])

we can also approximate the partition function in time polynomial in N and $1/\epsilon$.

If $\epsilon \geq 4e^{-c\beta n^{d-1}}$, then by Lemma 4.1, $Z_{\text{rest}} = Z_{\text{dis}} + Z_{\text{ord}}$ is an $\epsilon/2$-relative approximation to Z, so it suffices to find an $\epsilon/4$-relative approximation to both Z_{dis} and Z_{ord}. This can be done in time polynomial in N and $1/\epsilon$ by Lemma 5.1.

Next we consider the case $\beta > \beta_c$. Again there are two subcases. Let c be the constant from Lemma 4.1 as before, and let b_{dis} be the constant from Lemma 4.3. If $\epsilon < 4e^{-c\beta n^{d-1}} + 4e^{-b_{\text{dis}} n^{d-1}}$, then again $e^{O(n^{d-1})}$ is polynomial in N and $1/\epsilon$ and we can approximately count and sample by using the Glauber dynamics.

If $\epsilon \geq 4e^{-c\beta n^{d-1}} + 4e^{-b_{\text{dis}} n^{d-1}}$, then by Lemma 4.1 and Lemma 4.3, Z_{ord} is an $\epsilon/2$-relative approximation to Z and so it suffices to give an $\epsilon/2$-relative approximation to Z_{ord}. This can be done in time polynomial in N and $1/\epsilon$ by Lemma 5.1.

Lastly, consider $\beta < \beta_c$. The case $\beta \leq \beta_h$ was completed in Section 2. The case $\beta_h < \beta < \beta_c$ is done exactly as the case $\beta > \beta_c$ with the roles of ord and dis reversed; see [7] for details. □

6 CONCLUSIONS

In this paper we have given efficient approximate counting and sampling algorithms for the random cluster and q-state Potts models on $\mathbb{Z}^d$ at all inverse temperatures $\beta \geq 0$, provided $q \geq q_0(d)$ and $d \geq 2$. We believe the ideas of this paper will, however, allow for approximate counting and sampling algorithms to be developed for a much broader class of statistical mechanics models.

Sufficient conditions to implement Pirogov–Sinai theory for a given model are that there are only finitely many ground states and that there is 'sufficient τ-functionality', see [9]; the ideas of this paper show that these conditions also suffice for the development of efficient algorithms. In particular, our methods allow for the presence of unstable ground states, a significant improvement compared to the algorithms in [26].

Our results suggest that the algorithmic tasks of counting and sampling may be performed efficiently for a fairly broad class of statistical mechanics models with first-order phase transitions, but we leave a fuller investigation of this for future work. A related interesting question is the existence of efficient algorithms for all $\beta \geq \beta_c$ in the presence of a second-order transition; we are not aware of any results in this direction with the exception of the Ising model, i.e., the $q = 2$ state Potts model [25, 28]. To conclude we list some further open questions related to this paper.

(1) Our algorithms are restricted to $q \geq q_0(d)$ with $q_0(d) > \exp(25d \log d)$. Do efficient algorithms exist that avoid this constraint? Since the physical phenomena behind our results are believed to hold for $q \geq 3$ when $d \geq 3$, there is likely room for improvement.
(2) On the torus, we obtained an FPRAS (as opposed to an FPTAS) for the partition function because of the estimate on Z_{tunnel} from Lemma 4.1: the contribution of Z_{tunnel} cannot be ignored when $\epsilon \leq \exp(-\Omega(n^{d-1}))$. Fortunately, it is exactly when ϵ is this small that the Glauber dynamics mix in time polynomial in $1/\epsilon$, but of course Markov Chain Monte Carlo is a randomized algorithm. A method for systematically accounting for the interfaces that contribute to Z_{tunnel} would likely enable the development of an FPTAS. We leave this as an open problem.
(3) Our algorithms have at least two other features that could be improved. The first is the running time: while our algorithms are polynomial time, the degree of the polynomial is not small. The second is that our algorithms rely on *a priori* knowledge of whether or not $\beta = \beta_c$.
Both of these deficiencies have the potential to be addressed by Glauber-type dynamics as described in [13]; see also [26, Section 7.2]. Proving the efficiency of these proposed algorithms would be very interesting.
(4) Our deterministic algorithms for $\beta > \beta_c$ (and $\beta < \beta_c$) have diverging running times as $\beta \downarrow \beta_c$ ($\beta \uparrow \beta_c$). Are there *deterministic* algorithms that do not suffer from this dependence?
(5) The algorithmic adaptation of other sophisticated contour-based methods, e.g., [37], would be interesting, particularly for applications to problems such as counting the number of proper q-colorings of a graph.

REFERENCES

[1] Dimitris Achlioptas and Amin Coja-Oghlan. 2008. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 793–802.
[2] Kenneth S Alexander. 2004. Mixing properties and exponential decay for lattice systems in finite volumes. *The Annals of Probability* 32, 1A (2004), 441–487.
[3] Alexander Barvinok. 2017. Combinatorics and Complexity of Partition Functions. *Algorithms and Combinatorics* 30 (2017).
[4] Alexander Barvinok and Guus Regts. 2019. Weighted counting of solutions to sparse systems of equations. *Combinatorics, Probability and Computing* 28, 5 (2019), 696–719.
[5] Antonio Blanca and Alistair Sinclair. 2017. Random-cluster dynamics in $\mathbb{Z}^2$. *Probability Theory and Related Fields* 168, 3-4 (2017), 821–847.
[6] Magnus Bordewich, Catherine Greenhill, and Viresh Patel. 2016. Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Structures & Algorithms* 48, 1 (2016), 21–52.
[7] C Borgs, J Chayes, T Helmuth, W Perkins, and P Tetali. 2019. Efficient sampling and counting algorithms for the Potts model on Zd at all temperatures. *arXiv preprint arXiv:1909.09298* (2019).
[8] Christian Borgs, Jennifer T Chayes, and Prasad Tetali. 2012. Tight bounds for mixing of the Swendsen–Wang algorithm at the Potts transition point. *Probability Theory and Related Fields* 152, 3-4 (2012), 509–557.
[9] Christian Borgs and John Z Imbrie. 1989. A unified approach to phase diagrams in field theory and statistical mechanics. *Communications in mathematical physics* 123, 2 (1989), 305–328.
[10] Christian Borgs, Roman Kotecký, and Salvador Miracle-Solé. 1991. Finite-size scaling for Potts models. *Journal of Statistical Physics* 62, 3-4 (1991), 529–551.
[11] Sarah Cannon and Will Perkins. 2019. Counting independent sets in unbalanced bipartite graphs. *arXiv preprint arXiv:1906.01666* (2019).
[12] Katrin Casel, Philipp Fischbeck, Tobias Friedrich, Andreas Göbel, and JA Lagodzinski. 2019. Zeros and approximations of Holant polynomials on the complex plane. *arXiv preprint arXiv:1905.03194* (2019).
[13] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. 2019. Fast algorithms at low temperatures via Markov chains. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[14] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. 2011. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E* 84, 6 (2011), 066106.
[15] Hugo Duminil-Copin. 2017. Lectures on the Ising and Potts models on the hypercubic lattice. *arXiv preprint arXiv:1707.00520* (2017).
[16] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. 2019. Sharp phase transition for the random-cluster and Potts models via decision trees. *Annals of Mathematics* 189, 1 (2019), 75–99.
[17] Sacha Friedli and Yvan Velenik. 2017. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press.
[18] Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. 2011. Improved inapproximability results for counting independent sets in the hard-core model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 567–578.

[19] Andreas Galanis, Daniel Stefankovic, Eric Vigoda, and Linji Yang. 2016. Ferromagnetic Potts model: Refined #-BIS-hardness and related results. *SIAM J. Comput.* 45, 6 (2016), 2004–2065.

[20] David Gamarnik and Madhu Sudan. 2017. Performance of sequential local algorithms for the random NAE-K-SAT problem. *SIAM J. Comput.* 46, 2 (2017), 590–619.

[21] Reza Gheissari and Eyal Lubetzky. 2018. Mixing Times of Critical Two-Dimensional Potts Models. *Communications on Pure and Applied Mathematics* 71, 5 (2018), 994–1046.

[22] Reza Gheissari and Eyal Lubetzky. 2020. Quasi-polynomial mixing of critical two-dimensional random cluster models. *Random Structures & Algorithms* 56, 2 (2020), 517–556.

[23] Leslie Ann Goldberg and Mark Jerrum. 2012. Approximating the partition function of the ferromagnetic Potts model. *J. ACM* 59, 5 (2012), 25.

[24] C Gruber and H Kunz. 1971. General properties of polymer systems. *Communications in Mathematical Physics* 22, 2 (1971), 133–161.

[25] Heng Guo and Mark Jerrum. 2018. Random cluster dynamics for the Ising model is rapidly mixing. *Ann. Appl. Probab.* 28, 2 (04 2018), 1292–1313.

[26] Tyler Helmuth, Will Perkins, and Guus Regts. to appear. An extended abstract appeared at STOC 2019.. Algorithmic Pirogov-Sinai theory. *Probability Theory and Related Fields* (to appear. An extended abstract appeared at STOC 2019.).

[27] Matthew Jenssen, Peter Keevash, and Will Perkins. 2019. Full version at http://arxiv.org/abs/1807.04804v2. Algorithms for #BIS-hard problems on expander graphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2019)*. SIAM, 2235–2247.

[28] Mark Jerrum and Alistair Sinclair. 1993. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on computing* 22, 5 (1993), 1087–1116.

[29] Roman Kotecký and David Preiss. 1986. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics* 103, 3 (1986), 491–498.

[30] Roman Kotecký and SB Shlosman. 1982. First-order phase transitions in large entropy lattice models. *Communications in Mathematical Physics* 83, 4 (1982), 493–515.

[31] Florent Krzakała, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. 2007. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences* 104, 25 (2007), 10318–10323.

[32] Lahoussine Laanait, Alain Messager, Salvador Miracle-Solé, Jean Ruiz, and Senya Shlosman. 1991. Interfaces in the Potts model I: Pirogov-Sinai theory of the Fortuin-Kasteleyn representation. *Communications in Mathematical Physics* 140, 1 (1991), 81–91.

[33] Tsung-Dao Lee and Chen-Ning Yang. 1952. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Physical Review* 87, 3 (1952), 410.

[34] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. 2019. Counting independent sets and colorings on random regular bipartite graphs. *arXiv preprint arXiv:1903.07531* (2019).

[35] Fabio Martinelli, Enzo Olivieri, and Roberto H Schonmann. 1994. For 2-d lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics* 165, 1 (1994), 33–47.

[36] Viresh Patel and Guus Regts. 2017. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.* 46, 6 (2017), 1893–1919.

[37] Ron Peled and Yinon Spinka. 2018. Rigidity of proper colorings of $\mathbb{Z}^d$. *arXiv preprint arXiv:1808.03597* (2018).

[38] Sergey Anatol'evich Pirogov and Ya G Sinai. 1975. Phase diagrams of classical lattice systems. *Theoretical and Mathematical Physics* 25, 3 (1975), 1185–1192.

[39] Allan Sly. 2010. Computational transition at the uniqueness threshold. In *Proceedings of the Fifty-first Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)*. IEEE, 287–296.

[40] Allan Sly and Nike Sun. 2014. Counting in two-spin models on d-regular graphs. *The Annals of Probability* 42, 6 (2014), 2383–2416.

[41] Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. 2009. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *Journal of the ACM (JACM)* 56, 3 (2009), 18.

[42] Mario Ullrich. 2013. Comparison of Swendsen-Wang and heat-bath dynamics. *Random Structures & Algorithms* 42, 4 (2013), 520–535.

[43] Dror Weitz. 2006. Counting independent sets up to the tree threshold. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing (STOC 2006)*. ACM, 140–149.