New Conditions via Markov Chains: Approximating Partition Functions of Abstract Polymer Models without Cluster Expansion

Neue Bedingungen via Markov-Ketten: Approximation der Zustandssummen abstrakter Polymermodelle ohne Cluster-Expansion

Marcus Pappik

Master's Thesis submitted in partial fulfillment
of the requirements for the degree

Master of Science
(M. Sc.)

in IT-Systems Engineering

at the Research Group of Algorithm Engineering
Digital Engineering Faculty
University of Potsdam

March 6th, 2020

Supervisor Prof. Dr. Tobias Friedrich
Advisors Dr. Andreas Göbel
Dr. Martin Krejca

Abstract

Abstract polymer models are combinatorial structures that consist of a set of weighted objects, called polymers, together with a reflexive and symmetric incompatibility relation that describes which polymers cannot occur together.

In this thesis we present the first Markov chain approach for sampling from the Gibbs distribution of abstract polymer models. Known Markov chains for polymer models from vertex and edge spin systems can be seen as special cases of this polymer Markov chain. We introduce the concept of polymer cliques and propose a generalized polymer mixing condition as a way to guarantee rapid mixing of our chain. The form of this condition is similar to conditions from cluster expansion approaches, such as the Kotecký-Preiss condition and the Fernández-Procacci condition, but it is less restrictive.

To obtain an efficient sampling scheme for the Gibbs distribution from our polymer Markov chain, we prove that it suffices to draw each step of the chain approximately according to its transition probabilities. As one way to approximate each transition of the chain, we suggest to truncate each polymer clique based on some notion of size. We introduce the clique truncation condition as a general tool to determine the maximum size of polymers that we have to consider for the steps of the chain.

We prove that our sampling scheme can be used to construct an FPRAS for the partition function. By this, we obtain the first Markov chain Monte Carlo method that works for abstract polymer models in a similar regime as cluster expansion approaches and beyond, while avoiding their complicated analytical assumptions and arguments.

Further, we illustrate how our approach can be applied to algorithmic problems like the hard-core model on bipartite α-expander graphs and the perfect matching polynomial to obtain new trade-offs between runtime and weight bounds. We emphasize that similar results can be obtained for a variety of other applications.

Zusammenfassung

Abstrakte Polymermodelle sind kombinatorische Strukturen, bestehend aus einer Menge gewichteter Objekte, genannt Polymere, und einer symmetrischen und reflexiven Relation, die bestimmte Polymerkombinationen verbietet.

In dieser Arbeit präsentieren wir das erste Markov-Ketten-Verfahren, um von der Gibbs-Verteilung abstrakten Polymermodelle zu ziehen. Bekannte Markov-Ketten, etwa für Knoten- und Kanten-Spin-Systeme, können als Spezialfälle dieser Polymer-Markov-Kette betrachtet werden. Wir stellen das Konzept der Polymer-Clique und eine verallgemeinerte Polymer-Mischungs-Bedingung vor. Letztere spielt eine entscheidende Rolle für die Mischungszeit der Markov-Kette. Die verallgemeinerte Polymer-Mischungs-Bedingung hat eine ähnliche Form wie Kriterien für die Konvergenz der Cluster-Expansion, etwa die Kotecký-Preiss- oder die Fernández-Procacci-Bedingung, ist jedoch weniger restriktiv als diese.

Zur algorithmischen Verwendung unserer Polymer-Markov-Kette beweisen wir, dass es genügt jeden Schritt der Kette näherungsweise entsprechen ihrer Übergangsverteilung zu vollziehen. Eine Möglichkeit die Übergänge zu approximieren, ist die Polymer-Cliquen auf kleine Polymere einzuschränken. Nützliches Werkzeug zur Bestimmung der maximalen benötigten Polymergröße ist hierbei die Cliquen-Einschränkungs-Bedingung.

Wir beweisen weiterhin, dass unsere Verfahren dazu benutzt werden können einen FPRAS für die Zustandssumme abstrakter Polymermodelle zu erhalten. Das Resultat ist das erste Markov-Ketten-Monte-Carlo-Verfahren, dass ähnliche Bereiche wie Cluster-Expansion abdeckt und sogar über diese hinausgeht. Hierfür benötigen wir weder komplizierten analytischen Annahmen, noch derartige analytische Argumente.

Schließlich wenden wir unseren Ansatz auf den Hard-core-Prozess auf bipartiten α-Expander-Graphen und einem Polynom auf perfekten Matchings an. Wir erhalten neue Trade-offs zwischen Laufzeit und erlaubten Polymergewichten. Ähnliche Resultate können für ein breites Feld von Anwendungen erzielt werden.

Contents

Abstract iii

1 Introduction 1

2 Preliminaries 9

3 Polymer Dynamics 13
- 3.1 Polymer Markov chain . . . 14
- 3.2 Coupling with exponential diameter . . . 17
- 3.3 Mixing Time of the polymer Markov chain . . . 25
- 3.4 Algorithmic aspects of generalized polymer mixing condition . 27

4 Markov Chains with Transition Error 31
- 4.1 Sample from Markov chains with transition errors . . . 31
- 4.2 Application to the polymer Markov chain . . . 35

5 Inner Sampling by Truncation 39
- 5.1 Conditions for truncation . . . 40
- 5.2 Algorithmic application of truncation . . . 44

6 Comparison & Application 47
- 6.1 Approximating the partition function . . . 47
- 6.2 Comparison to cluster expansion . . . 54
- 6.3 Graph polymers from vertex spin systems . . . 61
- 6.4 Graph polymers from edge spin systems . . . 74

7 Conclusions & Outlook 83

Bibliography 87

Declaration of Authorship 91

1 Introduction

In recent years, abstract polymer models gained growing attention in the area of counting complexity. Informally, an abstract polymer model is defined by a set of weighted objects, called polymers, and an incompatibility relation that describes which objects cannot occur together. Based on these ingredients, two computational tasks are commonly considered. One is to sample from the Gibbs distribution, which is drawing sets of pairwise compatible polymers proportionally to the product of their weights. The second task is to compute the so called partition function, which can be thought of as calculating the normalization constant of the Gibbs distribution.

Abstract polymer models were initially used in statistical physics, in order to investigate phase transitions in particle systems (see for example Borgs and Imbrie [BI89] and Laanait et al. [Laa+91]). However, more and more relations between these models and standard problems from theoretical computer science have been discovered. On the one hand, the computational tasks related to polymer models have some obvious similarities to a weighted version of sampling and counting independent sets. On the other hand, many other computational problems can be expressed using such polymer models, as shown by Casel et al. [Cas+19], Cannon and Perkins [CP19], and Jenssen et al. [JKP18]. For many of these applications, the number of polymers is not polynomially bounded in the input size of the original problem, imposing additional algorithmic challenges.

Most algorithms for approximating the partition function of abstract polymer models can be divided into two categories. The first category of algorithms uses an infinite series, called cluster expansion, which is a formal power series representation of the logarithm of the partition function. Under certain conditions, only a small number of terms of this series is needed for a sufficiently good approximation. Based on an approximation of the partition function, the Gibbs measure can usually be reconstructed. However, such approaches often suffer from requiring complicated analytical assumptions and having high runtime.

The second type of algorithms uses Markov chains to sample from the Gibbs distribution in the first place. Based on these sampling schemes, randomized

approximations of the partition functions can be derived. Although these algorithms usually have better runtimes, such Markov chains are only known for certain types of polymer models and they often lead to much worse bounds on the weights of the polymers than cluster expansion algorithms.

Before having a detailed look at recent results from both algorithmic fields, it will be helpful to formalize our notion of polymer models.

Abstract polymer models

Throughout this thesis, we use the following formal definition of a polymer model.

Definition 1.1 (Polymer model). A polymer model is a tuple $(\mathcal{C}, w, \asymp)$ where $\mathcal{C}$ is a finite set of objects called polymers, $w = \{w_\gamma \in \mathbb{R}_{>0} \mid \gamma \in \mathcal{C}\}$ is a set of positive weights and $\asymp \subseteq \mathcal{C} \times \mathcal{C}$ is a reflexive and symmetric relation. We call $\asymp$ the incompatibility relation. In addition, we say $\gamma, \gamma' \in \mathcal{C}$ are incompatible if $\gamma \asymp \gamma'$ and compatible otherwise.

The algorithmic tasks that are connected to polymer models are based on so called polymer families.

Definition 1.2 (Polymer families). For a polymer model $(\mathcal{C}, w, \asymp)$ we call a subset of polymers $\Gamma \subseteq \mathcal{C}$ a polymer family if all $\gamma, \gamma' \in \Gamma$ with $\gamma \neq \gamma'$ are compatible. Moreover, we denote the set of all polymer families by $\mathcal{F}_{\mathcal{C}, \asymp}$, or write $\mathcal{F}$ if the polymer model is clear from the context.

We would like to mention a helpful interpretation of polymer models and polymer families. If we consider $(\mathcal{C}, w, \asymp)$ as a graph with vertices $\mathcal{C}$ and edges defined by the incompatibility relation $\asymp$, then every polymer family $\Gamma \in \mathcal{F}_{\mathcal{C}, \asymp}$ corresponds to an independent set. We will sometimes refer to this graph interpretation as polymer graph.

We continue by formalizing the partition function and the Gibbs distribution, which were already informally introduced before. We start with the partition function of the polymer model.

Definition 1.3 (Partition function). Given a polymer model $(\mathcal{C}, w, \asymp)$ with polymer families $\mathcal{F}_{\mathcal{C}, \asymp}$, we define the partition function of the polymer model as

$$Z(\mathcal{C}, w, \asymp) = \sum_{\Gamma \in \mathcal{F}_{\mathcal{C}, \asymp}} \prod_{\gamma \in \Gamma} w_\gamma,$$

where the empty polymer family $\emptyset \in \mathcal{F}_{C,}$ contributes 1. If the polymer model is obvious, we might simply write Z for the value of the partition function. J

Based on this, the Gibbs distribution can be defined as follows.

I **Definition 1.4 (Gibbs distribution).** Given a polymer model (C, w, Ω) with polymer families $\mathcal{F}_{C,}$, we define the Gibbs measure or Gibbs distribution as a probability distribution $\mu_{C,w,}$ on $\mathcal{F}_{C,}$ with

$$\mu_{C,w,}(\Gamma) = \frac{\prod_{\gamma \in \Gamma} w_\gamma}{Z(C, w, \Omega)}$$

for every $\Gamma \in \mathcal{F}_{C,}$. Again, we assign $\mu_{C,w,}(\emptyset, \Omega) = \frac{1}{Z(C,w,\Omega)}$, and we might simply write μ if the polymer model becomes clear from the context. J

Using the polymer graph interpretation, computing the partition function and sampling from the Gibbs distribution can be seen as counting and sampling independent sets of the polymer graph, proportionally to the product of the weights of their vertices.

We would like to discuss some details about these definitions. First of all, we consider only finite polymer models. We are aware of the fact that there are some areas, for example in statistical physics, where also properties of infinite polymer models are of interest. However, most computational tasks can be modeled by using only finite polymer sets (see for example [Cas+19; CP19; JKP18]).

Secondly, we restrict the weights to be positive real values. There are other algorithmic approaches for approximating the partition function for complex-valued weights. However, we will mainly focus on the computational task of sampling from the Gibbs distribution. Thus, it makes sense to assume non-negative real values as weights, such that the Gibbs distribution is always well defined. The fact that we even assume them to be strictly positive does not add any additional restrictions, as polymers $\gamma \in C$ with $w_\gamma = 0$ neither contribute to the Gibbs distribution, nor to the partition function.

Cluster expansion algorithms

An interesting property of polymer models is that the logarithm of their partition function can be represented by an infinite formal power series, called cluster expansion, as for example observed by Kotecký and Preiss [KP86] and

Dobrushin [Dob96]. More precisely, if the cluster expansion is absolute convergent, than it is equal to the log partition function up to rearrangement of terms. Since this is known, many sufficient conditions for absolute convergence of this cluster expansion have been proposed. A comprehensive study was for example written by Fernández and Procacci [FP07].

Besides purely theoretical interests, absolute convergence of the Taylor series of the log partition function is especially important from the algorithmic perspective. Similar properties have been used in approximation algorithms for related problems, for example by Barvinok [Bar14] and Patel and Regts [PR17].

Using the connection between approximation and absolute convergence, Helmuth et al. [HPR18] were the first who proposed an approximation algorithm for the partition function of polymer models that arise from vertex spin systems on finite graphs with constant degree bound (see Section 6.3 for a formal definition). Their idea was to model the weights of the polymers as functions of some complex parameter z and to relate analytical properties of the weight function with a certain notion of size of a polymer. If the partition function is a polynomial in z without any root in an open disc of radius $\delta > 0$ around zero, then the cluster expansion converges absolutely in this region and only a small number of its terms is needed to approximate the log partition function. This leads to an efficient approximation on bounded degree graphs for every z such that $\frac{\delta}{\delta - |z|} \leq \Theta(1)$ (see Theorem 3.4 of Casel et al. [Cas+19] for this bound on $|z|$).

Similar algorithms have later been proposed for cases where the partition function cannot be expressed as a polynomial of some parameter z. They were applied to extend known approximation bounds for interesting problems, like the hard-core model on bipartite α-expander graphs and the Potts model on α-expander graphs by Jenssen et al. [JKP18], and the hard-core model on unbalanced bipartite graphs by Cannon and Perkins [CP19]. However, those algorithms still rely on convergence analysis of the cluster expansion and need very specific versions of know convergence criteria. Usually these criteria are very restricted versions of the so called Kotecký-Preiss condition as for example stated in Theorem 8 of Jenssen et al. [JKP18]. Moreover, for vertex spin systems on graphs with n vertices, a bounded degree Δ and $q + 1$ possible spin assignments for each vertex, such algorithms often require a runtime of $(\frac{n}{\varepsilon})^{O(1)\ln(\Delta q)^{O(1)}}$ for computing an ε-approximation of the partition function.

The approach of Helmuth et al. [HPR18] was originally limited to polymer models for vertex spin systems, but it was later generalized for abstract polymer

models by Casel et al. [Cas+19]. They applied this generalization for example to edges spin systems to approximate certain Holant polynomials, like the perfect matching polynomial. However, to the best of our knowledge, no generalization for abstract polymer models was stated for cases where the partition function cannot be expressed as a polynomial of a single parameter z.

Markov chain algorithms

Besides the deterministic approaches for approximation of the partition function, recently some Markov chain Monte Carlo algorithms have been proposed as well. The idea of these approaches is to approximately sample from the Gibbs distribution and use this sampling schema to construct a randomized approximation algorithm. Naturally, such algorithms are limited to real-valued weights, as they need the Gibbs distribution to be well defined. However, at the same time they circumvent any analytical assumptions about the partition function or convergence analysis of the cluster expansion.

One of the most remarkable results in this area was by Chen et al. [Che+19], who proposed a Markov chain to sample from the Gibbs distribution of vertex spin polymer models and proved the polymer mixing condition as a sufficient condition for rapid mixing of this chain. They were able to use this to construct a randomized approximation for the partition function of a polymer model with runtime $O\left(\frac{1}{\varepsilon}\right)^2 \log\left(\frac{1}{\varepsilon}\right)^2$, but for the price of much worse bounds on the weights than cluster expansion approaches need.

Whereas this Markov chain approach initially was introduced for vertex spin systems, a version for edge spin systems has been proposed by Casel et al. [Cas+19]. Similar to the vertex spin version, it significantly improves the runtime but only works for smaller positive real weights than the deterministic algorithms.

Contributions and outline

In this thesis, we propose a Markov chain Monte-Carlo framework for sampling from the Gibbs distribution and approximating the partition function of abstract polymer models.

We start in Chapter 2 with specifying necessary definitions and clarifying our notation. This will include certain properties of Markov chains as well as notions

of efficient approximation and asymptotic runtime that are used throughout this thesis.

In Chapter 3 we will present a Markov chain for sampling from the Gibbs distribution of abstract polymer models. In order to do so, we will introduce the notion of polymer cliques and polymer clique covers. Known Markov chains as proposed by Chen et al. [Che+19] for vertex spin systems and by Casel et al. [Cas+19] for edge spin systems can be seen as special cases of our polymer Markov chain.

In addition, we propose the generalized polymer mixing condition (see Definition 3.3) which has a similar form as known conditions for absolute convergence of the cluster expansion (see for example Fernández and Procacci [FP07]) but is less restrictive than those conditions. We will show that under mild assumptions the generalized polymer mixing condition is sufficient to guarantee rapid mixing of our polymer Markov chain. At the end of this chapter, we will discuss in Theorem 3.13 under which conditions our chain can be used to obtain an efficient sampling scheme for the Gibbs distribution.

The main weakness of our proposed polymer Markov chain is that it is not obvious how each transition can be done efficiently. A solution for this problem will be investigated in Chapter 4. Namely, we will present conditions under which each step of a Markov chain can be done according to an approximation of the desired transition probability distribution without losing the ability to use it for approximate sampling. We use a technique that might also be applicable to other sampling problems, as it is not limited to polymer models. This will result in Theorem 4.5, which under suitable conditions reduces approximate sampling from the Gibbs distribution to approximate sampling from a so called clique polymer distribution to obtain approximated transitions.

In Chapter 5 we will examine one way of doing this approximate sampling of transitions, namely truncating each polymer clique by some size-function. We will introduce the clique truncation condition (see Definition 5.5) as a tool to determine up to which size polymers have to be considered to get a sufficiently good approximation of each transition. Although this idea is similar to truncation in cluster expansion algorithms, there is a significant difference. We circumvent assumptions about analytical properties of the polymer model and do truncation solely on the basis of individual polymer cliques instead of involving any infinite power series. By this, our condition will be more flexible and the idea behind it will be easier to understand. Theorem 5.7 shows how the generalized polymer

mixing condition and the clique truncation condition can be combined to get an efficient approximate sampling scheme that works with only little structural assumptions about the polymer model.

We would like to suggest the following point of view on these to conditions: whereas the generalized polymer mixing condition ensures that we can handle the combinatorial complexity that arises from drawing pairwise compatible subsets of polymers, the clique truncation condition takes care of the complexity that arises from the sheer number of polymers that we have to deal with. We noticed that Galanis et al. [GGS19] already proposed a similar combination of truncation and Glauber dynamics. However, their approach was limited to a specific algorithmic problem, whereas we cover a much broader setting.

Chapter 6 will start with introducing an FPRAS for the partition function in Theorem 6.2. Our algorithm will work for any of our proposed sampling methods by applying self-reducibility based on polymer cliques. Thus, we can apply it even with an exponential number of polymers.

We will proceed with comparing the conditions for our approach with conditions for algorithms that use cluster expansion. First, we prove that under minor assumptions all conditions for cluster expansion that were discussed by Fernández and Procacci [FP07] can be used to obtain rapid mixing of the polymer Markov chain. Further, we will show that under mild conditions, certain convergence criteria directly imply both, the generalized polymer mixing condition and the clique truncation condition, meaning that we can obtain an efficient sampling scheme for the Gibbs distribution and an FPRAS for the partition function. Especially, this also holds for abstract polymer models with partition functions that cannot be expressed as polynomials of a single parameter z.

We state our conditions for an FPRAS of the partition function of edge and vertex spin systems in a convenient to use form in Theorem 6.15 and Theorem 6.26. As a main result, we can obtain a randomized approximation for the partition function in time $\left(\frac{n}{\varepsilon}\right)^{O(1)\ln^{1}(\Delta q)^{\infty}}$ for a variety of vertex spin polymer models if for every polymer $\gamma \in \mathsf{C}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|};$$

where $a \in \mathbb{R}_{>0}$ and $|\gamma|$ is the number of vertices in a polymer. This condition for an FPRAS applies to polymer models like those used for the Potts model on α-expander graphs and the hard-core model on bipartite α-expander graphs by

Jenssen et al. [JKP18] as well as the hard-core model on unbalanced bipartite graphs by Cannon and Perkins [CP19]. Our result is a improvement over the size-dependent versions of the Kotecký-Preiss condition that were originally used.

Further, in Theorem 6.16 and Theorem 6.27 we state conditions for a randomized approximation of the partition function of vertex and edge spin polymer models in time $\left(\frac{n}{\varepsilon}\right)^{c}$ for some absolute constant c independent of q and Δ.

We illustrate the application of our approach on the hard-core model on α-expander graphs and the perfect matching polynomial. Although we only slightly improve known bounds on the weights while having the same $\left(\frac{n}{\varepsilon}\right)^{O(1)}$ runtime as known algorithms, we will show that our approach is capable of improving the runtime dependency on Δ with only a minor decrease of the polymer weights.

Finally, in Chapter 7 we discuss our results, as well as possible directions for future work and implications for other lines of research.

2 Preliminaries

In this chapter, we would like to clarify some definitions and notation that we will use throughout this thesis, especially with respect to Markov chains and asymptotic runtime behaviour.

Markov chains

We assume the reader to be familiar with the definition of a Markov chain, as well as with basic properties like aperiodicity, irreducibility and recurrence. For a detailed reading, we would like to refer to standard literature, like [Bré13].

We will denote a Markov chain as a sequence of random variables $(X_t)_{t \in \mathbb{N}}$ on some state space Ω. A homogeneous Markov chain is mainly characterized by its transitions $P : \Omega \times \Omega \to [0, 1]$ with $P(x, y) = \Pr[X_{i+1} = y \mid X_i = x]$. We generalize this notation to the t-step transitions $P^t : \Omega \times \Omega \to [0, 1]$ by $P^t(x, y) = \Pr[X_{i+t} = y \mid X_i = x]$. For some properties like irreducibility, we might use the Markov chain $(X_t)_{t \in \mathbb{N}}$ and its transitions P interchangeably (e.g. say P is irreducible) as far as such properties are characterized solely by its transitions. In addition, for any $t \in \mathbb{N}$ we write $P^t(x, \cdot)$ for the probability distribution of X_{i+t} given that $X_i = x$, and we might simplify this to $P(x, \cdot)$ for $t = 1$.

Some of the algorithmically most interesting properties of a Markov chain are its stationary distributions and its convergence to such a distribution.

Definition 2.1 (Stationary distribution). For a Markov chain $(X_t)_{t \in \mathbb{N}}$ on a state space Ω with transitions P, a stationary distribution is a probability distribution π_P on Ω such that for all $y \in \Omega$

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

It is well known that for an irreducible Markov chain such a distribution exists if and only if the Markov chain is positive recurrent, and that in this case

the stationary distribution is unique and positive everywhere. A useful way to identify stationary distributions is the detailed balance condition.

Definition 2.2 (Detailed balance). Given a Markov chain on a state space Ω with transitions P. A distribution κ on Ω is said to fulfill detailed balance if for every $x, y \in \Omega$

$$\kappa(x)P(x,y) = \kappa(y)P(y,x).$$

Note that if κ fulfills detailed balance with respect to P, it holds for all $y \in \Omega$ that

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

Thus, such a distribution κ is also a stationary distribution for P.

Next, we want to characterize the convergence of a Markov chain to its stationary distribution. To have a notion of convergence, we need some metric for probability distributions on a state space. It is common to use the total variation distance as such a metric. There are many equivalent ways to define total variation distance, but for this thesis the following definition will suffice.

Definition 2.3 (Total variation distance). For two probability distributions κ_1, κ_2 on a countable state space Ω, the total variation distance is defined as

$$d_{\mathrm{TV}}(\kappa_1, \kappa_2) = \frac{1}{2}\sum_{x\in\Omega} |\kappa_1(x) - \kappa_2(x)|.$$

As a matter of fact, if an irreducible, positive recurrent Markov chain with transitions P and stationary distribution π_P on a state space Ω is aperiodic, then for every $x \in \Omega$ we have that $\lim_{t\to\infty} d_{\mathrm{TV}}(P^t(x,\cdot), \pi_P) = 0$. In this case, we say $P^t(x,\cdot)$ converges to π_P as $t \to \infty$. A simple way to see that an irreducible Markov chain is aperiodic is if the chain has at least one state with positive self-loop probability (i.e., there is a state $x \in \Omega$ such that $P(x,x) > 0$).

We can now use the total variation distance to define the mixing time of a Markov chain as a formal way to describe the speed of this convergence.

Definition 2.4 (Mixing time). Let $(X_t)_{t\in\mathbb{N}}$ be an irreducible, positive recurrent, aperiodic Markov chain on a state space Ω with transitions P. In addition,

let π_P be the unique stationary distribution of this Markov chain. We define the mixing time for any $\varepsilon \in [0, 1)$ as

$$t_P(\varepsilon) = \max_{x \in \Omega} \{\min\{t \in \mathbb{N}_{>0} \mid \forall t' \geq t : d_{\mathrm{TV}}(P^{t'}(x, \cdot), \pi_P) \leq \varepsilon\}\}. \qquad \square$$

Informally, the mixing time bounds for any starting state $x \in \Omega$ the necessary number of steps to get ε-close to its stationary distribution in terms of total variation distance.

Finally, we would like to give one more definition, which is not necessarily only used for analysing Markov chains, but which turned out to be very helpful, and which we will also need frequently for our proofs. Namely, this is the definition of a coupling of random variables.

Definition 2.5 (Coupling). Let X and Y be random variables on countable state spaces Ω_X and Ω_Y, respectively. A coupling between X and Y is a new random variable Z on $\Omega_X \times \Omega_Y$, distributed such that for any $S_X \subseteq \Omega_X$ and $S_Y \subseteq \Omega_Y$

$$\Pr[Z \in S_X \times \Omega_Y] = \Pr[X \in S_X] \text{ and } \Pr[Z \in \Omega_X \times S_Y] = \Pr[Y \in S_Y]. \qquad \square$$

Informally, this means a coupling is a joint distribution of X and Y, such that each of its marginals behaves according to its original distribution.

Asymptotic runtime, approximations and complexity

We would like to clarify some important points regarding our notation of runtime and our notion of efficient (randomized) approximation. For asymptotic bounds on runtime, we will mainly use the well known Landau notation. Moreover, we write $\mathrm{poly}(n)$ for the class of functions with polynomial upper bound in n. Thus, writing $f(n) \in \mathrm{poly}(n)$ means there exists a function h, polynomial in n, such that $f(n) \in O(h(n))$. With slight abuse of notation, we might also write $f(n) \leq \mathrm{poly}(n)$ for saying that f has a polynomial upper bound and $f(n) \geq \mathrm{poly}(n)$ if f has a polynomial lower bound in n (i.e., there is a polynomial function h such that $f(n) \in \Omega(h(n))$). In addition, we will also use this notation in exponential functions. Thus, for example $f(n) \leq e^{\mathrm{poly}(n)}$ means that there is a function $h(n) \in \mathrm{poly}(n)$ such that $f(n) \in O(e^{h(n)})$, and $f(n) \geq e^{\mathrm{poly}(n)}$ means that there is a function $h(n) \in \mathrm{poly}(n)$ such that $f(n) \in \Omega(e^{h(n)})$.

For the rest of the thesis, we might use phrases like *efficient (randomized) approximation*, depending on some input size n. We say that there is an efficient approximation algorithm if there is an FPTAS (fully polynomial time approximation scheme).

Definition 2.6 (FPTAS). Given a set of instances S and a function $f : S \to \mathbb{R}_{>0}$. We say there is an FPTAS for f if for every $\varepsilon > 0$ there is an algorithm A such that for every instance $s \in S$ encoded with input size n it holds that

$$(1 - \varepsilon) f(s) \leq A(s) \leq (1 + \varepsilon) f(s)$$

and A has runtime in poly $\left(\frac{n}{\varepsilon}\right)$.

Similarly, we say that there is an efficient randomized approximation algorithm if there is an FPRAS (fully polynomial randomized approximation scheme).

Definition 2.7 (FPRAS). Given a set of instances S and a function $f : S \to \mathbb{R}_{>0}$. We say there is an FPRAS for f if for every $\varepsilon > 0$ there is a randomized algorithm A such that for every instance $s \in S$ encoded with input size n it holds that

$$\Pr[(1 - \varepsilon) f(s) \leq A(s) \leq (1 + \varepsilon) f(s)] \geq \frac{3}{4}$$

and A has runtime in poly $\left(\frac{n}{\varepsilon}\right)$.

It is quite common to use the probability $\frac{3}{4}$ for the definition of an FPRAS. However, note that its choice is rather arbitrary. More precisely, as long as the success probability is strictly larger than $\frac{1}{2}$, any success probability $1 - \delta$ can be archived by taking the median of $\log(\frac{1}{\delta})$ independent runs (see for example Lemma 6.1 of Jerrum et al. [JVV86]).

Finally, it will be useful to have some understanding of complexity classes for decision problems and their randomized extensions, like P, NP, RP and BPP. As discussions about proven and conjectured relations between these classes are outside the scope of this theses, we would like to refer to standard literature like [Pap03].

3 Polymer Dynamics

In this chapter, we will propose a Markov chain to sample from the Gibbs distribution of abstract polymer models, assuming only little knowledge about the structure of the polymer graph. Namely, this knowledge will be captured by so called polymer cliques. The main idea of our chain will be to first sample a polymer clique uniformly at random, and then sample a polymer from that clique with respect to a certain clique polymer distribution.

I **Definition 3.1 (Polymer clique).** For a set of polymers C and an incompatibility relation $\sim$, we define a subset $\Lambda \subseteq \mathsf{C}$ as polymer clique, if $\gamma \not\sim \gamma'$ for all $\gamma, \gamma' \in \Lambda$. J

For readers who are familiar with cluster expansion approaches for approximating partition functions, we would like to point out that this is different from the definition of a polymer cluster. Commonly, the definition of a polymer cluster does not assume pairwise incompatibility, but only connectedness in the polymer graph, which is crucial for the inclusion-exclusion principal of cluster expansion.

Based on polymer cliques, we can now define a polymer clique cover as follows.

I **Definition 3.2 (Polymer clique cover).** For a set of polymers C and an incompatibility relation $\sim$, a polymer clique cover of size m is a collection of polymer cliques $\Lambda_1, \ldots, \Lambda_m$, such that $\dot{\bigcup}_{1 \le i \le m} \Lambda_i = \mathsf{C}$. J

Note that this definition does not demand the polymer cliques $\Lambda_1, \ldots, \Lambda_m$ to be disjoint, in order to form a polymer clique cover. The only requirement is that every polymer is in at least one polymer clique.

For using such a cover to construct a Markov chain, the following condition will play a key role.

I **Definition 3.3 (Generalized polymer mixing condition).** We say a polymer model $(\mathcal{C}, w, \circ)$ fulfills the generalized polymer mixing condition if there is a function $f : \mathcal{C} \to \mathbb{R}_{>0}$ such that for all polymers $\gamma \in \mathcal{C}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'} \leq f(\gamma). \qquad \text{J}$$

In the following, we will use a given polymer clique cover for a polymer model $(\mathcal{C}, w, \circ)$ which fulfills the generalized polymer mixing condition to construct a Markov chain that converges to the Gibbs measure μ. In addition, we will bound the mixing time of this chain using the size of the polymer clique cover m and the function f from the generalized polymer mixing condition.

3.1 Polymer Markov chain

Given a polymer model $(\mathcal{C}, w, \circ)$ we will now construct a Markov chain $(X_t)_{t \in \mathbb{N}}$ on the set of polymer families $\mathcal{F}$. To state this chain, we will need the notion of a clique polymer distribution, defined as follows.

I **Definition 3.4 (Clique polymer distribution).** Let $(\mathcal{C}, w, \circ)$ be a polymer model that fulfills the generalized polymer mixing condition and let $\Lambda_1, \dots, \Lambda_m$ be a polymer clique cover. For each polymer clique Λ_i, we define the clique polymer distribution ν_i on $\Lambda_i \cup \{\emptyset\}$ as

$$\begin{aligned} \nu_i(\gamma) &= w_\gamma \text{ for all } \gamma \in \Lambda_i \\ \nu_i(\emptyset) &= 1 \sum_{\gamma \in \Lambda_i} w_\gamma. \end{aligned} \qquad \text{J}$$

Obviously, we should argue that this is a valid probability distribution.

I **Lemma 3.5.** Given a polymer model $(\mathcal{C}, w, \circ)$ fulfilling the generalized polymer mixing condition for f as in Definition 3.3. For all polymer clique covers $\Lambda_1, \dots, \Lambda_m$ and every clique Λ_i the clique polymer distribution ν_i, defined as in Definition 3.4, is a probability distribution on $\Lambda_i \cup \{\emptyset\}$. J

Proof of Lemma 3.5. By definition, we know that $\nu_i(\gamma) = w_\gamma \geq 0$. It remains to show that $\prod_{\gamma \in \Lambda_i} w_\gamma \leq 1$. By the definition of a polymer clique, we know that all

polymers in Λ_i are pairwise incompatible. Thus, for every $\gamma \in \Lambda_i$ we have

$$\sum_{\gamma' \in \Lambda_i} f(\gamma') w_{\gamma'} \leq \sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'}.$$

In addition, the generalized polymer mixing condition gives us

$$\sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'} \leq f(\gamma).$$

Now let $\gamma_{\min} \in \Lambda_i$ be such that for all $\gamma' \in \Lambda_i$ it holds that $f(\gamma_{\min}) \leq f(\gamma')$. Such a $\gamma_{\min}$ does always exist, as our polymer cliques are subsets of a finite set of polymers. We obtain that

$$\sum_{\gamma' \in \Lambda_i} f(\gamma') w_{\gamma'} \leq f(\gamma_{\min}).$$

As f is strictly positive, we can divide by $f(\gamma_{\min})$ and get

$$\sum_{\gamma' \in \Lambda_i} w_{\gamma'} \sum_{\gamma' \in \Lambda_i} \frac{f(\gamma')}{f(\gamma_{\min})} w_{\gamma'} \leq 1;$$

which concludes our proof.

With this given, we can construct the desired Markov chain. The chain starts at some state $X_0 \in \Gamma$ (e.g. $X_0 = \emptyset$). Given we have $X_t = \Gamma$ at time t, we construct X_{t+1} as follows.

1. Choose a polymer clique Λ_i uniformly at random.
2. Uniformly at random, proceed either with a) or with b).
   a) If there is $\gamma \in \Gamma \setminus \Lambda_i$, set $X_{t+1} = \Gamma \cap \{\gamma\}$.
   Otherwise, set $X_{t+1} = X_t$.
   b) Draw γ from ν_i.
   If $\gamma = \emptyset$ or if there is a $\gamma' \in \Gamma$ such that $\gamma \not\sim \gamma'$, set $X_{t+1} = X_t$.
   Otherwise, set $X_{t+1} = \Gamma \cup \{\gamma\}$.

We call this Markov chain the *polymer Markov chain* and will denote its transitions by P. Their are some similarities to insert/delete chains, as for example

used for sampling independent set (see Dyer and Greenhill [DG00]). However, the main difference is that this chain does not choose single polymers (i.e., vertices of the polymer graph) uniformly at random. Instead, it makes use of an underlying clique cover and the fact that at any time at most one polymer from each clique can be in the current polymer family. This for example guarantees that in a) the polymer $\gamma \in \Gamma \setminus \Lambda_i$, if it exists, is uniquely defined.

First of all, we have to prove that this Markov chain converges to the desired stationary distribution μ.

Lemma 3.6 (Convergence to μ). Given the polymer Markov chain with transitions P, defined as above. For any $\Gamma \in \mathcal{F}$ the distribution $P^t(\Gamma, \cdot)$ converges to the unique stationary distribution μ as $t \to \infty$.

Before we start to prove this lemma, let us characterize the transitions P of the polymer Markov chain. For this, the following definition will be useful.

Definition 3.7. Given a polymer model $(\mathcal{C}, w, \cdot)$ with a polymer clique cover $\Lambda_1, \dots, \Lambda_m$. For every polymer $\gamma \in \mathcal{C}$ we define m_γ to be the number of polymer cliques Λ_i in the polymer clique cover such that $\gamma \in \Lambda_i$.

Let $\triangle$ denote the symmetric set difference. Note that for every pair of polymer families $\Gamma, \Gamma' \in \mathcal{F}$ with $|\Gamma \triangle \Gamma'| > 1$ it holds that $P(\Gamma, \Gamma') = 0$, because we add or remove at most one polymer at a time.

Now, assume $|\Gamma \triangle \Gamma'| = 1$ and without loss of generality let $\Gamma = \Gamma' \cup \{\gamma\}$ for some $\gamma \in \Gamma'$. By the definition of the polymer Markov chain, we have that $P(\Gamma, \Gamma') = \frac{m_\gamma}{2m} w_\gamma$ and $P(\Gamma', \Gamma) = \frac{m_\gamma}{2m}$. Moreover, both transition probabilities are strictly positive as we have $w_\gamma > 0$ and by definition of a polymer clique cover $m_\gamma > 0$.

Finally, we consider the self-loop probability $P(\Gamma, \Gamma)$ for any $\Gamma \in \mathcal{F}$. Assume that the chain chooses a polymer clique Λ_i with $\Gamma \cap \Lambda_i = \emptyset$, then there is probability of $\frac{1}{2}$ that we try to remove a polymer from Λ_i. As such a polymer does not exist in Γ, this does not change the state of the chain. Otherwise, if we choose a polymer clique Λ_i such that $\Gamma \cap \Lambda_i = \{\gamma\}$, then there is a probability of $\frac{1}{2}$ that we try to add a polymer from Λ_i, which fails as such a polymer is incompatible to γ. This results in a self-loop probability of $P(\Gamma, \Gamma) \geq \frac{1}{2}$. We now use these observations to prove that the polymer Markov chain converges to the Gibbs distribution μ.

Proof of Lemma 3.6. First we prove that the polymer Markov chain converges to a unique stationary distribution. For this, it is sufficient to argue that the chain is aperiodic, irreducible and positive recurrent. The chain is aperiodic as every state has a positive self-loop probability of at least $\frac{1}{2}$. In addition, for every polymer family $\Gamma \in \mathcal{F}$ we can go to the empty family $\emptyset$ by removing all $\gamma \in \Gamma$ with positive probability in a finite number of steps. Similarly, we can go from the empty polymer family to any other $\Gamma' \in \mathcal{F}$ by adding every $\gamma \in \Gamma'$ with positive probability. This proves irreducibility. Finally, the state space is finite and irreducible, and thus every state is positive recurrent.

It remains to show that μ is the stationary distribution of the polymer Markov chain. For this, we prove that μ fulfills detailed balance with respect to P, which implies the desired result. This means, it suffices to show that for every $\Gamma, \Gamma' \in \mathcal{F}$,

$$\mu(\Gamma) P(\Gamma, \Gamma') = \mu(\Gamma') P(\Gamma', \Gamma).$$

For $\Gamma = \Gamma'$, this is trivially true. The same holds for $|\Gamma \triangle \Gamma'| > 1$, as in this case $P(\Gamma, \Gamma') = P(\Gamma', \Gamma) = 0$. Let us assume $\Gamma = \Gamma' \cup \{\gamma\}$ for some $\gamma \in \Gamma'$. By definition of μ and P, we have that

$$\mu(\Gamma) P(\Gamma, \Gamma') = \frac{\prod_{\gamma' \in \Gamma} w_{\gamma'}}{Z} \frac{m_\gamma}{2m} w_\gamma = \frac{\prod_{\gamma' \in \Gamma'} w_{\gamma'}}{Z} \frac{m_\gamma}{2m} = \mu(\Gamma') P(\Gamma', \Gamma).$$

This concludes our proof.

We now know that our polymer Markov chain converges to the desired stationary distribution. However, for using it algorithmically, its mixing time is of special interest. In order to bound the mixing time of this chain, we will apply a coupling argument, which will be stated more precisely in the following section.

3.2 Coupling with exponential diameter

In this section, we will investigate a useful lemma for bounding the mixing time of a Markov chain. The method will be very similar to a well know coupling method, stated for example in Theorem 2.1 of Dyer and Greenhill [DG98]. The main idea is to define for a Markov chain with state space Ω some non-negative integer-valued function $\delta : \Omega^2 \to \mathbb{N}$ such that $\delta(x, y) = 0$ if and only if $x = y$.

If it is possible to construct a coupling of the transitions of two versions of the Markov chain such that δ is in expectation strictly decreasing by a constant factor for all pairs of states $x; y \in \Omega$, then the mixing time can be upper bounded logarithmically in $\frac{1}{\varepsilon}$ and the diameter $D = \max_{x;y \in \Omega} \delta(x;y)$. However, if it is only possible to show that δ is non-increasing in expectation, we still can bound the mixing time polynomially in the diameter and logarithmically in $\frac{1}{\varepsilon}$, as long as for all pairs of states $x; y \in \Omega$ with $x \neq y$ there is at least a constant probability that δ is changing with each step of the chain.

This tends to be a useful method for bounding the mixing time, as long as the diameter is polynomial in the size of our input and if δ only takes integer values. However, for our polymer chain, we will apply a function δ that depends on the function f, used in the generalized polymer mixing condition. In order not to restrict f to integer-valued functions with an upper bound polynomial in the input size, we need a version of this coupling argument that can also handle exponential and real-valued functions δ. Such a version was proposed by Greenberg et al. [GPR] in Theorem 3.3. We decide to restate and prove a different version of their argument. This has two reasons. First of all, their theorem assumes that δ takes no values in the interval $(0; 1)$, which we might want to relax to an arbitrary interval $(0; d)$ for some $d > 0$. In addition, their theorem is incorrectly stated.

Lemma 3.8 (Coupling with exponential diameter). Given an irreducible, positive recurrent, aperiodic Markov chain on a state space Ω with transitions P, such that for all $x \in \Omega$ it holds that $P(x;x) > 0$. For some $d > 0$, let $\delta : \Omega^2 \to \{0\} \cup \mathbb{R}_{\geq d}$ be a function, taking only finitely many values in $\{0\} \cup [d; D]$ and let $\delta(x;y) = 0$ if and only if $x = y$. Assume their is a coupling between the transitions of two copies of the chain $(X_t)_{t \in \mathbb{N}}; (Y_t)_{t \in \mathbb{N}}$, such that for every pair of states $x; y \in \Omega$ it holds that

$$\mathrm{E}[\delta(X_{t+1}; Y_{t+1}) \mid X_t = x; Y_t = y] \leq \delta(x;y):$$

If there are $\kappa; \eta \in (0; 1)$ such that for the same coupling and all $x; y \in \Omega$ with $x \neq y$ it holds that

$$\Pr[|\delta(X_{t+1}; Y_{t+1}) - \delta(x;y)| \geq \eta\delta(x;y) \mid X_t = x; Y_t = y] \leq \kappa;$$

then the mixing time can be bounded by

$$t_P(\varepsilon) \leq \frac{3e \ln(D \cdot d^2)}{\ln(1+\eta^2\kappa)} \ln\frac{1}{\varepsilon} \; . \qquad \square$$

Besides the fact that we added the lower bound d to the function δ and included its effect on the mixing time bound, the main difference to Greenberg et al. [GPR] is that they use some adjacency structure $U \subseteq \Omega^2$ and state that, assuming δ is a metric, it is sufficient that $\mathbb{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \leq \delta(x,y)$ and $\Pr[\delta(X_{t+1}, Y_{t+1}) \neq \delta(x,y)] \geq \eta\delta(x,y)$ hold for adjacent pairs of states $(x,y) \in U$. However, this is false, and a simple counterexample can be constructed by considering the mixing time of a random walk on a cycle with self-loops.

To circumvent this, we decide to state Lemma 3.8 using a coupling for all pairs of states and assuming that all necessary properties hold for all pairs of states in the first place. As a side effect of this, δ is not restricted to be a metric anymore.

These differences aside, most of the proof is similar to Greenberg et al. [GPR]. However, we decide to reprove our version of this coupling lemma to have a consistent version, which we can refer to.

For proving Lemma 3.8, it will be useful to have the definition of a coupling time.

Definition 3.9 (Coupling time). Let $(X_t)_{t\in\mathbb{N}}, (Y_t)_{t\in\mathbb{N}}$ be two copies of a Markov chain on a state space Ω. For initial states $x, y \in \Omega$ we define the coupling time with respect to x, y as the following random variable

$$T_{xy} = \inf\{t \in \mathbb{N} \mid X_t = Y_t \text{ given that } X_0 = x, Y_0 = y\}.$$

The coupling time of the Markov chain is defined as $T_P = \max_{x,y\in\Omega} \mathbb{E}[T_{xy}]$. $\square$

In the proof of Lemma 3.8 we will focus on bounding the coupling time and apply the following lemma.

Lemma 3.10 (Aldous [Ald83], 3.13). Given an irreducible, positive-recurrent, aperiodic Markov chain with transitions P and coupling time T_P on a state space Ω. For any coupling, the mixing time can be upper bounded by

$$t_P(\varepsilon) \leq eT_P \ln\frac{1}{\varepsilon} \; . \qquad \square$$

In order to bound the coupling time, we will apply the following lemma about the first hitting time of a stochastic process, which was also used by Greenberg et al. [GPR].

Lemma 3.11 (Greenberg et al. [GRS17], Lemma 3.5). Given a bounded stochastic process $(S_t)_{t \in \mathbb{N}}$ with $d \leq S_t \leq D$ for some $d, D \in \mathbb{R}$ and all $t \geq 0$. In addition, let q be some stopping value and $T = \inf\{t \in \mathbb{N} \mid S_t = q\}$. If $\mathbb{E}[S_{t+1} \mid S_t = s] \leq s$ and $\mathbb{E}[(S_{t+1} - s)^2 \mid S_t = s] \geq Q$ for all $t \leq T$, then it holds that

$$\mathbb{E}[T] \leq \frac{D^2 + q^2 - 2Dq}{Q}.$$

With this, we can finally start the proof.

Proof of Lemma 3.8. First of all, we define $\delta'(x, y) = \frac{\delta(x,y)}{d}$. Note that δ' now takes finitely many values in $\{0\} \cup [1, D/d]$. It holds that

$$X_t = Y_t \Rightarrow \delta(X_t, Y_t) = 0 \Rightarrow \delta'(X_t, Y_t) = 0.$$

In addition, by linearity of expectation and the conditions of the lemma it holds for every $x, y \in \Omega$ that

$$\begin{aligned}
&\mathbb{E}[\delta'(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \\
&= \frac{1}{d}\mathbb{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \\
&\leq \frac{1}{d}\delta(x, y) \\
&= \delta'(x, y)
\end{aligned}$$

and also that

$$\begin{aligned}
&\Pr[|\delta'(X_{t+1}, Y_{t+1}) - \delta'(x, y)| \geq \eta\delta'(x, y) \mid X_t = x, Y_t = y] \\
&= \Pr\left[\tfrac{1}{d}|\delta(X_{t+1}, Y_{t+1}) - \delta(x, y)| \geq \eta\tfrac{1}{d}\delta(x, y) \;\middle|\; X_t = x, Y_t = y\right] \\
&= \Pr[|\delta(X_{t+1}, Y_{t+1}) - \delta(x, y)| \geq \eta\delta(x, y) \mid X_t = x, Y_t = y] \\
&\leq \kappa.
\end{aligned}$$

Thus, these properties of δ also hold for δ^0. From this point, the proof is mainly the same as by Greenberg et al. [GPR], except that it holds for all pairs of states in the first place.

We define a random process $(\varphi_t^{xy})_{t\in\mathbb{N}}$ with $\varphi_t^{xy} = \delta^0(X_t, Y_t)$ given $X_0 = x, Y_0 = y$. As we cannot assume this process to exhibit any stochastic drift, we need to transform it to some other process. This transformation is done by the function $\Phi : [0, D\cdot d] \to \mathbb{R}$ with

$$\Phi(x) = \begin{cases} 2\ln(2) & \text{if } x \in [0, 1] \\ \ln(x) & \text{if } x \in [1, D\cdot d] \end{cases}.$$

Based on that, we define a new stochastic process $\psi_t^{xy} = \Phi(\varphi_t^{xy})$. Now, we have

$$\psi_t^{xy} = 2\ln(2), \quad \varphi_t^{xy} = 0, \quad X_t = Y_t \text{ for } X_0 = x, Y_0 = y.$$

Thus, if we can upper bound the maximum expected $t \in \mathbb{N}$ such that $\psi_t^{xy} = 2\ln(2)$ for all $x, y \in \Omega$, this will give us a bound on the coupling time T_P as in Definition 3.9, which results in a bound on the mixing time by using Lemma 3.10.

To get such a bound, we will apply a stopping time argument from Lemma 3.11. For this, we need to show that $\mathrm{E}[\psi_{t+1}^{xy} - \psi_t^{xy} = s] \leq s$ and we need a lower bound on $\mathrm{E}[(\psi_{t+1}^{xy} - s)^2 \mid \psi_t^{xy} = s]$ for all $s > 0$.

Note that $\psi_t^{xy} = s > 0$ implies that $\varphi_t^{xy} > 1$. As Φ is a concave function on the entire interval $[0, D\cdot d]$, we can use Jensen's inequality for expected values to get

$$\begin{aligned} \mathrm{E}[\psi_{t+1}^{xy} - \psi_t^{xy} = s] &= \mathrm{E}[\Phi(\varphi_{t+1}^{xy}) - \Phi(\varphi_t^{xy}) = s] \\ &\leq \Phi(\mathrm{E}[\varphi_{t+1}^{xy} - \Phi(\varphi_t^{xy}) = s]) \\ &= \Phi(\mathrm{E}[\varphi_{t+1}^{xy} - \varphi_t^{xy} = e^s] - \Phi(e^s)) \\ &\leq s. \end{aligned}$$

Next we lower bound $\mathrm{E}[(\psi_{t+1}^{xy} - s)^2 \mid \psi_t^{xy} = s]$ for $s > 0$. First, we rewrite this in terms of conditional expectations. To do so, let A be the event, that the process jumps from $\psi_t^{xy} = s > 0$ directly to $\psi_{t+1}^{xy} = 2\ln(2)$ (i.e., from $\varphi_{t+1}^{xy} > 1$

to $\varphi_{t+1}^{xy} = 0$). By the law of total expectation, we get

$$\begin{aligned}\mathrm{E}\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s\right] &= \mathrm{E}\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s, A\right] \Pr[A] \\ &\quad + \mathrm{E}\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s, \bar{A}\right] 1 \quad \Pr[\bar{A}].\end{aligned}$$

Because $\psi_t^{xy} \leq 0$, we have

$$\mathrm{E}\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s, A\right] \Pr[A] = \left(2 \ln(2) \psi_t^{xy}\right) \Pr[A] \leq 4 \ln(2) \Pr[A].$$

In addition, we know that the process $(\psi_t^{xy})_{t \in \mathbb{N}}$ takes only finitely many different values because δ^0 does so. This also implies that $\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s\right]$ conditioned on $\psi_t^{xy} = s$ takes only finitely many different values. Let R be the set of all such values. Now we can rewrite the second conditional expectation as

$$\begin{aligned}\mathrm{E}\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = s, \bar{A}\right] &= \sum_{r \in R} r \Pr\left[\psi_{t+1}^{xy} \mid \psi_t^{xy} = r \quad \psi_t^{xy} = s, \bar{A}\right] \\ &\quad \ln(1+\eta) \Pr\left[\psi_{t+1}^{xy} \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right]\end{aligned}$$

Next, we want to lower bound this probability. Note that

$$\Pr\left[\psi_{t+1}^{xy} \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right] = \Pr\left[\psi_{t+1}^{xy} \neq s \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right].$$

This can further be decomposed as

$$\begin{aligned}\Pr\left[\psi_{t+1}^{xy} \neq s \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right] &= \Pr\left[\psi_{t+1}^{xy} > s \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right] \\ &\quad + \Pr\left[\psi_{t+1}^{xy} < s \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right].\end{aligned}$$

We rewrite the first probability as

$$\begin{aligned}\Pr\left[\psi_{t+1}^{xy} > s \mid \ln(1+\eta) \quad \psi_t^{xy} = s, \bar{A}\right] &= \Pr\left[\ln\left(\frac{\varphi_{t+1}^{xy}}{e^s}\right) > \ln(1+\eta) \quad \varphi_t^{xy} = e^s, \bar{A}\right] \\ &= \Pr\left[\frac{\varphi_{t+1}^{xy}}{e^s} > 1+\eta \mid \varphi_t^{xy} = e^s, \bar{A}\right] \\ &= \Pr\left[\varphi_{t+1}^{xy} > e^s + \eta e^s \mid \varphi_t^{xy} = e^s, \bar{A}\right].\end{aligned}$$

Furthermore, note that $\ln(1+x) \leq \ln(1) + x$ for $x \in [0,1)$. Thus, we can bound the second probability by

$$\begin{aligned}
\Pr\left[\psi_{t+1}^{xy} \geq s \cdot \ln(1+\eta) \cdot \psi_t^{xy} = s, \overline{A}\right] &\leq \Pr\left[\psi_{t+1}^{xy} \geq s \cdot \ln(1) + \eta \cdot \psi_t^{xy} = s, \overline{A}\right] \\
&= \Pr\left[\ln\left(\frac{\varphi_{t+1}^{xy}}{e^s}\right) \geq \ln(1) + \eta \cdot \varphi_t^{xy} = e^s, \overline{A}\right] \\
&= \Pr\left[\frac{\varphi_{t+1}^{xy}}{e^s} \geq 1 + \eta \cdot \varphi_t^{xy} = e^s, \overline{A}\right] \\
&= \Pr\left[\varphi_{t+1}^{xy} \geq e^s + \eta e^s \cdot \varphi_t^{xy} = e^s, \overline{A}\right].
\end{aligned}$$

Altogether, this gives us

$$\begin{aligned}
\Pr\left[|\psi_{t+1}^{xy} - s| \geq \eta s \cdot |\psi_t^{xy} = s, \overline{A}\right] &= \Pr\left[|\psi_{t+1}^{xy} - s| \geq \eta s \cdot |\psi_t^{xy} = s, \overline{A}\right] \\
&\quad \Pr\left[|\varphi_{t+1}^{xy} - e^s| \geq \eta e^s \cdot \varphi_t^{xy} = e^s, \overline{A}\right].
\end{aligned}$$

By assumption, we have a bound on the probability that $(\varphi_t^{xy})_{t \in \mathbb{N}}$ takes steps of at least size $\eta s'$ given $\varphi_t^{xy} = s'$ 1. We rewrite this by the law of total probability as

$$\begin{aligned}
\Pr\left[|\varphi_{t+1}^{xy} - s'| \geq \eta s' \cdot \varphi_t^{xy} = s'\right] &= \Pr\left[|\varphi_{t+1}^{xy} - s'| \geq \eta s' \cdot \varphi_t^{xy} = s', A\right] \Pr[A] \\
&\quad + \Pr\left[|\varphi_{t+1}^{xy} - s'| \geq \eta s' \cdot \varphi_t^{xy} = s', \overline{A}\right] (1 - \Pr[A]).
\end{aligned}$$

Now, remember that A is the event that we go from φ_t^{xy} 1 to $\varphi_{t+1}^{xy} = 0$. Thus, we now that for every $\eta \in [0,1)$

$$\Pr\left[|\varphi_{t+1}^{xy} - s'| \geq \eta s' \cdot \varphi_t^{xy} = s', A\right] = 1.$$

In addition, we know that our chain has a positive self-loop probability. Thus, we know that starting from $\varphi_t^{xy} = s'$ 0 the probability of going to $\varphi_{t+1}^{xy} = 0$ is strictly less than 1. This implies $\Pr[A] < 1$ and $1 - \Pr[A] > 0$. Knowing this, we

can write

$$\Pr\left[\left|\varphi_{t+1}^{xy} - s^0\right| \geq \eta s^0 \;\middle|\; \varphi_t^{xy} = s^0; \overline{A}\right] = \frac{\Pr\left[\left|\varphi_{t+1}^{xy} - s^0\right| \geq \eta s^0 \;\middle|\; \varphi_t^{xy} = s^0 \wedge \Pr[A]\right]}{1 - \Pr[A]} \cdot \frac{\kappa - \Pr[A]}{1 - \Pr[A]}.$$

By setting $s^0 = e^s$, this gives us the bound

$$\mathbb{E}\left[\psi_{t+1}^{xy} - s^0 \;\middle|\; \psi_t^{xy} = s; \overline{A}\right] \leq \ln(1+\eta^0) \Pr\left[\left|\psi_{t+1}^{xy} - s^0\right| \geq \ln(1+\eta^0) \;\middle|\; \psi_t^{xy} = s; \overline{A}\right] \cdot \ln(1+\eta^0) \frac{\kappa - \Pr[A]}{1 - \Pr[A]}.$$

Finally, by $\eta < 1$ we know that $\ln(1+\eta^0) \leq \ln(2^0)$. So, we get

$$\begin{aligned}\mathbb{E}\left[\psi_{t+1}^{xy} - s^0 \;\middle|\; \psi_t^{xy} = s\right] &\leq 4\ln(2^0)^2 \Pr[A] + 1 \cdot \Pr[A]^0 \ln(1+\eta^0) \frac{\kappa - \Pr[A]}{1 - \Pr[A]} \\ &= 4\ln(2^0)^2 \Pr[A] + \ln(1+\eta^0)\kappa \cdot \ln(1+\eta^0) \Pr[A] \\ &\leq 3\ln(2^0)^2 \Pr[A] + \ln(1+\eta^0)\kappa \\ &\leq \ln(1+\eta^0)\kappa.\end{aligned}$$

As we now know that

$$\begin{aligned}\mathbb{E}\left[\psi_{t+1}^{xy} \;\middle|\; \psi_t^{xy} = s\right] &\leq s \quad \text{and} \\ \mathbb{E}\left[\left|\psi_{t+1}^{xy} - s^0\right| \;\middle|\; \psi_t^{xy} = s\right] &\leq \ln(1+\eta^0)\kappa\end{aligned}$$

and as $(\varphi_t^{xy})_{t \in \mathbb{N}}$ is a bounded stochastic process, we can apply Lemma 3.11 to get for every $x, y \in \Omega$

$$\mathbb{E}[T_{xy}] \leq \frac{2\ln(D \cdot d^0)^2 + 4\ln(2^0)^2 + 4\ln(2^0)\ln(D \cdot d^0)}{\ln(1+\eta^0)\kappa} \leq \frac{3\ln(D \cdot d^0)^2}{\ln(1+\eta^0)\kappa}.$$

Together with Lemma 3.10, this gives the desired mixing time bound of

$$t_P(\varepsilon^0) \leq \frac{3e\ln(D \cdot d^0)^2}{\ln(1+\eta^0)\kappa} \ln\left(\frac{1}{\varepsilon}\right).$$

3.3 Mixing Time of the polymer Markov chain

We are now going to use Lemma 3.8 to bound the mixing time of the polymer Markov chain.

Lemma 3.12 (Mixing time of polymer Markov chain). Let $(\mathcal{C}, w, \phi)$ be a polymer model that fulfills the generalized polymer mixing condition for a function f as in Definition 3.3, and let $\Lambda_1, \dots, \Lambda_m$ be a given polymer clique cover of size m as described in Definition 3.2. The resulting polymer Markov chain with transition matrix P is mixing in time

$$t_P(\varepsilon) \in O\left(m^2 \ln\left(m^2 \frac{\max_{\gamma \in \mathcal{C}} f(\gamma)}{\min_{\gamma \in \mathcal{C}} f(\gamma)}\right)^2 \ln\left(\frac{1}{\varepsilon}\right)\right).$$

Proof of Lemma 3.12. We already know that the polymer Markov chain is irreducible, positive-recurrent and aperiodic, and that it has a self-loop probability of at least $\frac{1}{2}$. For every $\Gamma, \Gamma' \in \mathcal{F}$ we define

$$\delta(\Gamma, \Gamma') = \sum_{\gamma \in \Gamma \triangle \Gamma'} \frac{f(\gamma)}{m_\gamma}.$$

Obviously, we have $\delta(\Gamma, \Gamma') = 0$ if and only if $\Gamma = \Gamma'$ because $f(\gamma) > 0$ for all $\gamma \in \mathcal{C}$. Now we set

$$d = \min_{\Gamma, \Gamma'} \delta(\Gamma, \Gamma') = \min_{\gamma \in \mathcal{C}} \frac{f(\gamma)}{m_\gamma}$$
$$D = \max_{\Gamma, \Gamma'} \delta(\Gamma, \Gamma') = \max_{\Gamma, \Gamma'} \sum_{\gamma \in \Gamma \triangle \Gamma'} \frac{f(\gamma)}{m_\gamma}.$$

It holds that $d > 0$, and as there are only finitely many pairs of polymer families, δ takes only finitely many values in $\{0\} \cup [d, D]$. Thus, δ fulfills all conditions from Lemma 3.8.

In the next step, we need to construct a coupling of every pair states. Let $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$ be two copies of the polymer Markov chain. We couple every transition of these chains by letting both chains try to do the same update. That means both chains choose the same polymer clique Λ_i and either both try

to remove a polymer $\gamma \in \Lambda_i$ or both try to add the same polymer. This is a valid coupling, as all marginal transition probabilities are preserved.

As soon as $X_t = Y_t$, they remain equal for all $t' \geq t$. To apply Lemma 3.8, we now have to bound $\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = \Gamma, Y_t = \Gamma']$ for all pairs $|\Gamma \setminus \Gamma'| \leq 1$. Let us define $M = \Gamma \setminus \Gamma'$. Note that δ decreases by $\frac{f(\gamma)}{m_\gamma}$ whenever we try to remove a polymer $\gamma \in M$. This happens at least with probability $\frac{m_\gamma}{2m}$ for each such γ. In addition, δ increases if and only if we add a polymer γ' to only one of both chains. If this is the case, we know that there is $\gamma \in M$ such that $\gamma \not\sim \gamma'$ (we will write this as $\gamma' \not\sim M$) because otherwise we either could have added γ' to Γ and Γ', or to none of both. For each such polymer $\gamma' \not\sim M$, this happens with probability at most $w_{\gamma'} \frac{m_{\gamma'}}{2m}$, and it increases δ by $\frac{f(\gamma')}{m_{\gamma'}}$. From this, we get the following bound

$$\begin{aligned}
\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = \Gamma, Y_t = \Gamma'] &= \delta(\Gamma, \Gamma') \\
&\quad - \Bigg(\sum_{\gamma \in M} \frac{f(\gamma)}{m_\gamma} \frac{m_\gamma}{2m}\Bigg) + \Bigg(\sum_{\gamma' \not\sim M} \frac{f(\gamma')}{m_{\gamma'}} \frac{m_{\gamma'}}{2m} w_{\gamma'}\Bigg) \\
&= \delta(\Gamma, \Gamma') - \frac{1}{2m}\Bigg(\sum_{\gamma \in M} f(\gamma) - \sum_{\gamma' \not\sim M} f(\gamma') w_{\gamma'}\Bigg) \\
&\leq \delta(\Gamma, \Gamma') - \frac{1}{2m}\Bigg(\sum_{\gamma \in M} f(\gamma) - \sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'}\Bigg).
\end{aligned}$$

Now we can use the generalized polymer mixing condition to see that

$$\sum_{\gamma \in M} \Bigg(f(\gamma) - \sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'}\Bigg) \geq f(\gamma) - \sum_{\gamma' \not\sim \gamma} f(\gamma') w_{\gamma'} \geq 0.$$

Thus, we have

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = \Gamma, Y_t = \Gamma'] \leq \delta(\Gamma, \Gamma').$$

This shows that δ is in expectation non-increasing. Next, we need to lower bound the probability that δ changes at least by some constant fraction. Again,

we assume $X_t = \Gamma$, $Y_t = \Gamma^0$ with Γ^0 , Γ and set $M = \Gamma \quad \Gamma^0$. We know that

$$\mathrm{j}M\mathrm{j} \quad \mathrm{j}\Gamma \ [\ \Gamma^0\mathrm{j} \quad \mathrm{j}\Gamma\mathrm{j} + \mathrm{j}\Gamma^0\mathrm{j} \quad 2m$$

as every polymer family contains at most m polymers. By definition, we have that $\delta^1\Gamma;\Gamma^{0}\mathrm{o} = \prod_{\gamma 2M} \frac{f^1\gamma^{\mathrm{o}}}{m_\gamma}$. Thus, there is as least one $\gamma_{\max} \ 2 \ M$ such that $\frac{f^1\gamma_{\max}{}^{\mathrm{o}}}{m_{\gamma_{\max}}} \quad \frac{\delta^1\Gamma;\Gamma^{0}\mathrm{o}}{2m}$. The probability that this polymer is removed in the next step is at least $\frac{m_{\gamma_{\max}}}{2m} \quad \frac{1}{2m}$. Thus, for $\kappa = \frac{1}{2m}$ and $\eta = \frac{1}{2m}$ we have

$$\Pr\text{»}\mathrm{j}\delta^1X_{t+1};Y_{t+1}{}^{\mathrm{o}} \quad \delta^1X_t;Y_t{}^{\mathrm{o}}\mathrm{j} \quad \eta\delta^1X_t;Y_t{}^{\mathrm{o}}\text{¼} \quad \kappa\text{:}$$

It remains to bound d and D. Note that it holds that

$$d = \min_{\gamma 2\mathrm{C}} \frac{f^1\gamma^{\mathrm{o}}}{m_\gamma} \quad \frac{\min_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}}{m}$$

$$D = \max_{\Gamma,\Gamma^0} \tilde{\mathrm{O}}_{\gamma 2\Gamma \quad \Gamma^0} \frac{f^1\gamma^{\mathrm{o}}}{m_\gamma} \quad 2m \max_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}$$

Applying Lemma 3.8 and the fact that $\ln^1 1 + 1\bullet 2m^{\mathrm{o}} \quad \frac{1}{4m}$ results in

$$t_P{}^1\varepsilon^{\mathrm{o}} \quad \frac{6em}{\ln^1 1 + 1\bullet 2m^{\mathrm{o}}} \ln\ 2m^2 \frac{\max_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}}{\min_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}}^{\ 2} \ln\ \frac{1}{\varepsilon}$$

$$2\ \mathrm{O}\ m^2 \ln\ m^2 \frac{\max_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}}{\min_{\gamma 2\mathrm{C}} f^1\gamma^{\mathrm{o}}}^{\ 2} \ln\ \frac{1}{\varepsilon} \quad :$$

Although such results on the mixing time of single polymer updates (also called Glauber dynamics) are of separate theoretical interest, we will mainly focus on algorithmic aspects. Such algorithmic implications will be discussed in the following section.

3.4 Algorithmic aspects of generalized polymer mixing condition

We will now use the described polymer Markov chain for our first efficient sampling scheme for the Gibbs distribution.

I **Theorem 3.13.** Let $^1C; w;\ ^o$ be a polymer model encoded with input size n and let $\Lambda_1; :::; \Lambda_m$ be a given polymer clique cover.
Given that:

(1) $m\ 2\ \text{poly}^1n^o$ and we can draw a polymer clique Λ_i uniformly at random in time poly^1m^o

(2) for every polymer clique Λ_i and every $\gamma\ 2\ C$ we can check whether $\gamma\ 2\ \Lambda_i$ in time poly^1n^o

(3) for every $\gamma; \gamma^0$ we can check whether $\gamma\quad\gamma^0$ in poly^1n^o

(4) $^1C; w;\ ^o$ fulfills the generalized polymer mixing condition as stated in Definition 3.3, and $e^{\text{poly}^1n^o}\quad f^1\gamma^o\quad e^{\text{poly}^1n^o}$ for every $\gamma\ 2\ C$

(5) we can sample exactly from each clique polymer distribution ν_i in time poly^1n^o

Then we can ε-approximately sample from the Gibbs distribution μ in time $\text{poly}\ \frac{n}{\varepsilon}$. J

Proof of Theorem 3.13. Because the polymer model fulfills the generalized polymer mixing condition, we can construct the polymer Markov chain with transitions P using the given polymer clique cover. Because of Lemma 3.6 and Lemma 3.12 we know that this chain has the Gibbs distribution μ as its stationary distribution and mixing time

$$t_P{}^1\varepsilon^o\ 2\ O\ m^2 \ln\ m^2 \frac{\max_{\gamma 2C} f^1\gamma^o}{\min_{\gamma 2C} f^1\gamma^o}^{\ 2} \ln\ \frac{1}{\varepsilon}\ !\ :$$

By assumption (1) we know that $m\ 2\ \text{poly}^1n^o$ and by assumption (4) we know that $\frac{\max_{\gamma 2C} f^1\gamma^o}{\min_{\gamma 2C} f^1\gamma^o}\ 2\ e^{\text{poly}^1n^o}$. Thus we have we have $t_P{}^1\varepsilon^o\ 2\ \text{poly}\ \frac{n}{\varepsilon}$. It remains to argue that each step of the polymer Markov chain can be computed in time $\text{poly}\ \frac{n}{\varepsilon}$.

According to (1) we can draw a polymer clique uniformly at random in time poly^1m^o, which is also in poly^1n^o.

Now, assume our current polymer family is $\Gamma\ 2\ F$. If we want to know whether there is a polymer $\gamma\ 2\ \Gamma\ \setminus\ \Lambda_i$, we can simply go over all $\gamma\ 2\ \Gamma$ and

check whether $\gamma \in \Lambda_i$. There can be at most $m \in \text{poly}^1 n^0$ polymers in every family Γ and according to (2) we can check each of them in time $\text{poly}^1 n^0$.

In addition, according to (5) we can sample a polymer γ from any ν_i in time $\text{poly}^1 n^0$. If we now want to check if we can add γ to Γ, we iterate over all $\gamma' \in \Gamma$ and check if $\gamma \not\sim \gamma'$. Again, there are at most $m \in \text{poly}^1 n^0$ polymers $\gamma' \in \Gamma$, and because of (3), we can check each of them in time $\text{poly}^1 n^0$. This results in an algorithm with the desired run time bounds.

For the rest of this section, we will have a closer look at the assumptions of Theorem 3.13. On the one hand, we need efficient access to some polymer clique cover $\Lambda_1, \dots, \Lambda_m$ of polynomial size m such that we can efficiently sample a polymer clique Λ_i uniformly at random. On the other hand, we need an efficient way to sample from each clique polymer distribution ν_i. Note that the number of polymers can be bounded by $|C| \leq m \max_{1 \leq i \leq m} |\{\Lambda_i\}|$. Thus, if the number of polymers is exponential in n, at least m or $\max_{1 \leq i \leq m} |\{\Lambda_i\}|$ has to be exponential in n as well. This leaves two ways for getting an efficient algorithm with our Markov chain.

The first one is to choose our polymer cliques small, such that we can sample from each ν_i efficiently. However, this implies that m is exponential in n. As our mixing time bound is polynomial in m and we need to draw a polymer clique Λ_i uniformly at random in each step, it might be necessary to show that we can ignore all but polynomial many polymer cliques, without diverging too much from our desired Gibbs distribution μ.

The second way for getting an efficient approximate sampling algorithm would be to use a polymer clique cover of polynomial size in the first place. However, this not only leads to the question how to obtain such a polymer clique cover, but it also implies that there might be polymer cliques of exponential size, making it harder to sample from each clique polymer distribution ν_i.

For the rest of this thesis, we will investigate the second algorithmic idea. The reason is that many applications of polymer models give us a polymer clique cover of polynomial size in a very natural way. Examples for this will be shown in the applications in Chapter 6. Thus, from now on the main question remains how to sample efficiently from each clique polymer distribution ν_i. In the next chapter, we will relax this condition, by showing that it is sufficient to sample approximately from each clique polymer distribution.

4 Markov Chains with Transition Error

In this chapter we will give some general tools for bounding the divergence of a Markov chain from its stationary distribution after introducing small errors in its transition probabilities. We will apply this to our polymer Markov chain.

More precisely, instead of sampling from the clique polymer distribution ν_i for a polymer clique Λ_i, we approximate this distribution by another distribution ξ_i. This results in different transitions, say Q. Our goal is now to find a bound $d_{\mathrm{TV}}(\xi_i, \nu_i) \leq \varepsilon_{\mathrm{in}}$, such that we can still use Q in order to approximately sample from the Gibbs measure μ.

4.1 Sample from Markov chains with transition errors

In this section, we will state a general lemma about sampling from a Markov chain with bounded transition error. For this, we will use the following formal notion of transition error.

Definition 4.1 (Bounded transition error). Given two Markov chains with transitions P and Q on a state space Ω. We say that the transition error between P and Q is bounded by $\varepsilon_{\mathrm{tr}} \geq 0$, if $d_{\mathrm{TV}}(P(x,\cdot), Q(x,\cdot)) \leq \varepsilon_{\mathrm{tr}}$ for all $x \in \Omega$.

In other words, by saying that the transition error of P and Q is bounded by $\varepsilon_{\mathrm{tr}}$, we mean that we can bound the error of the one-step transition distributions from any state $x \in \Omega$. The following lemma shows how such a bound on the one-step transitions can be used to show that after a certain number of steps t the t-step distribution $Q^t(x,\cdot)$ is close the stationary distribution of P.

Lemma 4.2 (Bound on sampling error). Given two Markov chains $(X_t)_{t\in\mathbb{N}}$ and $(Y_t)_{t\in\mathbb{N}}$ with transitions P and Q, respectively, on a state space Ω. In addition, let P converge to a unique stationary distribution π_P with mixing time $t_P(\varepsilon)$. For some $\varepsilon \in (0, \frac{1}{4}]$ let $t \geq t_P\left(\frac{\varepsilon}{2}\right)$. If the one-step transition error between P and Q is bounded by $\frac{\varepsilon}{2t}$, then it holds that $d_{\mathrm{TV}}(Q^t(x,\cdot), \pi_P) \leq \varepsilon$ for every $x \in \Omega$.

The proof of this lemma is mainly based on using a coupling. More precisely, we will use the so called coupling inequality and the existence of an optimal coupling, as stated in the following lemma.

Lemma 4.3 (Hollander [Hol12], Theorem 2.4 and 2.12). Let X and Y be random variables on a countable state space Ω with distribution κ_X and κ_Y, respectively. For any coupling of X and Y it holds that

$$d_{\mathrm{TV}}(\kappa_X, \kappa_Y) \leq \Pr(X \neq Y).$$

In addition, there is an optimal coupling of X and Y such that

$$d_{\mathrm{TV}}(\kappa_X, \kappa_Y) = \Pr(X \neq Y). \qquad \square$$

We will now use Lemma 4.3 to prove Lemma 4.2.

Proof of Lemma 4.2. The first step to prove this lemma is to apply the fact that the total variation distance is a metric. Thus, we can apply the triangle inequality to decompose the total variation distance we actually want to bound. For every $x \in \Omega$ and $t \in \mathbb{N}$, we have

$$d_{\mathrm{TV}}(Q^t(x, \cdot), \pi_P) \leq d_{\mathrm{TV}}(Q^t(x, \cdot), P^t(x, \cdot)) + d_{\mathrm{TV}}(P^t(x, \cdot), \pi_P).$$

We will continue by bounding each of these distances separately.

First, we look at $d_{\mathrm{TV}}(P^t(x, \cdot), \pi_P)$. By setting $t \geq t_P(\frac{\varepsilon}{2})$ as stated in the lemma and using the definition of a mixing time, we directly get $d_{\mathrm{TV}}(P^t(x, \cdot), \pi_P) \leq \frac{\varepsilon}{2}$.

In order to bound $d_{\mathrm{TV}}(Q^t(x, \cdot), P^t(x, \cdot))$, we apply the Lemma 4.3, which gives us

$$d_{\mathrm{TV}}(Q^t(x, \cdot), P^t(x, \cdot)) \leq \Pr(Y_t \neq X_t \mid Y_0 = x, X_0 = x)$$

for any coupling between $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$. We construct such a coupling as follows:

- if at any step $i \in \mathbb{N}$ it holds that $Y_i = X_i = y$, then select X_{i+1}, Y_{i+1} according to an optimal coupling of $P(y, \cdot)$ and $Q(y, \cdot)$,
- otherwise, each chain proceeds independently.

This results in a valid coupling because it preserves the marginal distributions of each step for each of the chains.

We proceed with upper bounding $\Pr[Y_t \neq X_t \mid Y_0 = x, X_0 = x']$ according to this coupling. For this, we instead lower bound the probability of the complementary event $\Pr[Y_t = X_t \mid Y_0 = x, X_0 = x']$. This probability is at least as high as the probability that both chains agree on every state $1 \leq i \leq t$. Using the Markov property of $(X_t)_{t \in \mathbb{N}}, (Y_t)_{t \in \mathbb{N}}$, this can be decomposed as follows

$$\begin{aligned}
&\Pr[Y_t \neq X_t \mid Y_0 = x, X_0 = x'] \\
&\quad \leq 1 - \Pr\left[\bigwedge_{1 \leq i \leq t} Y_i = X_i \;\middle|\; Y_0 = x, X_0 = x'\right] \\
&\quad = \Pr[Y_1 = X_1 \mid Y_0 = x, X_0 = x'] \prod_{1 < i \leq t} \Pr[Y_i = X_i \mid Y_{i-1} = X_{i-1}]
\end{aligned}$$

For each one-step transition, we know that

$$\Pr[Y_i = X_i \mid Y_{i-1} = X_{i-1}] = 1 \leq \Pr[Y_i \neq X_i \mid Y_{i-1} = X_{i-1}]$$

and because we coupled every single step optimally according to Lemma 4.3 if $Y_{i-1} = X_{i-1}$, we get

$$\Pr[Y_i \neq X_i \mid Y_{i-1} = X_{i-1}] = d_{\mathrm{TV}}(Q(Y_{i-1}, \cdot), P(X_{i-1}, \cdot)) \leq \frac{\varepsilon}{2t}.$$

This gives us

$$\Pr[Y_i = X_i \mid Y_{i-1} = X_{i-1}] \geq 1 - \frac{\varepsilon}{2t}.$$

Because we know that this holds for every single step, we can lower bound the probability that it holds for all steps by using Bernoulli's inequality and get

$$\Pr[Y_t = X_t \mid Y_0 = x, X_0 = x'] \geq 1 - \frac{\varepsilon}{2t} \cdot t \geq 1 - t\frac{\varepsilon}{2t} = 1 - \frac{\varepsilon}{2}.$$

Returning to our initial bound, this gives us

$$\begin{aligned}
d_{\mathrm{TV}}(Q^t(x, \cdot), P^t(x, \cdot)) &\leq \Pr[Y_t \neq X_t \mid Y_0 = x, X_0 = x'] \\
&\leq 1 - \Pr[Y_t = X_t \mid Y_0 = x, X_0 = x'] \\
&\leq \frac{\varepsilon}{2}.
\end{aligned}$$

$$\begin{aligned} & 1 \quad {}^{1}1 \quad \frac{\varepsilon}{2}{}^{o} \\ = & \frac{\varepsilon}{2}: \end{aligned}$$

Putting everything together, we obtain the desired result

$$\begin{aligned} d_{\mathrm{TV}}{}^{1}Q^{t}{}^{1}x;\ {}^{o};\pi_P{}^{o} \quad & d_{\mathrm{TV}}{}^{1}Q^{t}{}^{1}x;\ {}^{o};P^{t}{}^{1}x;\ {}^{oo} + d_{\mathrm{TV}}{}^{1}P^{t}{}^{1}x;\ {}^{o};\pi_P{}^{o} \\ & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ = & \varepsilon; \end{aligned}$$

which concludes our proof.

The above lemma is especially interesting in cases were we want to approximately sample from the stationary distribution π_P of a chain P with mixing time $t_P{}^{1}\varepsilon{}^{o}$ 2 poly $\frac{n}{\varepsilon}$, but where no efficient way is known for drawing each transition exactly according to P. Assuming that we can $\varepsilon_{\mathrm{tr}}$-approximate each single transition in time poly $\frac{n}{\varepsilon_{\mathrm{tr}}}$, we might as well run this approximated chain for t_P $\frac{\varepsilon}{2}$ steps with $\varepsilon_{\mathrm{tr}}$ $\frac{\varepsilon}{2t_P{}^{1}\frac{\varepsilon}{2}{}^{o}}$. This would result in an overall runtime of poly $\frac{n}{\varepsilon}$ again and Lemma 4.2 tells us that the resulting distribution is an ε-approximation of π_P. Thus, this lemma can also be seen as a way of reducing one approximate sampling problem on another.

We also noted that there is a variety of literature dealing with a very similar problem as Lemma 4.2, namely to bound the difference between stationary distributions for a given transition error. However, we want to point out two main reasons, why we used our lemma instead.

First of all, most of the literature as Cho and Meyer [CM01], Ferré et al. [FHL12], Hunter [Hun06], Mitrophanov [Mit05], and Solan and Vieille [SV03] either need a very detailed understanding of algebraic properties of the Markov chain or end up with including terms linear in the size of the state space. Both will most likely not be feasible for algorithmic applications, where little about detailed algebraic properties is known and the state space tends to be exponential in the input size.

Secondly, bounding distance between stationary distributions is more restrictive than required. For our purpose, it is sufficient that the approximated dynamics are close to the stationary distribution after a certain amount of time.

However, we do not need the approximated chain to converge to anything near this stationary distribution or even to converge to any stationary distribution at all. This has many important implications, like that we not need to assume any notion of ergodicity for the approximated chain.

4.2 Application to the polymer Markov chain

In this section, we will apply our bound from Lemma 4.2 to our polymer Markov chain, which will also result in one of our main theorems. As a reminder: the idea of this section is to replace the sampling from each clique polymer distribution ν_i by approximately sampling from it. The approximated distribution will be modeled by an alternative distribution ξ_i for each polymer clique. Our first step is to bound the transition error of the resulting Markov chain in terms of $d_{\mathrm{TV}}(\nu_i;\xi_i)$.

Lemma 4.4 (Bound on transition error). Given a polymer model $(\mathcal{C};w;\ldots)$ fulfilling the generalized polymer mixing condition and a polymer clique cover $\Lambda_1;\ldots;\Lambda_m$. Let P be the transitions of the resulting polymer Markov chain and let Q be the transitions, resulting from sampling from ξ_i instead of ν_i whenever needed for a transition. If for all polymer cliques Λ_i of the polymer clique cover $d_{\mathrm{TV}}(\xi_i;\nu_i) \leq \varepsilon_{\mathrm{in}}$, then the transition error between P and Q is bounded by $\frac{\varepsilon_{\mathrm{in}}}{2}$.

Proof of Lemma 4.4. Let $\Gamma \in \mathcal{F}$ be any polymer family. By definition, the total variation distance between $Q(\Gamma;\ldots)$ and $P(\Gamma;\ldots)$ can be written as

$$d_{\mathrm{TV}}(Q(\Gamma;\ldots);P(\Gamma;\ldots)) = \frac{1}{2}\sum_{\Gamma'\in\mathcal{F}} |Q(\Gamma;\Gamma') - P(\Gamma;\Gamma')|.$$

Replacing ν_i by ξ_i does only influence transitions that add a polymer $\gamma \in \Lambda_i$ to the current polymer family Γ. Moreover, it does not add any new transitions between polymer families. Thus, if $P(\Gamma;\Gamma') = 0$ it also holds that $Q(\Gamma;\Gamma') = 0$. For simplicity of notation, let $\mathcal{C}_{;} = \mathcal{C}\setminus\{\gamma\}$ and $\Lambda_{i;;} = \Lambda_i\setminus\{\gamma\}$. This leads to

$$\frac{1}{2}\sum_{\Gamma'} |Q(\Gamma;\Gamma') - P(\Gamma;\Gamma')| \quad \frac{1}{2}\sum_{\gamma\in\mathcal{C}_{;}} \left(\prod_{\Lambda_{i;;}\ni\gamma} \frac{\xi_i(\gamma)}{2m}\right) \left(\prod_{\Lambda_{i;;}\not\ni\gamma} \frac{\nu_i(\gamma)}{2m}\right)$$

$$= \frac{1}{2m} \sum_{\gamma \in C_i} \frac{1}{2} \left(\sum_{\Lambda_i : \gamma} \xi_i(\gamma) \right)^2 \left(\sum_{\Lambda_i : \gamma} \nu_i(\gamma) \right)^2$$
$$\frac{1}{2m} \sum_{\gamma \in C_i} \frac{1}{2} \sum_{\Lambda_i : \gamma} |\xi_i(\gamma) \nu_i(\gamma)|.$$

The latter sum is over all Λ_i that contain γ and i is simply treated like a normal polymer. Now, note that iterating over all $\gamma \in C_i$ and over all Λ_i that contain γ is the same as iterating over all Λ_i and then over all $\gamma \in \Lambda_i$. Because all sums here are finite, we can use this observation to exchange them and obtain

$$\frac{1}{2m} \sum_{\gamma \in C_i} \frac{1}{2} \sum_{\Lambda_i : \gamma} |\xi_i(\gamma) \nu_i(\gamma)| = \frac{1}{2m} \sum_{1 \le i \le m} \frac{1}{2} \sum_{\gamma \in \Lambda_i} |\xi_i(\gamma) \nu_i(\gamma)|$$
$$= \frac{1}{2m} \sum_{1 \le i \le m} d_{\mathrm{TV}}(\xi_i, \nu_i).$$

Finally, we know that $\sum_{1 \le i \le m} d_{\mathrm{TV}}(\xi_i, \nu_i) \le m\varepsilon_{\mathrm{in}}$ and thus

$$d_{\mathrm{TV}}(Q(\Gamma), P(\Gamma)) \le \frac{\varepsilon_{\mathrm{in}}}{2},$$

which concludes our proof.

With this bound, we can prove our main theorem of this chapter.

Theorem 4.5. Let $(C, w, \dots)$ be a polymer model encoded with input size n and let $\Lambda_1, \dots, \Lambda_m$ be a given polymer clique cover.
Given that:

(1) $m \in \mathrm{poly}(n)$ and we can draw a polymer clique Λ_i uniformly at random in time $\mathrm{poly}(m)$

(2) for every polymer clique Λ_i and every $\gamma \in C$ we can check whether $\gamma \in \Lambda_i$ in time $\mathrm{poly}(n)$

(3) for every γ, γ' we can check whether $\gamma \not\sim \gamma'$ in time $\mathrm{poly}(n)$

(4) $(C, w, \dots)$ fulfills the generalized polymer mixing condition as stated in Definition 3.3, and $e^{-\mathrm{poly}(n)} \le f(\gamma) \le e^{\mathrm{poly}(n)}$ for every $\gamma \in C$

(5) we can ε_{in}-approximately sample from each clique polymer distribution ν_i in time poly $\frac{n}{\varepsilon_{\text{in}}}$

Then we can ε-approximately sample from the Gibbs distribution μ in time poly $\frac{n}{\varepsilon}$. J

Proof of Theorem 4.5. Most of the proof of this theorem is similar to the one of Theorem 3.13. Because of (1) and (4), we can use Lemma 3.6 and Lemma 3.12 to construct a Markov chain with transitions P that converges the stationary distribution μ with mixing time $t_P 1\varepsilon^0$ 2 poly $\frac{n}{\varepsilon}$. In addition, by (1), (2) and (3) we can do each step of the Markov chain, except for sampling from the clique polymer distribution ν_i. Instead, we take some t_P $\frac{\varepsilon}{2}$ t 2 poly $\frac{n}{\varepsilon}$ and approximately sample from ν_i with a total variation distance of at most $\frac{\varepsilon}{t}$. According to assumption (5), this can be done in time poly $\frac{nt}{\varepsilon}$, which is also polynomial in $\frac{n}{\varepsilon}$. The result is a Markov chain with transitions Q. By Lemma 4.4, we know that the transition error between P and Q is at most $\frac{\varepsilon}{2t}$. According Lemma 4.2, we can run this chain with transitions Q for t steps to sample ε-approximately from μ, which results in an overall runtime in poly $\frac{n}{\varepsilon}$.

The result stated in Theorem 4.5 can be seen as a theoretical framework for reducing approximate sampling from a polymer model to approximate sampling from a single polymer clique. This has the advantage that the combinatorial complexity imposed by the incompatibility relation is completely handled by the Markov chain. There might be different ways on how the approximate sampling from each clique polymer distribution can be done. In the next chapter, we will investigate one of those ways, which we call truncation.

5 Inner Sampling by Truncation

In this chapter we want to investigate truncation as a way to approximately sample from each clique polymer distribution. Informally speaking, the idea is to prove conditions under which only a small number of polymers have most of the probability mass of the clique polymer distribution. By assuming that we can draw from this small subset efficiently enough, this will give us an approximate sampler for each clique polymer distribution. A valid way to sample from this smaller subset could for example be to enumerate it and calculate each polymer's weight.

This procedure has some similarity with truncation as done in cluster expansion approaches, as those methods use the fact that the first terms of the cluster expansion only depend on small sets of polymers. However, our arguments will be much more direct, as we will truncate the polymer model, instead of using analytical properties of an infinite series. More precisely, we do not even have to argue about truncation of the entire polymer model but only about the truncation of single polymer cliques, ignoring polymer interactions that are caused by the incompatibility relations.

Usually it is useful to talk about truncation in terms of some notion of size of the polymers. We will stick to this way of thinking and define size as follows.

I **Definition 5.1 (Size of a polymer in a subset).** Given a polymer model $^1\mathsf{C}; w;\ ^\circ$ and some subset of polymers $\mathsf{C}^0 \quad \mathsf{C}$, a size-function on C^0 is a function $\mathsf{j}\,\mathsf{j}_{\mathsf{C}^0} : \mathsf{C}^0 \ ! \ \mathbb{R}_{>0}$. For a fixed size-function $\mathsf{j}\,\mathsf{j}_{\mathsf{C}^0}$ and some polymer $\gamma \ 2 \ \mathsf{C}^0$ we call $\mathsf{j}\gamma\mathsf{j}_{\mathsf{C}^0}$ the size of γ in C^0. J

In the following sections, we will assume that for a given polymer clique cover each polymer clique Λ_i is equipped with a fixed size function $\mathsf{j}\,\mathsf{j}_i$. In most applications, one size-function $\mathsf{j}\,\mathsf{j} : \mathsf{C} \ ! \ \mathbb{R}_{>0}$ will be defined on the entire set of polymers C. In this case, we can obtain a size-function for each polymer clique by setting $\mathsf{j}\,\mathsf{j}_i$ to be the restriction of $\mathsf{j}\,\mathsf{j}$ to the set Λ_i. However, the reason that we defined size-functions for subsets of polymers is that our results do not require the same size-function for each polymer clique. This means we could even assign different sizes to the same polymer in different cliques.

Given a notion of size, we can now define the truncation of a polymer clique.

Definition 5.2 (Truncation of polymer clique). Given a polymer model $(\mathcal{C}, w, \circ)$ with a polymer clique cover $\Lambda_1, \dots, \Lambda_m$ and let each polymer clique Λ_i be equipped with a size-function $|\cdot|_i$. For any $k \in \mathbb{R}$ we call $\Lambda_i^{\leq k} = \{\gamma \in \Lambda_i \mid |\gamma|_i \leq k\}$ the truncation of Λ_i to size k. In addition, we define $\Lambda_i^{>k} = \Lambda_i \setminus \Lambda_i^{\leq k}$.

With this general definitions of size and truncation for polymer cliques, we can now start to investigate sufficient conditions for using truncation for approximate sampling from clique polymer distributions.

5.1 Conditions for truncation

As said before, the main idea is to only consider a small subset of each polymer clique Λ_i, namely $\Lambda_i^{\leq k}$ for some $k \in \mathbb{R}$, and sample from this set, for example by enumerating all polymers $\gamma \in \Lambda_i^{\leq k}$.

In order to use this for each step of our polymer Markov chain, we have to express the resulting error in terms of a total variation distance to the actual clique polymer distribution ν_i. The first step for doing so is to formally define the probability distribution that results from such a truncation.

Definition 5.3 (Truncated clique polymer distribution). Given a polymer model $(\mathcal{C}, w, \circ)$ fulfilling the generalized polymer mixing condition as in Definition 3.3. In addition, let $\Lambda_1, \dots, \Lambda_m$ be a polymer clique cover with a size-function $|\cdot|_i$ for each polymer clique. For a polymer clique Λ_i we define the clique polymer distribution ξ_i^k under truncation to $k \in \mathbb{R}$ as

$$\xi_i^k(\gamma) = \begin{cases} w_\gamma & \text{if } \gamma \in \Lambda_i^{\leq k}; \\ 0 & \text{o.w.} \end{cases}$$
$$\xi_i^k(\circ) = 1 \Big/ \sum_{\gamma \in \Lambda_i^{\leq k}} w_\gamma.$$

Note that this is a valid probability distribution over $\Lambda_i \cup \{\circ\}$ because $w_\gamma > 0$, and for all $k \in \mathbb{R}$ we have $\sum_{\gamma \in \Lambda_i^{\leq k}} w_\gamma \leq \sum_{\gamma \in \Lambda_i} w_\gamma$ and by Lemma 3.5 we know that $\sum_{\gamma \in \Lambda_i} w_\gamma = \sum_{\gamma \in \Lambda_i} \nu_i(\gamma) \leq 1$.

Our goal is now to truncate Λ_i to some k such that $d_{\mathrm{TV}}(\xi_i^k, \nu_i) \leq \varepsilon_{\mathrm{in}}$. The following lemma will help us by giving a simple connection between a truncation size k and the resulting total variation distance $d_{\mathrm{TV}}(\xi_i^k, \nu_i)$.

Lemma 5.4 (Total variation distance of truncation). Given a polymer model (C, w, Ω) fulfilling the generalized polymer mixing condition and a polymer clique cover $\Lambda_1, \ldots, \Lambda_m$, each equipped with a size-function $|\cdot|_i$. For a polymer clique Λ_i the total variation distance between the clique polymer distribution ν_i and the distribution ξ_i^k under truncation to $k \in \mathbb{R}$ is

$$d_{\mathrm{TV}}(\xi_i^k, \nu_i) = \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma. \qquad \square$$

Proof of Lemma 5.4. Let $\Lambda_{i;\gamma} = \Lambda_i \setminus \{\gamma\}$. We write out the definition of the total variation distance and split up the sum as

$$\begin{aligned} d_{\mathrm{TV}}(\xi_i^k, \nu_i) &= \frac{1}{2} \sum_{\gamma' \in \Lambda_{i;\gamma}} |\xi_i^k(\gamma') - \nu_i(\gamma')| \\ &= \frac{1}{2} \Bigg(|\xi_i^k(\gamma) - \nu_i(\gamma)| + \sum_{\gamma' \in \Lambda_i^{\leq k}} |\xi_i^k(\gamma') - \nu_i(\gamma')| + \sum_{\gamma' \in \Lambda_i^{>k}} |\xi_i^k(\gamma') - \nu_i(\gamma')| \Bigg). \end{aligned}$$

We now calculate each term separately.

First, note that for all $\gamma' \in \Lambda_i^{\leq k}$ we have $\xi_i^k(\gamma') = w_{\gamma'} = \nu_i(\gamma')$. Thus we know that

$$\sum_{\gamma' \in \Lambda_i^{\leq k}} |\xi_i^k(\gamma') - \nu_i(\gamma')| = 0.$$

In addition, for every $\gamma' \in \Lambda_i^{>k}$ we have $\xi_i^k(\gamma') = 0$ and $\nu_i(\gamma') = w_{\gamma'}$. This gives us

$$\sum_{\gamma' \in \Lambda_i^{>k}} |\xi_i^k(\gamma') - \nu_i(\gamma')| = \sum_{\gamma' \in \Lambda_i^{>k}} w_{\gamma'} = \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma.$$

Finally, we get for the remaining part

$$\begin{aligned}\xi_i^{k}{}^{1};^{o} \quad \nu_i{}^{1};^{o} &= \Big(\sum_{\gamma \in \Lambda_i{}^{k}} w_\gamma\Big)^{-1}\Big(\sum_{\gamma \in \Lambda_i} w_\gamma\Big)^{-1} \\ &= \sum_{\gamma \in \Lambda_i} w_\gamma \sum_{\gamma \in \Lambda_i{}^{k}} w_\gamma \\ &= \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma :\end{aligned}$$

This gives us

$$\begin{aligned}d_{\mathrm{TV}}{}^{1}\xi_i^{k};\nu_i{}^{o} &= \frac{1}{2}\Big(\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma + 0 + \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma\Big) \\ &= \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma ;\end{aligned}$$

which proves the lemma.

We can now use this connection to bound the total variation distance between ν_i and ξ_i^k for any $k \in \mathbb{R}$. We proceed by introducing the clique truncation condition as a simple way to relate the size and the weight of polymers.

Definition 5.5 (Clique truncation condition). Let Λ_i be a polymer clique of a polymer model $^{1}\mathsf{C};w;\ {}^{o}$. In addition let n be the input size which encodes the polymer model and $|\cdot|_i$ is a size-function on Λ_i. We say that this polymer clique fulfills the clique truncation condition for some monotonic increasing, invertible function $\partial : \mathbb{R} \to \mathbb{R}_{>0}$ if

$$\sum_{\gamma \in \Lambda_i} \partial^{1}|\gamma|_i{}^{o} w_\gamma \leq h^{1}n^{o}$$

for some positive function $h^{1}n^{o} \in \mathrm{poly}^{1}n^{o}$.

Now we can state which truncation bounds we can obtain under this clique truncation condition.

I **Lemma 5.6 (Clique truncation).** Let Λ_i be a polymer clique of a polymer model $(\mathsf{C}, w, \circ)$ that fulfills the generalized polymer mixing condition. In addition let n be the input size which encodes the polymer model and $|\cdot|_i$ is a size-function on Λ_i. If Λ_i fulfills the clique truncation condition for two functions δ, h as in Definition 5.5, then for every $\varepsilon_{\text{in}} \in (0, 1]$ we have $d_{\text{TV}}(\xi_i^k, \nu_i^\circ) \leq \varepsilon_{\text{in}}$ for any $k \geq \delta^{-1}\left(\frac{h(n)}{\varepsilon_{\text{in}}}\right)$. J

Proof of Lemma 5.6. By Lemma 5.4 we know that $d_{\text{TV}}(\xi_i^k, \nu_i^\circ) = \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma$. Because of the clique truncation condition and the monotonicity of δ, we observe that

$$\delta(k) \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \sum_{\gamma \in \Lambda_i^{>k}} \delta(|\gamma|_i) w_\gamma \leq \sum_{\gamma \in \Lambda_i} \delta(|\gamma|_i) w_\gamma \leq h(n).$$

As δ is positive, we can divide by $\delta(k)$ to get

$$\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \frac{h(n)}{\delta(k)}.$$

Now we set $k \geq \delta^{-1}\left(\frac{h(n)}{\varepsilon_{\text{in}}}\right)$, which is possible as $\frac{h(n)}{\varepsilon_{\text{in}}} > 0$ and δ is invertible. We obtain

$$\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \frac{h(n)}{\delta(\delta^{-1}(\frac{h(n)}{\varepsilon_{\text{in}}}))} = \varepsilon_{\text{in}},$$

which concludes our proof.

The key idea behind Lemma 5.6 will be the following. If the time for drawing a polymer according to ξ_i^k (e.g. by enumerating Λ_i^k completely) can be bounded by some function $t_i(k)$, we only need to ensure $t_i(\delta^{-1}(x)) \leq \text{poly}(x)$ to get an ε_{in}-approximate sampler in time $\text{poly}\left(\frac{n}{\varepsilon_{\text{in}}}\right)$.

5.2 Algorithmic application of truncation

We can now start to use our observations to construct an algorithm for sampling from the Gibbs distribution using truncation. Before doing so, we want to add a comment on the clique truncation condition from Definition 5.5. Although it first seems like the algorithmic use of this bound would require to choose two functions $\partial; h$, in general it will suffice to select a function ∂ and show the existence of any polynomial upper bound h on this sum.

We now state our main theorem of this chapter in its most general form.

Theorem 5.7. Let $^{1}C; w;$ ^{o} be a polymer model encoded with input size n. In addition, let $\Lambda_1; :::; \Lambda_m$ be a given polymer clique cover with a fixed size-function $j\ j_i$ for each polymer clique.
Given that:

(1) $m\ 2\ \text{poly}^{1}n^{o}$ and we can draw a polymer clique Λ_i uniformly at random in time $\text{poly}^{1}m^{o}$

(2) for every polymer clique Λ_i and every $\gamma\ 2\ C$ we can check whether $\gamma\ 2\ \Lambda_i$ in time $\text{poly}^{1}n^{o}$

(3) for every $\gamma; \gamma^{0}$ we can check whether $\gamma\quad\gamma^{0}$ in time $\text{poly}^{1}n^{o}$

(4) $^{1}C; w;$ ^{o} fulfills the generalized polymer mixing condition as stated in Definition 3.3, and $e^{\text{poly}^{1}n^{o}}\quad f^{1}\gamma^{o}\quad e^{\text{poly}^{1}n^{o}}$ for every $\gamma\ 2\ C$

(5) we can sample exactly from each truncated clique polymer distribution ξ_i^k in time $t_i{}^{1}k^{o}$ (e.g. enumerate $\Lambda_i{}^{k}$ and calculate each weight)

(6) each polymer clique Λ_i fulfills the clique truncation condition for functions $\partial_i; h_i$ as in Definition 5.5 and $t_i{}^{1}\partial_i{}^{1}{}^{1}x^{oo}\ 2\ \text{poly}^{1}x^{o}$

Then we can ε-approximately sample from the Gibbs distribution μ in time poly $\frac{n}{\varepsilon}$.

Proof of Theorem 5.7. To prove this theorem, we simply show that its assumptions imply the assumptions from Theorem 4.5. First, note that assumptions (1) to (4) are exactly the same as (1) to (4) of Theorem 4.5. Thus, it is sufficient to show that (5) and (6) of Theorem 5.7 imply (5) of Theorem 4.5. Namely, we want to show that we can sample from each ν_i at least ε_{in}-approximately in time

poly $\frac{n}{\varepsilon_{\text{in}}}$. Because of (6) and Lemma 5.6, we know that $d_{\text{TV}}(\xi_i^k; \nu_i) \leq \varepsilon_{\text{in}}$ for any $k \geq \partial_i(1)\frac{h_i(n)}{\varepsilon_{\text{in}}}$. For our choice of k and because of (5), sampling from ξ_i^k takes time $t_i(\partial_i(1)\frac{h_i(n)}{\varepsilon_{\text{in}}})$. According to (6) we know that $t_i(\partial_i(1)\frac{h_i(n)}{\varepsilon_{\text{in}}}) \leq \text{poly}\frac{h_i(n)}{\varepsilon_{\text{in}}}$. Noting that by definition $h_i(n) \leq \text{poly}(n)$ concludes our proof.

Because Theorem 5.7 is very general, there are many interesting and more specific versions for different cases that can be derived. In many applications, enumerating all $\gamma \in \Lambda_i$ and calculating each weight w_γ might be the best known way to sample from ξ_i^k. Often, this will result in $t_i(k) \in e^{O(k)}$ and imply that ∂_i has to be an exponential function. An interesting corollary of Theorem 5.7 for such cases and under mild assumptions on the size-function is the following.

Corollary 5.8. Let $(C; w; \emptyset)$ be a polymer model encoded with input size n. In addition let $\Lambda_1; \dots; \Lambda_m$ be a given polymer clique cover and let the set of all polymers C be equipped with a size function $|\cdot|$.
Given that:

(1) $m \in \text{poly}(n)$ and we can draw a polymer clique Λ_i uniformly at random in time $\text{poly}(m)$

(2) for every polymer clique Λ_i and every $\gamma \in C$ we can check whether $\gamma \in \Lambda_i$ in time $\text{poly}(n)$

(3) for every $\gamma; \gamma'$ we can check whether $\gamma \sim \gamma'$ in time $\text{poly}(n)$

(4) we can sample exactly from each truncated clique polymer distribution ξ_i^k in time $e^{O(k)}$

(5) for every $\gamma \in C$ the size is bounded by $|\gamma| \leq \text{poly}(n)$, and for every polymer clique Λ_i there is a polymer $\gamma \in \Lambda_i$ with $|\gamma| \geq O(\log(n))$

(6) for every polymer $\gamma \in C$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|}$$

for some constant $a > 0$

Then we can ε-approximately sample from the Gibbs distribution μ in time poly $\frac{n}{\varepsilon}$. J

Proof of Corollary 5.8. We show that this corollary is a special case of Theorem 5.7. First, note that (1) to (3) in Corollary 5.8 are the same as (1) to (3) in Theorem 5.7. In addition (6) implies the generalized polymer mixing condition for $f(\gamma) = e^{a|\gamma|}$. Using the upper bound for the size in (5) and the fact that a size-function is positive by definition, we get $1 \quad e^{a|\gamma|} \quad e^{\mathrm{poly}(n)}$. Thus, (4) from Theorem 5.7 is fulfilled as well.

For every polymer clique Λ_i we can obtain a size-function $|\cdot|_i$ by restricting $|\cdot|$ to Λ_i. Note that (4) in Corollary 5.8 is essentially (5) of Theorem 5.7 with $t_i(k) \geq e^{O(k)}$. By observing that all polymers in Λ_i are by definition pairwise incompatible, we get for every $\gamma \in \Lambda_i$

$$\tilde{O}_{\gamma' \in \Lambda_i} e^{a|\gamma'|_i} w_{\gamma'} \quad \tilde{O}_{\gamma' \quad \gamma} e^{a|\gamma'|} w_{\gamma'} \quad e^{a|\gamma|}:$$

We now set $\partial_i(k) = e^{ak}$, which is monotonically increasing and invertible for $a > 0$. In addition, we can choose $\gamma_{\min} \in \Lambda_i$ such that for all $\gamma' \in \Lambda_i$ it holds that $|\gamma_{\min}| \quad |\gamma'|$, as we have only a finite set of polymers. Now, (5) gives us $|\gamma_{\min}| \geq O(\log(n))$ and thus $e^{a|\gamma_{\min}|} \geq \mathrm{poly}(n)$. This proofs that the clique truncation condition holds for every polymer clique Λ_i.

It remains to show that $t_i(\partial_i^{-1}(x)) \in \mathrm{poly}(x)$. By definition, we have $\partial_i^{-1}(x) = \frac{\log(x)}{a}$. Thus, for $t_i(k) \geq O(e^k)$, we get $t_i(\partial_i^{-1}(x)) \in \mathrm{poly}(x)$. This finally shows that also (6) of Theorem 5.7 is fulfilled. The corollary follows directly.

6 Comparison & Application

In this chapter, we are going to compare our results to those that can be found in existing literature. Most of this literature focuses on approximating the partition function in the first place. In order to make our results more comparable, we will first prove that we can efficiently ε-approximate the partition function Z under the same conditions that we need for sampling from μ.

6.1 Approximating the partition function

In this section, we will prove that the conditions from Theorem 3.13, Theorem 4.5 or Theorem 5.7 are also sufficient for a randomized approximation of the partition function. It is well known that polymer models fulfill the self-reducibility property, under which efficient randomized generation in some sense implies efficient randomized approximation (see for example Jerrum et al. [JVV86]). However, this self-reducibility in general refers to the deletion of single polymers. That means using this kind of self-reducibility might result in an exponential sequence of reductions if the number of polymers is exponential.

Whenever truncation is applied, one can argue that after truncation the number of polymers is polynomial in n and $\frac{1}{\varepsilon}$ and thus, self-reducibility on the basis of single polymers suffices to get an efficient approximation algorithm. However, in the setting of Theorem 3.13 and especially Theorem 4.5 this argument does not apply, as these theorems do not involve truncation.

In this section, we will show that under the generalized polymer mixing condition it is sufficient to have a polymer clique cover of polynomial size in order to derive a randomized approximation within a polynomial number of self-reduction steps. Both conditions are crucial for our sampling theorems anyway.

For the rest of this section, we will assume that we have a polymer model $^1\mathsf{C}; w; \ ^\circ$ with a polymer clique cover $\Lambda_1; \dots; \Lambda_m$. We define a sequence of polymer models $^1\mathsf{C}_j; w; \ ^\circ$ for $0 \quad j \quad m$ with $\mathsf{C}_0 = \ ;$ and $\mathsf{C}_j = \mathsf{C}_{j\ 1} [\Lambda_j$ for $1 \quad j \quad m$. We denote the corresponding sets of polymer families with F_j,

partition functions with Z_j and Gibbs distributions with μ_j. Note that we have $Z_0 = 1$ and $Z_m = Z$. Moreover, we define $\sigma_j = \frac{Z_j}{Z_{j-1}}$ for $1 \leq j \leq m$. Now it holds that

$$Z = \prod_{1 \leq j \leq m} \frac{Z_j}{Z_{j-1}} = \left(\prod_{1 \leq j \leq m} \sigma_j \right)^{-1}.$$

We will now try to approximate each σ_j. For this, let $X_j(\Gamma)$ be a random variable with the values

$$X_j(\Gamma) = \begin{cases} 1 & \text{if } \Gamma \in \mathsf{F}_{j-1}; \\ 0 & \text{o.w.} \end{cases}$$

for Γ drawn according to μ_j. Note that if we can efficiently check if $\gamma \in \Lambda_i$ for any polymer clique Λ_i and any $\gamma \in \mathsf{C}$, then we can also calculate $X_j(\Gamma)$ efficiently by doing at most $m(j-1)^2 \leq m^2$ such checks. In addition, it holds that

$$\mathrm{E}[X_j(\Gamma)] = \sum_{\Gamma \in \mathsf{F}_j} \mu_j(\Gamma) X_j(\Gamma) = \sum_{\Gamma \in \mathsf{F}_{j-1}} \mu_j(\Gamma) = \sum_{\Gamma \in \mathsf{F}_{j-1}} \frac{\prod_{\gamma \in \Gamma} w_\gamma}{Z_j} = \frac{Z_{j-1}}{Z_j} = \sigma_j.$$

Because $Z_j \leq Z_{j-1}$ we know that $\sigma_j \leq 1$. Moreover, note that every polymer family $\Gamma \in \mathsf{F}_j \setminus \mathsf{F}_{j-1}$ contains exactly one polymer $\gamma \in \mathsf{C}_j \setminus \mathsf{C}_{j-1} \subseteq \Lambda_j$. If it would contain no such polymer, the family would have been in F_{j-1} and if it contains more than one, it would not be a polymer family as all polymers in Λ_j are pairwise incompatible. Thus, we have

$$\begin{aligned} Z_j &= Z_{j-1} + \sum_{\Gamma \in \mathsf{F}_j \setminus \mathsf{F}_{j-1}} \prod_{\gamma \in \Gamma} w_\gamma \\ &= Z_{j-1} + \sum_{\gamma \in \mathsf{C}_j \setminus \mathsf{C}_{j-1}} \sum_{\substack{\Gamma \in \mathsf{F}_j \\ \text{s.t. } \gamma \in \Gamma}} \prod_{\gamma' \in \Gamma} w_{\gamma'} \\ &= Z_{j-1} + \sum_{\gamma \in \mathsf{C}_j \setminus \mathsf{C}_{j-1}} w_\gamma \left(\sum_{\substack{\Gamma \in \mathsf{F}_j \\ \text{s.t. } \gamma \in \Gamma}} \prod_{\gamma' \in \Gamma \setminus \{\gamma\}} w_{\gamma'} \right) \end{aligned}$$

$$
\begin{aligned}
Z_{j-1} &+ \sum_{\gamma \in C_j \cap C_{j-1}} w_\gamma \cdot \sum_{\Gamma \in F_{j-1}} \sum_{\gamma' \in \Gamma} w_{\gamma'} \\
&= Z_{j-1} + Z_{j-1} \sum_{\gamma \in C_j \cap C_{j-1}} w_\gamma \\
&\leq Z_{j-1} + Z_{j-1} \sum_{\gamma \in \Lambda_j} w_\gamma \\
&\leq 2Z_{j-1}.
\end{aligned}
$$

The last inequality comes from the fact that $\sum_{\gamma \in \Lambda_j} w_\gamma \leq 1$, as we know from Lemma 3.5. Together, this gives us $\frac{1}{2} \leq \sigma_j \leq 1$.

We go on with approximating each σ_j. For this, we need to approximately sample from each μ_j. The following lemma will help us doing so, given we can approximately sample from μ.

Lemma 6.1. Give a polymer model $(C, w, \circ)$ with encoding size n and a polymer clique cover $\Lambda_1, \dots, \Lambda_m$. In addition, define the sequence of polymer models $(C_j, w, \circ)$ for $0 \leq j \leq m$ as above. Given that:

(1) $(C, w, \circ)$ fulfills the conditions of Theorem 3.13, Theorem 4.5 or Theorem 5.7

(2) for every $1 \leq j \leq m$ we can draw a polymer clique from $\Lambda_1, \dots, \Lambda_j$ uniformly at random in at most the time that is needed to draw one uniformly from $\Lambda_1, \dots, \Lambda_m$

Then the time for sampling ε-approximately from μ_j for every $\varepsilon \in (0, 1)$ and $1 \leq j \leq m$ is upper bounded by the time that is needed to ε-approximately sample from μ.

Proof of Lemma 6.1. First, note that it holds that $C_j \subseteq C_m = C$. In addition, all other conditions of Theorem 3.13, Theorem 4.5 and Theorem 5.7 imply that the same conditions are also fulfilled when only considering a subset of the original polymer cliques. Thus, we can now apply the same sampling schema to approximately sample from μ_j that we apply to sample from μ, except that the Markov chain only draws a polymer clique from $\Lambda_1, \dots, \Lambda_j$ uniformly at random. Note that $(C_j, w, \circ)$ has a polymer clique cover of size j. Bys Lemma 3.12, we know that the mixing time of the polymer Markov chain is monotonically

increasing in the number of polymer cliques in the polymer clique cover. Thus, we know that for every $j \quad m$ the mixing time for sampling from μ_j is upper bounded by the mixing time for sampling from $\mu_m = \mu$. Arguing that the time for every other step of the sampling algorithm for $^1C; w; \ ^o$ is upper bounded by the time needed on $^1C; w; \ ^o$ concludes our proof.

Now that we know that under mild assumptions a sampler for μ also gives us a sampler for μ_j with at most the same runtime, we can us this to construct an FPRAS for Z. The ides will be to approximate each $\sigma_j = \frac{Z_{j\ 1}}{Z_j}$ by some $\hat{\sigma_j}$ and use them to get an approximation of Z as shown in Algorithm 1.

Algorithm 1: FPRAS for the partition function Z

Data: error bound ε, polymer model $^1C; w; \ ^o$, polymer clique cover of size m

Result: ε-approximation of Z with probability at least $\frac{3}{4}$

$\varepsilon_m = \frac{\varepsilon}{10m}$;
$l = \frac{200m}{\varepsilon^2} + 1$;
for $1 \quad j \quad m$ **do**
  for $1 \quad i \quad l$ **do**
    ε_m-approximately sample $\Gamma_j^{1i^o}$ from μ_j;
  $\hat{\sigma_j} = \frac{1}{l} \sum_{1 \ i \ l} X_j(\Gamma_j^{1i^o})$;
$\hat{\sigma} = \prod_{1 \ j \ m} \hat{\sigma_j}$;
return $\frac{1}{\hat{\sigma}}$;

I **Theorem 6.2.** Given a polymer model $^1C; w; \ ^o$ with encoding size n and a polymer clique cover $\Lambda_1; :::; \Lambda_m$. Given that:

(1) $^1C; w; \ ^o$ fulfills the conditions of Theorem 3.13, Theorem 4.5 or Theorem 5.7

(2) for every $1 \quad j \quad m$ we can draw a polymer clique from $\Lambda_1; :::; \Lambda_j$ uniformly at random in at most the time that is needed to draw one uniformly from $\Lambda_1; :::; \Lambda_m$

Then Algorithm 1 is an FPRAS for the partition function Z. Moreover, the runtime can be bounded by the time needed for drawing $O\left(\frac{m^2}{\varepsilon^2}\right)$ samples from an $\frac{\varepsilon}{10m}$-approximation of μ. J

The proof uses similar arguments as other self-reducibility methods, such as for example Proposition 3.4 of Jerrum [Jer03].

Proof of Theorem 6.2. We start by showing that our algorithm computes the desired result, given that we can sample approximately from each μ_j. The prove this, we show two things. First, we argue that $\mathrm{E}[\hat{\sigma}]$ is a sufficiently good approximation for $\frac{1}{Z}$. Then we show that $\hat{\sigma}$ is close to $\mathrm{E}[\hat{\sigma}]$ with a sufficiently high probability. Finally, we conclude that $\frac{1}{\hat{\sigma}}$ ε-approximates Z.

For each $1 \quad j \quad m$, we have

$$\sigma_j \quad \frac{\varepsilon}{10} \quad \mathrm{E}\ \hat{\sigma}_j \quad \sigma_j + \frac{\varepsilon}{10}:$$

Because $\frac{1}{2} \quad \sigma_j \quad 1$, this implies that

$$1 \quad \frac{\varepsilon}{5}\ \sigma_j \quad \mathrm{E}\ \hat{\sigma}_j \quad 1 + \frac{\varepsilon}{5}\ \sigma_j:$$

In addition, we know that all $\hat{\sigma}_j$ are independent. Thus, we have $\mathrm{E}[\hat{\sigma}] = \prod_{1 \quad j \quad m} \mathrm{E}\ \hat{\sigma}_j$. By further using the fact that $\frac{1}{Z} = \prod_{1 \quad j \quad m} \sigma_j$, this gives us

$$1 \quad \frac{\varepsilon}{5m}^{m} \frac{1}{Z} \quad \mathrm{E}[\hat{\sigma}] \quad 1 + \frac{\varepsilon}{5m}^{m} \frac{1}{Z}:$$

We now have a bound on $\mathrm{E}[\hat{\sigma}]$ in terms of $\frac{1}{Z}$. We continue by deriving a bound on $\hat{\sigma}$ in terms of $\mathrm{E}[\hat{\sigma}]$. To do so, we apply Chebyshev's inequality to get

$$\Pr\left[\,|\hat{\sigma} - \mathrm{E}[\hat{\sigma}]| \quad \frac{\varepsilon}{5}\mathrm{E}[\hat{\sigma}]\right] \quad \frac{25}{\varepsilon^2}\frac{\mathrm{Var}[\hat{\sigma}]}{\mathrm{E}[\hat{\sigma}]^2} = \frac{25}{\varepsilon^2}\frac{\mathrm{E}\ \hat{\sigma}^2}{\mathrm{E}[\hat{\sigma}]^2} \quad 1 \Bigg)$$

Again, by independence of the $\hat{\sigma}_j$ we have $\mathrm{E}[\hat{\sigma}]^2 = \prod_{1 \quad j \quad m} \mathrm{E}\ \hat{\sigma}_j\ ^2$ and $\mathrm{E}\ \hat{\sigma}^2 = \prod_{1 \quad j \quad m} \mathrm{E}\ \hat{\sigma}_j^{\,2}$. Thus, we can rewrite our bound as

$$\Pr\left[\,|\hat{\sigma} - \mathrm{E}[\hat{\sigma}]| \quad \frac{\varepsilon}{5}\mathrm{E}[\hat{\sigma}]\right] \quad \frac{25}{\varepsilon^2}\frac{\mathrm{E}\ \hat{\sigma}^2}{\mathrm{E}[\hat{\sigma}]^2} \quad 1 \Bigg)$$

$$= \frac{25}{\varepsilon^2} \sum_{1 \le j \le m} \left(\frac{\mathrm{E}\left[\hat{\sigma}_j^2\right]}{\mathrm{E}\left[\hat{\sigma}_j\right]^2} - 1 \right)$$
$$= \frac{25}{\varepsilon^2} \sum_{1 \le j \le m} \left(1 + \frac{\mathrm{Var}\left[\hat{\sigma}_j\right]}{\mathrm{E}\left[\hat{\sigma}_j\right]^2} \right) - 1 \Bigg).$$

Next, we need to bound the variance of $\hat{\sigma}_j$ for $1 \le j \le m$. For this, remember that $\hat{\sigma}_j = \frac{1}{l} \sum_{1 \le i \le l} X_j(\Gamma^{(1,i)}, \ldots, \Gamma^{(l,i)})$ where $\Gamma^{(1,1)}, \ldots, \Gamma^{(l,l)}$ are independently drawn from an $\frac{\varepsilon}{10}$-approximation of μ_j. Thus, we have

$$\mathrm{Var}\left[\hat{\sigma}_j\right] = \frac{1}{l^2} \mathrm{Var}\left[\sum_{1 \le i \le l} X_j(\Gamma^{(i,1)}, \ldots, \Gamma^{(i,l)}) \right] = \frac{1}{l} \mathrm{Var}\left[X_j(\Gamma)\right]$$

for $\Gamma \in \mathcal{F}_j$ drawn from an approximation of μ_j. Now, note that $X_j(\Gamma)$ is Bernoulli distributed with $\Pr\left[X_j(\Gamma) = 1\right] = \mathrm{E}\left[\hat{\sigma}_j\right]$. This gives us

$$\mathrm{Var}\left[\hat{\sigma}_j\right] = \frac{1}{l} \mathrm{E}\left[\hat{\sigma}_j\right] \left(1 - \mathrm{E}\left[\hat{\sigma}_j\right]\right).$$

By using the fact that

$$\mathrm{E}\left[\hat{\sigma}_j\right] \ge \left(1 - \frac{\varepsilon}{5}\right) \sigma_j \ge \left(1 - \frac{1}{5}\right) \frac{1}{2} \ge \frac{1}{3}$$

we get the bound

$$\frac{\mathrm{Var}\left[\hat{\sigma}_j\right]}{\mathrm{E}\left[\hat{\sigma}_j\right]^2} \le \frac{1}{l \mathrm{E}\left[\hat{\sigma}_j\right]} \le \frac{1}{l} \le \frac{2}{l}.$$

Now that we have this bound on the variance of $\hat{\sigma}_j$, we obtain

$$\begin{aligned} \Pr\left[\left|\hat{\sigma} - \mathrm{E}[\hat{\sigma}]\right| \ge \frac{\varepsilon}{5} \mathrm{E}[\hat{\sigma}]\right] &\le \frac{25}{\varepsilon^2} \left(1 + \frac{2}{l}\right)^m - 1 \\ &\le \frac{25}{\varepsilon^2} e^{\frac{2m}{l}} - 1 \\ &\le \frac{25}{\varepsilon^2} \frac{2m}{l - 1}. \end{aligned}$$

The last inequality comes from the fact that $e^{\frac{x}{k+1}} \leq 1 + \frac{x}{k}$ for $0 \leq x \leq 1$ and $k \in \mathbb{N}_{>0}$. For $l \geq \frac{200m}{\varepsilon^2} + 1$ as in our algorithm, this implies that with probability at least $1 - \frac{25}{\varepsilon^2}\frac{\varepsilon^2}{100} = \frac{3}{4}$ that we have

$$\left(1 - \frac{\varepsilon}{5}\right)\mathbb{E}[\hat{\sigma}] \leq \hat{\sigma} \leq \left(1 + \frac{\varepsilon}{5}\right)\mathbb{E}[\hat{\sigma}].$$

We now have both bounds that we need. We rewrite them in a slightly weaker form by using the fact that $e^{\frac{x}{k}} \geq 1 + \frac{x}{k+1}$ for $0 \leq x \leq 1$ and $k \in \mathbb{N}_{>0}$. By this we get

$$e^{-\frac{\varepsilon}{4}}\mathbb{E}[\hat{\sigma}] \leq \hat{\sigma} \leq e^{\frac{\varepsilon}{4}}\mathbb{E}[\hat{\sigma}]$$
$$e^{-\frac{\varepsilon}{4}}\frac{1}{Z} \leq \mathbb{E}[\hat{\sigma}] \leq e^{\frac{\varepsilon}{4}}\frac{1}{Z}.$$

Note that this directly implies

$$e^{-\frac{\varepsilon}{2}}\frac{1}{Z} \leq \hat{\sigma} \leq e^{\frac{\varepsilon}{2}}\frac{1}{Z}.$$

Finally, for $\varepsilon \in [0, 1]$ this results in

$$(1 - \varepsilon)Z \leq e^{-\frac{\varepsilon}{2}}Z \leq \frac{1}{\hat{\sigma}} \leq e^{\frac{\varepsilon}{2}}Z \leq (1 + \varepsilon)Z$$

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

Now that we know that our algorithm produces the desired output, it remains to argue that the runtime is as stated in the theorem. Because the polymer model $(\mathcal{C}, w, \Gamma)$ fulfills the conditions of Theorem 3.13, Theorem 4.5 or Theorem 5.7, we know that we can ε_S-approximately sample from μ in time polynomial in n and $\frac{1}{\varepsilon_S}$. Because $m \in \text{poly}(n)$ this allows us to sample with an error of at most $\varepsilon_S = \frac{\varepsilon}{10m}$ in time polynomial in n and $\frac{1}{\varepsilon}$. By Lemma 6.1, this also upper bounds the runtime for approximately sampling from each μ_j for $1 \leq j \leq m$. In addition, we have to draw $(200m/\varepsilon^2 + 1)m \in O(m^2/\varepsilon^2)$ samples. This results in an overall runtime in poly (n/ε).

With this equivalence of approximately sampling from μ and approximating Z, we can start comparing our conditions for sampling with the existing literature.

6.2 Comparison to cluster expansion

As already discussed in the related work, using the cluster expansion is one of the most common ways for approximating the partition function. One of the main advantages of this technique is that it also works for complex-valued weights. A necessary (although not sufficient) condition for the application of such approaches is to show that the cluster expansion converges absolutely for a certain set of weights w. Many conditions to guarantee absolute convergence have been proposed so far. We will mainly refer to the conditions that are stated by Fernández and Procacci [FP07], as this is the most comprehensive study that we found. We will only state the real-valued version of the conditions, as only those are of interest for our approach. However, the complex-valued version can be obtained by taking the modulus of the weights in each condition.

Our discussion will focus on two convergence conditions. The first one is the Kotecký-Preiss condition, as it turned out to be most convenient to use for many applications of polymer models.

Definition 6.3 (Kotecký-Preiss condition, [KP86]). We say that a polymer model $(\mathcal{C}, w, \circ)$ fulfills the Kotecký-Preiss condition if there are two functions $\partial, f : \mathcal{C} \to \mathbb{R}_{\geq 0}$ such that for all $\gamma \in \mathcal{C}$

$$\sum_{\gamma' \not\sim \gamma} e^{f(\gamma')\|\gamma'\| + g(\gamma')\|\gamma'\|} w_{\gamma'} \leq f(\gamma). \qquad \square$$

Note that Fernández and Procacci [FP07] discussed the least restrictive (weakest) version of the Kotecký-Preiss condition, namely the case that $\partial(\gamma) = 0$.

The second condition we would like to discuss is the one that was introduced by Fernández and Procacci [FP07] themselves. We will refer to this condition as Fernández-Procacci condition. To state it, the definition of incompatible polymer families will be needed.

Definition 6.4 (Incompatible polymer families). For a polymer model $(\mathcal{C}, w, \circ)$ with polymer families $\mathcal{F}$, we define for every polymer $\gamma \in \mathcal{C}$ the set of incompatible polymer families as

$$\mathcal{F}_\gamma = \{\Gamma \in \mathcal{F} \mid \exists \gamma' \in \Gamma : \gamma' \not\sim \gamma\}. \qquad \square$$

Note that especially $\emptyset \in \mathcal{F}_\gamma$ for all $\gamma \in \mathcal{C}$. Based on the notion of incompatible polymer families, the Fernández-Procacci condition can be stated as follows.

I **Definition 6.5 (Fernández-Procacci condition, [FP07]).** We say that a polymer model $(C; w; \circ)$ fulfills the Fernández-Procacci condition if there is a function $f : C \to \mathbb{R}_0$ such that for all $\gamma \in C$

$$\prod_{\Gamma \in F_\gamma} \prod_{\gamma' \in \Gamma} f(\gamma') w_{\gamma'} \leq f(\gamma);$$

where we assign the value 1 to the empty product for $\Gamma \in F_\gamma$. J

The reader might notice that Fernández and Procacci [FP07] stated a slightly different version, namely

$$w_\gamma^{\alpha} \prod_{\Gamma \in F_\gamma} \prod_{\gamma' \in \Gamma} f(\gamma')^{\alpha} \leq f(\gamma):$$

The equivalence of our version of the condition can be seen by setting $f(\gamma) = f(\gamma) w_\gamma$ or $f(\gamma) = \frac{f(\gamma)}{w_\gamma}$, respectively. The reason for choosing the Fernández-Procacci condition is that it is to the best of our knowledge the least restrictive characterization for convergence of the cluster expansion of abstract polymer models. Moreover, it is directly implied by most other criteria, like the Dobrushin's condition (see [Dob96]) or the Kotecký-Preiss condition.

Mixing time of the polymer Markov chain

We will start with investigating how the above conditions relate to the mixing time of the polymer Markov chain. More precisely, the core of this investigation will be a comparison between the generalized polymer mixing condition and the conditions for convergence of the cluster expansion stated above.

I **Proposition 6.6.** Given a polymer model $(C; w; \circ)$ with polymer clique cover $\Lambda_1; \ldots; \Lambda_m$. If the polymer model fulfills the Fernández-Procacci condition for some function f, then it also fulfills the generalized polymer mixing condition for the same function and the polymer Markov chain is mixing in time

$$t_P(\varepsilon) \in O\left(m^2 \ln\left(m^2 \frac{\max_{\gamma \in C} f(\gamma)}{\min_{\gamma \in C} f(\gamma)}\right)^2 \ln\left(\frac{1}{\varepsilon}\right)\right):$$ J

Proof of Proposition 6.6. To prove this, we first derive a lower bound on the left-hand side of the Fernández-Procacci condition as stated in Definition 6.5. First, note that for all $\gamma \in \mathsf{C}$ and $\gamma' \not\sim \gamma$ it holds that $\{\gamma'\} \in \mathsf{F}_\gamma$. This especially includes the family $\{\gamma\}$ itself. Together with the fact that $\emptyset \in \mathsf{F}_\gamma$, this yields

$$1 + \sum_{\gamma' \not\sim \gamma} f(\gamma')|w_{\gamma'}| \leq \sum_{\Gamma \in \mathsf{F}_\gamma} \prod_{\gamma' \in \Gamma} f(\gamma')|w_{\gamma'}| \leq f(\gamma).$$

The second inequality comes directly from the Fernández-Procacci condition. By subtracting 1, we get

$$\sum_{\gamma' \not\sim \gamma} f(\gamma')|w_{\gamma'}| \leq 1 - \frac{1}{f(\gamma)} < f(\gamma) < f(\gamma),$$

which shows that the Fernández-Procacci condition for a function f implies the generalized polymer mixing condition for the same function. By application of Lemma 3.12 we get the desired mixing time and conclude our proof.

Note that the lower bound for the left-hand side of the Fernández-Procacci condition in the above proof is precisely the best case scenario for this condition. Namely, this is the case when $\not\sim$ is an equivalence relation, meaning that it is not only reflexive and symmetric but also transitive. In terms of the polymer graph interpretation, this means that the maximum connected subgraphs of the polymer graph are polymer cliques.

The result from Proposition 6.6 directly extends to the those conditions that imply the Fernández-Procacci condition, including Dobrushins condition and the Kotecký-Preiss condition. Further, assume our polymer model is encoded with input size n. If we have $m \in \mathrm{poly}(n)$ and if it holds that $e^{-\mathrm{poly}(n)} \leq f(\gamma) \leq e^{\mathrm{poly}(n)}$, Proposition 6.6 directly implies the existence of a Markov chain for sampling from μ with mixing time polynomial in n and $\ln \frac{1}{\varepsilon}$.

The generalized polymer mixing condition alone does not always give us an efficient approximate sampler or an FPRAS for the partition function. Additional assumptions as in Theorem 3.13, Theorem 4.5 or Theorem 5.7 are required. However, neither does convergence of the cluster expansion directly imply an efficient approximation algorithm. We will proceed by investigating these additional assumptions.

Existence of an efficient algorithm

We would like to discuss how the additional assumptions that are needed for our approach compare to those from cluster expansion algorithms. Before doing so, we would like to give a formal argument that both, the generalized polymer mixing condition and cluster expansion are presumably insufficient for the existence of an FPTAS or an FPRAS for the partition function of abstract polymer models. For this, we will use the decision problem Unambiguous Sat.

I **Definition 6.7 (Unambiguous Sat).** Given a Boolean formula Φ in conjunctive normal form with n variables, such that there is at most one assignment to those variables that satisfies Φ, decide whether there exists such a satisfying assignment. J

A well known result by Valiant and Vazirani [VV86] shows that the existence of any (randomized) polynomial time algorithm for deciding Unambiguous Sat would imply NP = RP. This even holds for randomized algorithms with a two-sided error, as this would imply NP BPP, which again would lead to NP = RP (see for example Papadimitriou [Pap03], problem 11.5.18).

I **Proposition 6.8.** Unless NP = RP, the Kotecký-Preiss condition as stated in Definition 6.3 is not a sufficient condition for the existence of an FPTAS or FPRAS for the partition function of a polymer model. J

Proof Proposition 6.8. Given an instance Φ of Unambiguous Sat with n variables. We treat every possible assignment to those variables as a polymer $\gamma \in C$ and define the incompatibility to be the complete graph of these polymers $= C^2$. Note that this does not require to actually enumerate all such assignments, as the set of polymers has not to be given explicitly. Next we define a weight for every assignment $\gamma \in C$ as

$$w_\gamma = \begin{cases} \frac{1}{8} & \text{if } \gamma \text{ satisfies } \Phi \\ \frac{1}{2^{n+4}} & \text{o.w.} \end{cases}$$

First, we prove that this polymer model satisfies the Kotecký-Preiss condition. We set $f(\gamma) = 1$ and $\partial(\gamma) = 0$ for all $\gamma \in C$. There are at most 2^n non-satisfying assignments and at most 1 satisfying assignment. We get for all $\gamma \in C$ that

$$\sum_{\gamma' \not\sim \gamma} e^{f(\gamma')} w_{\gamma'} = e \sum_{\gamma' \in C} w_{\gamma'} \quad e \; 2^n \frac{1}{2^{n+4}} + \frac{1}{8} \quad = e \frac{3}{16} < f(\gamma).$$

Now assume we can deterministically approximate the partition function Z by Z' such that $(1-\varepsilon)Z \le Z' \le (1+\varepsilon)Z$ for all $\varepsilon \in (0,1]$ in time polynomial in n and $\frac{1}{\varepsilon}$. We choose some constant $\varepsilon \le \frac{1}{128}$. Note that if Φ is satisfiable, we have $Z \ge 1+\frac{1}{8}$. Thus, we know that

$$Z' \ge (1-\varepsilon)Z \ge \left(1-\frac{1}{128}\right)\left(1+\frac{1}{8}\right) > 1+\frac{1}{8}-\frac{2}{128} = 1+\frac{7}{64}.$$

If Φ is not satisfiable, we have $Z \le 1+\frac{2^n}{2^{2+4}} = 1+\frac{1}{16}$. Thus, we get that

$$Z' \le (1+\varepsilon)Z \le \left(1+\frac{1}{128}\right)\left(1+\frac{1}{16}\right) < 1+\frac{1}{16}+\frac{2}{128} = 1+\frac{5}{64}.$$

By checking if $Z' \le \frac{5}{64}$ or $Z' \ge \frac{7}{64}$, we could decide if Φ is satisfiable in polynomial time, which would imply NP = RP.

Now let us assume we could get such an approximation of Z with probability of at least $\frac{3}{4}$. By the same arguments, we could decide if Φ is satisfiable in polynomial time with a two-sided failure probability of at most $\frac{1}{4}$. Again, this implies NP = RP.

Note that Proposition 6.8 not only shows that the Kotecký-Preiss condition alone is not sufficient, but it also extends to Fernández-Procacci condition and convergence of the cluster expansion in general, as they are implied by the Kotecký-Preiss condition.

For the same reason, Proposition 6.8 shows that the generalized polymer mixing condition is not sufficient for an efficient randomized approximation. Further, the polymer model from the proof has a polymer clique cover of constant size, and $\gamma \not\sim \gamma'$ and $\gamma \in \Lambda_i$ are trivial to decide. Thus, we see that even the assumptions (1)–(4) of Theorem 3.13, Theorem 4.5 and Theorem 5.7 are not sufficient for an approximation of the partition function or for approximate sampling from the Gibbs distribution. The intuitive reason is that those conditions only ensure that we can handle the combinatorial complexity that comes from sampling an independent set with respect to the incompatibility relation and the weights of each polymer. However, the sheer number of polymers causes additional difficulty, especially without any further structural knowledge about the polymer model. We encapsulate this in sampling from the polymer clique distribution ν_i for each clique Λ_i. From this point of view, Theorem 3.13, Theorem 4.5 and

Theorem 5.7 can be seen as different ways to solve this problem, depending on how hard it is to sample from each polymer clique. In general, the combination of truncation and enumeration as proposed in Theorem 5.7 seems to be the best known way to do this sampling from each clique distribution.

This truncation and enumeration approach is similar to what is done in cluster expansion algorithms. If each polymer weight can be expressed as a function of some parameter $z \in \mathbb{C}$ such that the partition function is a polynomial in z, Theorem 3.4 of Casel et al. [Cas+19] shows a general setting under which convergence of the cluster expansion can be used to obtain an efficient approximation. To the best of our knowledge, this is also the only approximation algorithm for the partition function that is proposed for abstract polymer models.

Unfortunately, a general comparison between Theorem 5.7 and cluster expansion is hard, because our approach is based on the interplay of the generalized polymer mixing condition and the clique truncation condition and less on analytical properties. However, some more insights can be gained when assuming that the functions in the Kotecký-Preiss condition and the Fernández-Procacci condition take a specific form. Fernández and Procacci [FP07] stated that the following size-dependent versions of both conditions are common.

I **Definition 6.9 (Size-dependent Kotecký-Preiss condition).** We say a polymer model $(\mathcal{C}, w, |\cdot|)$ with a size-function $|\cdot|$ fulfills the size-dependent Kotecký-Preiss condition if there is some $a \in \mathbb{R}_{>0}$ and a function $\partial : \mathcal{C} \to \mathbb{R}_{\geq 0}$ such that for all $\gamma \in \mathcal{C}$

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|+g(\gamma')} w_{\gamma'} \leq a|\gamma|. \qquad \text{J}$$

This is the Kotecký-Preiss condition from Definition 6.3 with $f(\gamma) = a|\gamma|$. Again, note that Fernández and Procacci [FP07] discuss the version where $\partial(\gamma) = 0$ for all $\gamma \in \mathcal{C}$, which is obviously optimal if ∂ can be chosen freely.

Similarly, the Fernández-Procacci condition has a size-dependent version.

I **Definition 6.10 (Size-dependent Fernández-Procacci condition).** We say a polymer model $(\mathcal{C}, w, |\cdot|)$ with a size-function $|\cdot|$ fulfills the size-dependent Fernández-Procacci condition if there is some $a \in \mathbb{R}_{>0}$ such that for all $\gamma \in \mathcal{C}$

$$\sum_{\Gamma \in \mathcal{F}_\gamma} \prod_{\gamma' \in \Gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|}. \qquad \text{J}$$

For these size-dependent conditions, there are some very interesting relations to our approach.

I **Proposition 6.11.** Let $^1C; w;\ ^o$ be a polymer model encoded with input size n. In addition let $\Lambda_1; :::; \Lambda_m$ be a given polymer clique cover and let the set of all polymers C be equipped with a size-function j j.
Given that:

(1) $m\ 2\ \text{poly}^1 n^o$ and for every $1\quad j\quad m$ we can draw a polymer clique Λ_i from $\Lambda_1; :::; \Lambda_j$ uniformly at random in time $\text{poly}^1 m^o$

(2) for every polymer clique Λ_i and every $\gamma\ 2\ C$ we can check whether $\gamma\ 2\ \Lambda_i$ in time $\text{poly}^1 n^o$

(3) for every $\gamma; \gamma^0$ we can check whether $\gamma\quad \gamma^0$ in time $\text{poly}^1 n^o$

(4) we can sample exactly from each truncated clique polymer distribution ξ_i^k in time $e^{O^1 k^o}$ (e.g., enumerate $\Lambda_i{}^k$ and calculate weights)

(5) for every $\gamma\ 2\ C$ the size is bounded by $j\gamma j\ 2\ \text{poly}^1 n^o$

Then the following holds.

1) If there is a polymer $\gamma\ 2\ \Lambda_i$ with $j\gamma j\ 2\ O^1 \log^1 n^{oo}$ for every polymer clique Λ_i, then size-dependent Fernández-Procacci condition is sufficient for the existence of an FPRAS for the partition function.

2) Otherwise, the size-dependent Kotecký-Preiss condition is sufficient for the existence of an FPRAS for the partition function. J

Proof of Proposition 6.11. We start with case 1). From Proposition 6.6 we know that the Fernández-Procacci condition implies the generalized polymer mixing condition for the same function f. Thus, we have

$$\tilde{O}_{\gamma^0\ \gamma} e^{a j \gamma^0 j} w_{\gamma^0}\quad e^{a j \gamma j}:$$

We can now apply Corollary 5.8 and Theorem 6.2, which proves this case of the proposition.

Next, we consider case 2). As the function ∂ from the size-dependent Kotecký-Preiss condition is non negative, it holds that

$$\sum_{\gamma^0 \not\sim \gamma} e^{a|\gamma^0|} w_{\gamma^0} \sum_{\gamma^0 \not\sim \gamma} e^{a|\gamma^0| + g(\gamma^0)} w_{\gamma^0} \leq a|\gamma| \leq e^{a|\gamma|}.$$

Thus, the polymer model fulfills the generalized polymer mixing condition for $f(\gamma) = e^{a|\gamma|}$. In addition, we have for every polymer clique Λ_i and $\gamma \supseteq \Lambda_i$ that

$$\sum_{\gamma^0 \supseteq \Lambda_i} e^{a|\gamma^0|} w_{\gamma^0} \sum_{\gamma^0 \not\sim \gamma} e^{a|\gamma^0|} w_{\gamma^0} \sum_{\gamma^0 \not\sim \gamma} e^{a|\gamma^0| + g(\gamma^0)} w_{\gamma^0} \leq a|\gamma|.$$

By assumption (5) this implies that every polymer clique Λ_i fulfills the clique truncation condition with $\partial_i(k) = e^{-ak}$. Moreover, note that $\partial_i^{-1}(x) = \frac{\ln(x)}{a}$, and by assumption (4) we have that $t_i(\partial_i^{-1})^{(\infty)}(x) \in \text{poly}(x)$. We conclude the proof by application of Theorem 5.7 and Theorem 6.2.

Proposition 6.11 shows that under mild assumptions the size-dependent Fernández-Procacci condition or the size-dependent Kotecký-Preiss condition are sufficient conditions for an FPRAS for the partition function, even if the partition function of the polymer model might not be a polynomial of a single parameter z. Such polymer models have recently been applied to many algorithmic tasks, for example by Jenssen et al. [JKP18], or by Cannon and Perkins [CP19]. However, in contrast to our approach, their proposed algorithms only work with a specific versions of the size-dependent Kotecký-Preiss condition and are specialized on graph polymer models based on vertex spin systems, as discussed in the next section.

6.3 Graph polymers from vertex spin systems

Graph polymer models arise from certain graph-theoretical counting problems. In vertex spin systems, these counting problems are based on spin assignments, representing different configurations of the graph. Formally, polymer models that are derived from vertex spin systems can be characterized in the following way.

Definition 6.12 (Polymer model from vertex spin systems). Given an undirected graphs $G = (V, E)$, a set of spins $[q] = \{0, \dots, q\}$ and a ground state $\partial_v \in [q]$ for every vertex $v \in V$. A vertex spin polymer model is a polymer model $(\mathcal{C}, w, \cdot)$ with the following properties.

1. Every polymer $\gamma \in \mathcal{C}$ is defined by some non-empty vertex set $\overline{\gamma} \subseteq V$ such that $\overline{\gamma}$ induces a connected subgraph of G, and a spin assignment $\varphi_\gamma : \overline{\gamma} \to [q]$ such that $\varphi_\gamma(v) \neq \partial_v$ for every $v \in \overline{\gamma}$.
2. Two polymers $\gamma, \gamma' \in \mathcal{C}$ are incompatible if their graph distance is less than 2 (i.e., $\gamma \not\sim \gamma'$ if there are $v \in \overline{\gamma}, u \in \overline{\gamma'}$ such that $v = u$, or there is an edge $e \in E$ which is incident to v and u). ◻

For simplicity, we might as well write γ instead of $\overline{\gamma}$ for the set of vertices if the spin assignment does not matter (e.g., $v \in \gamma$ instead of $v \in \overline{\gamma}$).

In order to apply any of our theorems, we need a polymer clique cover. A natural choice is the set $\{\Lambda_v \mid v \in V\}$ where $\Lambda_v = \{\gamma \in \mathcal{C} \mid v \in \gamma\}$. Obviously, for all Λ_v and $\gamma, \gamma' \in \Lambda_v$ it holds that $\gamma \not\sim \gamma'$, and there are $|V|$ such cliques, which is polynomial in the input size. In addition, drawing a polymer clique uniformly at random is the same as drawing a vertex $v \in V$ uniformly and can be done in polynomial time.

As, we will mainly use Theorem 5.7 and Corollary 5.8 in this section, we also need a size-function. A simple choice is to define this size on the entire set of polymers $\mathcal{C}$ as $|\gamma| = |\overline{\gamma}|$. It holds that $|\gamma| \leq |V|$ and thus, the size-function has a polynomial upper bound. An interesting property of this choice is that the size of a polymer $|\gamma|$ coincides with m_γ, the number cliques which contain γ.

We note that for our choice of size-function and polymer cliques, the polymer Markov chain is precisely the same as the one proposed in Chen et al. [Che+19]. However, our generalized polymer mixing condition can be seen as a much more flexible condition to bound the mixing time of that chain and our truncation condition allows a wider range of weights, for the prize of higher runtime.

For all our applications, the following notion of computational feasibility for vertex spin polymer models will be required.

Definition 6.13 (Computationally feasible vertex spin model). Given an undirected graph $G = (V, E)$ with $|V| = n$. We say a vertex spin polymer model $(\mathcal{C}, w, \cdot)$ on G with spins $[q] = \{0, \dots, q\}$ is computationally feasible if the following holds:

1. For every vertex set $V^0 \subseteq V$ with $|V^0| = k$ and every spin assignment $\varphi : V^0 \to [q]$ we can decide in time at most $O(e^k)$ whether V^0 and φ form a valid polymer in $\mathcal{C}$.

2. For every polymer $\gamma \in \mathcal{C}$ with $|\gamma| = k$ we can compute w_γ in time at most $O(e^k)$. ◻

Similar feasibility assumptions are commonly used for vertex spin systems, for example by Chen et al. [Che+19] or Jenssen et al. [JKP18].

To apply the clique truncation condition, we need to bound the time to enumerate polymers in a polymer clique up to some size k. For our choice of size-function and polymer clique cover, the following lemma about subgraph enumeration will be helpful.

Lemma 6.14 (Patel and Regts [PR17], Lemma 3.7). Given an undirected graph $G = (V, E)$ with bounded degree Δ and some $v \in V$. There is an algorithm that enumerates all connected, vertex-induced sugraphs of G that contain v and have at most $k \in \mathbb{N}_{>0}$ vertices in time $e^{O(k \ln(\Delta))}$. ◻

This implies for $q + 1$ spins that we can enumerate all candidates for polymers up to size k that contain v in time $e^{O(k \ln(\Delta q))}$ by enumerating all subgraphs that contain v up to size k and all possible spin assignments for the vertices of these subgraphs. If the model is computationally feasible, this ensures that we can also enumerate $\Lambda_v^{\,k}$ in time $e^{O(k \ln(\Delta q))}$.

We can now state under which conditions we can give an FPRAS for the partition function of vertex spin polymer models in a convenient to use form.

Theorem 6.15. Given an undirected graph $G = (V, E)$ with $|V| = n$ and constant degree bound Δ. Let $(\mathcal{C}, w, \circ)$ be a computationally feasible vertex spin polymer model with spins $[q]$ for some constant q. Assume for every polymer $\gamma \in \mathcal{C}$ it holds that

$$\sum_{\gamma^0 \nsim \gamma} e^{a|\gamma^0|} w_{\gamma^0} \leq e^{a|\gamma|}$$

for some constant $a \in \mathbb{R}_{>0}$. Further, assume one of the following conditions is true:

1) For every $v \in V$ either $\Lambda_v = \emptyset$ or there is a polymer $\gamma \in \Lambda_v$ such that $|\gamma| \geq \log(n)$.

2) For every $v \in V$ it holds that

$$\tilde{O}\left(\sum_{\gamma \in \Lambda_v} e^{b|\gamma|} w_\gamma\right) \leq h(n)$$

for some constant $b \in \mathbb{R}_{>0}$ and $h(n) \in \text{poly}(n)$.

Then there is an FPRAS for the partition function which has runtime at most $\left(\frac{n}{\varepsilon}\right)^{O(1)\ln(1)\Delta q^{\Theta}}$. ∎

Proof of Theorem 6.15. First note that our polymer clique cover $\{\Lambda_v \mid v \in V\}$ has at most size n. Let $v_1, \ldots, v_n$ be some enumeration of V. For any $1 \leq j \leq n$ we can draw a polymer clique from $\Lambda_{v_1}, \ldots, \Lambda_{v_j}$ uniformly at random by simply drawing a vertex from $v_1, \ldots, v_j$ uniformly, which can be done efficiently. By Theorem 6.2, this means that we can construct an FPRAS with the desired runtime if we can sample from μ in time $\left(\frac{n}{\varepsilon}\right)^{O(1)\ln(1)\Delta q^{\Theta}}$.

To show this, we will use Corollary 5.8 for case 1) and Theorem 5.7 for case 2). Moreover, the following observations will be useful for both cases.

- we can efficiently draw a polymer clique uniformly at random by drawing a vertex uniformly
- we can decide $\gamma \in \Lambda_v$ by checking $v \in \overline{\gamma}$ efficiently
- we can check $\gamma \subseteq \gamma'$ efficiently
- $|\gamma|$ is at most n

Furthermore, we can sample from the truncated clique distribution ξ_v^k by enumerating all connected subgraphs that contain v up to size k, enumerating all possible spin assignments for these subgraphs, checking which of them result in valid polymers and calculating weights of these polymers. By Lemma 6.14 and the assumption that the model is computationally feasible, this can be done in time $t_v(k) \leq e^{O(k \ln(1)\Delta q^{\Theta})}$.

We start by considering case 1). Because of our observations and the fact that every non-empty polymer clique contains a polymer of at most logarithmic size, we can apply Corollary 5.8 to get an efficient approximate sampling schema for μ. For the precise runtime, note that the generalized polymer mixing condition is fulfilled for $f(\gamma) = e^{a|\gamma|}$ and thus $1 \leq f(\gamma) \leq e^{an}$, which yields a mixing

time of the polymer Markov chain of $t_P(\varepsilon) = O(n^3 \ln(n) \ln \frac{1}{\varepsilon})$. Moreover, we know that the clique truncation conditions is fulfilled for $\delta(k) = e^{-ak}$ and $h(n) = e^{O(\log(n))} = \text{poly}(n)$. Thus, we have $\delta^{-1}(x) = \frac{\ln(x)}{a}$ and by Lemma 5.6, it is sufficient to truncate at size

$$k = \delta^{-1}\left(\frac{h(n)t_P(\frac{\varepsilon}{2})}{\varepsilon}\right) = O\left(\ln \frac{n}{\varepsilon}\right).$$

This means, approximately sampling from each clique polymer distribution can be done in time at most $(\frac{n}{\varepsilon})^{O(\ln(\Delta q))}$. As we need to do this $O(n^3 \ln(n) \ln \frac{1}{\varepsilon})$ times to sample ε-approximately from μ, this proves the desired runtime.

We continue with 2). Again, we know that the generalized polymer mixing condition is fulfilled for $f(\gamma) = e^{-a|\gamma|}$. Moreover, by assumption we have that the clique truncation condition is fulfilled for $\delta(k) = e^{-bk}$ and some polynomial function $h(n)$. Together with our previous observations, this already implies that we can apply Theorem 5.7. For a more precise runtime bound, note that the mixing time of the polymer chain is the same as for case 1). Moreover, we have $\delta^{-1}(x) = \frac{\ln(x)}{b}$, which means that we again truncate at size $k = O(\ln \frac{n}{\varepsilon})$. Thus, analogously to case 1) we can ε-approximately sample from μ in time $(\frac{n}{\varepsilon})^{O(\ln(\Delta q))}$.

Graph polymer models based on vertex spin systems have recently been applied to multiple algorithmic problems, such as the Potts model on α-expander graphs or the hard-core model on bipartite α-expander graphs by Jenssen et al. [JKP18], and the hard-core model on unbalanced bipartite graphs by Cannon and Perkins [CP19]. All these algorithmic applications use polymer models for which the set Λ_v is either empty or contains a polymer of at most constant size. This means, that we can apply case 1) of Theorem 6.15 to all of them. Thus, a sufficient condition for an FPRAS for these applications is that for every $\gamma \in \Gamma$ and some constant $a > \mathbb{R}_{>0}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|}.$$

This condition is even less restrictive than the size-dependent Fernández-Procacci condition, but as convenient to use as the size-dependent Kotecký-Preiss condition.

Moreover, our result also holds if the partition function is not a polynomial in a single parameter, which is especially important for both versions of the hard-core model. Although the algorithm for vertex spin polymer models proposed by Jenssen et al. [JKP18] also works without this assumption, their cluster expansion approach needs the size-dependent Kotecký-Preiss condition to be fulfilled for $a = 1$ and $\partial^1\gamma^\circ \quad \rho|\gamma|$ for some constant $\rho \in \mathbb{R}_{>0}$. Note that our condition is not only more flexible on the left-hand side but also exponentially larger on the right-hand side.

Besides this result, we can further show some general conditions to remove the heavy runtime dependency on Δ and q.

Theorem 6.16. Given an undirected graph $G = (V, E)$ with $|V| = n$ and constant degree bound Δ. Let $(\Omega, w, \circ)$ be a computationally feasible vertex spin polymer model with spins $\leq q$ for some constant q. Assume for every polymer $\gamma \in \Omega$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \quad e^{a|\gamma|}$$

for some constant $a \in \mathbb{R}_{>0}$. Further, assume for every $v \in V$ it holds that

$$\sum_{\gamma \in \Lambda_v} e^{b \ln(\Delta q)|\gamma|} w_\gamma \quad h(n)$$

for some $b \in \mathbb{R}_{>0}$ and $h(n) \in \mathrm{poly}(n)$. Then there is an FPRAS for the partition function with runtime at most $\left(\frac{n}{\varepsilon}\right)^{c_0 + \frac{c_1}{b}}$ for some absolute constants c_0, c_1, which are independent of Δ and q.

Proof of Theorem 6.16. Most of the proof of this theorem is similar to the proof of case 2) of Theorem 6.15. The key ingredient is to note that the $\ln(\Delta q)$ in the exponent of the runtime comes from the time that is needed sample from the truncated clique polymer distribution ξ_v^k for some given truncation size k. However, by changing the function that is used in the clique truncation from $\partial(|\gamma|) = e^{b|\gamma|}$ to $\partial(|\gamma|) = e^{b \ln(\Delta q)|\gamma|}$, we can truncate at a sufficiently small size.

More precisely, note that $\partial^{-1}(x) = \frac{\ln(x)}{b \ln(\Delta q)}$. As the mixing time of the polymer chain is the same as in the proof of Theorem 6.15, this results in a truncation

size of

$$k \geq \frac{\ln(n/\varepsilon)}{b \ln(\Delta q)}.$$

For this truncation size, we can sample from the truncated clique polymer distribution in time $e^{O(k \ln(\Delta q))} = e^{O(\ln(n/\varepsilon) \cdot b)}$. This can be bounded by $(n/\varepsilon)^{c_1/b}$ for some absolute constant c_1. Further, the runtime of every other step and the mixing time of the polymer Markov chain are bounded by $(n/\varepsilon)^{c_0}$ for some absolute constant c_0. Multiplying both yields the desired runtime.

Again, this theorem can be applied to all algorithmic problems mentioned above. In the following section, we will illustrate the application of Theorem 6.15 and Theorem 6.16 for the hard-core model on bipartite α-expander graphs.

Hard-core model on bipartite expander graphs

The hard-core model can be seen as a weighted generalization of counting independent sets.

Definition 6.17 (Hard-core model). Given an undirected graph G, let $\mathcal{I}$ be the set of all its independent sets. For a weight $\lambda \geq \mathbb{R}_{>0}$, called fugacity, the hard-core partition function is defined as

$$Z_G(\lambda) = \sum_{I \in \mathcal{I}} \lambda^{|I|}.$$

The corresponding hard-core Gibbs distribution is a probability distribution on $\mathcal{I}$ such that for every $I \in \mathcal{I}$

$$\mu_{G,\lambda}(I) = \frac{\lambda^{|I|}}{Z_G(\lambda)}. \tag{J}$$

The hard-core model is well studied for general bounded degree graphs, and especially along the positive real axis there are known algorithmic upper bounds on the fugacity λ which are known to be tight unless NP = RP (see Sly and Sun [SS12] and Weitz [Wei06]).

Obviously, the hard-core model can be seen as a vertex spin graph polymer model with spins $\{0, 1\}$, ground state $\partial_v = 0$ for all $v \in V$ and every single vertex

together with spin assignment 1 being a polymer of weight $w_\gamma = \lambda$. Usually, general polymer approaches perform worse than methods that are specialized on the hard-core model, due to the price of generality.

Aside general bounded-degree graphs, there are different polymer-based algorithms for some graph classes that extend the regime in which the partition function can be approximated. One such graph class are bipartite α-expander graphs.

I **Definition 6.18 (Bipartite α-expander graph).** Given a bipartite graph $G = (V, E)$ with partitions $V = V_L \dot\cup V_R$. In addition, for any set of vertices $S \subseteq V$ let $N_G(S)$ be the neighborhood of S including S itself. For any $\alpha \in (0, 1)$ we call G a bipartite α-expander graph if for every $i \in \{L, R\}$ and $S \subseteq V_i$ with $|S| \leq \frac{|V_i|}{2}$ it holds that $|N_G(S)| \geq (1 + \alpha)|S|$. J

The way that we will construct the polymer model is as proposed by Jenssen et al. [JKP18]. For a bipartite α-expander graph G with bonded degree Δ we consider the graph G^2, which is the graph with vertices V and an edge between $v, u \in V$ if v, u have at most distance 2 in G. We define the vertex spin polymer model $(\mathcal{C}^L, w^L, \circ)$ for graph G^2 and spins $\{0, 1\}$ by:

- $\partial_v = 0$ if $v \in V_L$ and $\partial_u = 1$ if $u \in V_R$
- every $\gamma \in \mathcal{C}^L$ is defined by a set of vertices $\overline{\gamma} \subseteq V_L$ with $|\overline{\gamma}| \geq \frac{|V_L|}{2}$ that induces a connected subgraph in G^2 and spin assignment $\varphi_\gamma(v) = 1$ for every $v \in \gamma$
- $w^L_\gamma = \frac{\lambda^{|\overline{\gamma}|}}{(1+\lambda)^{|N_G(\overline{\gamma})|}}$ where $N_G(\overline{\gamma})$ is the neighborhood of $\overline{\gamma}$ in G

Similarly we can define $(\mathcal{C}^R, w^R, \circ)$ by swapping the ground states and constructing the polymers from subset of vertices from V_R. Let for $i \in \{L, R\}$ denote μ_i and Z_i the Gibbs distribution and the partition function of $(\mathcal{C}^i, w^i, \circ)$. Based on these two polymer models, Jenssen et al. [JKP18] proved the following statement.

I **Lemma 6.19 (Jenssen et al. [JKP18], Lemma 19).** Given a bipartite α-expander graph $G = (V, E)$ with $|V| = n$ and vertex partitions $V = V_L \dot\cup V_R$. Assume that $\lambda \geq e^{\frac{11}{\alpha}}$ and set

$$\hat{Z} = (1 + \lambda^{|V_L|})Z_R + (1 + \lambda^{|V_R|})Z_L.$$

For the hard-core partition function $Z_G(\lambda)$ it holds that

$$(1 - e^{-n} Z_G(\lambda)) \hat{Z} (1 + e^{-n} Z_G(\lambda)) = 1. \quad \square$$

This means, that a sufficiently good approximation of Z_L and Z_R yields also a good approximation for the hard-core partition function on G.

As we aim for applying Theorem 6.15 in order to bound the weights for which our approach gives us an FPRAS, it will be useful to have a bound on the number of polymers that are in a given polymer clique. Such an upper bound can be derived from the number of vertex-induced subgraphs of G^2 that contain a certain vertex v multiplied with the number of possible spin assignments. For the number of vertex-induced subgraphs it is common to use the following lemma.

Lemma 6.20 (Borgs et al. [Bor+10], Lemma 2.1). For an undirected graph $G = (V, E)$ with bounded degree Δ and any $v \in V$, the number of vertex-induced sugraphs that contain v and have at most $k \in \mathbb{N}_{>0}$ vertices can be upper bounded by $\frac{(e\Delta)^k}{2}$. $\square$

To investigate for which regime of λ we can do such an approximation, we will use similar arguments as Jenssen et al. [JKP18]. However, instead of using the size-dependent Kotecký-Preiss condition to prove that cluster expansion can be applied, we will justify our bounds on λ with Theorem 6.15. In addition, we will get a randomized approximation, as our approximations are based on sampling from μ_L and μ_R. For handling the error probability, we will use similar arguments as Chen et al. [Che+19].

Proposition 6.21. Given a bipartite α-expander graph $G(V, E)$ with $|V| = n$ and degree bounded by some constant Δ. For $\lambda < \max\{(\frac{9}{2}\Delta^{2})^{\frac{1}{\alpha}}, e^{\frac{11}{\alpha}}\}$ there is an FPRAS for $Z_G(\lambda)$ with runtime in $(\frac{n}{\varepsilon})^{O(\ln(\Delta))}$. $\square$

Proof of Proposition 6.21. First, note that we actually only have to consider cases where $\varepsilon \in \omega(e^{-n})$ because for $\varepsilon \in O(e^{-n})$ we can simply brute-force all independent sets. There are at most 2^n such sets, which is obviously polynomial in $\frac{1}{e^{-n}}$. Thus, we focus on the case $\varepsilon > 4e^{-n}$.

Knowing this, we can apply Lemma 6.19, which tells us that $Z_G(\lambda)$ can be e^{-n} approximated by Z_L and Z_R. Let A_L and A_R be $\frac{\varepsilon}{4}$-approximations of Z_L and Z_R, each with probability of at least $\sqrt{\frac{3}{4}}$, and set $\hat{A}$ to be the resulting approximation

of $\hat{Z}$. Now we have with probability at least $1 \quad \varepsilon$ that

$$1 \quad \frac{\varepsilon}{4}\,\hat{Z} \quad \hat{A} \quad 1+\frac{\varepsilon}{4}\,\hat{Z}:$$

This directly gives us

$$^{1}1 \quad \varepsilon^{o}Z_{G}{}^{1}\lambda^{o} \quad 1 \quad \frac{\varepsilon}{4}\,{}^{1}1 \quad e^{\ n o}Z_{G}{}^{1}\lambda^{o} \quad 1 \quad \frac{\varepsilon}{4}\,\hat{Z} \quad \hat{A}$$

and similarly

$$\hat{A} \quad 1+\frac{\varepsilon}{4}\,\hat{Z} \quad 1+\frac{\varepsilon}{4}{}^{o}1 1+e^{\ n}\ Z_{G}{}^{1}\lambda^{o} \quad {}^{1}1+\varepsilon^{o}Z_{G}{}^{1}\lambda^{o}:$$

Thus, $\hat{A}$ is an ε-approximation of $Z_{G}{}^{1}\lambda^{o}$ with probability at least $\frac{3}{4}$.

Let $i \ 2 \ \mathrm{f}L;R\mathrm{g}$. We now know that it is sufficient if A_i is an $\frac{\varepsilon}{4}$-approximation of Z_i with probability at least $\frac{\sqrt{3}}{2}$. To do so, we take the median of $\mathrm{O} \ \ln \ \frac{2}{3} \quad = \mathrm{O}^{1}1^{o}$ independent trials with constant failure probability at most $\frac{1}{4}$.

To obtain each of those independent $\frac{\varepsilon}{4}$-approximations of Z_i for $i \ 2 \ \mathrm{f}L;R\mathrm{g}$ with failure probability at most $\frac{1}{4}$, we can apply Theorem 6.15.

First, note that the degree of G^2 is bounded by Δ^2 and that we only need two spins, thus $q = 1$. Further, the polymer model is computationally feasible as calculating weights and deciding whether a subgraph is a polymer can be done efficiently with respect to the number of vertices in the subgraph. Every polymer clique Λ_{v} is empty if $v < V_i$ or otherwise contains a polymer of constant size, namely $\overline{\gamma} = \mathrm{f}v\mathrm{g}$ with $\varphi_{\gamma}{}^{1}v^{o} = 1$. Thus, we can apply case 1) of Theorem 6.15. For this, it remains to show that for the stated range of λ the polymer model $^{1}\mathrm{C}_i; w_i; \ ^{o}$ fulfills

$$\tilde{\mathrm{O}}_{\gamma^0 \ \gamma} e^{a\mathrm{j}\gamma^0\mathrm{j}} w_{\gamma^0}^{i} \quad e^{a\mathrm{j}\gamma\mathrm{j}}$$

for some $a > 0$ and every $\gamma \ 2 \ \mathrm{C}^{i}$. We can upper bound the left-hand side by

$$\tilde{\mathrm{O}}_{\gamma^0 \ \gamma} e^{a\mathrm{j}\gamma^0\mathrm{j}} w_{\gamma^0}^{i} \qquad \tilde{\mathrm{O}}_{\mathrm{v}2N_{G^2}{}^{1}\overline{\gamma}^{o}} \ \tilde{\mathrm{O}}_{\gamma^0 2\Lambda_{\mathrm{v}}} e^{a\mathrm{j}\gamma^0\mathrm{j}} w_{\gamma^0}^{i}:$$

Now, observe that if G is a bipartite α-expander we can bound each polymer's

weight by

$$w^i_{\gamma^0} = \frac{\lambda^{|\gamma^0|}}{1 + \lambda^{|N_G(\gamma^0)|}} \quad \frac{1}{\lambda^{|\gamma|}}:$$

Moreover, for each $v \in V$ we can apply Lemma 6.20 to get

$$\sum_{\gamma^0 \ni \Lambda_v} e^{a|\gamma^0|} w^i_{\gamma^0} \quad \sum_{k \ge 1} \sum_{\substack{\gamma^0 \ni \Lambda_v \\ |\gamma^0| = k}} e^{a|\gamma^0|} w^i_{\gamma^0}$$
$$\sum_{k \ge 1} \frac{(e\Delta^2)^k}{2} \frac{e^{ak}}{\lambda^{\alpha k}}$$
$$\frac{1}{2e\Delta^2} \sum_{k \ge 1} \frac{e^{1+a}\Delta^2}{\lambda^\alpha}^k :$$

In addition, note that for every $\gamma \in C^i$ we have that $|N_G(\gamma)| \quad \Delta^2|\gamma| = \Delta^2|\gamma|$. This leads to the following upper bound

$$\sum_{\gamma^0 \ni \gamma} e^{a|\gamma^0|} w^i_{\gamma^0} \quad \frac{|\gamma|}{2e} \sum_{k \ge 1} \frac{e^{1+a}\Delta^2}{\lambda^\alpha}^k :$$

Now, we set $a = \frac{1}{4}$ and $\lambda \quad (\frac{9}{2}\Delta^2)^{\frac{1}{\alpha}}$. This gives us

$$\sum_{\gamma^0 \ni \Lambda_v} e^{a|\gamma^0|} w^i_{\gamma^0} \quad \frac{|\gamma|}{2e} \sum_{k \ge 1} \frac{2e^{1+\frac{1}{4}}}{9}^k = \frac{|\gamma|}{2e} \frac{2e^{1+\frac{1}{4}}}{9 - 2e^{1+\frac{1}{4}}} \quad e^{\frac{1}{4}|\gamma|}:$$

This proves that we can apply Theorem 6.15 and because $q = 1$ for this polymer mode, we get an FPRAS with the desired runtime.

We would like to add an additional remark on our bound. Note that the last inequality that was used to prove that we can apply Theorem 6.15, namely

$$\frac{|\gamma|}{2e} \sum_{k \ge 1} \frac{2e^{1+\frac{1}{4}}}{9}^k \quad e^{\frac{1}{4}|\gamma|}$$

is tightest for $|\gamma| = 4$. In general, for any $a > 0$ the worst case is $|\gamma| = \frac{1}{a}$. Knowing this, we can even derive a slightly better but more complex bound $\lambda \quad \lambda_{\Delta;\alpha} \quad {}^{1}4{:}4278 \quad \Delta^{2\circ\frac{1}{\alpha}}$. However, for simplicity of the proposition, we decide to state a slightly weaker bound.

By applying a modified version of the cluster expansion algorithm of Helmuth et al. [HPR18], together with a special version of the Kotecký-Preiss condition, Jenssen et al. [JKP18] proved a bound of $\lambda \quad \max\{{}^{1}2e^{3}\Delta^{4\circ\frac{1}{\alpha}}; e^{\frac{11}{\alpha}}\}$ for the same asymptotic runtime. However, their estimates of the subgraph counts were quite rough. By using Lemma 6.20 with their algorithm, a bound of $\lambda \quad e + \frac{1}{2}\, e^{2}\Delta^{2\ \frac{1}{\alpha}}$ can be derived. Still, our approach improves their result by a factor of more than $5^{\frac{1}{\alpha}}$.

The reason that Theorem 6.15 only performs better by a constant factor, is that the existence of polymers with constant size reduces the worst case of the right-hand site of the condition

$$\tilde{\mathrm{O}}_{\gamma\ \gamma^{0}} e^{a|\gamma^{0}|} w_{\gamma^{0}} \quad e^{a|\gamma|}$$

to a constant. Although the same also holds for the size-dependent Kotecký-Preiss condition, the theoretical advantage of our approach vanishes. However, note that our condition gets more powerful compared to the size-dependent Kotecký-Preiss condition the larger the smallest polymers in the model are. Especially, if for every polymer clique Λ_{v} it holds that the smallest polymer $\gamma \in \Lambda_{\mathrm{v}}$ fulfills $|\gamma| \in \Theta(\log(n))$, then the right-hand site of our condition in is in $\Theta(\mathrm{poly}(n))$, whereas the right-hand site of the size-dependent Kotecký-Preiss condition is in $\Theta(\log(n))$.

Besides this result, Chen et al. [Che+19] proposed a randomized approach for approximating the hard-core partition function for $\lambda \quad \max\{{}^{1}3\Delta^{\circ\frac{6}{\alpha}}; e^{\frac{11}{\alpha}}\}$ in time $\mathrm{O}\left({}^{1}\frac{n}{\varepsilon}{}^{\circ 2}\log\left({}^{1}\frac{n}{\varepsilon}{}^{\circ 3}\right)\right.$. A comparison is hard as their approach aims at an improved runtime but needs much more restrictive conditions for doing so. However, the following proposition shows that our approach is capable to give results in an interesting regime in between both trade-offs.

Proposition 6.22. Given a bipartite α-expander graph $G=(V,E)$ with $|V|=n$ and degree bounded by some Δ. If for some constant $d>0$ we have

$$\lambda \geq \max\left(3\Delta^{2+d}+\tfrac{1}{\alpha}, \left(\frac{9}{2}\Delta^2\right)^{\frac{1}{\alpha}}, e^{\frac{11}{\alpha}}\right),$$

then there is an FPRAS for $Z_G(\lambda)$ with runtime in $\left(\frac{n}{\varepsilon}\right)^{c_0+\frac{c_1}{d}}$ for some absolute constants c_0, c_1, independent of Δ.

Proof of Proposition 6.22. Most of the arguments are similar to the proof of Proposition 6.21, except that we apply Theorem 6.16 this time. First, note that $\lambda \geq e^{\frac{11}{\alpha}}$. Thus, we can again apply Lemma 6.19. Now, let $i \in \{L,R\}$. Because $\lambda \geq \frac{9}{2}\Delta^{2+\frac{1}{\alpha}}$, we know that for all polymers $\gamma \in \mathcal{C}^i$ it holds that

$$\sum_{\gamma^0 \subseteq \gamma} e^{a|\gamma^0|} w^i_{\gamma^0} \leq e^{a|\gamma|}$$

for $a=\frac{1}{4}$.

Note that in our case $q=1$, as we have only two spins and that the degree of G^2 is bounded by Δ^2. It remains to prove that for all $v \in V$ it holds that

$$\sum_{\gamma \ni \Lambda_v} e^{b\ln(\Delta^2 q^2)|\gamma|} \leq \sum_{\gamma \ni \Lambda_v} e^{2b\ln(\Delta^2)|\gamma|} \leq h(n^2)$$

for some $b \in \mathbb{R}_{>0}$ and $h(n^2) \in \mathrm{poly}(n^2)$. We do this by setting $b=\frac{d}{2}>0$. Because $\lambda \geq 3\Delta^{2+d}+\frac{1}{\alpha}$, it now holds for every $v \in V$ and $i \in \{L,R\}$ that

$$\begin{aligned}
\sum_{\gamma \ni \Lambda_v} e^{d\ln(\Delta^2)|\gamma|} w^i_\gamma &\leq \sum_{k\geq 1} \frac{(e\Delta^{2+k}+1)}{2} e^{d\ln(\Delta^2)k} \frac{1}{\lambda^{\alpha k}} \\
&\leq \frac{1}{2e\Delta^2} \sum_{k\geq 1} \left(\frac{e\Delta^{2+d}}{\lambda^\alpha}\right)^k \\
&\leq \frac{1}{2e\Delta^2} \sum_{k\geq 1} \left(\frac{e}{3}\right)^k \\
&= \frac{1}{2e\Delta^2}\cdot\frac{e}{3-e} \in \Theta(1/n^2).
\end{aligned}$$

Thus, we can apply Theorem 6.16 and get an FPRAS with the desired runtime.

Naturally speaking, this means that by making the bound on λ worse by a factor of only $\Delta^{\frac{d}{\alpha}}$ for any constant $d > 0$, the heavy runtime dependency on Δ can be removed.

6.4 Graph polymers from edge spin systems

A different type of graph polymer models are those that arise from edge spin systems. Formally, polymer models from edge spin systems can be characterized in the following way.

Definition 6.23 (Polymer models from edge spin systems). Given an undirected graph $G = (V, E)$, a set of spins $[q] = \{0, \dots, q\}$ and a ground state $\partial_e \in [q]$ for every edge $e \in E$. An edge spin polymer model is a polymer model $(\mathcal{C}, w, \dots)$ with the following properties.

1. Ever polymer $\gamma \in \mathcal{C}$ is defined by some non-empty set of edges $\overline{\gamma} \subseteq E$ that induces a connected subgraph in G, and a spin assignment $\varphi_\gamma : \overline{\gamma} \to [q]$ such that for every $e \in \overline{\gamma}$ it holds that $\varphi_\gamma(e) \neq \partial_e$.

2. Two polymers $\gamma, \gamma' \in \mathcal{C}$ are incompatible if their induced subgraphs share a vertex (i.e., $\gamma \not\sim \gamma'$ if there are $e \in \overline{\gamma}, f \in \overline{\gamma'}$ such that there is a vertex $v \in V$ which is incident to e and f).

For simplicity, we might write γ to denote the edge set $\overline{\gamma}$ if the spin assignment is not important (e.g., $e \in \gamma$ instead of $e \in \overline{\gamma}$). Furthermore, we will also write $v \in \gamma$ for some vertex $v \in V$ if there is an edge $e \in \gamma$ which is incident to v.

Again, we have to argue that such polymer models can be covered with a polynomial number of polymer cliques. As for the vertex spin systems, we choose $\{\Lambda_v \mid v \in V\}$ with $\Lambda_v = \{\gamma \in \mathcal{C} \mid v \in \gamma\}$ as polymer clique cover. Thus, drawing a polymer clique uniformly at random again boils down to choosing a vertex uniformly at random. In addition, whenever we need a size-function we will choose the number of edges in a polymer $|\gamma| = |\overline{\gamma}|$. Note that in contrast to our choice for vertex spin polymer models, this time the size of a polymer is not equal to the number of cliques which contain it (i.e., $|\gamma| \neq m_\gamma$).

For all our applications, the following notion of computational feasibility for edge spin polymer models will be required.

I **Definition 6.24 (Computationally feasible edge spin model).** Given an undirected graph $G = (V, E)$ with $|V| = n$. We say an edge spin polymer model $(\mathcal{C}, w, \circ)$ on G with spins $[q]$ is computationally feasible if the following holds:

1. For every edge set $E' \subseteq E$ with $|E'| = k$ and every spin assignment $\varphi : E' \to [q]$ we can decide in time at most $O(e^k)$ whether E' and φ form a valid polymer in $\mathcal{C}$.

2. For every polymer $\gamma \in \mathcal{C}$ with $|\gamma| = k$ we can compute w_γ in time at most $O(e^k)$. J

This is very similar to our definition of computational feasibility for vertex spin systems.

To apply our sampling schema, we will need the following lemma about the enumeration of edge induced subgraphs.

I **Lemma 6.25 (Casel et al. [Cas+19], part of Theorem 1.1).** Given an undirected graph $G = (V, E)$ with bounded degree Δ and some $v \in V$. There is an algorithm that enumerates all connected, edge-induced sugraphs of G that contain v and have at most $k \in \mathbb{N}_{>0}$ edges in time $e^{O(k \ln(\Delta))}$. J

As for the vertex spin polymer models, this implies that we can enumerate all candidates for Λ_v^k in time at most $e^{O(k \ln(\Delta q))}$ by enumerating all edge-induced subgraphs that contain v up to size k and all spin assignments for their edges.

With this, we can use Theorem 5.7 and Corollary 5.8 to prove the following theorem about the existence of an FPRAS for the partition function of edge spin polymer models.

I **Theorem 6.26.** Given an undirected graph $G = (V, E)$ with $|V| = n$ and constant degree bound Δ. Let $(\mathcal{C}, w, \circ)$ be a computationally feasible edge spin polymer model with spins $[q]$ for some constant q. Assume for every polymer $\gamma \in \mathcal{C}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|}$$

for some constant $a \in \mathbb{R}_{>0}$. Further, assume one of the following conditions is true:

1) For every $v \in V$ either $\Lambda_v = \emptyset$ or there is a polymer $\gamma \in \Lambda_v$ such that $|\gamma| \geq \log^1(n^0)$.

2) For every $v \in V$ it holds that

$$\sum_{\gamma \in \Lambda_v} e^{b|\gamma|} |w_\gamma| \leq h^1(n^0)$$

for some constant $b \in \mathbb{R}_{>0}$ and $h^1(n^0) \in \text{poly}^1(n^0)$.

Then there is an FPRAS for the partition function which has runtime at most $\left(\frac{n}{\varepsilon}\right)^{O^1(\ln^1(\Delta q^{\infty}))}$. ⌟

We omit the proof, because for our choice of polymer clique cover and due to Lemma 6.25, it is identical to the proof of Theorem 6.15.

Just as for the vertex spin polymer models, we can also state a version without runtime dependence on $\ln^1(\Delta q^0)$.

⌞ **Theorem 6.27.** Given an undirected graph $G = (V, E^0)$ with $|V| = n$ and constant degree bound Δ. Let $(\mathcal{C}, w, \emptyset^0)$ be a computationally feasible edge spin polymer model with spins $\leq q^{1/4}$ for some constant q. Assume for every polymer $\gamma \in \mathcal{C}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} |w_{\gamma'}| \leq e^{a|\gamma|}$$

for some constant $a \in \mathbb{R}_{>0}$. Further, assume for every $v \in V$ it holds that

$$\sum_{\gamma \in \Lambda_v} e^{b \ln^1(\Delta q^0)|\gamma|} |w_\gamma| \leq h^1(n^0)$$

for some $b \in \mathbb{R}_{>0}$ and $h^1(n^0) \in \text{poly}^1(n^0)$. Then there is an FPRAS for the partition function with runtime at most $\left(\frac{n}{\varepsilon}\right)^{c_0 + \frac{c_1}{b}}$ for some absolute constants c_0, c_1, which are independent of Δ and q. ⌟

Again, the proof is identical to the proof of Theorem 6.16 and thus omitted.

Edge spin polymer models as described above have for example been used by Casel et al. [Cas+19] in order to approximate Holant polynomials. They applied this modeling to problems which naturally can be encoded in this Holant framework and used the size-dependent Kotecký-Preiss condition to give bounds

for zero-free regions in the complex plain, which are closely related to bounds for efficient approximation via cluster expansion. We will investigate approximation bounds using the same modeling but with our approach instead of cluster expansion. Opposed to Casel et al. [Cas+19], we are again restricted to positive real weights for our polymer models, as we start from sampling the Gibbs distribution.

Perfect matching polynomial

The question of the computational complexity of counting perfect matchings remains an unsolved problem in theoretical computer science, related to many approximation problems. Besides counting perfect matchings, attempts have been made to approximate other statistics over the set of all perfect matching. One of them is the so called perfect matching polynomial.

I **Definition 6.28 (Perfect matching polynomial).** Given an undirected graph $G = (V, E)$ with a non-empty set of perfect matchings $\mathcal{M}$ and some fixed $M_0 \in \mathcal{M}$. The perfect matching polynomial for a weight $z \in \mathbb{R}_{>0}$ is defined as

$$P_{G;M_0}(z) = \sum_{M \in \mathcal{M}} z^{|M_0 \triangle M|};$$

where $M_0 \triangle M$ denotes the symmetric difference of the edge sets of both perfect matchings. J

Similar to Casel et al. [Cas+19], we will first translate this problem into a polymer model. Note that the case $M = M_0$ contributes 1 to the polynomial. In addition, for any perfect matching $M \in \mathcal{M}$ with $M \neq M_0$, the symmetric difference between M and M_0 can be decomposed into vertex-disjoint cycles, alternating between matching and non-matching edges with respect to M_0. Moreover, any combination of vertex-disjoint alternating cycles corresponds to a perfect matching, different from M_0.

Based on these observations, we define our polymer model for a given graph $G = (V, E)$ and a perfect matching M_0. We use the spins $\{0, 1\}$ and set $\partial_e = 0$ for every $e \in E$. Our polymers $\mathcal{C}$ are all alternating cycles in G with respect to M_0, together with spin 1 for every $e \in \gamma$. In addition, we set the weight $w_\gamma = z^{|\overline{\gamma}|} = z^{|V| - |\gamma|}$ for every polymer $\gamma \in \mathcal{C}$. According to our definition of for

edge spin polymer models, we have a one to one correspondence between the set of polymer families F and the set of perfect matching M. Let $\Gamma_M \in \mathsf{F}$ be the polymer family that represents $M \in \mathsf{M}$. Obviously, we have $\Gamma_{M_0} = \emptyset$ and for all M, M_0 it holds that

$$z^{|M_0 \triangle M|} = \prod_{\gamma \in \Gamma_M} z^{\gamma} = \prod_{\gamma \in \Gamma_M} w_\gamma .$$

Thus, we get

$$Z_{G;M_0}(z) = \sum_{\Gamma_M \in \mathsf{F}} \prod_{\gamma \in \Gamma_M} w_\gamma = \sum_{M \in \mathsf{M}} z^{|M_0 \triangle M|} = P_{G;M_0}(z).$$

We can now apply our approach for approximating the partition function $Z_{G;M_0}(z)$ as an approximation of $P_{G;M_0}(z)$. In order to do so, the following lemma will be useful.

Lemma 6.29 (Casel et al. [Cas+19], part of Theorem 7.7). Given an undirected graph $G = (V, E)$ with a perfect matching M_0. For every edge $e \in E$ and every $k \geq 2$, there are at most $(\Delta - 1)^{2k}$ alternating cycles of length $2k$ with respect to M_0. In addition, for every alternating cycle with length l, there are at most $\frac{1}{2}l(\Delta - 1)^{2k}$ alternating cycles of length $2k$ that share at least one vertex.

With this bound, we can prove the following result.

Proposition 6.30. Given an undirected graphs $G = (V, E)$ with bounded degree Δ and $|V| = n$. In addition, let M_0 be some perfect matching of G. For $z < \frac{1}{3(\Delta-1)^3}$ there is an FPRAS for the perfect matching polynomial $P_{G;M_0}(z)$ with runtime in $\left(\frac{n}{\varepsilon}\right)^{O(\log(\Delta))}$.

Proof of Proposition 6.30. We want to apply Theorem 6.26. Note that the polymer model is computationally feasible as calculating weights and identifying which subgraphs are valid polymers can be done efficiently.

First, we prove that for every $\gamma \in \mathsf{C}$ and some constant $a \in \mathbb{R}_{>0}$ it holds that

$$\sum_{\gamma' \not\sim \gamma} e^{a|\gamma'|} w_{\gamma'} \leq e^{a|\gamma|}.$$

For this, we set $a = \frac{1}{5}$. Note that every alternating cycle has an even length, which implies that there are only polymers with even size. By application of

Lemma 6.29 we get the following bound

$$\sum_{\gamma^0 \not\sim \gamma} e^{\frac{1}{5}|\gamma^0|} w_{\gamma^0} \leq \frac{1}{2}|\gamma| \sum_{k \geq 1} (1+\Delta)^k (1+z^2 e^{\frac{2}{5}})^k = \frac{1}{2}|\gamma| \sum_{k \geq 1} \left((1+\Delta)(1+z^2 e^{\frac{2}{5}})\right)^k.$$

By our choice of z, we can further see that

$$\sum_{k \geq 1} (1+\Delta)^k (1+z^2 e^{\frac{2}{5}})^k = \sum_{k \geq 1} \left(\frac{e^{\frac{2}{5}}}{3}\right)^k = \frac{e^{\frac{2}{5}}}{3 - e^{\frac{2}{5}}} < 1$$

and so we have

$$\sum_{\gamma^0 \not\sim \gamma} e^{\frac{1}{5}|\gamma^0|} w_{\gamma^0} < \frac{1}{2}|\gamma| \cdot e^{\frac{1}{5}|\gamma|}.$$

For the given polymer model and our choice of the polymer clique cover, we can not assume that for every non-empty polymer clique Λ_v contains a polymer of logarithmic size. Thus, we need to apply case 2) of Theorem 6.26, meaning that we also have to show for every $v \in V$ that

$$\sum_{\gamma \in \Lambda_v} e^{b|\gamma|} w_\gamma \leq h(n)$$

for some constant $b \in \mathbb{R}_{>0}$ and $h(n) \in \text{poly}(n)$. A simple way to ensure this is to set $b = a = \frac{1}{5}$. Now, for any vertex $v \in V$ and any polymer $\gamma \in \Lambda_v$ we have that

$$\sum_{\gamma^0 \in \Lambda_v} e^{\frac{1}{5}|\gamma^0|} w_{\gamma^0} \sum_{\gamma^0 \not\sim \gamma} e^{\frac{1}{5}|\gamma^0|} w_{\gamma^0} < \frac{1}{2}|\gamma|.$$

The last inequality follows directly from our calculations above.

This shows that we can apply Theorem 6.26 and because $q = 1$ it concludes our proof.

We want to add one more comment at this point. The bound on z that we used in Proposition 6.30 is actually not optimal for our approach. When looking

at the proof, we see that the main restriction is imposed by the condition

$$\frac{1}{2}\mathrm{j}y\mathrm{j}\tilde{\mathrm{O}}_{k\ 1}{}^{1}\Delta\quad 1^{\mathrm{o}}z^{2}e^{\frac{2}{5}}{}^{k}\quad e^{\frac{1}{5}\mathrm{j}y\mathrm{j}}:$$

We ensured this by setting z such that

$$\tilde{\mathrm{O}}_{k\ 1}{}^{1}\Delta\quad 1^{\mathrm{o}}z^{2}e^{\frac{2}{5}}{}^{k}<1:$$

This is actually more restrictive than necessary. By observing that $\mathrm{j}y\mathrm{j}$ also only takes even positive integers, some calculations show that

$$z\quad \text{q}\frac{1}{\frac{3+e^{6\bullet 5}}{e^{4\bullet 5}}{}^{1}\Delta\quad 1^{\mathrm{o}}}\quad \text{p}\frac{1}{2{:}8399^{1}\Delta\quad 1^{\mathrm{o}}}$$

would also be sufficient. Further improvements can be achieved combining this with an optimized value for a. However, for simplicity of the proposition, we decided to state a slightly weaker version.

For the complex case Theorem 7.7 of Casel et al. [Cas+19] states a convergence radius of the cluster expansion of approximately $\delta = \text{p}\frac{1}{4{:}85718^{1}\Delta\ \ 1^{\mathrm{o}}}$, leading to an approximation for any constant $\mathrm{j}z\mathrm{j} < \delta$. Obviously, Proposition 6.30 suggest that a randomized approximation can be done for slightly larger values of z along the positive real axis.

Besides this improvement of parameter bounds, the following result is interesting in terms of runtime.

I **Proposition 6.31.** Given an undirected graphs $G = {}^{1}V;E^{\mathrm{o}}$ with bounded degree $\Delta \quad 3$ and $\mathrm{j}V\mathrm{j} = n$. In addition, let M_0 be some perfect matching of G. If for some constant $d > 0$ we have

$$z\quad \min\left(\text{p}\frac{1}{2^{1}\Delta\quad 1^{\mathrm{o}1+d}};\ \text{p}\frac{1}{3^{1}\Delta\quad 1^{\mathrm{o}}}\right);$$

then there is an FPRAS for the perfect matching polynomial $P_{G;M_0}{}^{1}z^{\mathrm{o}}$ with runtime in $\frac{n}{\varepsilon}{}^{c_0+\frac{c_1}{a}}$ for some absolute constants $c_0;c_1$ independent of Δ. J

Proof of Proposition 6.31. We will prove this by applying Theorem 6.27. First,

note that for $z \geq \frac{1}{3^1\Delta - 1^0}$ we know from the proof of Proposition 6.30 that for every $\gamma \in C$

$$\prod_{\gamma^0 \ni \gamma} e^{\frac{1}{5}|\gamma^0|} w_{\gamma^0} \leq e^{\frac{1}{5}|\gamma|}.$$

Further, because $q = 1$ and $\Delta \geq 3$, we know that for every $v \in V$ it holds that

$$\sum_{\gamma \in \Lambda_v} e^{b \ln^1\Delta q^0|\gamma|} w_\gamma \leq \sum_{\gamma \in \Lambda_v} e^{2b \ln^1\Delta - 1^0|\gamma|} w_\gamma.$$

We set $b = \frac{d}{2} > 0$, and because $z \geq \frac{1}{2^1\Delta - 1^{0a+1}}$ we get that

$$\sum_{\gamma \in \Lambda_v} e^{d \ln^1\Delta - 1^0|\gamma|} w_\gamma \leq \prod_{k \geq 1} \left({}^1\Delta - 1^{01+d} z^2 \right)^k \frac{{}^1\Delta - 1^{01+d} z^2}{1 - {}^1\Delta - 1^{01+d} z^2} \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

Thus, we can apply Theorem 6.27, which yields the desired runtime and concludes our proof.

Again, with only slightly worse bounds on the parameter, namely by a factor of roughly ${}^1\Delta - 1^{0\frac{d}{2}}$, we can remove the runtime dependency on Δ.

7 Conclusions & Outlook

In this thesis, we introduced a framework to use Markov chains for sampling from the Gibbs distribution and approximating the partition function of abstract polymer models. The concept of polymer cliques, which we introduced for this, can be seen as the core property that existing chains for vertex spin systems by Chen et al. [Che+19] or edge spin systems by Casel et al. [Cas+19] implicitly used.

Moreover, we introduced the generalized polymer mixing condition as a way for bounding the mixing time of our chain. It shows interesting relations between commonly known conditions for convergence of the cluster expansion and the mixing time of the polymer Markov chain, as for example stated in Proposition 6.6. Namely, all conditions that are given by Fernández and Procacci [FP07] can be shown to imply the generalized polymer mixing condition. This gives evidence that for positive real valued polymer models the generalized polymer mixing condition tends to be a less restrictive condition than convergence of the cluster expansion. Although we were not able to prove such a statement in general, we see this as an interesting problem for future research and such a relation might be of separate theoretical interest.

When turning our Markov chain into an efficient sampling schema, we faced the problem that it is not obvious how to do each transition of the Markov chain efficiently. The reason is that it involves drawing from the clique polymer distribution. We doubt that there are general conditions, under which this can be done efficiently, as detailed knowledge about the structure of the polymers would be needed.

To give a more general setting under which our polymer Markov chain can be used to construct an efficient sampler for the Gibbs distribution, we relaxed the condition of exact sampling from the clique polymer distribution to efficient approximate sampling from each of them. The technique that we used for this can be seen as a general way to use Markov chains with transition errors for sampling. We emphasize that this might be of separate interest for other algorithmic applications.

Further we proposed truncation and enumeration of each polymer clique as one way for approximate sampling from the clique polymer distribution. Although this seems similar to what is done in cluster expansion algorithms, our arguments for truncation directly involve the polymer model instead of using absolute convergence of some infinite series. Moreover, the resulting clique truncation condition only has to ensure that we can bound the resulting error for each polymer clique, as the combinatorial complexity of constructing polymer families is already handled by the polymer Markov chain.

This decoupling of conditions for truncation and conditions for constructing families of polymers is of special interest for the algorithmic application of vertex and edge spin polymer models. Here, the choice of truncation size heavily influences the runtime of approximation algorithms.

We showed that by carefully choosing the function that is used for the clique truncation condition, the exponential runtime dependency on $\ln^1\Delta^o$ can be removed for only slightly worse bounds on the weights. This can be seen as a trade-off between fast Markov chain approaches as proposed by Chen et al. [Che+19] and slow cluster expansion approaches as introduced by Helmuth et al. [HPR18]. We would like to point out that such improvements are not limited to the examples that we have given, but rather work for most applications of graph polymer models.

Although truncation imposes an additional condition beside the generalized polymer mixing condition to do efficient sampling, constructions like in the proof of Proposition 6.8 show that such additional conditions are necessary unless NP = RP. More such examples would be interesting to get a better characterization under which conditions efficient approximation algorithms can be expected.

The reader might have noticed that we used Theorem 5.7 and Corollary 5.8, which both rely on truncation, in all applications. This rises the question why we stated Theorem 3.13 and Theorem 4.5 as separate theorems in the first place.

Typical applications for Theorem 3.13, namely exact sampling from each clique polymer distribution, are for instance polymer models with a number of polymers that is polynomial in the input size. In such cases, each polymer can be seen as its own clique, making exact sampling from each clique polymer distribution a trivial task. Examples for this are the hard-core model on general bounded degree graphs or the monomer dimer model (i.e., counting and sampling weighted matchings).

In contrast to that, it is less obvious for which algorithmic problems Theorem 4.5 can be applied without relying on truncation and enumeration. Although one can think of other methods, like Markov chains, to approximately sample from each polymer clique, such Markov chains need a more detailed structural understanding of the polymers (e.g., how can we get from one polymer in a clique to another polymer from the same clique). As such structural properties heavily depend on the concrete polymer model, we consider this question to be outside the scope of this thesis. However, we want to emphasise that such alternative approaches could especially lead to better runtime results, whenever there is a more efficient sampling method than plain enumeration. Thus, we think that Theorem 4.5 has its own right for future research.

Finally, we would like to add some notes regarding implications of our results for other lines of research, which we have not discussed in this thesis. For example Chen et al. [Che+19] studied the mixing time of restricted Glauber dynamics. They showed that if their Markov chain for vertex spin polymer models is rapidly mixing and polymers are of at most logarithmic size, those restricted Glauber dynamics are rapidly mixing as well. As our polymer Markov chain is a generalization of their chain, it would be interesting to see how generalized polymer mixing condition and clique truncation condition can be used to extend the region of rapid mixing of those restricted Glauber dynamics.

Bibliography

[Ald83] David Aldous. "Random walks on finite groups and rapidly mixing Markov chains." In: *Séminaire de Probabilités XVII 1981/82*. Springer, 1983, 243–297 (see page 19).

[Bar14] Alexander Barvinok. *Computing the permanent of (some) complex matrices*. 2014. arXiv: 1405.1303 [math.CO] (see page 4).

[BI89] Christian Borgs and John Z. Imbrie. **A unified approach to phase diagrams in field theory and statistical mechanics**. *Comm. Math. Phys.* 123:2 (1989), 305–328. URL: https://projecteuclid.org:443/euclid.cmp/1104178765 (see page 1).

[Bor+10] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. *Left and right convergence of graphs with bounded degree*. 2010. arXiv: 1002.0115 [math.CO] (see page 69).

[Bré13] Pierre Brémaud. **Markov chains: Gibbs fields, Monte Carlo simulation, and queues**. Vol. 31. Springer Science & Business Media, 2013 (see page 9).

[Cas+19] Katrin Casel, Philipp Fischbeck, Tobias Friedrich, Andreas Göbel, and J. A. Gregor Lagodzinski. *Zeros and approximations of Holant polynomials on the complex plane*. 2019. arXiv: 1905.03194 [cs.DS] (see pages 1, 3–6, 59, 75–78, 80, 83).

[Che+19] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. **Fast algorithms at low temperatures via Markov chains**. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2019 (see pages 5, 6, 62, 63, 69, 72, 83–85).

[CM01] Grace E Cho and Carl D Meyer. **Comparison of perturbation bounds for the stationary distribution of a Markov chain**. *Linear Algebra and its Applications* 335:1-3 (2001), 137–150 (see page 34).

[CP19] Sarah Cannon and Will Perkins. *Counting independent sets in unbalanced bipartite graphs*. 2019. arXiv: 1906.01666 [cs.DS] (see pages 1, 3, 4, 8, 61, 65).

[DG00] Martin Dyer and Catherine Greenhill. **On Markov Chains for Independent Sets**. *Journal of Algorithms* 35:1 (2000), 17–49. ISSN: 0196-6774. DOI: https://doi.org/10.1006/jagm.1999.1071. URL: http://www.sciencedirect.com/science/article/pii/S0196677499910714 (see page 16).

[DG98] Martin E. Dyer and Catherine S. Greenhill. **A more rapidly mixing Markov chain for graph colorings**. *Random Struct. Algorithms* 13 (1998), 285–317 (see page 17).

[Dob96] RL Dobrushin. **Estimates of semi-invariants for the Ising model at low temperatures**. *Translations of the American Mathematical Society-Series 2* 177 (1996), 59–82 (see pages 4, 55).

[FHL12] Déborah Ferré, Loïc Hervé, and James Ledoux. **Regular Perturbation of V-Geometrically Ergodic Markov Chains**. *Journal of Applied Probability* 50 (June 2012), 184–194. DOI: 10.1239/jap/1363784432 (see page 34).

[FP07] Roberto Fernández and Aldo Procacci. **Cluster Expansion for Abstract Polymer Models. New Bounds from an Old Approach**. *Communications in Mathematical Physics* 274:1 (June 2007), 123–140. ISSN: 1432-0916. DOI: 10.1007/s00220-007-0279-2. URL: http://dx.doi.org/10.1007/s00220-007-0279-2 (see pages 4, 6, 7, 54, 55, 59, 83).

[GGS19] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. **A note on the high-fugacity hard-core model on bounded-degree bipartite expander graphs**. *ArXiv* abs/1901.06653 (2019) (see page 7).

[GPR] Sam Greenberg, Amanda Pascoe, and Dana Randall, 76–85. In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. DOI: 10.1137/1.9781611973068.9. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611973068.9. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611973068.9 (see pages 18–21).

[GRS17] Sam Greenberg, Dana Randall, and Amanda Pascoe Streib. *Sampling Biased Monotonic Surfaces using Exponential Metrics*. 2017. arXiv: 1704.07322 [cs.DS] (see page 20).

[Hol12] Frank den Hollander. **Probability theory: The coupling method**. *Lecture notes* (2012). URL: http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf (see page 32).

[HPR18] Tyler Helmuth, Will Perkins, and Guus Regts. *Algorithmic Pirogov-Sinai theory*. 2018. arXiv: 1806.11548 [cs.DS] (see pages 4, 72, 84).

[Hun06] Jeffrey J Hunter. **Mixing times with applications to perturbed Markov chains**. *Linear Algebra and its Applications* 417:1 (2006), 108–123 (see page 34).

[Jer03] Mark Jerrum. **Counting, sampling and integrating: algorithms and complexity**. Springer Science & Business Media, 2003 (see page 51).

[JKP18] Matthew Jenssen, Peter Keevash, and Will Perkins. **Algorithms for #BIS-hard problems on expander graphs**. *CoRR* abs/1807.04804 (2018). arXiv: 1807.04804. URL: http://arxiv.org/abs/1807.04804 (see pages 1, 3, 4, 8, 61, 63, 65, 66, 68, 69, 72).

[JVV86] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. **Random generation of combinatorial structures from a uniform distribution**. *Theoretical Computer Science* 43 (1986), 169–188 (see pages 12, 47).

[KP86] Roman Kotecký and David Preiss. **Cluster expansion for abstract polymer models**. *Communications in Mathematical Physics* 103:3 (1986), 491–498 (see pages 3, 54).

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

[Mit05] A. Yu. Mitrophanov. **Sensitivity and Convergence of Uniformly Ergodic Markov Chains**. *Journal of Applied Probability* 42:4 (2005), 1003–1014. ISSN: 00219002. URL: http://www.jstor.org/stable/30040824 (see page 34).

[Pap03] Christos H Papadimitriou. **Computational complexity**. John Wiley and Sons Ltd., 2003 (see pages 12, 57).

[PR17] Viresh Patel and Guus Regts. **Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials**. *SIAM Journal on Computing* 46:6 (2017), 1893–1919 (see pages 4, 63).

[SS12] Allan Sly and Nike Sun. *The computational hardness of counting in two-spin models on d-regular graphs*. 2012. arXiv: 1203.2602 [math.PR] (see page 67).

[SV03] Eilon Solan and Nicolas Vieille. **Perturbed Markov Chains**. *Journal of Applied Probability* 40:1 (2003), 107–122. ISSN: 00219002. URL: http://www.jstor.org/stable/3215844 (see page 34).

[VV86] L.G. Valiant and V.V. Vazirani. **NP is as easy as detecting unique solutions**. *Theoretical Computer Science* 47 (1986), 85–93. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(86)90135-0. URL: http://www.sciencedirect.com/science/article/pii/0304397586901350 (see page 57).

[Wei06] Dror Weitz. **Counting Independent Sets up to the Tree Threshold**. In: *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '06. Seattle, WA, USA: Association for Computing Machinery, 2006, 140–149. ISBN: 1595931341. DOI: 10.1145/1132516.1132538. URL: https://doi.org/10.1145/1132516.1132538 (see page 67).

Declaration of Authorship

I hereby declare that this thesis is my own unaided work. All direct or indirect sources used are acknowledged as references.

Potsdam, March 5, 2020

Marcus Pappik