# Perfect Sampling from the Anti-Ferromagnetic Potts Model


1040555

University of Oxford

A thesis submitted for the degree of

*M.Sc. in Mathematics and Foundations of Computer Science*

Trinity 2020

# Abstract

Sampling algorithms provide a numerical means of computing the marginal distribution of a collection of random variables when the joint distribution cannot be determined efficiently. In this thesis, we consider the problem of sampling from the Gibbs distribution for the $q$-state Potts model, a prototypical graphical model with applications in fields such as Statistical Physics, Biology, Computer Science and Machine Learning. We focus on the problem of perfect sampling - sampling exactly from the Gibbs distribution - in the anti-ferromagnetic regime of the model.

Our primary contribution is the generalization of a recent perfect sampling algorithm for the uniform distribution over proper $q$-colourings of graphs with maximum degree $\Delta$. Specifically, for integers $q, \Delta \geq 3$ with $q > 2\Delta$ and real parameter $B \in (1 - \frac{q-2\Delta}{\Delta}, 1)$, we construct a randomized algorithm which takes an $n$-vertex graph $G$ with maximum degree $\Delta$ and outputs a sample from the Gibbs distribution for the $q$-state Potts model on $G$ with parameter $B$. The expected run-time of our algorithm is $\mathrm{O}(n^2)$.

## Acknowledgement

I would like to thank my supervisor, Andreas Galanis, whose guidance was indispensable.

# Contents

# 1 Introduction

The $q$-state Potts model is a graphical model introduced to study the macroscopic properties of materials as these properties arise from microscopic interactions between parts. In the model, atoms or molecules are modelled as random variables indexed by the nodes of a graph $G = (V, E)$. Each random variable may be in one of $q \geq 2$ states, where the conditional probability that one of the random variables is in some state is dependent only on the state of the neighbouring variables. The strength of the interaction between neighbouring variables is given by a real parameter $B > 0$ known in Statistical Physics as the inverse-temperature of the model: if $x : V \to [q] \coloneqq \{1, \cdots, q\}$ assigns to each vertex of the graph a state, and $m(x, v, c)$ denotes the number of neighbours of a vertex $v$ which are in state $c \in [q]$ under the assignment $x$, then the probability that the random variable indexed by $v$ is in state $c$ is proportional to $B^{m(x,v,c)}$.

The joint distribution of this collection of random variables is called the Gibbs distribution for the $q$-state Potts model on $G$ with parameter $B$. If $m(x)$ denotes the number of edges in $G$ connecting to vertices which are in the same state as one another in the assignment $x : V \to [q]$, then the Gibbs distribution for the $q$-state Potts model with parameter $B$ is given by $\mu_G(x) \propto B^{m(x)}$. Notwithstanding the relative simplicity of the proportional description of the Gibbs distribution, we will see that the problem of computing the normalizing constant - known as the partition function - for graphs with maximum degree less than $\Delta \geq 3$ is #P-complete [8]. This means that the problem is at least as difficult as counting the number of accepting paths of any non-deterministic Turing machine.

The Potts model may be viewed as a generalization of the proper $q$-colourings model more commonly studied in Computer Science and Mathematics. A proper $q$-colouring is an assignment $x : V \to [q]$ mapping vertices to colours in such a way that no vertex is assigned the same colour as any of its neighbours. When a graph $G$ admits a proper $q$-colouring, we may extend the Potts model to include the parameter $B = 0$. In this case, the Gibbs distribution is uniform over proper $q$-colourings, and the partition function counts the number of proper $q$-colourings. Many of the algorithms that we study for the $q$-state Potts model were first introduced in the context of the proper $q$-colourings model. For this reason, we will say that assignments $x : V \to [q]$ are $q$-colourings of $G$.[1]

[1] In Statistical Physics, maps $x : V \to [q]$ are more typically known as $q$-spin assignments of $G$.

Despite its origins as a model in Statistical Mechanics, the Potts model has since gained popularity as a general framework for studying the behaviour of systems of random variables governed by nearest neighbour interactions. Notable examples of the application of the Potts model include the study of phase transitions in the macroscopic properties of materials [1, 6], the dynamics of opinion formation [17, 3, 34], image reconstruction and segmentation [35, 27], biological modelling [18, 36, 37] and clustering [31, 33].

In such applications, there is often a need to evaluate the marginal distribution of subsets of the random variables, a need which is complicated significantly by the fact that the partition function cannot be computed efficiently. To avoid computing the partition function of the model, we work with the conditional distribution at each vertex, generating a colouring using a greedy algorithm and then updating the colour of the vertices repeatedly according to the conditional Gibbs distribution at each vertex. Using this approach, we will see that it is possible to produce samples distributed according to $\mu_G$ without knowledge of the partition function. We can then evaluate any marginal distribution we would like using the mean weight of sufficiently many samples.

Our focus in this thesis is on the problem of sampling from the Gibbs distribution for the $q$-state Potts model on graphs with maximum degree $\Delta$. We focus on the regime $q \geq 3$, excluding the case $q = 2$ which has itself been well-studied as the Ising model. We will only consider the model in the anti-ferromagnetic regime $B \in (0, 1)$, where the probability that a vertex is assigned a colour is a decreasing function of the number of neighbours assigned that colour. While we will primarily consider algorithms which generate samples from the Gibbs distribution itself, we will give a brief treatment of algorithms which generate samples from distributions which approximate the Gibbs distribution to an arbitrary multiplicative factor. Our treatment of these algorithms will provide us with the most general setting to outline the reduction from the problem of approximating the partition function to the problem of sampling in Chapter 2, and will form the basis of our introduction to 'perfect sampling' algorithms in Chapter 3.

## 1.1 Approximate Sampling

An approximate sampling algorithm for the Gibbs distribution for the $q$-state Potts model is an algorithm which takes an $n$-vertex graph $G$ with maximum degree $\Delta$ and an $\epsilon$ error tolerance and returns a sample drawn from a distribution which is within an $\epsilon$ multiplicative factor of Gibbs distribution on $G$. An approximate sampling algorithm is efficient, or *fully-polynomial*, if the expected

run-time of the algorithm is polynomial in $n$ and $\log(\epsilon^{-1})$.

The Markov chain Monte Carlo framework provides a very popular method for approximately sampling from the Gibbs distribution on the Potts model. This method relies on a finite, irreducible, aperiodic Markov chain with the Gibbs distribution as its limiting distribution. To draw a sample, we determine an upper bound on $\tau(\epsilon)$, the length of time required before $X_{\tau(\epsilon)}$ is guaranteed to approximate the limiting distribution within an $\epsilon$ factor, independently of the state $X_0$.

Most of the work of constructing an approximate sampling algorithm is concerned with bounding the value of $\tau(\epsilon)$ in terms of the size of the graph and $\log(\epsilon^{-1})$. When $\tau(\epsilon)$ is polynomial in these components, as in the definition of a fully-polynomial approximate sampler, we say that the underlying Markov chain is rapidly mixing. In one of the earliest major contributions to the field of approximate sampling, Jerrum [25] demonstrated that a well-known Markov chain known as the Glauber dynamics is rapidly mixing on the proper $q$-colourings model when $q > 2\Delta$. This result was later improved by Vigoda [38] to $q > \frac{11}{6}\Delta$, who used a different Markov chain called the 'flip dynamics', Currently, the best known approximate sampling algorithm for the proper $q$-colourings model on graphs of maximum degree less than $\Delta$ is due to Chen et al. [5], who improve Vigoda's analysis of the flip dynamics to the range $q > (\frac{11}{6} - \epsilon_0)\Delta$ for $\epsilon_0 \sim 10^{-4}$.

## 1.2 Perfect Sampling

Perfect sampling algorithms improve upon approximate samplers by eliminating the approximation error. More precisely, for graphs with maximum degree $\Delta$, perfect sampling algorithms are randomized algorithms which generate a sample from a specified sampling distribution with expected run-time polynomial in the size of the graph. Even when a small amount of approximation error is permissible, the stronger conditions on perfect samplers provide a number of improvements over approximate sampling. By eliminating the dependence of the run-time on $\log(\epsilon^{-1})$, they may return samples significantly faster than approximate samplers when the application requires that $\epsilon$ is small. Moreover, all of the perfect sampling algorithms considered in this thesis terminate in response to a well-defined stopping condition. Thus, unlike approximate sampling techniques which rely on an analysis of the mixing time of the chain to determine when to stop, perfect sampling algorithms may terminate significantly sooner than the upper bounds on their run-time suggest and may be used to draw samples exactly from the desired sampling distribution even in the absence of such upper bounds on their run-time.

A wide variety of approaches to the problem of perfect sampling have been proposed in the literature. Propp and Wilson were the first to introduce a perfect sampling algorithm with their coupling from the past framework [32], which was followed by another framework due to Fill [11]. More recently, Guo et al. have proposed partial rejection sampling, which relies on the Lovasz Local lemma [19] and Fill and Huber have introduced the randomness recycler, which iteratively generates perfect samples from distributions that approach the desired sampling distribution [10]. In a more general setting, Feng et al. have introduced a sampler which works on a broad class of graphical models and admits dynamical updates of the form which one might expect to see in statistical inference [9].

The perfect sampling algorithms which we consider here are based on Propp and Wilson's coupling from the past framework [32]. Given $\Delta, q \geq 3$ and a graph $G$ of maximum degree $\Delta$, we sample endomorphisms $\phi$ of the set of $q$-colourings of $G$. Each sample $\phi$ simulates the behaviour of a Markov chain that is stationary with respect to the Gibbs measure, in the sense that $\phi(x)$ is distributed according to the sampling distribution whenever $x$ is. Using the samples $\phi$, we may define a chain over the function space which, alongside a secondary chain that detects when this chain is a constant function, provides the basis for a perfect sampler.

The first perfect sampling algorithm for the anti-ferromagnetic Potts model was introduced by Huber [22], who demonstrated[2]that it is possible to perfectly sample in the colour-temperature regime described by Table 1.

| # of colours | inverse-temperature |
|---|---|
| $q \geq \Delta(\Delta+2)$ | $B > 0$ |
| $\Delta < q < \Delta(\Delta+2)$ | $B^{\Delta} > 1 - \frac{q}{\Delta(\Delta+2)}$ |
| $q \leq \Delta$ | $B^{\Delta} > 1 - \frac{1}{\Delta q}$ |

Table 1: Colour-parameter regime of Huber's perfect sampling algorithm [22].

Bhandari and Chakraborty [2] have recently improved upon Huber's results in the special case of proper $q$-colourings. They use a modified version of the Glauber dynamics to sample from the uniform distribution over proper $q$-colourings when $q > 3\Delta$. This result has no counterpart for the anti-ferromagnetic Potts model, where the bounds produced by Huber have not been improved in nearly two decades.

[2]See the footnote in Section 3.6 for an explanation of the discrepancy between the results recorded here and those described in [22].

## 1.3 Overview

This thesis is devoted to the problem of perfect sampling from the Gibbs distribution on the anti-ferromagnetic Potts model over graphs with bounded degree. In Chapter 2, we motivate this problem by treating its relationship with the problem of computing the partition function for the Gibbs distribution. In particular, after introducing the relevant computational structures, we provide in Theorem 10 a reduction from the problem of approximating the partition function to the sampling problem. Since the problem of determining the partition function exactly has been shown to belong to the class #P of the counting versions of NP problems, this establishes that sampling is computationally 'hard' in a well-defined sense.

Chapter 3 contains an introduction to the topic of perfect sampling by way of approximate sampling. In Section 3.1, we review the use of ergodic Markov chains for approximate sampling, concluding with a proof of the coupling lemma (c.f. Lemma 19). We then use this lemma in Section 3.4 to motivate the coupling from the past approach to perfect sampling. After introducing seeding and bounding chains [22] in Section 3.5 to deal with some of the computational challenges posed by perfect sampling, we conclude the third chapter with the perfect sampling algorithm used by Huber to sample the regime described in Table 1.

Finally, in Chapter 4 we extend Bhandari and Chakraborty's recent algorithm [2] for sampling from the uniform distribution over proper $q$-colourings when $q > 3\Delta$ to produce the first linear bounds on the number of colours required to sample perfectly from the anti-ferromagnetic Potts model. Our extension formalizes the arguments presented in [2], replacing two of the Markov chains used in the coupling from the past framework with a bivariate stochastic process. Our results may be stated as follows:

**Theorem 1.** *Suppose $q, \Delta \geq 3$ are integers with $q > 2\Delta$, and let $B$ be a real in the interval $(1 - \frac{q-2\Delta}{\Delta}, 1)$. Then there exists a randomized algorithm which, on input an $n$-vertex graph $G$ of maximum degree $\Delta$, outputs in expected time $O(n^2)$ a sample from the Gibbs distribution for the $q$-state Potts model on $G$ with parameter $B$.*

## 2 Counting and Sampling

Let $G = (V, E)$ be an undirected graph and fix an integer $q \geq 2$. Recall that a *q-colouring* of $G$ is an assignment $x : V \to [q]$ of a colour in the set $[q] \coloneqq \{1, \cdots, n\}$ to each vertex of the graph.[3] We will denote the set of $q$-colourings of $G$ by $\Omega_q$.

**Definition 2.** The *q-state Potts model* with parameter $B > 0$ assigns to every $q$-colouring $x \in \Omega_q$ a weight $w(x) = B^{m(x)}$, where $m(x)$ denotes the number of monochromatic edges in $x$. The *Gibbs distribution* $\mu_G$ has

$$\mu_G(x) \coloneqq \frac{w(x)}{Z_G}$$

where $Z_G \coloneqq \sum_{y \in \Omega_q} w(y)$ is a normalizing constant called the *partition function*.

The parameter $B$ corresponds to what is known in Statistical Physics as the inverse-temperature of the model. When $B > 1$, the Gibbs distribution assigns higher probability to $q$-colourings with many monochromatic edges and we say that the model is *ferromagnetic*. Otherwise, more weight is put on $q$-colourings with fewer monochromatic edges and the model is *anti-ferromagnetic*.

We will compare the $q$-state Potts model with another model known as the proper $q$-colourings model. If $N : V \to 2^V$ maps a vertex to its set of neighbours, then a $q$-colouring $x$ is *proper* if $x(v) \notin x(N(v))$ for all $v \in V$. The proper $q$-colourings model has configurations the set $\Omega_{\text{prop}}$ of proper $q$-colourings of $G$; the Gibbs distribution for the proper $q$-colourings model is the uniform distribution supported over $\Omega_{\text{prop}}$. Under the assumption that the graph $G$ admits a proper $q$-colouring, we may extend the $q$-state Potts model to include $B = 0$. In this case, the Gibbs distribution for the proper $q$-colourings model corresponds to the Gibbs distribution for the $q$-state Potts model.

The critical property of the Gibbs distribution is that the conditional distribution at each vertex is proportional to the number of monochromatic edges involving that vertex. Specifically, if $v$ is a vertex of the graph and $m(x, v, c)$ is the number of neighbours of vertex $v$ with colour $c$ in a colouring $x$, then

$$\mu_G\Big(x \,\Big|\, x|_{V \setminus \{v\}}\Big) = \frac{B^{m(x,v,x(v))}}{\sum_{c' \in [q]} B^{m(x,v,c')}}$$

The normalizing constant of the conditional distribution at any vertex may be computed in constant time, in stark contrast to the partition function of the Gibbs distribution itself, which we will see

[3] We will occasionally treat $x$ as a function on $2^{V(G)}$, i.e. if $S \subseteq V(G)$ then $x(S) = \{x(s) \,\big|\, s \in S\}$.

cannot be computed efficiently.

In the first section of this chapter, we introduce the relevant computational structures and review the body of research which demonstrates that the problem of computing the partition function $Z_G$ is computationally infeasible. We then explore in Section 2.2 how randomized algorithms which generate samples from the Gibbs distribution for the anti-ferromagnetic Potts model using the simple form of the conditional distribution at each vertex may be used to approximate the partition function. The key result is Theorem 10, which establishes a reduction from the problem of approximate counting to the problem of approximate sampling.

## 2.1 The Complexity of Counting

For $q \geq 3$, the decision problem of determining whether an unrestricted graph $G$ admits a proper $q$-colouring is well known to be NP-complete [15]. Indeed, this problem remains NP-complete when restricted to graphs of maximum degree $\Delta$, as long as $q$ is sufficiently small compared to $\Delta$ - for example, the 3-colourability problem remains NP-complete when restricted to graphs with maximum degree four [15]. On the other hand, if $q > \Delta$ then a greedy algorithm shows that $G$ is proper $q$-colourable. For this range of $q$, it is natural to ask whether it is possible to count the number of proper $q$-colourings which are admitted by $G$.

This problem belongs to the class #P of counting versions of NP problems; to introduce this class, we recall the following definitions of the classes P and NP.

**Definition 3.** Fix a finite alphabet $\Sigma$ and encode problem instances as strings in $\Sigma^*$. A *decision problem* is a predicate $\rho : \Sigma^* \to \{0, 1\}$. If a deterministic Turing Machine can compute $\rho$ in polynomial time in the size of the input then $\rho$ belongs to the class P of *polynomial-time* decision problems. A decision problem is in NP if there exists a polynomial $p$ and a polynomial-time checkable[4] predicate $\psi : \Sigma^* \times \Sigma^* \to \{0, 1\}$ such that

$$\rho(x) = 1 \iff \exists w \in \Sigma^* : |w| \leq p(x) \wedge \psi(x, w) = 1$$

for all $x \in \Sigma^*$. The word $w$ is known as a *witness* because $w$ witnesses to the *witness-checking predicate* $\psi$ that $\rho(x) = 1$.

[4] A predicate is polynomial-time checkable if it may be computed in polynomial time on a deterministic Turing Machine.

We will typically ignore the role of the encoding, assuming only that we have a reasonable encoding of the graphs and colourings. In the proper $q$-colourings setting, the witness-checking predicate takes an $n$-vertex graph $G = (V, E)$ with maximum degree $\Delta$ and colouring $x : V \to [q]$ and accepts if and only if $x$ is a proper $q$-colouring of $G$. We can check $\psi$ in polynomial-time by iterating over the vertices and establishing the condition $x(v) \notin x(N(v))$. There is no analogous decision problem for the Potts model, because the model admits every $q$-colouring.

Of more interest for the Potts model are counting problems. Counting problems replace $\rho$ with a function $f$ which takes an encoding and returns a positive integer. The relevant computational class #P consists of those counting problems which correspond to the number of accepting paths of a non-deterministic Turing Machine.[5]

**Definition 4.** A *counting problem* is a function $f : \Sigma^* \to \mathbb{N}$. A counting problem $f$ is in #P if there exists a polynomial $p$ and a polynomial-time checkable predicate $\psi$ such that, for all $x \in \Sigma^*$,

$$f(x) = |\{w \in \Sigma^* : |w| \leq p(x) \wedge \psi(x, w) = 1\}|$$

The counting problem for the proper $q$-colourings model is a function which assigns to a graph $G$ with maximum degree $\Delta \geq 3$ the number of proper $q$-colourings $|\Omega_{\text{prop}}|$. Bubley et al. [4] were the first to show that this problem was #P-complete. In the more general setting of the Potts model, the relevant counting problem is the function $f(G) = Z_G$ mapping a graph to its partition function. Analogous to the result in Bubley et al. [4], Greenhill and Dyer [8] showed that the counting problem for the anti-ferromagnetic Potts model on graphs with maximum degree $\Delta$ is #P-complete.

## 2.2 From Counting to Sampling

In light of the fact that the counting problem for the Potts model on graphs with maximum degree $\Delta$ is #P-complete, polynomial-time approximation of the partition function is the best we can hope for. The relevant approximating notion is an FPRAS, due to Karp, Luby and Madras [28].

**Definition 5.** Fix integers $q, \Delta \geq 3$ and real parameter $B > 0$. A *randomized approximation scheme* (RAS) for approximating the partition function of the $q$-state Potts model with parameter $B$ is a randomized algorithm which, given a graph $G$ with maximum degree $\Delta$ and an error tolerance

[5] For more on the class #P, see [12].

$\epsilon \in (0,1)$, outputs a random variable $\hat{Z}_G(\epsilon)$ satisfying

$$\mathbb{P}(|\hat{Z}_G(\epsilon) - Z_G| \leq \epsilon Z_G) \geq \frac{3}{4}$$

A RAS is a *fully-polynomial randomized approximation scheme* (FPRAS) if the running time of the algorithm is polynomial in $|G|$ and $\epsilon^{-1}$.

It is not always possible to approximate the partition function in the sense of an FPRAS. Indeed, for all even $q \geq 4, \Delta \geq 3$ and $B < \frac{\Delta - q}{\Delta}$, Galanis et al. [14] have shown that there cannot exist an FPRAS for approximating the partition function of the $q$-state Potts model (unless NP = RP); it is believed this result holds for odd $q$ as well. For $q, \Delta$ and $B$ outside of this regime, however, the problem of constructing a FPRAS remains an active area of research.

We will construct an FPRAS for approximating the partition function of the Potts model using a reduction from the approximate counting problem to the problem of sampling from the Gibbs distribution. The class of algorithms that we are primarily concerned with draw samples exactly from the Gibbs distribution for the anti-ferromagnetic Potts model.

**Definition 6.** Fix $\Delta, q \geq 3$ and let $\mu_G$ denote the Gibbs distribution for the $q$-state Potts model with parameter $B \in (0,1)$ on a graph $G$. A *perfect sampling algorithm* for the anti-ferromagnetic $q$-state Potts model with parameter $B$ is a randomized algorithm which, on input an $n$-vertex graph $G$ with maximum degree $\Delta$, outputs a sample colouring distributed according to $\mu_G$ in expected-time polynomial in $n$.

The reduction from approximate counting to sampling permits a small amount of error in the sampling procedure. We will prove this reduction in its more general form; for this, we must establish a metric on the space of probability distributions.

**Definition 7.** Let $\pi_1$ and $\pi_2$ be two probability distributions on a finite set $S$. The *total-variation distance* between $\pi_1$ and $\pi_2$ is defined

$$\|\pi_1 - \pi_2\|_{\mathrm{TV}} \coloneqq \frac{1}{2} \sum_{s \in S} |\pi_1(s) - \pi_2(s)|$$

A careful reader will notice that the total-variation distance is half the $L_1$ norm. In the context of probability distributions, we have the following well-known alternate characterization.

**Lemma 8.** *Suppose $\pi_1$ and $\pi_2$ are distributions on a finite set $S$. Then*

$$\|\pi_1 - \pi_2\|_{TV} = \max_{U \subseteq S} |\pi_1(U) - \pi_2(U)|$$

*Proof.* We will partition the set $S = S^+ \sqcup S^-$ where $S^+ = \{s \in S \,|\, \pi_1(s) \geq \pi_2(s)\}$. Both $\pi_1$ and $\pi_2$ are probability distributions, so $\pi_1(S^+) + \pi_1(S^-) = 1 = \pi_2(S^+) + \pi_2(S^-)$. Thus

$$\begin{aligned}
\|\pi_1 - \pi_2\|_{\mathrm{TV}} &= \frac{1}{2} \sum_{x \in S^+} (\pi_1(x) - \pi_2(x)) + \frac{1}{2} \sum_{x \in S^-} (\pi_2(x) - \pi_1(x)) \\
&= \frac{1}{2}\Big((\pi_1(S^+) - \pi_2(S^+) + \pi_1(S^-) - \pi_2(S^-)\Big) \\
&= \pi_1(S^+) - \pi_2(S^+)
\end{aligned}$$

The set $S^+$ maximizes $\pi_1(U) - \pi_2(U)$ over subsets $U$ of $S$, since

$$\pi_1(U) - \pi_2(U) = \sum_{s \in U} \pi_1(s) - \pi_2(s) \leq \sum_{s \in S^+} \pi_1(s) - \pi_2(s) = \pi_1(S^+) - \pi_2(S^+)$$

As a consequence, $\|\pi_1 - \pi_2\|_{\mathrm{TV}} = \max_{U \subseteq S} |\pi_1(U) - \pi_2(U)|$. □

With the definition of total-variation distance, we are now ready to introduce the class of approximate sampling algorithms.

**Definition 9.** Fix integers $\Delta, q \geq 3$ and real $B \in (0, 1)$. Let $\mu_G$ denote the Gibbs distribution over the $q$-state Potts model with parameter $B$ on a graph $G$ with maximum degree $\Delta$. An *approximate sampler* for the Gibbs distribution on the $q$-state Potts model with parameter $B$ is an algorithm which takes as input an $n$-vertex graph $G$ with maximum degree $\Delta$ and an error tolerance $\epsilon > 0$, and samples its output from a distribution $\hat{\mu}_{G,\epsilon}$ with

$$\|\hat{\mu}_{G,\epsilon} - \mu_G\|_{\mathrm{TV}} \leq \epsilon$$

The algorithm is a *fully-polynomial approximate sampler* (FPAS) if the expected run-time of the algorithm is polynomial in $n$ and $\log(\epsilon^{-1})$.

We are now ready to prove the reduction from the problem of approximate counting to the problem of sampling from the Gibbs distribution for the anti-ferromagnetic Potts model. This will establish the hardness of the problem of approximate sampling and, by extension, the hardness of the problem of perfect sampling. Our argument is based on a reduction for the equivalent problem

regarding independent sets which may be found in [16]; for more on the reduction in the case of the $q$-state Potts model, see [7, 26]. The proof relies critically on Lemma 11, which is a consequence of Chernoff's bound.[6]

**Theorem 10.** *Fix integers $\Delta, q \geq 3$ and real $B \in (0,1)$. For the $q$-state Potts model with parameter $B$, an FPAS for the Gibbs distribution on graphs of maximum degree $\Delta$ implies an FPRAS for the partition function on graphs of maximum degree $\Delta$.*

**Lemma 11.** *Suppose $\{X_i\}_{1 \leq i \leq N}$ are independent, identically distributed indicator random variables and let $X = \sum_{1 \leq i \leq N} X_i$. If $\epsilon, \delta \in (0,1)$ and $N \geq 3 \log \frac{2}{\delta} \backslash (\epsilon^2 E[X_i])$ then*

$$\mathbb{P}\Big(\Big|\frac{X}{N} - E[X_i]\Big| \geq \epsilon E[X_i]\Big) \leq \delta$$

*Proof.* We use a slightly weaker version of Chernoff's bound which may be found in [16]:

$$\mathbb{P}(X \geq (1 \pm \epsilon) E[X_i]) \leq e^{-E[X_i]\epsilon^2 \backslash 3}$$

This implies $\mathbb{P}(|X - NE[X_i]| \geq \epsilon N E[X_i]) \leq 2e^{-N\epsilon^2 E[X_i] \backslash 3}$ from which the desired result follows immediately. □

*Proof of Theorem 11.* Given a graph $G$ with $m$ edges and $n$ vertices, order the edges $e_1, \cdots, e_m$ and let $G_i$ denote the graph which includes only the edges $e_1, \cdots, e_i$. We will use the notation $m_i(x)$ to denote the number of monochromatic edges in colouring $x$ on $G_i$. To estimate $Z_G$, note that $G_m = G$ and use the telescoping product

$$Z_G = \frac{Z_{G_m}}{Z_{G_{m-1}}} \frac{Z_{G_{m-1}}}{Z_{G_{m-2}}} \cdots \frac{Z_{G_1}}{Z_{G_0}} Z_{G_0}$$

The last factor $Z_{G_0}$ is simply $q^n$ because $G_0$ has no edges, so we may develop an estimate using only the ratios $\varrho_i \coloneqq \frac{Z_{G_i}}{Z_{G_{i-1}}}$.

If $\epsilon \in (0,1)$, we claim that we may approximate $Z_G$ with an $\epsilon$ multiplicative factor using estimates $\hat{\varrho}_i$ of $\varrho_i$ with the property that $\mathbb{P}(|\hat{\varrho}_i - \varrho_i| \geq \frac{\epsilon}{2m}\varrho_i) \leq \frac{1}{4m}$. If $\hat{\varrho}_i$ all satisfy this condition, the union bound gives that $|\hat{\varrho}_i - \varrho_i| \leq \frac{\epsilon}{2m}\varrho_i$ for all $1 \leq i \leq m$ with probability at least $\frac{3}{4}$. This, in turn, implies that

$$1 - \epsilon \leq e^{-\epsilon} \leq \Big(1 - \frac{\epsilon}{2m}\Big)^m \leq \prod_{i=1}^{m} \frac{\hat{\varrho}_i}{\varrho_i} \leq \Big(1 + \frac{\epsilon}{2m}\Big)^m \leq e^{\epsilon/2} \leq 1 + \epsilon$$

[6] See [30] for a number of detailed proofs of Chernoff's bound and related tail bounds.

with probability at least $\frac{3}{4}$ as desired.

We now turn to the problem of approximating the ratios $\{\varrho_i\}_{1\leq i\leq m}$ using an approximate sampler. Let $1 \leq i \leq m$ and suppose $e_i = (u, v)$. The ratio $\varrho_i$ can be expanded

$$\varrho_i = \frac{B\Big(\sum_{\substack{x\in\Omega_q \\ x(u)=x(v)}} B^{m_{i-1}(x)}\Big) + \sum_{\substack{x\in\Omega_q \\ x(u)\neq x(v)}} B^{m_{i-1}(x)}}{Z_{G_{i-1}}} \tag{1}$$

If $\eta_i$ denotes the probability that a colouring $x$ drawn from the Gibbs distribution for the $q$-state Potts model on $G_{i-1}$ has $x(u) \neq x(v)$, we can write Equation (1) as $\varrho_i = B(1-\eta_i) + \eta_i$. Rather than estimate $\eta_i$ directly, we estimate the probability $\nu_i$ that a sample drawn from the approximate sampler assigns $u$ and $v$ different colours. We may treat this estimate of $\nu_i$ as an estimate of $\eta_i$ using Lemma 8, which states that $|\nu_i - \eta_i| \leq \frac{\epsilon B^\Delta}{4m(1+B^\Delta)}$ if we use an FPAS with error tolerance less than $\frac{\epsilon B^\Delta}{4m(1+B^\Delta)}$.

The estimate $\hat{\nu_i}$ of $\nu_i$ that we use is the proportion of the first $N_i$ samples which assign $u$ and $v$ different colours. As a consequence of Lemma 11, we may ensure that $|\hat{\nu_i} - \nu_i| \leq \frac{\epsilon B}{4m(1+B)}$ with probability at least $1 - \frac{1}{4m}$ by setting $N_i \geq \Big(3\log(8m)\Big)/\Big(\Big[\frac{B\epsilon}{4m(1+B)}\Big]^2 \nu_i\Big)$. Now $N_i$ is polynomial in $n$ and $\log(\epsilon^{-1})$ if $\nu_i$ is bounded above 0 by a constant. Given a colouring $x$ which assigns $u$ and $v$ the same colour, we may define a unique colouring $f(x)$ by recolouring $u$ with the 'lowest' numbered colour which is not $x(u)$. This gives us an injective mapping

$$f : \{x \in \Omega \,\Big|\, x(u) = x(v)\} \to \{x \in \Omega \,\Big|\, x(u) \neq x(v)\}$$

with $w(f(x)) \geq B^\Delta w(x)$, from which we may deduce $\eta_i > \frac{B^\Delta}{1+B^\Delta}$. Alongside the fact that $|\nu_i - \eta_i| \leq \frac{\epsilon B^\Delta}{4m(1+B^\Delta)}$, this implies that $\nu_i > \frac{B^\Delta}{1+B^\Delta}\Big(1 - \frac{\epsilon}{4}\Big)$, so $|\hat{\nu_i} - \nu_i| \leq \frac{\epsilon B}{4m(1+B)}$ with probability at least $1 - \frac{1}{4m}$ when $N_i \geq \Big(3\log(8m)\Big)/\Big(\frac{B^{\Delta+2}}{(1+B^\Delta)^3}\frac{\epsilon^2}{m^2}\frac{4-\epsilon}{4^3}\Big)$.

It remains to derive an estimate $\hat{\varrho_i}$ of $\varrho_i$ from $\hat{\nu_i}$. To this end, let $\hat{\varrho_i} = B(1-\hat{\nu_i}) + \hat{\nu_i}$. Since $|\hat{\nu_i} - \nu_i| \leq \frac{\epsilon B}{4m(1+B)}$ with probability at least $1 - \frac{1}{4m}$, we have that $|\hat{\varrho_i} - \varrho_i| \leq (1+B)|\hat{\nu_i} - \mu_i| \leq \frac{\epsilon}{2m}B$ with probability at least $1 - \delta$. By Equation (1), using the fact that $B \in (0, 1)$, we have that $B \leq \varrho_i \leq 1$, from which we may conclude that $|\hat{\varrho_i} - \varrho_i| \leq \frac{\epsilon}{2m}\varrho_i$ with probability at least $1 - \delta$. □

In our context, the key observation of Theorem 10 is that the problem of sampling from the Gibbs distribution on the anti-ferromagnetic Potts model is at least as difficult as approximating a

#P-complete problem.[7]

[7] In the proper $q$-colourings setting, an algorithm which deterministically approximates the counting problem within an $\epsilon$ factor may be used to construct a perfect sampler, and an FPAUS and FPRAS are inter-reducible, see [26].

# 3 Introduction to Sampling using Markov Chains

In this chapter, we provide an overview of the construction of a sampling algorithm. The first half of the chapter is concerned with necessary background material and the theory of approximate sampling. In Section 3.1, we review Markov chains and introduce the Glauber dynamics, a natural Markov chain which is stationary with respect to the Gibbs distribution for the Potts model. In Section 3.2, we briefly review partial rejection sampling, a common subroutine used to produce colours distributed according to the conditional distribution at a vertex. Finally, we conclude the first half of the chapter with a discussion of the basics of coupling in Section 3.3.

In the second half of the chapter, we turn to the construction of a perfect sampling algorithm as outline by the coupling from the past framework [32]. Coupling from the past is introduced in Section 3.4, where it is motivated on the basis of the coupling lemma. In Section 3.5, we see how we can overcome some of the problems posed by the size of the space of colourings and we introduce bounding chains as a method for detecting when coupling has occurred. Finally, in Section 3.6 we reproduce Huber's perfect sampling algorithm for the anti-ferromagnetic Potts model [22] and prove that the algorithm is fully-polynomial.

## 3.1 Markov Chains and the Glauber Dynamics

We will construct sampling algorithms which starts from an arbitrary colouring and updates the colour of the vertices repeatedly using random choices. The relevant probabilistic structure is a Markov chain.

**Definition 12.** A finite, discrete-time *Markov chain* is a sequence of random variables $(X_t)_{t\in\mathbb{N}}$ all over a finite set $\Omega$ with the *Markov property*

$$\mathbb{P}(X_{t+1} = x_{t+1} \,|\, X_t = x_t, \cdots, X_0 = x_0) = \mathbb{P}(X_{t+1} = x_{t+1} \,|\, X_t = x_t)$$

If $t \in \mathbb{N}$, the random variable $X_t$ describes the state of the chain at time $t$, after the chain has taken $t$ steps from the starting state $X_0$. We will typically suppress the index set entirely and denote the chain by $(X_t)$. The following are important properties of Markov chains:

- The *transition matrix* of $(X_t)$ is a row-stochastic matrix $P$ with rows and columns indexed by $\Omega$, where the entry at $(x, y) \in \Omega^2$ is the probability of transitioning from $x$ to $y$.

- The chain $(X_t)$ is *irreducible* if, for all $x, y \in \Omega$, there exists a $t > 0$ such that $P^t_{x,y} > 0$.
- The *period* of a state $x \in \Omega$ is $\gcd\{t > 0 : P^t_{x,x} > 0\}$; if the period of $x$ is 1, we say $x$ is *aperiodic*. If all of the states of a Markov chain are aperiodic, then we say that the chain is *aperiodic*.
- A probability distribution $\pi \coloneqq \{\pi_x\}_{x\in\Omega}$ is a *stationary distribution* of $(X_t)$ if $\pi = \pi P$. It is the *limiting distribution* of the chain if $\lim_{t\to\infty} \mathbb{P}(X_t = y \,\big|\, X_0 = x) = \pi_y$ for all $x, y \in \Omega$.

The condition that the transition matrix is row-stochastic ensures that the rows of $P$, denoted $P_{x,\cdot}$ for $x \in \Omega$, may be interpreted as probability distributions over the space $\Omega$. If $y \in \Omega$ then $(P^2)_{x,y} = \sum_{z\in\Omega} P_{x,z}P_{z,y}$, so the $x$ row of $P^t$ is a distribution $P^t_{x,\cdot}$ over $\Omega$ describing the probability of finding the chain in a state if the chain is simulated for $t$ steps starting from the state $x$. Under this interpretation, the limiting distribution $\pi$ is the weak limit as $t \to \infty$ of the distributions $P^t_{x,\cdot}$, i.e. $\pi_y = \lim_{t\to\infty} P^t_{x,y}$ for all $y \in \Omega$.

Whenever the limiting distribution of a Markov chain exists, it is also the only stationary distribution of the chain. To see this, suppose $\pi$ is a stationary distribution of $(X_t)$. Then

$$\lim_{t\to\infty} P^t_{x,y} = \lim_{t\to\infty} \mathbb{P}(X_t = y \,\big|\, X_0 \sim \pi) = \pi_y \tag{2}$$

where equality the first equality follows from the fact that the limiting distribution is independent of initial state and the second equality follows from the fact that $\pi$ is a stationary distribution of $(X_t)$.

It is often the case that finding a stationary distribution of $(X_t)$ is easier than determining the limiting distribution, so Equality 2 is a useful tool for determining the limiting distribution when we know that it exists. However, not all Markov chains have a limiting distribution. To outline the problem, consider a Markov chain with state space $\Omega = \{a, b, c, d\}$ and transition matrix

$$P_{x,y} \coloneqq \begin{cases} 1 & (x, y) \in \{(a, b), (b, a), (c, d), (d, c)\} \\ 0 & \text{else} \end{cases} \tag{3}$$

There are two issues with establishing the existence of a limiting distribution for this chain. First, $P^t_{x,y}$ depends on the whether $x \in \{a, b\}$ or $\{c, d\}$ and thus cannot be independent of $x$ as in the definition of limiting distribution. To eliminate the possibility of the initial state determining the future behaviour of the chain, we must require that the chain is irreducible. Practically, this amounts

to a minor condition on the space of chains, as reducible Markov chains decompose into multiple smaller irreducible components defined on subsets of the state space.

Even when we restrict our attention to the irreducible subsets of the chain defined in (3), we still cannot determine a unique stationary distribution. The second issue that we must address is that the smaller chains are irreducible: their transition matrix is $Q = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, so $Q^t = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ for odd $t$ and the identity for even $t$. All of the probability mass alternates between states $a$ and $b$ (respectively $c$ and $d$) without approaching a limiting distribution because the chain is periodic. Somewhat surprisingly, for finite Markov chains, the following well-known theorem says that the assumptions of irreducibility and aperiodicity suffice to guarantee the existence of a limiting distribution.

**Theorem 13** (Fundamental Theorem of Markov Chains, [29])**.** *If $(X_t)$ is a finite, irreducible and aperiodic Markov chain with state space $\Omega$ and transition matrix $P$ then $\lim_{t\to\infty} P^t_{x,y}$ exists for all $x, y \in \Omega$.*

When a Markov chain has a limiting distribution, we may approximate this limiting distribution by simulating the chain from any state. For the Potts model with parameter $B > 0$, Algorithm 1 gives the standard Markov chain with limiting distribution the Gibbs distribution, known as Glauber dynamics. Given a set $S$, we will use the notation $s \in_{\mathrm{U}} S$ to mean $s$ sampled uniformly from the set $S$.

**Algorithm 1:** glauberUpdate$(G, B, q, x)$

**Input :** a graph $G = (V, E)$ with maximum degree $\Delta$, parameter $B > 0$, number of colours $q \geq 2$ and a colouring $x \in \Omega_q$
**Output:** a colouring $y \in \Omega_q$

**Function** *glauberUpdate*$(G, B, q, x)$**:**
  sample $v \in_{\mathrm{U}} V$ and $c \in [q]$ with probability proportional to $B^{m(x,v,c)}$
  $y|_{V\setminus\{v\}} \longleftarrow x|_{V\setminus\{v\}}$
  $y(v) \longleftarrow c$
  **return** $y$

**Lemma 14.** *Suppose $q \geq 2$ and $\Delta \geq 3$ are integers and $B > 0$ is a real number. If $G$ is a graph with degree less than or equal to $\Delta$, then the limiting distribution of the Glauber dynamics (defined by Algorithm 1) is the Gibbs distribution of the $q$-state Potts model.*

*Proof.* Suppose $x, y \in \Omega_q, v \in V$ and let $m(y|_{V\setminus\{v\}})$ denote the number of monochromatic edges not adjacent to $v$. Then

$$\sum_{x\in\Omega_q} \pi_x P_{x,y} = \sum_{v\in V} B^{m(y|_{V\setminus\{v\}})} \sum_{c\in[q]} B^{m(y,v,c)} \frac{B^{m(y,v,y(v))}}{\sum_{c'\in[q]} B^{m(y,v,c')}} = \pi_y$$

In particular, $\pi$ is a stationary distribution for the Glauber dynamics. Since the limiting distribution equals the stationary distribution when the limiting distribution exists, the Fundamental theorem of Markov chains implies that $\pi$ is the limiting distribution of the Glauber dynamics if the Glauber dynamics is irreducible and aperiodic.

The Glauber dynamics is aperiodic because $P_{x,x} > 0$ for all $x \in \Omega_q$ and $t \geq 0$. To see that the Glauber dynamics is irreducible, order the vertices $v_1, \cdots, v_n$. If $x_i$ denotes the colouring which agrees with $y$ on the first $i$ vertices and $x$ on the remaining, then $P_{x_i,x_{i+1}} > 0$ for all $0 \leq i < n$. In particular, $P^n_{x,y} \geq \prod_{0 \leq i < n} P_{x_i,x_{i+1}} > 0$. □

## 3.2 Partial Rejection Sampling Subroutine

Suppose we wish to draw sample colourings for updating the colour of a vertex from the conditional distribution at a vertex, but we wish to generate the colourings from a different distribution first. Rejection sampling is a well known method for constructing a sampler for a target distribution $f$ over a space $\Omega$ using a sampler for a distribution $g$ over $\Omega$, under the condition that the ratio $\sup_{x \in \Omega} \frac{f(x)}{g(x)}$ is bounded.

**Algorithm 2:** rejectionSample$(f, g, M)$

**Input :** a target distribution $f$ over finite $\Omega$, a method for sampling from a strictly positive distribution $g$ over $\Omega$ and $M \geq \sup_{x \in \Omega} \frac{f(x)}{g(x)}$

**Output:** a sample distributed according to $f$

**Function** *rejectionSample*$(f, g, M)$**:**
  **repeat**
    sample $x$ according to distribution $g$
    sample $\gamma \in_{\mathrm{U}} [0, 1]$
  **until** $\gamma \leq \frac{f(x)}{Mg(x)}$
  **return** $x$

**Lemma 15.** *Let $g$ be a strictly positive distribution over finite set $\Omega$, and let $f$ by a target distribution over $\Omega$. If $M \geq \sup_{x \in \Omega} \frac{f(x)}{g(x)}$ then the output of Algorithm 2 is distributed according to $f$.*

*Proof.* Let $X$ denote a random variable distributed according to $g$ and $U$ denote a uniform random variable on $[0, 1]$. The algorithm accepts a realization of $X$ whenever $\gamma \leq \frac{f(x)}{Mg(x)}$, so to prove the statement of the lemma we must establish that $\mathbb{P}\Big(X = x \,\Big|\, U \leq \frac{f(X)}{Mg(X)}\Big) = f(x)$. To do so, we use the law of total probability to deduce that

$$\mathbb{P}\Big(U \leq \frac{f(X)}{Mg(X)}\Big) = \sum_{x \in \Omega} \mathbb{P}\Big(U \leq \frac{f(X)}{Mg(X)} \,\Big|\, X = x\Big)\mathbb{P}(X = x) = M^{-1}$$

and therefore, by Bayes' rule, that

$$\mathbb{P}\Big(X = x \,\Big|\, U \leq \frac{f(X)}{Mg(X)}\Big) = \frac{\mathbb{P}(U \leq \frac{f(x)}{Mg(x)})\mathbb{P}(X = x)}{\mathbb{P}(U \leq \frac{f(X)}{Mg(X)})} = \frac{f(x)}{M\mathbb{P}(U \leq \frac{f(X)}{Mg(X)})} = f(x)$$

□

## 3.3 Mixing Times and Couplings

To use the Glauber dynamics to construct an approximate sampler for the Gibbs distribution, we have to determine for each $\epsilon \in (0,1)$ an upper bound on $\tau(\epsilon) \in \mathbb{N}$, the number of iterations of Algorithm 1 required before the Glauber dynamics $t$-step distribution of the Glauber dynamics starting in an arbitrary state $x$ approximates the Gibbs distribution within a multiplicative factor of $\epsilon$.

**Definition 16.** Suppose $(X_t)$ is a finite, irreducible, aperiodic Markov chain with state space $\Omega$, stationary distribution $\pi$ and transition matrix $P$. For $t \in \mathbb{N}$ and $x \in \Omega$, define $d_{\mathrm{TV}_x}(t) \coloneqq \|P^t_{x,\cdot} - \pi\|_{\mathrm{TV}}$ and $d_{\mathrm{TV}}(t) \coloneqq \max_{x\in\Omega} d_{\mathrm{TV}_x}(t)$. The *mixing time starting from* $x$ is a function $\tau_x : (0,1) \to \mathbb{N}$ with $\tau_x(\epsilon) \coloneqq \min\{t \,|\, d_{\mathrm{TV}_x}(t) \leq \epsilon\}$ for all $\epsilon \in (0,1)$. The $\epsilon$ *mixing time* $\tau(\epsilon)$ of $(X_t)$ is the first time $t$ such that $d_{\mathrm{TV}_x}(t) \leq \epsilon$ for all $x \in \Omega$, i.e. $\tau(\epsilon) \coloneqq \max_{x\in\Omega} \tau_x(\epsilon)$. The chain $(X_t)$ is *rapidly mixing* if $\tau(\epsilon)$ is polynomial in the size of the problem and $\log(\epsilon^{-1})$ for all $\epsilon \in (0,1)$.

**Lemma 17.** *Suppose $(X_t)$ is a finite, irreducible, aperiodic Markov chain with state space $\Omega$, stationary distribution $\pi$ and transition matrix $P$. Let $x \in \Omega$ and $0 \leq t \leq t'$. Then $d_{TV_x}(t') \leq d_{TV_x}(t)$.*

*Proof.* It suffices to show that $d_{\mathrm{TV}_x}(t) \leq d_{\mathrm{TV}_x}(t+1)$ for all $t \in \mathbb{N}$, as an inductive argument will extend this to the desired inequality. Now

$$d_{\mathrm{TV}_x}(t+1) = \frac{1}{2}\sum_{y\in\Omega} |P^{t+1}_{x,y} - \pi_y| = \frac{1}{2}\sum_{y\in\Omega} |\sum_{z\in\Omega} P_{z,y}(P^t_{x,z} - \pi_z)|$$

Using the triangle inequality, we conclude

$$\begin{aligned}\frac{1}{2}\sum_{y\in\Omega}|\sum_{z\in\Omega}P_{z,y}(P^t_{x,z}-\pi_z)| &\geq \frac{1}{2}\sum_{y\in\Omega}\sum_{z\in\Omega}P_{z,y}|P^t_{x,z}-\pi_z| \\ &= \frac{1}{2}\sum_{z\in\Omega}|P^t_{x,z}-\pi_z|\sum_{y\in\Omega}P_{z,y} \\ &= \frac{1}{2}\sum_{z\in\Omega}|P^t_{x,z}-\pi_z|\end{aligned}$$

□

Definition 16 has been chosen to coincide with the time constraints on a fully-polynomial approximate sampler (c.f. Definition 9). Indeed, if $G$ is a graph with maximum degree $\Delta$, $x$ is a colouring generated using the greedy algorithm and $\epsilon \in (0,1)$, then the algorithm which initializes the Glauber dynamics with $X_0 = x$ and outputs $X_{\tau(\epsilon)}$ is a fully-polynomial approximate sampler. We focus on one of the most basic and popular methods for bounding the mixing time, the method of coupling.

**Definition 18.** Suppose $\mathcal{M}$ is a Markov chain over $\Omega_{\mathcal{M}}$ with transition matrix $P$ and $\mathcal{N}$ is a Markov chain over $\Omega_{\mathcal{N}}$ with transition matrix $Q$. A *coupling* of $\mathcal{M}$ and $\mathcal{N}$ is a bivariate Markov chain $(X_t, Y_t)$ over $\Omega_{\mathcal{M}} \times \Omega_{\mathcal{N}}$ with the property that

$$\mathbb{P}(X_{t+1} = x' \,|\, (X_t, Y_t) = (x, y)) = P_{x,x'}$$

$$\mathbb{P}(Y_{t+1} = y' \,|\, (X_t, Y_t) = (x, y)) = Q_{y,y'}$$

for all $(x,y),(x',y') \in \Omega_{\mathcal{M}} \times \Omega_{\mathcal{N}}$. The transition matrix $R$ of the coupling is a matrix with rows and columns indexed by $\Omega_{\mathcal{M}} \times \Omega_{\mathcal{N}}$, so $R_{(x,y),(x',y')}$ is the probability of transitioning from $(x,y)$ to $(x',y')$ in one step and $R^t_{(x,y),(x',y')}$ is the $t$-step transition probability from $(x,y)$ to $(x',y')$.

Informally, couplings allow us to correlate the behaviour of two Markov chains. When the chains $\mathcal{M}$ and $\mathcal{N}$ are the same, we will additionally require that $X_t = Y_t \implies X_{t+1} = Y_{t+1}$. Such a coupling is called a *simple coupling*, and may be used to bound the mixing time of the chain as in the following foundational lemma.

**Lemma 19** (Coupling Lemma)**.** *Suppose $\mathcal{M} = (M_t)$ is a finite, irreducible, aperiodic Markov chain with state space $\Omega$, transition matrix $P$ and unique stationary distribution $\pi$, and let $(X_t, Y_t)$ be a*

*simple coupling of $\mathcal{M}$. If $\tau_{\mathcal{M}}(\epsilon)$ denotes the $\epsilon$-mixing time of $\mathcal{M}$ and $t : [0,1] \to \mathbb{N}$ satisfies*

$$\mathbb{P}(X_{t(\epsilon)} \neq Y_{t(\epsilon)} \,|\, (X_0, Y_0) = (x, y)) \leq \epsilon$$

*for all $x, y \in \Omega$ and $\epsilon \in (0,1)$ then $\tau_{\mathcal{M}}(\epsilon) \leq t(\epsilon)$.*

*Proof.* We take advantage of two techniques we have used before. First, we will use the alternate form of the total-variation distance given by Lemma 8, i.e. if $\pi_1$ and $\pi_2$ are distributions over a finite set $\Omega$, then $\|\pi_1 - \pi_2\|_{\text{TV}} = \max_{U \subseteq \Omega} |\pi_1(U) - \pi_2(U)|$. Second, for $x \in \Omega$ and $\epsilon \in (0,1)$, we will use the stationarity of $\pi$ to conclude that $\|P^{t(\epsilon)}_{x,\cdot} - \pi\|_{\text{TV}} \leq \epsilon$ when $\|P^{t(\epsilon)}_{x,\cdot} - P^{t(\epsilon)}_{y,\cdot}\|_{\text{TV}} \leq \epsilon$ for all $y \in \Omega$. Specifically, if $z \in \Omega$ then

$$\pi_z = \mathbb{P}(X_t = z \,\Big|\, X_0 \sim \pi) = \sum_{y \in \Omega} \mathbb{P}(X_t = z \,\Big|\, X_0 = y)\pi_y \leq \max_{y \in \Omega} \mathbb{P}(X_t = z \,\Big|\, X_0 = y)$$

The proof that $\min_{y \in \Omega} \mathbb{P}(X_t = z \,\Big|\, X_0 = y) \leq \pi_z$ is similar. Together, these imply that $\|P^{t(\epsilon)}_{x,\cdot} - \pi\|_{\text{TV}} \leq \max_{y \in \Omega} \|P^{t(\epsilon)}_{x,\cdot} - P^{t(\epsilon)}_{y,\cdot}\|_{\text{TV}} \leq \epsilon$ as desired.

It remains to show that $\max_{U \subseteq \Omega} |P^{t(\epsilon)}_{x,\cdot}(U) - P^{t(\epsilon)}_{y,\cdot}(U)| \leq \epsilon$, i.e.

$$\max_{U \subseteq \Omega} |\mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = x) - \mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = y)| \leq \epsilon \tag{4}$$

Suppose $x, y \in \Omega$ and $A \subseteq \Omega$. By the definition of coupling and the union bound,

$$\begin{aligned}
\mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = x) &= \mathbb{P}(X_{t(\epsilon)} \in U \,|\, (X_0, Y_0) = (x, y)) \\
&\geq \mathbb{P}(X_{t(\epsilon)} = Y_{t(\epsilon)} \wedge Y_{t(\epsilon)} \in U \,|\, (X_0, Y_0) = (x, y)) \\
&= 1 - \mathbb{P}(X_{t(\epsilon)} \neq Y_{t(\epsilon)} \vee Y_{t(\epsilon)} \notin U \,|\, (X_0, Y_0) = (x, y)) \\
&= 1 - \mathbb{P}(X_{t(\epsilon)} \neq Y_{t(\epsilon)} \,|\, (X_0, Y_0) = (x, y)) - \mathbb{P}(Y_{t(\epsilon)} \notin U \,|\, (X_0, Y_0) = (x, y))
\end{aligned}$$

By the definition of coupling, this is equal to $\mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = y) - \mathbb{P}(X_{t(\epsilon)} \neq Y_{t(\epsilon)} \,|\, (X_0, Y_0) = (x, y))$ which is bounded below by $\mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = y) - \epsilon$ because $\mathbb{P}(X_{t(\epsilon)} \neq Y_{t(\epsilon)} \,|\, (X_0, Y_0) = (x, y)) \leq \epsilon$. A symmetric argument gives that $\mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = y) \geq \mathbb{P}(M_{t(\epsilon)} \in U \,|\, M_0 = y) - \epsilon$ and Equation (4) follows immediately. □

Given a Markov chain $\mathcal{M}$ over a finite set $\Omega$ with transition matrix $P$, a basic version of a simple

coupling for $\mathcal{M}$ is the bivariate process $(X_t, Y_t)$ with transition matrix

$$P_{(x,y),(x',y')} \coloneqq \begin{cases} P_{x,x'}P_{y,y'} & x \neq y \\ P_{x,x'} & x = y \wedge x' = y' \\ 0 & x = y \wedge x' \neq y' \end{cases} \tag{5}$$

for all $x, y \in \Omega$. The chain $(X_t, Y_t)$ is a simple coupling of $\mathcal{M}$ which operates two independent copies of $\mathcal{M}$ until the two copies arrive at the same state. Once the components occupy the same state they have 'coupled' and transition together.

The basic simple coupling described by Definition 5 does not correlate the behaviour of the components of the coupling except after they have coupled, and will not, in general, be sufficient to demonstrate rapid mixing. The topic of simple couplings of the Glauber dynamics with the rapid mixing property, while outside the scope of this thesis, is itself a rich vein of research. We refer the interested reader to Jerrum [25], who is able to demonstrate the rapid mixing of the Glauber dynamics when $q > 2\Delta$.

## 3.4 Coupling from the Past

The coupling lemma raises the question of whether we may sample from the limiting distribution exactly by coupling many copies of the chain, one for each initial state, and then outputting the terminal colouring once we detect all of the components have coupled. While this does not work, with only a slight modification we arrive at the notion of coupling from the past as introduced by Propp and Wilson. Their approach begins with a complete coupling.

**Definition 20.** Suppose $\mathcal{M}$ is a finite Markov chain over $\Omega$ with transition matrix $P$. A *complete coupling* of $\mathcal{M}$ is a probability distribution $\mathcal{C}(P)$ over the space of functions $\Omega \to \Omega$ with the property that $\mathbb{P}(\phi(x) = y) = P_{x,y}$ for all $x, y \in \Omega$ and $\phi : \Omega \to \Omega$ drawn from the distribution. In the context of complete couplings, we will refer to the chain $\mathcal{M}$ as the *underlying chain* of the coupling.

Using a complete coupling, we may construct a Markov chain $(\Phi_0^t)$ over the space $\Omega^\Omega$ which is updated from time $t-1$ to time $t$ by setting $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$ for $\phi_t$ sampled from $\mathcal{C}(P)$. We will call $(\Phi_0^t)$ the *forward chain* for the complete coupling $\mathcal{C}(P)$. This chain has the property that $\mathbb{P}(\Phi_0^t(x_0) = y \,\big|\, \Phi_0^{t-1}(x_0) = x) = P_{x,y}$, so we may simulate the underlying chain $\mathcal{M}$ from state $x_0 \in \Omega$ as the sequence $\Phi_0^0(x), \Phi_0^1(x), \cdots$. Similarly, we may construct a simple coupling $(X_t, Y_t)$ starting

from state $(x, y) \in \Omega^2$ by evolving $(x, y)$ to $(\Phi_0^1(x), \Phi_0^1(y))$ and so on. If there exists a time $t_0 \in \mathbb{N}$ such that $|\Phi_0^{t_0}(\Omega)| = 1$ then, for all $t \geq t_0$, we know that $\Phi_0^t(x) = \Phi_0^{t_0}(y)$. In particular, for all $x, y \in \Omega$, we have the bound

$$\mathbb{P}(X_t \neq Y_t \,|(X_0, Y_0) = (x, y)) \leq \mathbb{P}(\Phi_0^{t_0} \text{ is not constant function}) \tag{6}$$

Comparing this to the coupling lemma (c.f. Lemma 19), we see that it provides a method to bound the mixing time of the chain.

Unfortunately, it is not the case that we may sample exactly from the distribution $\pi$ by computing $\Phi_0^t$ for increasingly large values of $t$ until $\Phi_0^t$ is a constant function. To see why, consider a Markov chain with a state $z_1$ that has a unique predecessor $z_0$. The proposed approach to sampling will never return this element because if $x, y \in \Omega$ then $\Phi_0^t(x) = z_1 = \Phi_0^t(y)$ implies $\Phi_0^{t-1}(x) = z_0 = \Phi_0^{t-1}(y)$.

To remove the dependence of the output on the stopping condition, we introduce a new chain based on the complete coupling $\mathcal{C}(P)$. The *reverse chain* for the complete coupling $\mathcal{C}(P)$ is the Markov chain $(\Psi_{-t}^0)$ with transitions $\Psi_{-t}^0 = \Psi_{-(t-1)}^0 \circ \phi_{-t}$ for $\phi_{-t}$ sampled from $\mathcal{C}(P)$. In contrast to the forward chain, if $(\Psi_{-t}^0)$ has completely coupled at time $t_0$ then $\Psi_{-t}^0 = \Psi_{-t_0}^0$ for all $t \geq t_0$. This allows us to move to the limit and recover the limiting distribution.

**Algorithm 3:** basicCFTP($\mathcal{C}(P)$)

**Input :** a complete coupling of a finite, irreducible, aperiodic Markov chain $(M_t)$ with transition matrix $P$

**Output:** a sample from the unique stationary distribution of $(M_t)$

```
Function basicCFTP(C(P)):
    t ⟵ 0
    Ψ_0^0 ⟵ the identity map
    do
        sample φ from C(P)
        Ψ^0_{-(t+1)} ⟵ Ψ^0_{-t} ∘ φ
        t ⟵ t + 1
    while Ψ^0_{-t} is not a constant map
    return the unique element in the image of Ψ^0_{-t}
```

**Lemma 21.** *Let $\mathcal{M}$ be a finite, irreducible, aperiodic Markov chain over $\Omega$ with transition matrix $P$, and let $\mathcal{C}(P)$ be a complete coupling of $\mathcal{M}$. Then the coupling from the past algorithm (Algorithm 3) returns a value with probability 1 and this value is distributed according to the limiting distribution $\lim_{t\to\infty} P_{x,\cdot}^t$.*

*Proof.* Let $T$ be a random variable denoting the first time that $|\Psi_{-t}^0(\Omega)| = 1$. We begin by checking

that $\mathbb{P}(T < \infty) = 1$. To see that this is the case, we use the irreducibility and aperiodicity of the chain to choose an integer $L > 0$ such that $\mathbb{P}(M_L = y \,|\, M_0 = x) > 0$ for all $x, y \in \Omega$. For integers $i > 0$, denote by $\Psi_{-iL}^{-(i-1)L}$ the map $\phi_{-(i-1)L} \circ \cdots \circ \phi_{-iL}$. The constant $L$ has been chosen to ensure that $\Psi_{-iL}^{-(i-1)L}$ is constant with probability greater than 0 for all $i > 0$, from which we may conclude that $\mathbb{P}(T < \infty) = 1$.

Having established that $T$ is finite with probability 1, we prove that the value output by the algorithm at time $T$ is distributed according to the limiting distribution of the chain. Note that the distribution of $\Psi_{-t}^{0}(x)$ is equal to the distribution of $\Phi_0^t(x)$. Alongside the fact that the unique element in the image does not change with further pre-composition, this implies

$$\mathbb{P}(\Psi_{-T}^{0}(x) = y) = \lim_{t \to \infty} \mathbb{P}(\Psi_{-t}^{0}(x) = y) = \lim_{t \to \infty} \mathbb{P}(\Phi_0^t(x) = y) = \pi(y).$$

□

We are interested in sampling efficiently, and therefore in maximizing the probability that functions $\phi$ drawn from the complete coupling are constant. In a complete coupling of the Glauber dynamics for the anti-ferromagnetic Potts model, the probability that $\phi$ is constant is zero. The following lemma establishes that if $T$ is an integer with $T \geq 1$ then we may replace the single step forward in each iteration in Algorithm 3 with a $T$-step walk of the forward chain $(\Phi_0^t)$ and retain the same limiting distribution.

**Lemma 22.** *Suppose $\mathcal{M}$ is a finite, irreducible, aperiodic Markov chain over $\Omega$ with transition matrix $P$ and fix an integer $T \geq 1$. Then the Markov chain $\mathcal{N}$ with transition matrix $P^T$ is finite, irreducible and aperiodic with the same limiting distribution as $\mathcal{M}$.*

*Proof.* Suppose $x, y \in \Omega$. By the definition of irreducibility of $\mathcal{M}$, there exists $t_0 > 0$ such that $P_{x,y}^{t_0} > 0$. If $k = \lceil \frac{t_0}{T} \rceil$ then $(P^T)_{x,y}^{k} > P_{x,y}^{t_0}(P_{y,y})^{\lceil kT \rceil - t_0}$, which is greater than 0 because $\mathcal{M}$ is aperiodic, so $\mathcal{N}$ is irreducible. Moreover, $P_{x,x}^{T} > (P_{x,x})^T > 0$ for all $x \in \Omega$ because $\mathcal{M}$ is aperiodic, so $\mathcal{N}$ is aperiodic.

The fact that the limiting distribution exists implies that there is a unique stationary distribution and that it is equal to the limiting distribution. Both $\mathcal{M}$ and $\mathcal{N}$ are stationary with respect to the same distribution, so they must have the same limiting distribution. □

If $\mathcal{M}$ is a finite, irreducible, aperiodic Markov chain with transition matrix $P$, Lemma 22 allows

us to apply Lemma 21 to a complete coupling $\mathcal{C}(P^T)$. Assuming we have a complete coupling $\mathcal{C}(P)$, the simplest way to do this is to sample $\phi_1, \cdots, \phi_T$ from the complete coupling $\mathcal{C}(P)$ and then view $\phi_T \circ \cdots \circ \phi_1$ as a sample from $\mathcal{C}(P^T)$. This produces Algorithm 4, a generalization of Algorithm 3.

**Algorithm 4:** $\underline{\text{CFTP}}(\mathcal{C}(P), T)$

**Input :** a complete coupling of a finite, irreducible, aperiodic Markov chain $(M_t)$ with transition matrix $P$ and integer $T > 0$

**Output:** a sample from the unique stationary distribution of $(M_t)$

**Function** $\underline{CFTP}(\mathcal{C}(P), T)$**:**
- $i \longleftarrow 0$
- $\Psi_0^0 \longleftarrow$ the identity map
- **do**
  - $i \longleftarrow i + 1$
  - sample $\phi_0, \cdots, \phi_T$ from $\mathcal{C}(P)$
  - $\Phi_0^T \longleftarrow \phi_T \circ \cdots \circ \phi_0$
  - $\Psi_{-iT}^0 \longleftarrow \Psi_{-(i-1)T}^0 \circ \Phi_0^T$
- **while** $\Phi_0^T$ is not a constant map
- **return** the unique element in the image of $\Psi_{-iT}^0$

The parameter $T$ determines the length of time that the forward simulation is run to generate $\Phi_0^T$. If the probability that $\Phi_0^T$ is not a constant is bounded away from 0 by a polynomial in the size of the graph $n$, then the expected number of samples $\Phi_0^T$ is polynomial in $n$. Later, we will set $T$ so that $\Phi_0^T$ is constant with probability greater than $\frac{1}{2}$.

## 3.5 Computational Considerations

We now turn to the first of two computational issues we are faced with when attempting to apply the coupling from the past framework to the problem of perfect sampling from the Gibbs distribution on the anti-ferromagnetic Potts model. If $q, \Delta \geq 3$, then the size of the state space $\Omega_q$ of the $q$-state Potts model on an $n$-vertex graph $G$ grows exponentially with $n$. This raises the problem of storing and computing samples from the complete coupling. In order to do this efficiently, we follow Bhandari and Chakraborty [2] and use random seeds $\alpha$ which may be decoded into samples from the complete coupling. The form and distribution of $\alpha$ will vary between the sampling algorithms that we construct, but $\alpha$ will generally be a tuple consisting of a vertex, one or more random numbers drawn i.i.d from from the uniform distribution over the unit interval, and a tuple of colours. The generation of the $\alpha$ is the only source of randomness in the construction of $\phi$, so we may determine the distribution of samples $\phi$ using only the properties of the seeding distribution and the decoding algorithm.

Even when $\phi$ can be computed and stored, checking when the forward chain $(\Phi_0^t)$ is constant is

itself a non-trivial task. For some models - including the ferromagnetic Ising model - the state space admits a partial ordering $\leq$ with greatest and least[8] elements $\hat{0}$ and $\hat{1}$. In these cases, a monotone Monte Carlo algorithm [32] is able to detect mixing by simulating only two chains. Monotone algorithms respect the partial order of the state space in the sense that for $\phi$ sampled from a monotone complete coupling, $\phi(x) \leq \phi(y)$ whenever $x \leq y$. A sampler with this property can detect mixing using only the predicate $\Phi_0^t(\hat{0}) = \Phi_0^t(\hat{1})$. For more on a monotonic formulation of the Potts model, see [13] or [20].

The anti-ferromagnetic Potts model is not known to admit a monotone Monte Carlo algorithm. Instead, this model requires a different approach, a type of coupling known as a bounding chain.

**Definition 23** (Huber, [22])**.** Fix $q, \Delta \geq 3$ and let $G = (V, E)$ be a graph with maximum degree $\Delta$. Suppose $\mathcal{M}$ is a Markov chain with state space $\Omega_q$ and transition matrix $P$, and let $\mathcal{C}(P)$ be a complete coupling of $\mathcal{M}$. Suppose further that $\mathcal{L} = (L_t)$ is a Markov chain with state space $(2^{[q]})^V$. We will say that $L_t$ bounds a colouring $x \in \Omega_q$ if $x(v) \in L_t(v)$ for all $v \in V$; in this case, we write $x \in L_t$. The chain $\mathcal{L}$ is a *bounding chain* for the forward chain $(\Phi_0^t)$ associated with $\mathcal{C}(P)$ if there exists a coupling $(\Phi_0^t, L_t)$ of $(\Phi_0^t)$ and $\mathcal{L}$ with the property that

$$\Phi_0^t(x) \in L_t \quad \forall x \in \Omega \implies \Phi_0^{t+1}(x) \in L_{t+1} \quad \forall x \in \Omega$$

The lists $L_t(v)$ are called *bounding lists* because if $L_0 = [q]^V$ then $\Phi_0^t(\Omega_q) \subseteq L_t$ for all $t$.

We may use the bounding chain to efficiently determine whether $\Phi_0^T$ is constant by checking the condition $\prod_{v \in V} |L_T(v)| = 1$. This is the last of the technical elements necessary to construct a perfect sampler.

## 3.6 Huber's Perfect Sampling Algorithm

Huber [22] was the first to publish a perfect sampling algorithm for the anti-ferromagnetic Potts model. His algorithm is based on the Glauber dynamics and relies on rejection sampling to construct a complete coupling. In what follows, we will work with a fixed $n$-vertex graph $G = (V, E)$ of maximum degree $\Delta \geq 3$, a fixed number of colours $q \geq 2$ and parameter $B \in (0, 1)$. Additional constraints on $B, \Delta$ and $q$ will be stated when required. If $L_t$ denotes the state of the bounding list, $v$ is a vertex and $c \in [q]$, we will use the term $m_Q(L_t, v, c)$ to denote the number of neighbours $w$ of

[8] That is $\hat{0} \leq x \leq \hat{1}$ for all $x \in \Omega$.

$v$ with $L_t(w) = \{c\}$ and $m_L(L_t, v, c)$ to denote the number of neighbours $w$ of $v$ with $c \in L_t(w)$.

**Algorithm 5:** glauber.seed()

**Output:** a seed $\alpha = (v, (c_k, \gamma_k)_{1 \le k \le l})$ where $l \ge \Delta + 1$

**Function** *glauber.seed*():
- sample $v \in_U V$
- $l \longleftarrow 0$
- **repeat**
  - $l \longleftarrow l + 1$
  - sample $c_l \in_U [q]$ and $\gamma_l \in_U [0, 1]$
- **until** $\{c_k\}_{1 \le k \le l}$ contains more than $\Delta$ distinct colours or $\gamma_l \le B^{\Delta}$
- **return** $\alpha = (v, (c_k, \gamma_k)_{1 \le k \le l})$

**Algorithm 6:** glauber.FC$(x, \alpha)$

**Input :** a seed $\alpha = (v, (c_k, \gamma_k)_{1 \le k \le l})$ produced by glauberSeed()

**Output:** a colouring $y \in \Omega_q$

**Function** *glauber.FC*$(x, \alpha)$:
- let $m(x, v, c)$ denote the number of neighbours $w$ with $x(w) = c$
- $y|_{V \setminus v} \longleftarrow x|_{V \setminus v}$
- $y(v) \longleftarrow c_k$ for $1 \le k \le l$ the smallest index such that $\gamma_k \le B^{m(x,v,c_k)}$
- **return** $y$

**Lemma 24.** *If $\alpha$ is generated using glauber.seed() then Algorithm 6 defines a complete coupling of the Glauber dynamics, i.e. if $P$ is the transition matrix of the Glauber dynamics, $x$ is a colouring and $\phi(x) =$ glauber.FC$(x, \alpha)$ then $\mathbb{P}(\phi(x) = y) = P_{x,y}$ for all $y \in \Omega_q$.*

*Proof.* Let $g$ denote the conditional distribution over $[q]$, $f(c) = \frac{B^{m(x,v,c)}}{\sum_{c' \in [q]} B^{m(x,v,c)}}$ for $c \in [q]$ and $M = \frac{q}{\sum_{c \in [q]} B^{m(x,v,c)}}$. Then Lemma 15 implies $y(v)$ is distributed according to $f$. Since the vertex $v$ is chosen uniformly at random, this implies $\mathbb{P}(\phi(x) = y) = P_{x,y}$. □

Algorithm 5 chooses a vertex to update and proposes new colours for this vertex according to a distribution shared by all the colourings in $\Omega_q$. Algorithm 6 is responsible for further refining the distribution in accordance with a specific colouring $x$. Using the information contained in the bounding chain, we are able to determine when how far into the seed we must progress before $\gamma_k \le B^{m(x,v,c_k)}$ for all $x \in \Omega_q$.

**Lemma 25.** *Suppose $L_{t-1} \in (2^{[q]})^V$ and $\alpha$ is generated using glauber.seed(). Recall that $x \in L_{t-1}$ if $x(v) \in L_{t-1}(v)$ for all $v \in V$. If $y =$ glauber.FC$(x, \alpha)$ and $L_t =$ glauber.BC$(L_{t-1}, \alpha)$ then $y \in L_t$. In particular, let $\phi_t(x) =$ glauber.FC$(x, \alpha_t)$ for all $x \in \Omega_q$. If $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$ then $\Phi_0^t(\Omega_q) \subseteq L_t$.*

*Proof.* Let $v$ denote the vertex to be updated, suppose $y(v) = c_k$ and ignore the condition on Line 8 for the moment. To see that $c_k$ will be added to $L_t(v)$, note that $m_Q(L_{t-1}, v, c_k) \le m(x, v, c_k)$. If

**Algorithm 7:** glauber.BC$(L_{t-1}, \alpha_t)$

**Input :** a seed $\alpha = (v, (c_k, \gamma_k)_{1 \le k \le l})$ produced by glauberSeed$(G, B, \Delta, q)$
**Output:** a state of the bounding chain $L_t$

1 **Function** *glauber.BC*$(L_{t-1}, \alpha)$**:**
2 $\quad L_t|_{V \setminus \{v\}} \longleftarrow L_{t-1}|_{V \setminus \{v\}}$
3 $\quad L_t(v) \longleftarrow \emptyset$
4 $\quad$ **for** $1 \le k \le l$ **do**
5 $\quad\quad$ **if** $\gamma_k \le B^{m_Q(L_{t-1}, v, c_k)}$ **then**
6 $\quad\quad\quad$ add $c_k$ to $L_t(v)$
7 $\quad\quad$ **end**
8 $\quad\quad$ **if** $\gamma_k \le B^{m_L(L_{t-1}, v, c_k)}$ **then**
9 $\quad\quad\quad$ break
10 $\quad\quad$ **end**
11 $\quad$ **end**
12 $\quad$ **return** $L_t$

$\gamma_k \le B^{m(x,v,c_k)}$ then $\gamma_k \le B^{m_Q(L_{t-1}, v, c_k)}$, so $c_k \in L_t(v)$.

To establish that the condition on Line 8 will not trigger too soon, note that $m_L(L_{t-1}, v, c_k) \ge m(x, v, c_k)$. If $\gamma_k \le B^{m_L(L_{t-1}, v, c_k)}$ then $\gamma_k \le B^{m(x,v,c_k)}$ so the condition on Line 8 will not trigger unless every possible chain in the complete coupling has found a new assignment for vertex $v$. □

It remains to set $T$ such that $\Phi_0^T$ is constant with probability at least $\frac{1}{2}$ (c.f. Section 3.4). To do so, we split $T$ into two periods $T_1 + T_2$ and consider different aspects of the behaviour of the algorithm in different periods. After the first phase of the algorithm, we will be able to ensure that with probability at least $\frac{3}{4}$ all of the bounding chains satisfy a certain condition. Using this condition, we then prove that with probability at least $\frac{2}{3}$ the second phase terminates with $\prod_{v \in V} |L_T| = 1$.

**Theorem 26.** *Suppose $\Delta, q \ge 3$ and $B \in (0, 1)$ and let $H_n$ denote the $n^{th}$ Harmonic number. Suppose further that $q, \Delta$ and $B$ satisfy the following constraints:*[9]

| *# of colours* | *inverse-temperature* |
|---|---|
| $q \ge \Delta(\Delta + 2)$ | $B > 0$ |
| $\Delta < q < \Delta(\Delta + 2)$ | $B^\Delta > 1 - \frac{q}{\Delta(\Delta+2)}$ |
| $q \le \Delta$ | $B^\Delta > 1 - \frac{1}{\Delta q}$ |

*There exists a constant $\beta \in (0, 1)$ such that $\mathbb{P}(\prod_{v \in V} |L_T(v)| \ne 1) \le \frac{1}{2}$ when Huber's perfect sampling algorithm (c.f. Algorithms 5, 6 and 7) is run on an $n$-vertex graph $G$ with maximum degree $\Delta$ for $T = 4nH_n + \lceil -\log_\beta 3n \rceil$ steps.*

[9] The statement of Theorem 26 differ from those in [22] by the addition of a power of $\Delta$ in the condition on $B$ and in the use of two phases. To the best of our knowledge, the stronger bounds on $B$ reflect the regime under which Algorithms 5, 6 and 7 provide a perfect sampling algorithm. Our proof differs slightly in the addition of a warm-up phase to ensure the condition that the bounding lists all have size at most $\Delta + 1$.

Before we prove this result, we establish a concavity result which allows us to bound the normalizing constant in the conditional distribution at each vertex.

**Lemma 27.** *Fix integers $q \geq d \geq 1$ and $B \in (0,1)$. If $m_1, \cdots, m_q \in \mathbb{N}$ with $\sum_{1 \leq i \leq q} m_i = d$ then*

$$q - d(1-B) \leq \sum_{1 \leq i \leq q} B^{m_i}$$

*Proof.* Let $f : \mathbb{N}^q \to \mathbb{R}$ with $f(n_1, \cdots, n_q) = \sum_{1 \leq i \leq q} B^{n_i}$. We wish to examine the minimum of $f$ over the subspace $\sum_{1 \leq i \leq q} n_i = d$. To this end, suppose $\{m_i\}_{1 \leq i \leq q}$ is in the subspace, and suppose further that there exist $1 \leq j, k \leq q$ with $m_j - m_k > 1$. Then

$$\sum_{\substack{1 \leq i \leq q \\ i \notin \{j,k\}}} B^{m_i} + B^{m_j} + B^{m_k} \geq \sum_{\substack{1 \leq i \leq q \\ i \notin \{j,k\}}} B^{m_i} + B^{m_j - 1} + B^{m_k + 1}$$

Thus, $f$ achieves its minimum over the subspace $\sum_{1 \leq i \leq q} n_i = d$ on a set $\{m_i\}_{1 \leq i \leq q}$ with $|m_j - m_k| \leq 1$ for all $1 \leq j, k \leq q$. If $q \geq n$ then there must exist $1 \leq j \leq q$ with $m_j = 0$, so $m_i \in \{0, 1\}$ for all $1 \leq i \leq q$. At most $n$ values $m_i$ may be 1, from which we conclude $\sum_{1 \leq i \leq q} B^{m_i} \geq q - d(1-B)$ for all sets of natural numbers $\{m_i\}_{1 \leq i \leq q}$ satisfying $\sum_{1 \leq i \leq q} m_i = d$. □

*Proof of Theorem 26.* At the end of the first phase of the algorithm, we would like to ensure that with probability at least $\frac{3}{4}$ the bounding lists all have size at most $\Delta + 1$. Algorithms 5 and 7 reduce the bounding list associated with a vertex to size at most $\Delta + 1$, so it suffices to ensure that every vertex is visited by the algorithm at least once before $T_1$. This is the coupon collector problem: if $V_1, \cdots, V_n$ are random variables denoting the amount of time taken to visit the $i$th vertex after the $i - 1$st vertex has been visited, then

$$E\Big[\sum_{1 \leq i \leq n} V_i\Big] = \sum_{1 \leq i \leq n} E[V_i] = \sum_{1 \leq i \leq n} \frac{n}{n - i + 1} = nH_n$$

$H_n$ here denotes the $n^{\text{th}}$ Harmonic number, which grows $\mathrm{O}(\log n)$. Using Markov's inequality, we have $\mathbb{P}\Big(\sum_{1 \leq i \leq n} V_i \geq 4nH_n\Big) \leq \frac{1}{4}$. Thus, it suffices to set $T_1 = 4nH_n$.

In the second period, we take advantage of the fact that the bounding lists have size at most $\Delta + 1$. Let $W_t$ denote the number of nodes with $|L_{T_1+t}(v)| > 1$. We begin with $W_0 \leq n$, and seek the expected time until $W_t = 0$. To do so, we first bound $E[W_{t+1} \,\big|\, L_{T_1+t}] \leq \beta W_t$ for some $\beta \in (0,1)$. This allows us to bound $E[W_{t+1}] = E[E[W_{t+1} \,\big|\, L_{T_1+t}]] \leq \beta E[W_t]$, which implies $E[W_t] \leq \beta^t E[W_0]$.

Using the bound $W_0 \leq n$, we may deduce that $E[W_t] \leq \beta^t n$ and therefore, by Markov's inequality, that $\mathbb{P}(W_t > 1) \leq \beta^t n$. In particular, conditional on the algorithm visiting every vertex in the first phase, it suffices to simulate the second phase for $\lceil -\log_\beta 3n \rceil$ steps to ensure that we detect coupling with probability at least $\frac{2}{3}$. Since every vertex in the first phase is visited with probability at least $\frac{3}{4}$, we visit every vertex in the first phase and then detect coupling with probability at least $\frac{1}{2}$.

The constant $\beta$ differs depending on whether $q > \Delta$ or $q \leq \Delta$. We consider the case $q > \Delta$ first, dealing with $q \leq \Delta$ at the end. Our analysis focusses on taking an expectation over one iteration of the algorithm; in particular, we follow the algorithm and reduce the size of the bounding list to zero when a vertex is chosen. We then consider the probability that more than one colour is added to the bounding list associated with the vertex chosen.

In order to produce a bounding list of size two or greater, the first colour that the algorithm proposes which is not rejected on the basis of Line 5 must be contained in a bounding list of size two or greater associated with a neighbour of $v$. There are at most $\sum_{\{w \in N(v) \,|\, |L_{T_1+t}(w)|>1\}} |L_{T_1+t}(w)|$ such colours.The probability that a colour $c$ is the first colour proposed by the algorithm which isn't rejected in Line 5 is $B^{m_Q(L_{t-1},v,c)} \setminus \sum_{c' \in [q]} B^{m_Q(L_{t-1},v,c')}$ (c.f. Lemma 15), which we may bound above by $(q - \Delta(1-B))^{-1}$ using Lemma 27. Assuming $c$ and $\gamma$ are the first pair which pass Line 5, the probability that they do not pass Line 8 is bounded above by $(1 - B^\Delta)$. Thus, we have

$$E[W_{t+1} \,\Big|\, L_{T_1+t}] \leq W_t - \frac{W_t}{n} + \sum_{v \in V} \frac{(1-B^\Delta)}{n(q-\Delta(1-B))} \sum_{\substack{w \in N(v) \\ |L_{T_1+t}(w)|>1}} |L_{T_1+t}(w)|$$

Assuming that every vertex was visited in the first stage of the algorithm, we can bound the right hand side above by $W_t - \frac{W_t}{n} + \frac{(\Delta+1)(1-B^\Delta)}{n(q-\Delta(1-B))} \sum_{v \in V} \sum_{\{w \in N(v): |L_{T_1+t}(w)|>1\}} 1$. Now the double sum is the number of neighbours of vertices in $W_t$, which is bounded above by $\Delta W_t$, so

$$E[W_{t+1} \,\Big|\, L_{T_1+t}] \leq W_t - \frac{W_t}{n} + \frac{W_t}{n}\Big[\frac{(\Delta+1)\Delta(1-B^\Delta)}{q-\Delta(1-B)}\Big] = \beta W_t$$

for $\beta = 1 - \frac{1-(\Delta+1)\Delta(1-B^\Delta)/(q-\Delta(1-B))}{n}$. This is in $(0,1)$ when $\frac{(\Delta+1)\Delta(1-B^\Delta)}{q-\Delta(1-B)} < 1$. Using the fact that $q - \Delta(1-B^\Delta) < q - \Delta(1-B)$, we have

$$\frac{(\Delta+1)\Delta(1-B^\Delta)}{q-\Delta(1-B)} < 1 - \frac{q-\Delta(1-B^\Delta)-(\Delta+1)\Delta(1-B^\Delta)}{q-\Delta(1-B)}$$

The right hand side is less than 1 when $q - (\Delta+2)\Delta(1-B^\Delta) > 0$, which we may rewrite as the

condition $B^\Delta > 1 - \frac{q}{\Delta(\Delta+2)}$. When $q \geq \Delta(\Delta+2)$, this further reduces to $B \geq 0$.

When $q \leq \Delta$, we bound $|L_t(w)|$ by $q$ rather than $\Delta + 1$ and we bound the probability that a colour is the first to be proposed and not rejected in Line 5 by one instead of $\frac{1}{q-\Delta(1-B)}$. This gives

$$E[W_{t+1} \Big| L_{T_1+t}] \leq W_t - \frac{W_t}{n} + \frac{W_t}{n}\Big[q\Delta(1-B^\Delta)\Big] = \beta W_t$$

for $\beta = 1 - \frac{1-q\Delta(1-B^\Delta)}{n}$, which is in $(0,1)$ when $B^\Delta > 1 - \frac{1}{\Delta q}$. □

# 4 Perfect Sampling from the Potts Model

The generalized form of coupling from the past defined in Algorithm 4 does not rely critically on the fact that the samples $\phi$ are drawn from a complete coupling of a Markov chain with the Gibbs stationary distribution. Indeed, when moving from the basic form of the coupling from the past algorithm to the general form in Algorithm 4, we relied only on the fact that the $T$-step Markov chain $\mathcal{N}$ with transition matrix $P^T$ has the same unique stationary distribution as the chain $\mathcal{M}$ with transition matrix $P$ and that the complete coupling of $\mathcal{M}$ could be used to make a complete coupling of $\mathcal{N}$. Bhandari and Chakraborty have recently re-instigated interest in the problem of perfect sampling from the uniform distribution over proper $q$-colourings for integers $q \geq 3$ by replacing both the forward chain and bounding chain with a bivariate process. By updating the state of the bivariate process using both the current state and the time and thereby violating the Markov property, they are able to derive tighter bounds on the image of the forward process $(\Phi_0^t)$ and thereby apply more sophisticated coupled updates.

In this chapter, we generalize their results to the $q$-state Potts model with parameter $B \in (0, 1)$. In the process we formalize their results into the language of stochastic processes and functions. Given an $n$-vertex graph $G = (V, E)$ with maximum degree $\Delta$, our algorithm samples a vertex $v$ on each iteration which is then passed to sub-algorithms used to generate the seeds $\alpha$. Seeds are then decoded into samples $\phi$ which will be used to maintain the forward chain. These samples themselves depend heavily on the state of the bounding list and the time, but conditional on $v$ being the vertex updated, the distribution from which they are drawn satisfies

$$\mathbb{P}\Big(\phi(x)(v) = c \,\Big|\, v \in \alpha\Big) = \frac{B^{m(x,v,c)}}{\sum_{c' \in [q]} B^{m(x,v,c')}} \tag{7}$$

Equation 7 says that the stochastic process which updates $x, \phi_1(x), \phi_2 \circ \phi_1(x), \cdots$ applies the Glauber dynamics to a vertex which is not necessarily chosen uniformly from the vertex set. This process, which we will denote by $\mathcal{M}$, is analogous to the underlying chain described in Definition 20, but it is not a Markov chain because the choice of vertex is time dependent. We can recover a complete coupling of a Markov chain $\mathcal{N}$ from our complete coupling of $\mathcal{M}$ by moving to the forward process for this complete coupling. This is the analogous stochastic process to the forward chain described after Definition 20; $\mathcal{N}$ uses $T$ samples $\phi$ to update in one step. The key difference between $\mathcal{M}$ and $\mathcal{N}$ is that $\mathcal{N}$ 'resets' the index of $\mathcal{M}$ after each step and therefore $\mathcal{N}$ has the Markov property. We

will show that $\mathcal{N}$ is irreducible and aperiodic, and that the limiting distribution of $\mathcal{N}$ is the Gibbs distribution. The samples $\phi$ may be used to construct a complete coupling of $\mathcal{N}$ (c.f. the discussion following Lemma 22), so we may apply Lemma 21 and use the coupling from the past framework to generate a perfect sample.

Let $P(t)$ denote the transition matrix of $\mathcal{M}$ at step $t$. To ensure that $\mathcal{N}$ is stationary with respect to the Gibbs measure, we require that $P(t)$ is stationary with respect to the Gibbs measure for all $t \geq 0$. In particular, we will require that $P(t)$ has the form of a Glauber update distributed uniformly across some subset of $V$. For $t \leq T_1 = n + |E|$, this set will be a singleton set; afterwards, we will choose $v$ from $V$.

**Lemma 28.** *Fix $q \geq 2, \Delta \geq 3$ and $B \in (0,1)$. Given a graph $G = (V, E)$ with maximum degree $\Delta$, let $\Omega_q$ denote the set of $q$-colourings of $G$. Suppose $\mathcal{M}$ is a finite stochastic process over $\Omega_q$ with transition matrix at time $t$ given by $P(t)$, and let*

$$Q(v)_{x,y} \coloneqq \begin{cases} 0 & x|_{V\setminus\{v\}} \neq y|_{V\setminus\{v\}} \\ \frac{B^{m(x,v,y(v))}}{\sum_{c\in[q]} B^{m(x,v,c)}} & x|_{V\setminus\{v\}} = y|_{V\setminus\{v\}} \end{cases}$$

*for $x, y \in \Omega_q$. If $P(t)$ has the form $\sum_{v\in S} p_{v,t} Q(v)$ for some set $S \subseteq V$ and $\{p_{v,t}\}_{v\in S}$ satisfying $\sum_{v\in V} p_{v,t} = 1$, then $P(t)$ is stationary with respect to the Gibbs measure $\mu_G$ for the $q$-state Potts model with parameter $B$.*

*Proof.* It suffices to show that $Q(v)$ is stationary with respect to the Gibbs measure, as linearity will imply $\mu_G P(t) = \sum_{v\in S} p_{v,t} \mu_G Q(v) = \mu_G$. The proof that $Q(v)$ is stationary with respect to the Gibbs measure is the same as Lemma 14. □

Lemma 28 establishes that $\mathcal{N}$ is stationary with respect to the Gibbs distribution, as long as $v_t$ is chosen according to a distribution which varies only with $t$ and Condition 7 is met. This provides us with a lot of freedom in the details of the updates that we perform. In Section 4.1, we outline the two types of updates that we will use in our algorithm and show that they satisfy these conditions. In Section 4.2, we describe how these updates are used in the two distinct phases of the algorithm, and we use Lemma 28 to prove the following result.

**Theorem 29.** *Suppose $q, \Delta \geq 3$ are integers with $q > 2\Delta$, and let $B$ be a real in the interval $(1 - \frac{q-2\Delta}{\Delta}, 1)$. Then there exists a randomized algorithm which, on input an $n$-vertex graph $G$ of*

*maximum degree $\Delta$, outputs in expected time $O(n^2)$ a sample from the Gibbs distribution for the $q$-state Potts model on $G$ with parameter $B$.*

## 4.1 Types of Updates

Let $G = (V, E)$ be a graph, $q > 2\Delta \geq 3$ and $\frac{3\Delta - q}{\Delta} < B$. We will construct a stochastic process $(\Phi_0^t, L_t)$ over $\Omega_q^{\Omega_q} \times (2^{[q]})^V$. In state $(\Phi_0^{t-1}, L_{t-1})$, this process updates by generating a seed $\alpha_t$. Unlike Huber's algorithm, the forward chain will be updated using two different forms of updates across two different phases. The first form of update - contract - is only used in the first phase of updates. When applied to state $(\Phi_0^t, L_t)$, the contract update reduces the bounding list associated of one vertex $v$ to the form $L_t(v) = A \cup \{c_1\}$ for a set $A$ of size $\Delta$. We will say that a colour $c$ is *fixed from the perspective of* $v$ if

$$c \in F_{L_t}(v) \coloneqq [q] \Big\backslash \bigcup_{\substack{w \in N(v) \\ |L_t(w)| > 1}} L_t(w)$$

By repeatedly applying compress to the neighbors of $v$, we ensure that at least $\Delta(1-B)$ colours are fixed by the bounding lists associated with the neighbors of $v$. Once sufficiently many colours are fixed from the perspective of $v$, we may apply a contract update to reduce the size of the bounding list $L_t(v)$ down to at most 2. Once all the vertices have $|L_t(v)| \leq 2$ the second phase of the algorithm starts. In this phase, the contract update is repeatedly applied to vertices chosen at random. We will show that the bounding chain detects complete coupling with probability at least $\frac{1}{2}$ when $q > 2\Delta$, $B > 1 - \frac{q-2\Delta}{\Delta}$ and $T = n + |E| + n^2\left(\frac{q-\Delta(1-B)}{q-\Delta(3-B)}\right)$.

The random seeds $\alpha_t$ will differ between the two types of updates. For each type of update, we will provide an algorithm for generating the seed, either compress.seed or contract.seed. We will also provide two algorithms for each type of update which are responsible for maintaining the forward process or the bounding process: compress.FP, compress.BP, contract.FP and contract.BP. These provide a means of decoding the seed $\alpha_t$ to determine the step $(\Phi_0^{t-1}, L_{t-1}) \to (\Phi_0^t, L_t)$.

In what follows, we will always assume that we are working with $\Delta, q \geq 3$, an $n$-vertex graph $G = (V, E)$ with maximum degree $\Delta$ and parameter $B \in (0, 1)$. We will omit these arguments from the algorithms, noting always when there are additional constraints on $\Delta, q$ and $B$.

### 4.1.1 Compress

Random seeds are generated for the compress update using Algorithm 8, which is shown in more generality than we will use.

**Algorithm 8:** compress.seed$(L_{t-1}, v, A, t)$

**Input :** a vertex $v$, a set $A$ of at least $\Delta(1-B)$ colours, a state $L_{t-1} \in (2^{[q]})^V$ and time $t$
**Output:** a seed $\alpha_t = (v, \gamma_t, \zeta_t, (A, c_1))$

**Function** *compress.seed*$(L_{t-1}, v, A, t)$**:**
  sample $\gamma_t, \zeta_t \in_{\mathrm{U}} [0,1]$
  sample $c_1 \in_{\mathrm{U}} [q] \backslash A$
  **return** $\alpha_t = (v, \gamma_t, \zeta_t, (A, c_1))$

Algorithm 8 selects the colour $c_1$ which the vertex $v$ will be updated to with probability $\frac{1}{q-|A|}$, while $\mathbb{P}(\phi_t(x)(v) = c_1 \,|\, v \in \alpha_t)$ must be $\frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}$. When $|A| \geq \Delta(1-B)$ Lemma 27 implies $\frac{1}{q-|A|} \geq \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}$, so we may use the random number $\gamma_t$ to further reduce the probability that $\phi_t(x)(v) = c_1$ is chosen to the appropriate probability.

**Algorithm 9:** compress.FP$(x, \alpha_t)$

**Input :** a $q$-colouring $x$ and a tuple $\alpha_t = (v, \gamma_t, \zeta_t, (A, c_1))$ from compress.seed$(L_{t-1}, v, A, t)$
**Output:** a $q$-colouring $y$

**Function** *compress.FP*$(x, \alpha_t)$**:**
  $y|_{V\setminus\{v\}} \longleftarrow x|_{V\setminus\{v\}}$
  **if** $\gamma_t \leq \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|)$ **then**
    $y(v) \longleftarrow c_1$
  **else**
    $y(v) \longleftarrow c_2$ for $c_2 \in A$ with

$$\frac{\sum_{\{c\in A \,|\, c<c_2\}} B^{m(x,v,c)}}{\sum_{c\in A} B^{m(x,v,c)}} \leq \zeta_t < \frac{\sum_{\{c\in A \,|\, c\leq c_2\}} B^{m(x,v,c)}}{\sum_{c\in A} B^{m(x,v,c)}}$$

  **end**
  **return** $y$

**Lemma 30.** *Suppose $q \geq \Delta \geq 3$ and $v \in V$. Suppose further that $A$ is a set of size greater than $\Delta(1-B)$. If $L_{t-1} \in (2^{[q]})^V$, $\alpha_t$ is generated using* compress.seed$(L_{t-1}, v, A, t)$ *and $y =$* compress.FP$(x, \alpha_t)$ *for $x \in L_{t-1}$ then $y(w) = x(w)$ for all $w \neq v$ and $\mathbb{P}(y(v) = c) = \frac{B^{m(x,v,c)}}{\sum_{c'\in[q]} B^{m(x,v,c')}}$ for all $c \in [q]$.*

*Proof.* Before we begin, we establish $\mathbb{P}\Big(\gamma_t \leq \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|)\Big) = \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|)$ for all $c_1 \in [q]$. By Lemma 27, $\sum_{c\in[q]} B^{m(x,v,c)} \geq q - \Delta(1-B) \geq q - |A|$, which implies that

$$\mathbb{P}\Big(\gamma_t \leq \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|)\Big) = \min\Big(\frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|), 1\Big) = \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}(q-|A|)$$

.

Algorithm 9 does not change the colour of any vertex other than $v$, so $y(w) = x(w)$ for all $w \neq v$. Now, consider the vertex $v$, and suppose $c \in [q]\backslash A$. Then

$$\mathbb{P}(y(v) = c) = \mathbb{P}(c_1 = c)\frac{B^{m(x,v,c)}}{\sum_{c' \in [q]} B^{m(x,v,c')}}(q - |A|) = \frac{B^{m(x,v,c)}}{\sum_{c' \in [q]} B^{m(x,v,c')}}$$

Otherwise, for $c \in A$,

$$\mathbb{P}(y(v) = c) = \mathbb{P}(c_2 = c)\Big(1 - \frac{\sum_{c' \in [q]\backslash A} B^{m(x,v,c')}}{\sum_{c' \in [q]} B^{m(x,v,c')}}\Big) = \frac{B^{m(x,v,c)}}{\sum_{c' \in [q]} B^{m(x,v,c')}}$$

□

Critically, Lemma 30 is independent of the choice of $A$ because $A$ is only used to correlate the update to different colourings. Associated with Algorithm 8 is Algorithm 10, which maintains the bounding chain.

**Algorithm 10:** $\underline{\text{compress.BP}}(L_{t-1}, \alpha_t)$

**Input :** a state $L_{t-1} \in (2^{[q]})^V$ and tuple $\alpha_t = (v, \gamma_t, \zeta_t, (A, c_1))$ from $\underline{\text{compress.seed}}(v, A, L_{t-1}, t)$

**Output:** $L_t \in (2^{[q]})^V$

**Function** $\underline{\textit{compress.BP}}(L_{t-1}, \alpha_t)$**:**

  $L_t|_{V\backslash\{v\}} \longleftarrow L_{t-1}|_{V\backslash\{v\}}$

  $L_t(v) \longleftarrow A \cup \{c_1\}$

  **return** $L_t$

**Lemma 31.** *Suppose $q \geq \Delta \geq 3$. Suppose further that $v \in V$ and $L_{t-1} \in (2^{[q]})^V$. If $\alpha_t =$ $\underline{\textit{compress.seed}}(L_{t-1}, v, A, t)$ and $\phi_t(x) = \underline{\textit{compress.FP}}(x, \alpha_t)$ for all $x \in \Omega_q$, then*

$$\Phi_0^t(\Omega_q) \subseteq L_t$$

*for $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$ and $L_t = \underline{\textit{compress.BP}}(L_{t-1}, \alpha_t)$.*

*Proof.* Neither the colour nor bounding list associated with any vertex other than $v$ are updated by either algorithm, so $y(v) \in L_t(v)$ for all $\neq w$. By construction, $y(w) \in A \cup \{c_1\} = L_t(v)$, so $y \in L_t$ whenever $x \in L_{t-1}$. □

### 4.1.2 Contract

The contract update may be applied to a vertex $v$ when at least $\Delta(1-B)$ colours are fixed from the perspective of $v$. More precisely, if $L \in (2^{[q]})^V$, then contract.seed generates a seed containing two colours, one from $[q]\backslash F_L(v)$ and one from $F_L(v)$. Denoting by $m_Q(L,v,c)$ the number of times the colour $c$ occurs in bounding lists of size one in the neighbours of $v$, the fact that $c_2 \in F_L(v)$ implies $B^{m(x,v,c)} = B^{m_Q(L,v,c)}$ for all $x \in L$. Since $B^{m_Q(L,v,c)}$ is independent of $x$, we may choose $c_2$ from $F_L(v)$ with probability proportional to $B^{m(x,v,c)}$ without knowing which colouring $x$ we are working with.

**Algorithm 11:** contract.seed$(L_{t-1}, v, t)$

**Input :** a vertex $v$, state $L_{t-1} \in (2^{[q]})^V$ with $|F_{L_{t-1}}(v)| > \Delta(1-B)$ and time $t$
**Output:** a seed $\alpha_t = (v, \gamma_t, (c_1, c_2))$
**Function** *contract.seed*$(L_{t-1}, v, t)$**:**
  sample $\gamma_t \in_U [0,1]$
  sample $c_1 \in_U [q]\backslash F_{L_{t-1}}(v)$
  sample $c_2$ from $F_{L_{t-1}}(v)$ with probability proportional to $B^{m_Q(L_{t-1},v,c_2)}$
  **return** $\alpha_t = (v, \gamma_t, (c_0, c_1))$

If $|F_{L_{t-1}}(v)| > \Delta(1-B)$ then $\frac{1}{q-|F_{L_{t-1}}(v)|} \geq \frac{B^{m(x,v,c_1)}}{\sum_{c\in[q]} B^{m(x,v,c)}}$, so we may use the same technique as in the compress update to reduce the probability that $c_1$ is the colour chosen for the update. When $c_1$ is not chosen, we simply choose $c_2$.

**Algorithm 12:** contract.FP$(x, \alpha_t)$

**Input :** a $q$-colouring $x$ and a tuple $\alpha_t = (v, \gamma_t, (c_1, c_2))$ from contract.seed$(L_{t-1}, v, t)$
**Output:** a $q$-colouring $y$
**Function** *contract.FP*$(x, \alpha_t)$**:**
  $y|_{V\backslash\{v\}} \longleftarrow x|_{V\backslash\{v\}}$
  **if** $\gamma_t \leq \frac{q-|F_{L_t}(v)|}{\sum_{c'\in[q]} B^{m(x,v,c')}} B^{m(x,v,c_1)}$ **then**
    $y(v) \longleftarrow c_1$
  **else**
    $y(v) \longleftarrow c_2$
  **end**
  **return** $y$

**Lemma 32.** *Suppose $q \geq \Delta \geq 3$ and $v \in V$. Suppose further that $t \geq 1$ and $L_{t-1} \in (2^{[q]})^V$ with $|F_{L_{t-1}}(v)| > \Delta(1-B)$. If $\alpha_t$ is generated using contract.seed$(L_{t-1}, v, t)$ and $y =$ contract.FP$(x, \alpha_t)$ for $x \in L_{t-1}$ then $y(w) = x(w)$ for all $w \neq v$ and $\mathbb{P}(y(v) = c) = \frac{B^{m(x,v,c)}}{\sum_{c'\in[q]} B^{m(x,v,c')}}$ for all $c \in [q]$.*

*Proof.* To begin, note that the condition $|F_{L_{t-1}}(v)| > \Delta(1-B)$ implies, alongside Lemma 27, that $\frac{q-|F_{L_{t-1}}(v)|}{\sum_{c'\in[q]} B^{m(x,v,c')}} B^{m(x,v,c_1)} \leq 1$ for all $c_1 \in [q]$, so $\mathbb{P}\Big(\gamma_t \geq \frac{q-|F_{L_{t-1}}(v)|}{\sum_{c'\in[q]} B^{m(x,v,c')}} B^{m(x,v,c_1)}\Big) = 1 -$

$\frac{q-|F_{L_{t-1}}(v)|}{\sum_{c\in[q]} B^{m(x,v,c)}} B^{m(x,v,c_1)}$. Suppose $c \in [q]\backslash F_{L_{t-1}}(v)$. Then

$$\mathbb{P}(y(v)=c) = \mathbb{P}(c_1=c)\frac{q-|F_{L_{t-1}}(v)|}{\sum_{c'\in[q]} B^{m(x,v,c')}} B^{m(x,v,c)} = \frac{B^{m(x,v,c)}}{\sum_{c'\in[q]} B^{m(x,v,c')}}$$

Otherwise, if $c \in F_{L_{t-1}}(v)$ then $B^{m_Q(L_{t-1},v,c)} = B^{m(x,v,c)}$, so

$$\begin{aligned}
\mathbb{P}(y(v)=c) &= \frac{B^{m_Q(L_{t-1},v,c)}}{\sum_{c'\in F_{L_{t-1}}(v)} B^{m_Q(L_{t-1},v,c')}}\Big(1 - \frac{\sum_{c_1\in[q]\backslash F_{L_{t-1}}(v)} B^{m(x,v,c_1)}}{\sum_{c'\in[q]} B^{m(x,v,c')}}\Big) \\
&= \frac{B^{m_Q(L_{t-1},v,c)}}{\sum_{c'\in F_{L_{t-1}}(v)} B^{m_Q(L_{t-1},v,c)}} \frac{\sum_{c'\in F_{L_{t-1}}(v)} B^{m(x,v,c')}}{\sum_{c'\in[q]} B^{m(x,v,c')}} \\
&= \frac{B^{m(x,v,c)}}{\sum_{c'\in[q]} B^{m(x,v,c')}}
\end{aligned}$$

□

To update $L_{t-1} \to L_t$ when contract.FP is applied, the contract.BP method detects when the parameter $\gamma_t$ is sufficiently large to recolour $v$ in all possible $x \in L_t$ to the colour $c_2$.

**Algorithm 13:** contract.BP$(L_{t-1}, \alpha_t)$

**Input :** a tuple $\alpha_t = (v, \gamma_t, (c_0, c_1))$ from contract.seed$(L_{t-1}, v, t)$ and a state $L_{t-1} \in (2^{[q]})^V$ with $|F_{L_{t-1}}(v)| > \Delta(1-B)$

**Output:** $L_t \in (2^{[q]})^V$

**Function** *contract.BP*$(L_{t-1}, \alpha_t)$**:**
- $L_t|_{V\backslash\{v\}} \longleftarrow L_{t-1}|_{V\backslash\{v\}}$
- **if** $\gamma_t > \frac{q-|F_{L_{t-1}}(v)|}{q-\Delta(1-B)}$ **then**
  - $L_t(v) \longleftarrow \{c_2\}$
- **else**
  - $L_t(v) \longleftarrow \{c_1, c_2\}$
- **end**
- **return** $L_t$

**Lemma 33.** *Suppose $q \geq \Delta \geq 3$ and $v \in V$. Suppose further that $L_{t-1} \in (2^{[q]})^V$ with $|F_{L_{t-1}}(v)| > \Delta(1-B)$. If $\alpha_t$ is generated using contract.seed$(L_{t-1}, v, t)$ and $\phi_t(x) =$ contract.FP$(x, \alpha_t)$ for all $x \in \Omega_q$, then*

$$\Phi_0^t(\Omega_q) \subseteq L_t$$

*for $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$ andå $L_t =$ contract.BP$(L_{t-1}, \alpha_t)$.*

*Proof.* Neither the colour nor bounding list associated with any vertex other than $v$ are updated by either algorithm, so $y(v) \in L_t(v)$ for all $\neq w$. If $\gamma_t \geq \frac{q-|F_{L_{t-1}}(v)|}{q-\Delta(1-B)}$ then $y(v) = c_2$ because

$\frac{q-|F_{L_{t-1}}(v)|}{\sum_{c\in[q]} B^{m(x,v,c)}} B^{m(x,v,c_1)} \leq \frac{q-|F_{L_{t-1}}(v)|}{q-\Delta(1-B)}$ (c.f. Lemma 27). Otherwise, $y(v) \in \{c_1, c_2\}$. In either case, $y \in L_t$. □

## 4.2 Structure of Algorithm

The two types of updates are used across two different phases which proceed for $T_1$ and $T_2$ steps respectively. The purpose of the first phase is to reduce the bounding lists to size at most two, after which the contract method may be applied to vertices chosen uniformly at random while maintaining the condition that all bounding lists have size at most two.

### 4.2.1 Phase One: Warm-Up

The first phase of the algorithm lasts for $T_1 = n + |E|$ steps and uses both compress and contract updates. We proceed over the vertices deterministically, using the compress update to ensure that $|F_{L_t}(v_k)| > \Delta(1-B)$ before using the contract method to reduce the size of the bounding list associated with $v$.

**Algorithm 14:** phaseOne$((v_i)_{1\leq i\leq n})$

**Input :** an ordering of the vertices $v_1, \cdots, v_n$
**Output:** a pair $(\Phi_0^{T_1}, L_{T_1})$

**Function** *phaseOne*$((v_i)_{1\leq i\leq n})$**:**
  let $N_>(v_k) = \{v_j \in N(v_k) \,|\, j > k\}$ and $N_<(v_k) = \{v_j \in N(v_k) \,|\, j < k\}$
  $t \longleftarrow 0$
  $\Phi_0^0 \longleftarrow$ the identity function
  $L_0(v) \longleftarrow [q]$ for all $v \in V$
  **for** $k \in [n]$ **do**
    choose $A$ to be the set of $\Delta$ 'least' colours maximally intersecting $\cup_{w\in N_<(v_k)} L_t(w)$
    **for** $w \in N_>(v_k)$ **do**
      $t \longleftarrow t+1$
      $\alpha_t \longleftarrow$ compress.seed$(L_{t-1}, w, A, t)$
      $\phi_t \longleftarrow$compress.FP$(\cdot, \alpha_t)$
      $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$
      $L_t \longleftarrow$ compress.BP$(L_{t-1}, \alpha_t)$
    **end**
    $t \longleftarrow t+1$
    `// Requires that` $|F_{L_{t-1}}(v)| > \Delta(1-B)$
    $\alpha_t \longleftarrow$ contract.seed$(L_{t-1}, v_k, t)$
    $\phi_t \longleftarrow$ contract.FP$(\cdot, \alpha_t)$
    $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$
    $L_t \longleftarrow$ contract.BP$(L_{t-1}, \alpha_t)$
  **end**
  **return** $(\Phi_0^{T_1}, L_{T_1})$

**Lemma 34.** *Suppose $q > 2\Delta \geq 6$ and $B > 1 - \frac{q-2\Delta}{\Delta}$. Let $L_t$ denote the state of the bounding chain at the start of iteration $k \in [n]$, and let $t'$ be the first time after the compress method has been applied to all the vertices in $N_>(v_k)$. If $|L_t(v_j)| \leq 2$ for all $j < k$ then $|F_{L_t}(v_k)| > \Delta(1-B)$.*

*Proof.* Let $S_{L_t}(v) = \cup_{v \in N(v)} L_t(v)$, so that $[q] \backslash F_L(v) \subseteq S_L(v)$. We will show that $|S_{L_{t'}}(v)| \leq 2\Delta$, from which we may deduce that $|F_{L_{t'}}(v)| \geq q - 2\Delta$. When we have done this, it will suffice to set $B > 1 - \frac{q-2\Delta}{\Delta}$, because then $\Delta(1-B) < q - 2\Delta$.

To see that $|S_{L_{t'}}(v)| \leq 2\Delta$, we decompose $|S_{L_{t'}}(v_k)| = a + b - c$ for

$$a = \Big| \bigcup_{w \in N_<(v_k)} L_{t'}(w) \Big|, \quad b = \Big| \bigcup_{w \in N_>(v_k)} L_{t'}(w) \Big|$$

$$c = \Big| \Big( \bigcup_{w \in N_<(v_k)} L_{t'}(w) \Big) \cap \Big( \bigcup_{w \in N_<(v_k)} L_{t'}(w) \Big) \Big|$$

The bounding lists $L_{t'}(w)$ associated with $w \in N_<(v_k)$ all have $|L_{t'}(w)| \leq 2$, so $a \leq 2|N_<(v_k)|$. The lists associated with $w \in N_>(v_k)$ all have the form $A \cup \{c_1\}$ for some $c_1 \notin A$, so $b \leq \Delta + |N_>(v_k)|$. Finally, $A$ is chosen to maximally intersect $\bigcup_{w \in N_<(v_k)} L_{t'}(w)$, so $c \geq \min\{a, \Delta\}$. Together, these bounds imply $\Big|S_{L_{t'}}(v_k)\Big| \leq a + b - \min\{a, \Delta\} \leq 2\Delta$. □

Lemma 34 does not benefit from the selection of $A$ smaller than $\Delta$, justifying this choice of $A$ in Algorithm 14.

**Corollary 35.** *Suppose $q > 2\Delta \geq 6$ and $B > 1 - \frac{q-2\Delta}{\Delta}$. If $T_1 = n + |E|$ is the first time after the warm-up phase has ended, then the bounding lists $L_{T_1}(v)$ all satisfy $|L_{T_1}(v)| \leq 2$.*

### 4.2.2 Phase Two: Coalescence

The coalescence phase lasts from time $T_1$ until time $T \coloneqq T_1 + T_2$, where $T_2$ is an integer to be specified later that depends on the graph $G, \Delta$ and $q$. In this phase, we repeatedly apply the contract method to vertices $v$ with $|L_t(v)| \leq 2$. The contract method maintains $|L_{t+1}(v)| \leq 2$, with a non-zero probability that $|L_{t+1}(v)| = 1$. Thus, using the contract method we can make progress towards the goal of $\prod_{v \in V} |L_T(v)| = 1$. When the bounding list associated with a vertex has size one and not all of the neighbours of $v$ are fixed there is a non-zero probability that the contract method increases the size of the bounding list. The following lemma, which we state without proof, allows us to manage this possibility.

**Algorithm 15:** phaseTwo$(\Phi_0^{T_1}, L_{T_1})$

**Input :** a pair $(\Phi_0^{T_1}, L_{T_1})$ with $\Phi_0^{T_1}(\Omega) \subseteq L_{T_1}$ and $|L_{T_1}| \leq 2$
**Output:** a pair $(\Phi_0^T, L_T)$
**Function** *phaseTwo*$(\Phi_0^{T_1}, L_{T_1})$**:**
  $t \longleftarrow T_1$
  **for** $t \leq T$ **do**
    $t \longleftarrow t+1$
    sample $v \in_{\mathrm{U}} V$
    $\alpha_t \longleftarrow$ contract.seed$(L_{t-1}, v, t)$
    $\phi_t \longleftarrow$contract.FP$(\cdot, \alpha_t)$
    $\Phi_0^t = \phi_t \circ \Phi_0^{t-1}$
    $L_t \longleftarrow$contract.BP$(L_{t-1}, \alpha_t)$
  **end**
  **return** $(\Phi_0^T, L_T)$

**Lemma 36** ([21, Theorem 4])**.** *Suppose $(X_t)$ is a random walk on $\{0, \cdots, n\}$ where 0 is a reflecting state and $n$ is an absorbing state, and let $e_i$ denote the expected number of times that $(X_t)$ hits state $i$, where $1 \leq i \leq n$. If $|X_{t+1} - X_t| \leq 1$ and $E[X_{t+1} - X_t \,|\, X_t] \geq \kappa_i > 0$ for all $X_t < n$ then*

$$\sum_{i \in [n]} e_i \leq \sum_{i \in [n]} \frac{1}{\kappa_i}$$

**Lemma 37.** *Suppose $q > 2\Delta \geq 6$ and $B > 1 - \frac{q-2\Delta}{\Delta}$. If $T = n + |E| + n^2\left(\frac{q-\Delta(1-B)}{q-\Delta(3-B)}\right)$ then the probability that the bounding process detects coupling in $T$ steps is greater than $\frac{1}{2}$.*

*Proof.* Recall that $T = T_1 + T_2$ where $T_1 = n + |E|$ is the run-time of the first phase and $T_2$ is the run-time of the second phase. The first phase of the algorithm is guaranteed to reduce the bounding lists to size at most two. We maintain this invariant throughout the second phase, so it suffices to set $T_2$ so that the bounding lists all have size one with probability at least one half if the second phase is allowed to run for $T_2$ iterations. Let $W_t = \{v \in V \,\big|\, |L_t(v)| = 1\}$. Then $|W_{t+1}| - |W_t| = -1$ whenever $v \in \overline{W_t}$ and $\gamma_t \geq \frac{q-|F_{L_t}(v)|}{q-\Delta(1-B)}$. Similarly, $|W_{t+1}| - |W_t| = 1$ whenever whenever $v \in W_t$ and

$\gamma_t \leq \frac{q-|F_{L_t}(v)|}{q-\Delta(1-B)}$. Thus

$$\begin{aligned}
E[|W_{t+1}| - |W_t| \,\Big|\, W_t] &= \frac{1}{n}\Big( \sum_{v\in\overline{W_t}} (1 - \frac{q-|F_{L_t}(v)|}{q-\Delta(1-B)}) - \sum_{v\in W_t} \frac{q-|F_{L_t}(v)|}{q-\Delta(1-B)}\Big) \\
&= \frac{1}{n}\Big(|\overline{W_t}| - \sum_{v\in V} \frac{q-|F_{L_t}(v)|}{q-\Delta(1-B)}\Big) \\
&\geq \frac{1}{n}\Big(|\overline{W_t}| - \sum_{v\in V} \frac{2|N(v)\cap\overline{W_t}|}{q-\Delta(1-B)}\Big) \\
&\geq \frac{|\overline{W_t}|}{n}\Big(1 - \frac{2\Delta}{q-\Delta(1-B)}\Big) > 0
\end{aligned}$$

It follows immediately from this, Lemma 36 and Markov's inequality that it suffices to set $T_2 = n^2\Big(\frac{q-\Delta(1-B)}{q-\Delta(3-B)}\Big)$. □

**Theorem 38.** *Suppose $q, \Delta \geq 3$ are integers with $q > 2\Delta$, and let $B$ be a real in the interval $(1 - \frac{q-2\Delta}{\Delta}, 1)$. Then there exists a randomized algorithm which, on input an $n$-vertex graph $G$ of maximum degree $\Delta$, outputs in expected time $O(n^2)$ a sample from the Gibbs distribution for the $q$-state Potts model on $G$ with parameter $B$.*

*Proof.* Lemmas 28, 30 and 32 establish that the Gibbs distribution is the stationary distribution of the underlying chain $\mathcal{N}$. To use Lemma 21 with the complete coupling of $\mathcal{N}$ to sample perfectly from the Gibbs distribution, we must establish that $\mathcal{N}$ is irreducible and aperiodic. To see that $\mathcal{N}$ is aperiodic, note that the Glauber dynamics always has a possibility of recolouring the vertex to be recoloured with the same colour. To see that $\mathcal{M}$ is irreducible, note that every vertex is selected for a Glauber update at least once by $\mathcal{N}$. Since any colour may be selected for vertex $v$ at any step, it is possible to move between any two colourings in one step.

Let $T = n + |E| + n^2\Big(\frac{q-\Delta(1-B)}{q-\Delta(3-B)}\Big)$. Lemmas 31 and 33 establish that we can detect when $\Phi_0^T$ is constant using the bounding process, and Lemma 37 establishes the value of $T$ sufficient to detect coupling with probability $\frac{1}{2}$. Using the coupling from the past framework, we may thereby construct a perfect sampling algorithm with expected run-time $O(n^2)$. □

# 5 Conclusion

At a high level, the major advancement of Bhandari and Chakraborty's perfect sampling algorithm for the proper $q$-colourings model is in the use of stochastic processes rather than Markov chains. In the first phase of the algorithm, the additional freedom given by the stochastic process formalism allows for a resampling of the neighbours of a vertex $v$ before resampling $v$ itself. This reduces the variance of the colourings in the neighbourhood, ensuring that there are at most $2\Delta$ colours in $S_L(v)$, the union of the bounding lists associated with the neighbours of $v$. To apply their version of the contract method and ensure polynomial expected run-time, they require at least $\Delta$ colours are not included in the bounding lists, resulting in the condition that $q > 3\Delta$ for $\Delta \geq 3$.

We generalize their result to the anti-ferromagnetic $q$-state Potts model, providing the first improvement over Huber's perfect sampling algorithm [22] and the first linear bounds on the number of colours $q$ as a function of the maximum degree of the graph $\Delta$. To account for the fact that $B^1 \neq B^m$ for $B \in (0,1)$ and $m > 1$, we use the set $F_L(v)$ of colours 'fixed' from the perspective of $v$.[10]The generalized contract algorithm we define is able to resample $v$ when $\Delta(1-B)$ colours are fixed from the perspective of $v$, while the compress algorithm ensures that $|S_L(v)| \leq 2\Delta$ when applied to the neighbours of $v$. All the colours which aren't fixed from the perspective of $v$ must be in a bounding list associated with a neighbour of $v$, so $[q] \backslash F_L(v) \subseteq S_L(v)$, from which we conclude that $|F_L(v)| \geq \Delta(1-B)$ when $q > 2\Delta \geq 6$ and $B \in (1 - \frac{q-2\Delta}{\Delta}, 1)$.

Jain et al. [24] have recently demonstrated that the condition $q > 2\Delta$ may be weakened when $\Delta$ is large. They follow Bhandari and Chakraborty in using a stochastic process rather than a Markov chain, but improve on the $3\Delta$ barrier for large $\Delta$ using an argument based on Jerrum's coupling of the Glauber dynamics [25]. Beyond extending Jerrum's analysis to bounding lists, one of the key advancements of their algorithm is the refinement of Lemma 34 through the use of a distinguished subset of *seeded vertices* which make up approximately a third of the vertices in any neighbourhood of the graph. Using these seeded vertices and the compress algorithm, they are able to ensure that $|S_L(v)| \leq \approx 5\Delta \backslash 3$. The requirement that the seeding set be sufficiently dense amongst the vertices of the graph, however, limits the application of their algorithm to $\Delta$ sufficiently large, and, by extension, to $q$ sufficiently large.

[10] A colour is fixed from the perspective of a vertex $v$ if it does not occur in a bounding list associated with a neighbour of $v$, i.e. $c \in [q]$ is fixed from the perspective of $v \in V$ if $c \in F_L(v) \coloneqq [q] \backslash \bigcup_{\substack{w \in N(v) \\ |L(w)|>1}} L(w)$.

As the range of parameters at which perfect sampling from the proper $q$-colourings and anti-ferromagnetic Potts models continues to expand towards the edge of the regime in which approximate sampling is possible, a number of questions come into view. Can perfect sampling algorithms for the proper $q$-colourings or anti-ferromagnetic $q$-state Potts models match the assumptions necessary for approximate sampling, or is there a collection of parameters at which only approximate sampling is possible? How close to the complexity threshold described by Galanis [14] at $\Delta \geq 3$, (even) $q \geq 3$ and $B = \frac{\Delta - q}{\Delta}$ can sampling algorithms get? One possible direction for the improvement of the arguments presented here is a probabilistic analysis of the intersection of the bounding lists. While Lemma 34 makes use of a set which is chosen to intersect with as many of the bounding lists in a neighbourhood as possible, nowhere in this lemma or elsewhere do we use the fact that it is unlikely that the bounding lists of size two produced by the contract method are entirely disjoint. This leads us to consider the possibility of an improved perfect sampling algorithm which considers the probability that the bounding lists in a neighbourhood have a non-trivial intersection. Such an algorithm would perform particularly when the number of colours $q$ is smaller than the maximum degree of the graphs considered, because the fewer colours are used in the model the higher the probability that the bounding lists in a neighbourhood have a large intersection.

Future research into perfect sampling using stochastic processes may look to extending the stochastic processes approach to sampling from the anti-ferromagnetic Potts model in two complimentary directions. First, the Glauber dynamics is not the only Markov chain which is stationary with respect to the Gibbs distribution, and is not currently the Markov chain which provides the best mixing time when used in a Markov chain Monte Carlo approximate sampling algorithm - the best known chain is a modification of Vigoda's 'flip dynamics' [38] proposed by [5]. It is reasonable, therefore, to consider the possibility of using a modified version of the flip dynamics, rather than the Glauber dynamics, as the underlying processes in a sampling algorithm for the Potts model. Second, the Potts model is only one prototypical model within the broader collection of graphical models, an important class of probabilistic models capable of encoding a variety of complicated dependencies between random variables using graphs. These models allow for distributions which depend on factors like the location of a vertex in the graph or the number of monochromatic edges of a specific colour, and are a natural generalized setting in which perfect sampling algorithms based on coupling from the past might benefit from the freedom provided by stochastic processes.

# References


[1] Jayanth R. Banavar, G. S. Grest, and David Jasnow. Ordering and phase transitions in antiferromagnetic potts models. *Phys. Rev. Lett.*, 45:1424–1428, Oct 1980.

[2] Siddharth Bhandari and Sayantan Chakraborty. Improved bounds for perfect sampling of k-colorings in graphs. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 631–642, New York, NY, USA, 2020. Association for Computing Machinery.

[3] Clelia M. Bordogna and Ezequiel V. Albano. Dynamic behavior of a social model for opinion formation. *Phys. Rev. E*, 76:061125, Dec 2007.

[4] Russ Bubley, Martin Dyer, Catherine Greenhill, and Mark Jerrum. On approximately counting colorings of small degree graphs. *SIAM Journal on Computing*, 29(2):387–400, 1999.

[5] S. Chen, M. Delcourt, A. Moitra, G. Perarnau, and L. Postle. Improved bounds for randomly sampling colorings via linear programming. In *SODA*, 2019.

[6] Eytan Domany, M. Schick, and J. S. Walker. Classification of order-disorder transitions in common adsorbed systems: Realization of the four-state potts model. *Phys. Rev. Lett.*, 38:1148–1151, May 1977.

[7] Martin Dyer, Leslie A. Goldberg, and Mark Jerrum. Counting and sampling h-colourings. In *Randomization and Approximation Techniques in Computer Science*, pages 51–67, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[8] Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Structures and Algorithms*, 17(34):260–289, 2000.

[9] Weiming Feng, Nisheeth K. Vishnoi, and Yitong Yin. Dynamic sampling from graphical models. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 1070?1081, New York, NY, USA, 2019. Association for Computing Machinery.

[10] J. A. Fill and M. Huber. The randomness recycler: A new technique for perfect sampling. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, page 503, USA, 2000. IEEE Computer Society.

[11] James Allen Fill. An interruptible algorithm for perfect sampling via markov chains. *Ann. Appl. Probab.*, 8(1):131–162, 02 1998.

[12] Lance Fortnow. Counting complexity. In *Complexity Theory Retrospective II*, pages 81–107. Springer-Verlag, 1997.

[13] C.M. Fortuin and P.W. Kasteleyn. On the random-cluster model: I. introduction and relation to other models. *Physica*, 57(4):536 – 564, 1972.

[14] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6), December 2015.

[15] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified np-complete problems. In *Proceedings of the Sixth Annual ACM Symposium on Theory of Computing*, STOC '74, pages 47–63, New York, NY, USA, 1974. Association for Computing Machinery.

[16] Leslie Ann Goldberg. Probability and Computing Course Notes, Oxford, 2019-20.

[17] A. Grabowski and R.A. Kosiński. Ising-based model of opinion formation in a complex network of interpersonal interactions. *Physica A: Statistical Mechanics and its Applications*, 361(2):651 – 664, 2006.

[18] Fran çois Graner and James A. Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Phys. Rev. Lett.*, 69:2013–2016, Sep 1992.

[19] Heng Guo, Mark Jerrum, and Jingcheng Liu. Uniform sampling through the lovász local lemma. *J. ACM*, 66(3), April 2019.

[20] Olle Haggstrom and Karin Nelander. On exact simulation of markov random fields using coupling from the past. *Scandinavian Journal of Statistics*, 26(3):395–411, 1999.

[21] Mark Huber. Exact sampling and approximate counting techniques. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 31–40, New York, NY, USA, 1998. Association for Computing Machinery.

[22] Mark Huber. Perfect sampling using bounding chains. *The Annals of Applied Probability*, 14, 06 2004.

[23] F. Jaeger, D. L. Vertigan, and D. J. A. Welsh. On the computational complexity of the jones and tutte polynomials. *Mathematical Proceedings of the Cambridge Philosophical Society*, 108(1):35–53, 1990.

[24] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. Perfectly sampling $k \geq (8/3+o(1))\delta$-colorings in graphs, 2020.

[25] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.

[26] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169 – 188, 1986.

[27] Timothy D Johnson, Zhuqing Liu, Andreas J Bartsch, and Thomas E Nichols. A bayesian non-parametric potts model with application to pre-surgical fmri data. *Statistical Methods in Medical Research*, 22(4):364–381, 2013. PMID: 22627277.

[28] Richard M Karp, Michael Luby, and Neal Madras. Monte-carlo approximation algorithms for enumeration problems. *Journal of Algorithms*, 10(3):429 – 448, 1989.

[29] George Maltese. A simple proof of the fundamental theorem of finite markov chains. *The American Mathematical Monthly*, 93(8):629–630, 1986.

[30] Wolfgang Mulzer. Five proofs of chernoff's bound with applications, 2018.

[31] Alejandro Murua, Larissa Stanberry, and Werner Stuetzle. On potts model clustering, kernel k-means, and density estimation. *Journal of Computational and Graphical Statistics*, 17(3):629–658, 2008.

[32] James Gary Propp and David Bruce Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9:223–252, 1996.

[33] Jörg Reichardt and Stefan Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Phys. Rev. Lett.*, 93:218701, Nov 2004.

[34] D. Stauffer. Social applications of two-dimensional ising models. *American Journal of Physics*, 76(4):470–473, 2008.

[35] Martin Storath, Andreas Weinmann, Jergen Frikel, and Michael Unser. Joint image reconstruction and segmentation using the potts model. *Inverse Problems*, 31(2):025003, Jan 2015.

[36] E.L. Stott, N.F. Britton, J.A. Glazier, and M. Zajac. Stochastic simulation of benign avascular tumour growth using the potts model. *Mathematical and Computer Modelling*, 30(5):183 – 198, 1999.

[37] Stephan Turner and Jonathan A. Sherratt. Intercellular adhesion and cancer invasion: A discrete simulation using the extended potts model. *Journal of Theoretical Biology*, 216(1):85 – 100, 2002.

[38] Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.