Polymer Dynamics via Cliques: New Conditions for Approximations

Tobias Friedrich* Andreas Göbel* Martin S. Krejca†
Marcus Pappik*

* Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
{tobias.friedrich, andreas.goebel, marcus.pappik}@hpi.de

† LIX, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France
martin.krejca@polytechnique.edu

Abstract

Abstract polymer models are systems of weighted objects, called polymers, equipped with an incompatibility relation. An important quantity associated with such models is the partition function, which is the weighted sum over all sets of compatible polymers. Various approximation problems reduce to approximating the partition function of a polymer model. Central to the existence of such approximation algorithms are weight conditions of the respective polymer model. Such conditions are derived either via complex analysis or via probabilistic arguments. We follow the latter path and establish a new condition—the clique dynamics condition—, which is less restrictive than the ones in the literature. We introduce a new Markov chain where the clique dynamics condition implies rapid mixing by utilizing cliques of incompatible polymers that naturally arise from the translation of algorithmic problems into polymer models. This leads to improved parameter ranges for several approximation algorithms, such as a factor of at least $2^{1/\alpha}$ for the hard-core model on bipartite α-expanders.

Keywords: Markov chain, partition function, Gibbs distribution, approximate counting, abstract polymer model

1 Introduction

Statistical physics models systems of interacting particles as probability distributions. This approach explains a variety of real-world phenomena, including ferromagnetism [21], segregation [35], and real-world network generation [10]. Characteristic of such systems is that they undergo phase transitions depending on some external parameter. Such phase transitions have been recently linked to the tractability of computational tasks, leading to a two-way exchange: tools from statistical physics are used to explain computational phenomena, and tools from computer science are used to explain physical phenomena. An established technique for investigating these phase transitions that involves translating the states of a spin system as perturbations from a ground state [16, Chapter 7] has been recently introduced to computer science as an algorithmic tool for computational tasks of spin systems [20].

To motivate the definition of the central mathematical object of this article, we give a high-level description of how to model a spin system in terms of perturbations from a ground state. Assume we study a q-state spin system on a graph G. The states of the spin system are usually mappings $\sigma\colon V(G) \to Q$ from the vertices of G to some finite set Q. Each such configuration σ has a weight $w(\sigma) \in \mathbb{R}_{\geq 0}$ and the sum of the weights of all the configurations $Z = \sum_\sigma w(\sigma)$ is called the *partition function*. For each configuration σ, the probability distribution μ that characterizes our system yields $\mu(\sigma) = w(\sigma)/Z$. Let σ_0 be the ground state we use in this translation. Given a configuration σ, we identify the set of vertices $D \subseteq V(G)$ where, for each $v \in D$, we have $\sigma_0(v) \neq \sigma(v)$. Observe that we can uniquely identify this configuration by a set Γ whose elements γ consist of a connected component of $G[D]$ together with the restriction of σ on this component. Furthermore, we assign a weight w_γ to each $\gamma \in \Gamma$, such that $\prod_{\gamma\in\Gamma} w_\gamma = w(\sigma)/w(\sigma_0)$. Thus, provided that all such sets of pairs Γ contain no two pairs γ, γ' that are incompatible, i.e. Γ cannot be uniquely decoded to an assignment because for example γ and γ' map the same vertex to a different element in Q, there is a bijection between the configurations σ and the sets Γ. Furthermore, the distribution μ is expressed as a distribution over the sets Γ, since it retains the property that the probability of Γ is proportional to its weight. Such a construction suggests the following definition.

A *polymer model* $\mathcal{P} = (C, w, \nsim)$ is a tuple consisting of a non-empty, countable[1] set C, a set $w = \{w_\gamma\}_{\gamma\in C}$ of positive real weights and a reflexive and symmetric relation $\nsim \subseteq C^2$. The elements $\gamma \in C$ are called *polymers*. The relation $\nsim$ is called the *incompatibility* relation and, for $\gamma, \gamma' \in C$, we say that γ and γ' are *incompatible* if $\gamma \nsim \gamma'$, and that they are *compatible* otherwise. We call a *finite* subset $\Gamma \subseteq C$ a *polymer family* if and only if all polymers of Γ are pairwise compatible. Given a polymer model $\mathcal{P}$, let $\mathcal{F}^{(\mathcal{P})}$ denote the set of *all* polymer families of $\mathcal{P}$. Note that $\mathcal{F}^{(\mathcal{P})}$ is countable since the set of finite subsets of a countable set is countable. The *partition function* of $\mathcal{P}$ is defined as

$$Z(\mathcal{P}) = \sum\nolimits_{\Gamma\in\mathcal{F}^{(\mathcal{P})}} \prod\nolimits_{\gamma\in\Gamma} w_\gamma, \tag{1}$$

which we require to be finite. Further, the *Gibbs distribution* of $\mathcal{P}$ is the probability distribution $\mu^{(\mathcal{P})}$

[1] In our applications, C is finite.

over $\mathcal{F}^{(\mathcal{P})}$ such that, for all $\Gamma \in \mathcal{F}^{(\mathcal{P})}$,

$$\mu^{(\mathcal{P})}(\Gamma) = \frac{\prod_{\gamma \in \Gamma} w_\gamma}{Z(\mathcal{P})}. \tag{2}$$

A helpful interpretation for understanding the definition of a polymer model is the following. Ignoring the reflexivity of $\nsim$, we view the pair $(C, \nsim)$ as a graph, which we call the *polymer graph*. We observe that the families of $\mathcal{F}^{(\mathcal{P})}$ correspond to the independent sets of $(C, \nsim)$. Thus, for the special case where $w_\gamma = \lambda \in \mathbb{R}$, for each $\gamma \in C$, the distribution μ is the hard-core model [34] on the polymer graph and $Z(\mathcal{P})$ is the independence polynomial [32].

We aim to maximize the parameter range for which the following two computational tasks can be done in polynomial time.

(1) Approximately sampling from the Gibbs distribution of a polymer model, i.e., return a random family Γ from a distribution with total-variation distance at most ε from $\mu^{(\mathcal{P})}$.

(2) Returning an estimate $\widetilde{Z}$, such that $(1-\varepsilon)Z(\mathcal{P}) \leq \widetilde{Z} \leq (1+\varepsilon)Z(\mathcal{P})$ with constant probability greater than $\frac{1}{2}$.

1.1 Known Algorithmic Results

There is an expanding list of results that utilize abstract polymer models to obtain efficient approximation and sampling algorithms for new parameter regimes for various spin systems on graphs. This line of research was initiated by Helmuth et al. [20], who used polymers to obtain polynomial-time approximation and sampling algorithms at a regime where the weight of the interactions of particles with an external field is low. For problems that are hard to approximate on general inputs, polynomial-time approximation algorithms are derived by utilizing restrictions upon the input graph of a spin system. Examples include spin systems on expander graphs [18, 22, 23, 27], the hard-core model on unbalanced bipartite graphs [6], and the ferromagnetic Potts model on d-dimensional lattices [4]. Furthermore, polymer models have been used to approximate and sample edge spin systems (holant problems) at low temperatures [7].

Translating a spin system on a graph G with n vertices into an abstract polymer model commonly results in an exponential number of polymers in terms of n—as can be observed in our initial example. However, we are interested in approximation and sampling algorithms with a runtime polynomial in n. There are two main approaches for such algorithms.

(i) Cluster Expansion This approach considers complex weights for the polymers and is based on an infinite series expansion of $\ln Z$ (the *cluster expansion*). The essential element for polynomial-time computation is a theorem of Kotecký and Preiss [26, Theorem 1], a condition for establishing absolute convergence of the cluster expansion. By satisfying this condition, the cluster expansion is truncated to its most significant terms, obtaining an ε-additive approximation for $\ln Z$. Computing the significant terms is achieved by enumerating connected induced subgraphs of the polymer

graph of small size. By an algorithm of Patel and Regts, the enumeration takes polynomial time in terms of the input graph of the spin system [31]. The ε-additive approximation of $\ln Z$ immediately gives a multiplicative ε-approximation for Z. The runtime of this approach is commonly $n^{O(\log \Delta)}$, where Δ is the maximum degree of the input graph G for the spin system and $n = |V(G)|$. Based on an approximation algorithm for Z, self-reducibility of the polymer model yields an approximate sampler for $\mu^{(\mathcal{P})}$ [20].

(ii) Markov Chain Monte Carlo Initiated by Chen et al. [8], this approach defines a Markov chain with state space $\mathcal{F}^{(\mathcal{P})}$ and stationary distribution $\mu^{(\mathcal{P})}$. The Markov chain requires the polymer model to have originated from a spin system on a graph G with n vertices. Iteratively, the chain samples a polymer γ with probability proportional to its weight w_γ and then adds or removes γ from its state if possible. When the *mixing condition* [8, Definition 1] is satisfied, the Markov chain converges to $\mu^{(\mathcal{P})}$ after $O(n \log n)$ iterations. The mixing condition matches a convergence condition arising from an analysis by Fernández et al. [14] of another stochastic process of polymers on lattices. An ε-approximate sampler for $\mu^{(\mathcal{P})}$ is obtained by simulating the Markov chain. The computational challenge for this approach is to sample the polymer γ in order to perform a transition of the Markov chain. As Chen et al. [8] show, this can be done in expected constant time provided the *sampling condition* [8, Definition 4] is satisfied. This results in an $O(n \log n)$ algorithm for sampling from the Gibbs distribution of a spin system. Using Simulated Annealing, Chen et al. convert this sampler into a randomized approximation scheme (FPRAS) for Z that runs in expected $O\left(n^2 \log n\right)$ time.

Comparison of Established Conditions for the Methods Above A number of conditions for the convergence of the cluster expansion have appeared in the literature [11, 15, 26]. The condition of Fernández and Procacci [15] is the least restrictive among them, i.e., the other conditions imply it. Thus, using the Fernández–Procacci condition, one could potentially obtain approximation algorithms for broader parameter ranges than the ones obtained by using, e.g., the Kotecký–Preiss condition [26]. However, the condition by Kotecký and Preiss is convenient to apply in polymer models of vertex spin systems and comes with implications on the rate of convergence of the cluster expansion used in algorithmic settings. When compared to cluster expansion conditions (restricted to non-negative real weights), the mixing condition of Chen et al. [8] is less restrictive than the Kotecký–Preiss condition, however, it is incomparable with the Fernández–Procacci condition. Note that the fast run-times of Chen et al. [8] are dependent on the sampling condition, which imposes the largest restriction on the parameter range of the applications it is used for.

1.2 Our Results

We study a new Markov chain $(X_t)_{t \in \mathbb{N}}$ for abstract polymer models with stationary distribution $\mu^{(\mathcal{P})}$, called the *polymer clique dynamics*. The dynamics of our Markov chain are based on a *clique cover*, i.e., a set $\Lambda = \{\Lambda_i\}_{i \in [m]}$ with $\bigcup \Lambda = C$ such that the polymers in each clique Λ_i are pairwise incompatible. Note that families of compatible polymers contain at most one polymer per

clique. At each step, our Markov chain chooses $i \in [m]$ uniformly at random and samples a family in Λ_i according to the distribution $\mu_{|\Lambda_i}$ defined as follows. For $\gamma \in \Lambda_i$, we have $\mu_{|\Lambda_i}(\{\gamma\}) = w_\gamma / Z_{|\Lambda_i}$ and, for the empty set, $\mu_{|\Lambda_i}(\emptyset) = 1/Z_{|\Lambda_i}$, where $Z_{|\Lambda_i} = 1 + \sum_{\gamma \in \Lambda_i} w_\gamma$. If the empty family is chosen and X_t contains a polymer from Λ_i, then the chain removes this polymer. If the family chosen contains a polymer, then, if possible, the chain adds this polymer to its state. For a detailed description of our chain, please refer to Definition 7.

Our chain can be viewed as a generalization of the single-site Glauber dynamics (cf. [13, insert/delete chain]). Any abstract polymer model has a trivial clique cover, where each clique contains exactly one polymer. The clique dynamics with the trivial clique cover coincides with the Glauber dynamics. Clique covers with a much smaller number of cliques than the number of polymers arise naturally from the translation of spin systems into polymer models. For example, the translation we discussed earlier in the introduction yields a clique cover with n cliques, one for each vertex in the original graph G. Since the number of polymers is commonly exponential in the size of G, our chain utilizes that a family of compatible polymers may contain at most one polymer from each polymer clique. The chain of Chen et al. [8] also utilizes this fact, however, in a more restricted setting and with a different sampling distribution for each clique.

Central to our mixing time analysis for this chain is the following condition.

▶ **Condition 1 (Clique Dynamics).** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, and let $f : C \to \mathbb{R}_{>0}$. We say that $\mathcal{P}$ *satisfies the clique dynamics condition with* f if and only if, for all $\gamma \in C$,

$$\sum_{\substack{\gamma' \in C \colon \gamma' \nsim \gamma, \\ \gamma' \neq \gamma}} f(\gamma') \frac{w_{\gamma'}}{1 + w_{\gamma'}} \leq f(\gamma) \,. \qquad \blacktriangleleft$$

We show that when the clique dynamics condition is satisfied, the mixing time of our Markov chain is polynomial in the number of cliques in the clique cover and logarithmic in f.

▶ **Theorem 2.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, and let $Z_{\max} = \max_{i \in [m]} \{Z_{|\Lambda_i}\}$. Further assume that $\mathcal{P}$ satisfies the clique dynamics condition with function f, and let $f_{\max} = \max_{\gamma \in C} \{f(\gamma)\}$ and $f_{\min} = \min_{\gamma \in C} \{f(\gamma)\}$.

Then the polymer clique dynamics $\mathcal{M}(\mathcal{P})$ has the unique stationary distribution $\mu^{(\mathcal{P})}$ and, for all $\varepsilon \in (0, 1]$, it holds that

$$\tau_{\mathcal{M}(\mathcal{P})}(\varepsilon) \in \mathrm{O}\left(m^3 Z_{\max} \ln\left(m^2 Z_{\max}^2 \frac{f_{\max}}{f_{\min}} \right)^2 \ln\left(\frac{1}{\varepsilon} \right) \right). \qquad \blacktriangleleft$$

When restricted to the setting of Chen et al. [8], the clique dynamics condition is implied by the mixing condition and thus less restrictive. Due to that weaker assumptions, our mixing time bound is slightly worse than the one proposed in [8]. The function f in our condition makes it easily comparable to conditions for cluster expansion. Unlike the Kotecký–Preiss condition [26], the clique dynamics condition allows for a more flexible choice of f in order to obtain a greater parameter range for the algorithmic applications (cf. [7, Remark 2.4]).

We further show that the clique dynamics condition is more general than the Fernández–Procacci [15] condition for the cluster expansion—and consequently more general than the Kotecký–Preiss condition [26] (see Section 3.1). An implication of our analysis is that cluster expansion conditions imply our condition for the mixing time of clique dynamics. To the best of our knowledge, this is the first connection between cluster expansion and mixing times of Markov chains for polymer models and might be of independent interest to the statistical-physics community. This is in line with the special case of the hard-core model, where the parameter range on the reals has a bigger radius than on the complex plane [32].

The clique dynamics condition allows us to prove the following theorem for approximately sampling from μ, which is our main algorithmic result.

▶ **Theorem 3.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model with polymer clique cover Λ of size m, assume $\mathcal{P}$ and Λ are computationally feasible as stated in Definition 12, and let $Z_{\max} = \max_{i\in[m]}\{Z_{|\Lambda_i}\}$. Further assume that

(a) $\mathcal{P}$ satisfies the clique dynamics condition for a function f such that, for all $\gamma \in C$, it holds that $\mathrm{e}^{-\mathrm{poly}(m)} \le f(\gamma) \le \mathrm{e}^{\mathrm{poly}(m)}$,

(b) $Z_{\max} \in \mathrm{poly}(m)$, and that,

(c) for all $i \in [m]$, we can sample from $\mu_{|\Lambda_i}$ in time $\mathrm{poly}(m)$.

Then, for all $\varepsilon \in (0, 1]$, we can ε-approximately sample from μ in time $\mathrm{poly}(m/\varepsilon)$. ◀

Additionally, as we discuss in Section 4.2, we use self-reducibility on the clique cover as well as Theorem 3 to obtain an ε-approximation algorithm for the partition function Z.

▶ **Theorem 4.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model with polymer clique cover Λ of size m such that $\mathcal{P}$ and Λ are computationally feasible as stated in Definition 12. Assume that $\mathcal{P}$ satisfies the conditions of Theorem 3. For all $\varepsilon \in (0, 1]$, there is a randomized ε-approximation of Z computable in time $\mathrm{poly}(m/\varepsilon)$. ◀

Since it is common for spin systems on graphs with n vertices to translate into polymer models with a clique cover of n cliques, Theorems 3 and 4 imply polynomial-time algorithms for their respective problems. Assumption (a) allows for a broad range in the choice of f from Condition 1, and assumption (b) is commonly satisfied by choosing the parameters such that they satisfy assumption (a). However, when considering spin systems previously studied in the literature, assumption (c) is not straightforward to satisfy, as the sizes of the cliques are commonly exponential in $n = |V(G)|$. We introduce the clique truncation condition as a way to deal with this requirement in a general setting.

▶ **Condition 5 (Clique Truncation).** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, and let $|\cdot|$ be a size function for $\mathcal{P}$ as in Definition 15. For all $i \in [m]$,

we say that Λ_i *satisfies the clique truncation condition* for a monotonically increasing, invertible function $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ and a bound $B \in \mathbb{R}_{>0}$ if and only if

$$\sum\nolimits_{\gamma \in \Lambda_i} g(|\gamma|) w_\gamma \leq B. \qquad \blacktriangleleft$$

We show that the clique truncation condition allows to reduce the size of each clique to a polynomial in n by removing low-weight polymers from the polymer model.[2] More precisely, Corollary 19 states that, for an ε-approximation, it is sufficient to consider only polymers γ with $|\gamma| \leq g^{-1}(Bm/\varepsilon)$. This allows us to use the algorithm of Patel and Regts [31] to sample from the Gibbs distribution of each clique by enumerating all its polymers. Table 1 illustrates the improved conditions for approximation for different algorithmic problems that can be obtained from combining the clique truncation condition with Theorem 4.

Problem	**Previous range**	**New range**
Hard-core model on bipartite α-expanders	$\lambda > (\mathrm{e}^2\Delta^2)^{\frac{1}{\alpha}}$ [22]	$\lambda \geq \left(\frac{\mathrm{e}}{0.8}\Delta^2\right)^{\frac{1}{\alpha}}$
q-state Potts model on α-expanders	$\beta > \frac{9/4+\ln(\Delta q)}{\alpha}$ [22]	$\beta \geq \frac{3/2+\ln(\Delta q)}{\alpha}$
Hard-core model on unbalanced bipartite graphs	$6\Delta_{\mathrm{L}}\Delta_{\mathrm{R}}\lambda_{\mathrm{R}} \leq (1+\lambda_{\mathrm{L}})^{\frac{\delta_{\mathrm{R}}}{\Delta_{\mathrm{L}}}}$ [6]	$3.3353\Delta_{\mathrm{L}}\Delta_{\mathrm{R}}\lambda_{\mathrm{R}} \leq (1+\lambda_{\mathrm{L}})^{\frac{\delta_{\mathrm{R}}}{\Delta_{\mathrm{L}}}}$
Perfect matching polynomial	$z \leq \left(\sqrt{4.8572\,(\Delta-1)}\right)^{-1}$ [7]	$z \leq \left(\sqrt{2.8399(\Delta-1)}\right)^{-1}$

Table 1: Improvement on the parameter ranges of our technique for problems with known approximation algorithms. Note that for a fair comparison, we refined the calculations of the bounds in [22] in a similar fashion as in Section 6.

Interplay of Conditions So far, we introduced two major conditions for approximating the partition function of abstract polymer models, namely the clique dynamics condition and the clique truncation condition. We briefly want to discuss their interplay.

In Section 5.1 we investigate to what extent an additional condition, besides being computationally feasible (Definition 12) and satisfying the clique dynamics condition, is required to ensure that the partition function of a polymer model can be approximated efficiently. To this end, we show that, unless NP = RP, those two conditions are not sufficient to ensure the existence of an algorithm that takes time polynomial in the size of the used clique cover. The intuition is that, even if rapid mixing of the clique dynamics is guaranteed, it might be computationally hard to sample efficiently from each polymer clique without additional assumptions. The clique truncation

[2] A similar idea was used for the hard-core model on bipartite expanders in the first arXiv version of [8].

condition presents a general way to deal with this problem. However, for specific applications of polymer models, it might be replaced with a different approach using more structural insights.

Besides this theoretical result, it should be noted that in all calculations we did to obtain Table 1, the parameter regime was mainly determined by the clique dynamics condition. More precisely, choosing the parameters such that the clique dynamics condition is satisfied always ensured that the clique truncation condition is also satisfied. Thus, from a practical point of view, the clique truncation condition usually does not impose any additional restriction. Note that for none of the problems in the table any computational hardness results are known. This means that there is currently no way to argue that any of the given bounds is tight, even in terms of the asymptotic dependency on problem specific parameters, such as the maximum degree Δ or the expansion parameter α. However, in all cases, our approach improves the previously known algorithmic results by at least a constant factor.

1.3 Recent Results, Other Applications, and Outlook

In this paper, we introduced clique dynamics and demonstrated how they can be applied to overcome the exponential blow-up resulting from encoding spin systems in polymer models. An interesting question for future work is whether there are other applications where clique dynamics can be used in a similar manner. In a subsequent work, we improve the parameter regime for efficient approximation of the partition function of the continuous hard-sphere model, using clique dynamics to handle the super-exponential sized graph that results from the applied discretization scheme [17]. This raises the question if clique dynamics can be applied similarly to other continuous models from statistical physics to obtain approximation results based on discretization.

Another direction for future research is to further improve the condition for rapid mixing of clique dynamics, which is crucial for the efficient algorithmic application. In [17], we prove, using a recent spectral independence method [1, 9], a version of Condition 1 that involves strict inequality to guarantee rapid mixing of clique dynamics. Note that such a result could also be obtained by directly using the path coupling method of Bubley and Dyer [5]. However, as discussed before, such a strict condition is not comparable to the best known conditions for convergence of the cluster expansion, such as the Fernández–Procacci condition. Furthermore, we prove in [17] that, for the case that all polymers have the same weight (which corresponds to a hard-core model on the polymer graph), clique dynamics are actually rapidly mixing up to the tree threshold of the polymer graph, which is in general considered a tight bound for local Markov chains [29]. Thus, in the univariate case (and in contrast to the multivariate case), the result based on spectral independence is strictly better than Condition 1. It would be interesting to get a more theoretical comparison between the power of both techniques in terms of rapid mixing results. Especially, it raises the question if a more sophisticated application of the spectral independence method is able to prove equivalent or even better conditions for rapid mixing of clique dynamics when weights are not uniform. This would directly lead to further improvements of the results on polymer models, presented in this work.

Finally, in a more recent work Blanca et al. [2] introduced a near-linear time perfect sampling algorithm for polymer models that are based on vertex spin systems. Their algorithm achieves similar parameter regimes as our approach by combining a coupling from the past argument with a new percolation based method for sampling polymers that contain a fixed vertex. Note that the latter could also be incorporated into our Markov chain as an efficient sampling method for the polymer clique distribution of such vertex spin polymer models, which would improve the running time that is given in our paper for several applications.

2 Preliminaries

We denote the set of all natural numbers, including 0, by $\mathbb{N}$ and the set of all real numbers by $\mathbb{R}$. For all $n \in \mathbb{N}$, let $[n] = [1, n] \cap \mathbb{N}$. If the polymer model $\mathcal{P}$ is clear from context, we may drop the index and write $\mathcal{F}$, Z, and μ instead of $\mathcal{F}^{(\mathcal{P})}$, $Z(\mathcal{P})$, and $\mu^{(\mathcal{P})}$, respectively.

We use the following formal notion of approximate sampling. Let ν be a probability distribution on a countable state space Ω. For $\varepsilon \in (0, 1]$, we say that a distribution ξ on Ω is an *ε-approximation of ν* if and only if $d_{\mathrm{TV}}(\nu, \xi) \leq \varepsilon$, where $d_{\mathrm{TV}}(\cdot, \cdot)$ denotes the total-variation distance. Further, we say that we can *ε-approximately sample from ν* if and only if we can sample from some distribution ξ such that ξ is an ε-approximation of ν.

For $x \in \mathbb{R}_{>0}$ and $\varepsilon \in (0, 1]$, we call a random variable X a *randomized ε-approximation for x* if and only if $\Pr[(1-\varepsilon)x \leq X \leq (1+\varepsilon)x] \geq 3/4$. Note that if x is the output to an algorithmic problem on an instance and independent samples of X can be obtained in polynomial time in the instance size and $1/\varepsilon$, this translates to the definition of an FPRAS.

2.1 Restricted Polymer Models

We base the transitions of our Markov chain for a polymer model $(C, w, \nsim)$ on restricted sets $\mathcal{B} \subseteq C$. We define the set of all polymer families *restricted to* $\mathcal{B}$ to be $\mathcal{F}_{|\mathcal{B}} = \mathcal{F} \cap 2^{\mathcal{B}}$. Further, we define the restricted partition function $Z_{|\mathcal{B}}$ to be equation (1) but with $\mathcal{F}^{(\mathcal{P})}$ replaced by $\mathcal{F}_{|\mathcal{B}}$. Similarly, we define the restricted Gibbs distribution $\mu_{|\mathcal{B}}$ to be a probability distribution over $\mathcal{F}_{|\mathcal{B}}$, i.e., equation (2) but with $Z(\mathcal{P})$ replaced by $Z_{|\mathcal{B}}$. Our restrictions are special sets of polymers, which we define next.

By definition, for a polymer model, a polymer family Γ cannot contain incompatible polymers. Thus, when considering a subset $\mathcal{B} \subseteq C$ where all polymers are pairwise incompatible, at most one polymer of $\mathcal{B}$ is in Γ. We call such a subset $\mathcal{B}$ a *polymer clique.*

Last, for an $m \in \mathbb{N}_{>0}$, we call a set $\Lambda = \{\Lambda_i\}_{i \in [m]}$ of polymer cliques a *polymer clique cover* if and only if $\bigcup \Lambda = C$, and we call m the *size* of Λ. Note that the elements of Λ need not be pairwise disjoint. Further note that, for each $i \in [m]$, we have

$$Z_{|\Lambda_i} = \sum_{\Gamma \in \mathcal{F}_{|\Lambda_i}} \prod_{\gamma \in \Gamma} w_\gamma = 1 + \sum_{\gamma \in \Lambda_i} w_\gamma,$$

as the polymers of Λ_i are pairwise incompatible and thus each non-empty family of Λ_i contains a single polymer. Similarly, the Gibbs distribution restricted to Λ_i simplifies to $\mu_{|\Lambda_i}(\emptyset) = 1/Z_{|\Lambda_i} = 1/(1 + \sum_{\gamma \in \Lambda_i} w_\gamma)$ and, for each $\gamma' \in \Lambda_i$, to $\mu_{|\Lambda_i}(\{\gamma'\}) = w_{\gamma'}/Z_{|\Lambda_i} = w_{\gamma'}/(1 + \sum_{\gamma \in \Lambda_i} w_\gamma)$.

2.2 Markov Chains

For a Markov chain $\mathcal{M}$ with a unique stationary distribution D and an $\varepsilon \in (0, 1]$, let $\tau_{\mathcal{M}}(\varepsilon)$ denote the mixing time of $\mathcal{M}$ (with error ε). That is, $\tau_{\mathcal{M}}(\varepsilon)$ denotes the first point in time $t \in \mathbb{N}$ such that, for every initial state, the total-variation distance between D and the distribution of $\mathcal{M}$ at time t is at most ε.

In order to bound mixing times, we use a theorem by Greenberg et al. [19, Theorem 3.4]. We state it in a form that better suits our purpose, providing a full proof in Appendix A.

▶ **Theorem 6.** Let $\mathcal{M}$ be an ergodic Markov chain with state space Ω and with transition matrix P such that, for all $x \in \Omega$, it holds that $P(x, x) > 0$. For $d, D \in \mathbb{R}_{>0}$, $d \leq D$, let $\delta \colon \Omega^2 \to \{0\} \cup [d, D]$ be such that $\delta(x, y) = 0$ if and only if $x = y$. Assume that there is a coupling between the transitions of two copies $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$ of $\mathcal{M}$ such that, for all $t \in \mathbb{N}$ and all $x, y \in \Omega$, it holds that

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \leq \delta(x, y). \tag{3}$$

Furthermore, assume that there are $\kappa, \eta \in (0, 1)$ such that, for the same coupling and all $t \in \mathbb{N}$ and all $x, y \in \Omega$ with $x \neq y$, it holds that

$$\Pr[|\delta(X_{t+1}, Y_{t+1}) - \delta(x, y)| \geq \eta\delta(x, y) \mid X_t = x, Y_t = y] \geq \kappa. \tag{4}$$

Then, for all $\varepsilon \in (0, 1]$, it holds that

$$\tau_{\mathcal{M}}(\varepsilon) \leq \frac{\left(\ln(D/d) + 2\ln(2)\right)^2}{\ln(1+\eta)^2\kappa} \ln\left(\frac{1}{\varepsilon}\right).$$

If $\ln(D/d) \in \Omega(1)$, then this bound simplifies to

$$\tau_{\mathcal{M}}(\varepsilon) \in \mathrm{O}\left(\frac{\ln(D/d)^2}{\ln(1+\eta)^2\kappa} \ln\left(\frac{1}{\varepsilon}\right)\right). \qquad \blacktriangleleft$$

3 Polymer Dynamics

We analyze the following Markov chain for a polymer model with a polymer clique cover.

▶ **Definition 7 (Polymer Clique Dynamics).** Let $\mathcal{P}$ be a polymer model, and let Λ be a polymer clique cover of $\mathcal{P}$ with size m. We define $\mathcal{M}(\mathcal{P})$ to be a Markov chain with state space $\mathcal{F}$. Let $(X_t)_{t \in \mathbb{N}}$ denote a (random) sequence of states of $\mathcal{M}(\mathcal{P})$, where X_0 is arbitrary.

Then, for all $t \in \mathbb{N}$, the transitions of $\mathcal{M}(\mathcal{P})$ are as follows:

choose $i \in [m]$ uniformly at random;
choose $\Gamma \in \mathcal{F}_{|\Lambda_i}$ according to $\mu_{|\Lambda_i}$;
if $\Gamma = \emptyset$ **then** $X_{t+1} = X_t \setminus \Lambda_i$;
else if $X_t \cup \Gamma$ is a valid polymer family **then** $X_{t+1} = X_t \cup \Gamma$;
else $X_{t+1} = X_t$; ◀

Given a polymer model $\mathcal{P} = (C, w, \nsim)$ and a polymer Markov chain $\mathcal{M}(\mathcal{P})$, let P denote the transition matrix of $\mathcal{M}(\mathcal{P})$, that is, for all $\Gamma, \Gamma' \in \mathcal{F}^{(\mathcal{P})}$, the entry $P(\Gamma, \Gamma')$ denotes the probability to transition from state Γ to state Γ' in a single step. Note that P is time-homogeneous and that, for all $\Gamma, \Gamma' \in \mathcal{F}^{(\mathcal{P})}$ with $P(\Gamma, \Gamma') > 0$, the symmetric difference of Γ and Γ' has a cardinality of at most 1, since the polymer families of a polymer clique are all singletons. Further note that $\mathcal{M}(\mathcal{P})$ has a positive self-loop probability, as the polymers from a polymer clique are pairwise incompatible.

The transition probabilities of two neighboring states of $\mathcal{M}(\mathcal{P})$ follow a simple pattern. In order to ease notation, for all $\gamma \in C$, let $z_\gamma = \sum_{i \in [m]\colon \gamma \in \Lambda_i} 1/Z_{|\Lambda_i}$. For all $\Gamma, \Gamma' \in \mathcal{F}^{(\mathcal{P})}$ such that there is a $\gamma \in C$, $\gamma \notin \Gamma$ with $\Gamma' = \Gamma \cup \{\gamma\}$, it holds that

$$P(\Gamma, \Gamma') = \frac{1}{m} \sum_{\substack{i \in [m]\colon \\ \gamma \in \Lambda_i}} \mu_{|\Lambda_i}(\{\gamma\}) = \frac{1}{m} \sum_{\substack{i \in [m]\colon \\ \gamma \in \Lambda_i}} \frac{w_\gamma}{Z_{|\Lambda_i}} = w_\gamma \frac{z_\gamma}{m} > 0 \quad \text{and that} \tag{5}$$

$$P(\Gamma', \Gamma) = \frac{1}{m} \sum_{\substack{i \in [m]\colon \\ \gamma \in \Lambda_i}} \mu_{|\Lambda_i}(\{\emptyset\}) = \frac{1}{m} \sum_{\substack{i \in [m]\colon \\ \gamma \in \Lambda_i}} \frac{1}{Z_{|\Lambda_i}} = \frac{z_\gamma}{m} > 0.$$

The polymer clique dynamics is suitable for sampling from the Gibbs distribution of a polymer model, since the limit distribution of the Markov chain converges to μ.

▶ **Lemma 8.** Let $\mathcal{P}$ be a polymer model. The polymer Markov chain $\mathcal{M}(\mathcal{P})$ is ergodic with stationary distribution $\mu^{(\mathcal{P})}$. ◀

Proof. First, note that $\mathcal{M}(\mathcal{P})$ is irreducible, as there is a positive probability to go from any polymer family $\Gamma \in \mathcal{F}$ to the empty polymer family $\emptyset$ in a finite number of steps by consecutively removing each polymer $\gamma \in \Gamma$. Similarly, there is a positive probability to go from $\emptyset$ to any polymer family $\Gamma' \in \mathcal{F}$ in a finite number of steps by consecutively adding all polymers $\gamma' \in \Gamma'$.

We proceed by proving that $\mu^{(\mathcal{P})}$, which we abbreviate as μ, is a stationary distribution of $\mathcal{M}(\mathcal{P})$. To this end, we show that $\mathcal{M}(\mathcal{P})$ satisfies the detailed-balance condition with respect to μ. That is, for all $\Gamma, \Gamma' \in \mathcal{F}$, it holds that

$$\mu(\Gamma) \cdot P(\Gamma, \Gamma') = \mu(\Gamma') \cdot P(\Gamma', \Gamma). \tag{6}$$

Note that it is sufficient to check equation (6) for all pairs of states with a symmetric difference of exactly one polymer.

Let $\Gamma, \Gamma' \in \mathcal{F}$ and assume without loss of generality that $\Gamma' = \Gamma \cup \{\gamma\}$ for some polymer $\gamma \notin \Gamma$. Note that, by equation (5), $P(\Gamma, \Gamma') = w_\gamma \cdot P(\Gamma', \Gamma)$. Further, by definition of the Gibbs distribution, we have $\mu(\Gamma') = w_\gamma \cdot \mu(\Gamma)$. Thus, we get

$$\mu(\Gamma) \cdot P(\Gamma, \Gamma') = \mu(\Gamma) \cdot w_\gamma \cdot P(\Gamma', \Gamma) = \mu(\Gamma') \cdot P(\Gamma', \Gamma),$$

which shows that μ is a stationary distribution of $\mathcal{M}(\mathcal{P})$.

Finally, we argue that $\mathcal{M}(\mathcal{P})$ is ergodic. Note that an irreducible Markov chain has a stationary distribution if and only if it is positive recurrent. In addition, every state of $\mathcal{M}(\mathcal{P})$ has a positive self-loop probability, which implies that the chain is aperiodic. This shows that $\mathcal{M}(\mathcal{P})$ is ergodic and concludes the proof. ■

Recall Condition 1 (clique dynamics) from the introduction. Assuming that the condition holds, we obtain the following bound on the mixing time of $\mathcal{M}(\mathcal{P})$.

▶ **Lemma 9.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model satisfying Condition 1 with function f, and let Λ be a polymer clique cover of $\mathcal{P}$ with size m. Then, for all $\varepsilon \in (0, 1]$, it holds that

$$\tau_{\mathcal{M}(\mathcal{P})}(\varepsilon) \in \mathrm{O}\left(\frac{m^3}{\min_{\gamma \in C}\{z_\gamma\}} \ln\left(m \frac{\max_{\gamma \in C}\left\{\frac{f(\gamma)}{z_\gamma(1+w_\gamma)}\right\}}{\min_{\gamma \in C}\left\{\frac{f(\gamma)}{z_\gamma(1+w_\gamma)}\right\}} \right)^2 \ln\left(\frac{1}{\varepsilon}\right) \right). \quad ◀$$

Proof. We aim to apply Theorem 6, which requires us to define a potential δ. We do so by utilizing the function $\delta' : C \to \mathbb{R}_{>0}$ with $\gamma \mapsto f(\gamma)/\big(z_\gamma(1+w_\gamma)\big)$. Let $\oplus$ denote the symmetric set difference. For all $\Gamma, \Gamma' \in \mathcal{F}$, we define

$$\delta(\Gamma, \Gamma') = \sum_{\gamma \in \Gamma \oplus \Gamma'} \delta'(\gamma).$$

Note that $\delta(\Gamma, \Gamma')$ only depends on the symmetric difference of Γ and Γ' and that $\delta(\Gamma, \Gamma') = 0$ if and only if $\Gamma \oplus \Gamma' = \emptyset$, which is the case only when $\Gamma = \Gamma'$.

We continue by constructing a coupling between two copies of $\mathcal{M}(\mathcal{P})$, namely between $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$. We couple these chains such that, for each transition,

- both choose the same index $i \in [m]$ and
- both draw the same polymer family $\Gamma \in \mathcal{F}_{|\Lambda_i}$ from $\mu_{|\Lambda_i}$.

This constitutes a valid coupling, as each chain transitions according to its desired marginal transition probabilities.

We now show for all $t \in \mathbb{N}$ and $\Gamma, \Gamma' \in \mathcal{F}$ that

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

Note that this trivially holds if $\Gamma = \Gamma'$, as the chains X and Y behave identically from then on. Thus, we are left with the case that $\Gamma \neq \Gamma'$, which implies that $|\Gamma \oplus \Gamma'| \geq 1$.

We introduce the following notation. For all $\gamma \in C$, let $N(\gamma) = \{\gamma' \in C \mid \gamma' \nsim \gamma, \gamma' \neq \gamma\}$ denote the neighborhood of γ. We extend this definition to arbitrary subsets of polymers $\mathcal{B} \subseteq C$ by $N(\mathcal{B}) = \bigcup_{\gamma \in \mathcal{B}} N(\gamma)$ (note that $N(\mathcal{B})$ might contain polymers from $\mathcal{B}$).

Let $\Delta = \Gamma \oplus \Gamma'$, and let $\gamma \in \Delta$. Assume without loss of generality that $\gamma \in \Gamma$. By equation (5), with probability z_γ/m, the chain X removes γ, causing $\delta(X_{t+1}, Y_{t+1})$ to decrease by at least[3] $\delta'(\gamma)$. Further, if $\gamma \in \Delta \setminus N(\Delta)$, with probability $w_\gamma z_\gamma/m$, the chain Y adds γ and the chain X remains in its state. Again, $\delta(X_{t+1}, Y_{t+1})$ decreases by $\delta'(\gamma)$. Note that the case $\gamma \in \Gamma'$ is exactly symmetric with X and Y swapped.

Let $\delta^-(\Gamma, \Gamma')$ denote the expected (conditional) decrease of δ. By the observations above, we see that

$$\delta^-(\Gamma, \Gamma') = \sum_{\gamma \in \Delta} \delta'(\gamma) \frac{z_\gamma}{m} + \sum_{\gamma \in \Delta \setminus N(\Delta)} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m} = \sum_{\gamma \in \Delta} \delta'(\gamma) \frac{z_\gamma}{m} (1 + w_\gamma) - \sum_{\gamma \in \Delta \cap N(\Delta)} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m}.$$

Moreover, δ increases whenever a polymer γ is added to only one of the chains. This only occurs if $\gamma \in N(\Delta) \setminus \Delta$ and has probability $w_\gamma z_\gamma/m$ for each such polymer. Similarly to the expected decrease, we denote the expected increase by $\delta^+(\Gamma, \Gamma')$. We bound

$$\begin{aligned}
\delta^+(\Gamma, \Gamma') &\leq \sum_{\gamma \in N(\Delta) \setminus \Delta} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m} \\
&= \sum_{\gamma \in N(\Delta)} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m} - \sum_{\gamma \in \Delta \cap N(\Delta)} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m} \\
&\leq \sum_{\gamma \in \Delta} \sum_{\gamma' \in N(\gamma)} \delta'(\gamma') w_{\gamma'} \frac{z_{\gamma'}}{m} - \sum_{\gamma \in \Delta \cap N(\Delta)} \delta'(\gamma) w_\gamma \frac{z_\gamma}{m},
\end{aligned}$$

where the first inequality comes from the fact that there might be polymers $\gamma \in N(\Delta) \setminus \Delta$ that cannot be added by any of both chains.

Together, we obtain

$$\begin{aligned}
&\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = \Gamma, Y_t = \Gamma'] = \delta(\Gamma, \Gamma') + \delta^+(\Gamma, \Gamma') - \delta^-(\Gamma, \Gamma') \\
&\quad \leq \delta(\Gamma, \Gamma') + \sum_{\gamma \in \Delta} \sum_{\gamma' \in N(\gamma)} \delta'(\gamma') w_{\gamma'} \frac{z_{\gamma'}}{m} - \sum_{\gamma \in \Delta} \delta'(\gamma) \frac{z_\gamma}{m} (1 + w_\gamma) \\
&\quad = \delta(\Gamma, \Gamma') + \sum_{\gamma \in \Delta} \left(\sum_{\gamma' \in N(\gamma)} \delta'(\gamma') w_{\gamma'} \frac{z_{\gamma'}}{m} - \delta'(\gamma) \frac{z_\gamma}{m} (1 + w_\gamma) \right).
\end{aligned}$$

We proceed by showing that, for each $\gamma \in \Delta$, the respective summand in the sum above is at

[3] A larger decrease is possible if Y also removes some polymer $\gamma' \in \Delta$. However, we only need a crude lower bound.

most zero. By the definition of δ', we get

$$\sum_{\gamma' \in N(\gamma)} \delta'(\gamma') w_{\gamma'} \frac{z_{\gamma'}}{m} - \delta'(\gamma) \frac{z_\gamma}{m} (1 + w_\gamma) = \frac{1}{m} \left(\sum_{\gamma' \in N(\gamma)} f(\gamma') \frac{w_{\gamma'}}{1 + w_{\gamma'}} - f(\gamma) \right).$$

By the definition of $N(\gamma)$ and since $\mathcal{P}$ satisfies the clique dynamics condition, we bound

$$\frac{1}{m} \left(\sum_{\gamma' \in N(\gamma)} f(\gamma') \frac{w_{\gamma'}}{1 + w_{\gamma'}} - f(\gamma) \right) = \frac{1}{m} \left(\sum_{\substack{\gamma' \in C \colon \gamma' \nsim \gamma \\ \gamma' \neq \gamma}} f(\gamma') \frac{w_{\gamma'}}{1 + w_{\gamma'}} - f(\gamma) \right) \leq 0.$$

Consequently, we get that

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

We now show that there are values $\eta, \kappa \in (0, 1)$ such that, for all $t \in \mathbb{N}$ and all $\Gamma, \Gamma' \in \mathcal{F}$ with $\Gamma \neq \Gamma'$, it holds that

$$\Pr\left[|\delta(X_{t+1}, Y_{t+1}) - \delta(\Gamma, \Gamma')| \geq \eta \delta(\Gamma, \Gamma') \,\middle|\, X_t = \Gamma, Y_t = \Gamma' \right] \geq \kappa. \tag{7}$$

Note that every polymer family in $\mathcal{F}$ has at most m polymers because it can have at most one polymer from each polymer clique. Thus, for $\Delta = \Gamma \oplus \Gamma'$, we bound $|\Delta| \leq 2m$. Consequently, there is at least one polymer $\gamma \in \Delta$ such that $\delta'(\gamma) \geq \delta(\Gamma, \Gamma')/(2m)$. Assume without loss of generality that $\gamma \in \Gamma$. With probability z_γ/m, chain X deletes γ, resulting in δ decreasing by at least $|\delta(X_{t+1}, Y_{t+1}) - \delta(\Gamma, \Gamma')| \geq \delta'(\gamma) \geq \delta(\Gamma, \Gamma')/(2m)$. Thus, equation (7) is true for $\eta = 1/(2m)$ and $\kappa = z_\gamma/m \geq \left(\min_{\gamma \in C}\{z_\gamma\}\right)/m$.

It remains is to determine $d, D \in \mathbb{R}_{>0}$ such that, for all $\Gamma, \Gamma' \in \mathcal{F}$ with $\Gamma \neq \Gamma'$, it holds that $\delta(\Gamma, \Gamma') \in [d, D]$. Let $\Delta = \Gamma \oplus \Gamma'$, noting again that $|\Delta| \leq 2m$. We choose

$$d \geq \min_{\gamma \in C}\{\delta'(\gamma)\} = \min_{\gamma \in C}\left\{ \frac{f(\gamma)}{z_\gamma(1 + w_\gamma)} \right\} \text{ and}$$
$$D \leq 2m \max_{\gamma \in C}\{\delta'(\gamma)\} = 2m \max_{\gamma \in C}\left\{ \frac{f(\gamma)}{z_\gamma(1 + w_\gamma)} \right\}.$$

Applying Theorem 6 and observing that $\ln\left(1 + \frac{1}{2m}\right)^2 \geq \frac{1}{4m^2}$ concludes the proof. ∎

By combining Lemmas 8 and 9 and observing that $1/Z_{\max} \leq z_\gamma \leq m$ and $1 \leq 1 + w_\gamma \leq Z_{\max}$ for all polymers $\gamma \in C$, we obtain Theorem 2.

3.1 Comparison to Conditions for Cluster Expansion

In order to set our clique dynamics condition in the context of existing conditions for absolute convergence of the cluster expansion, we compare it to the condition of Fernández and Procacci [15]. We choose it because it is, to the best of our knowledge, the least restrictive condition for absolute convergence of the cluster expansion of abstract polymer models. As Fernández and Procacci show, their condition is an improvement over other known conditions, including the Dobrushin condition [11] and the Kotecký–Preiss condition [26].

▶ **Definition 10 (From [15]).** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, and let $N(\gamma) = \{\gamma' \in C \mid \gamma' \nsim \gamma\}$. We say that $\mathcal{P}$ *satisfies the Fernández–Procacci condition* if and only if there is a function $f : C \to \mathbb{R}_{>0}$ such that, for all $\gamma \in C$, it holds that

$$\sum_{\Gamma \in \mathcal{F}^{(\mathcal{P})}_{|N(\gamma)}} \prod_{\gamma' \in \Gamma} f(\gamma') w_{\gamma'} \leq f(\gamma). \qquad \blacktriangleleft$$

Note that we state the condition slightly differently from the version of the original authors to ease comparison. The original form is recovered by setting $f : \gamma \mapsto \hat{f}(\gamma)/w_\gamma$ for some function $\hat{f} : C \to \mathbb{R}_{>0}$. Further, the original version allows f (or $\hat{f}$ respectively) to take the value 0. However, note that if $f(\gamma) = 0$ for any $\gamma \in C$, then the condition is trivially void because $\emptyset \in \mathcal{F}^{(\mathcal{P})}_{|N(\gamma)}$, which lower bounds the left hand side of the inequality by 1.

The following statement shows how our clique dynamics condition relates to the Fernández–Procacci condition as given in Definition 10.

▶ **Proposition 11.** If a polymer model $\mathcal{P} = (C, w, \nsim)$ satisfies the Fernández–Procacci condition for a function f, then it also satisfies the clique dynamics condition for the same function. ◀

Proof. Instead of directly working with the clique dynamics condition, we show that the Fernández–Procacci condition implies that

$$\sum_{\gamma' \in C \colon \gamma' \nsim \gamma} f(\gamma') w_{\gamma'} \leq f(\gamma),$$

for all polymers $\gamma \in C$. By $w_\gamma > \frac{w_\gamma}{1+w_\gamma}$, we can immediately conclude that clique dynamics condition is satisfied as well.

Note that $\emptyset \in \mathcal{F}^{(\mathcal{P})}_{|N(\gamma)}$ and, for all $\gamma' \in C$ with $\gamma' \nsim \gamma$, it holds that $\{\gamma'\} \in \mathcal{F}^{(\mathcal{P})}_{|N(\gamma)}$. Thus,

$$\sum_{\gamma' \in C \colon \gamma' \nsim \gamma} f(\gamma') w_{\gamma'} < 1 + \sum_{\gamma' \in C \colon \gamma' \nsim \gamma} f(\gamma') w_{\gamma'} \leq \sum_{\Gamma \in \mathcal{F}^{(\mathcal{P})}_{|N(\gamma)}} \prod_{\gamma' \in \Gamma} f(\gamma') w_{\gamma'} \leq f(\gamma),$$

which proves the claim. ■

Note that Proposition 11 implies that if a polymer model satisfies the Fernández–Procacci condition, then Theorem 2 bounds the mixing time of the polymer Markov chain for any given

clique cover. Further, Proposition 11 and its implied mixing time bounds for the polymer Markov chain carry over to all convergence conditions that are more restrictive than the Fernández–Procacci condition, such as the Dobrushin and the Kotecký–Preiss condition. We emphasize that, even though the Proposition 11 in fact proves that the Fernández–Procacci condition implies the clique dynamics condition with a strict inequality, it does not necessarily lead to an improved mixing time bound via a path coupling argument. The reason is that the actual contraction along a path coupling might not be sufficiently strong.

4 Algorithmic Results

We now discuss how the polymer Markov chain $\mathcal{M}$ of a polymer model $\mathcal{P}$ with a clique cover of size m is used to approximate $Z(\mathcal{P})$ in a randomized fashion. To this end, $\mathcal{M}$ is turned into an approximate sampler for $\mathcal{P}$ (Theorem 3). Then this sampler is applied in an algorithmic framework (Algorithm 1) that yields an ε-approximation of $Z(\mathcal{P})$ (Theorem 4). Under certain assumptions, such as that the restricted partition function of each polymer clique is in $\mathrm{poly}(m)$, the approximation is computable in time $\mathrm{poly}(m/\varepsilon)$.

In order to discuss the computation time of operations on a polymer model rigorously, we need to make assumptions about the operations we consider and their computational cost.

▶ **Definition 12 (Computationally Feasible).** We say that a polymer model $\mathcal{P} = (C, w, \nsim)$ with a polymer clique cover Λ of size m is *computationally feasible* if and only if all of the following operations can be performed in time $\mathrm{poly}(m)$:

(1) for all $i \in [m]$ and all $\gamma \in C$, we can check whether $\gamma \in \Lambda_i$,

(2) for all $\gamma, \gamma' \in C$, we can check whether $\gamma \nsim \gamma'$,

(3) for all $\gamma \in C$, we can compute w_γ. ◀

In addition to the operations above, we further assume for all $\gamma \in C$ and all $\Gamma \in \mathcal{F}$ that we can compute $\Gamma \setminus \{\gamma\}$ and $\Gamma \cup \{\gamma\}$ and that we can decide whether $\Gamma = \emptyset$ in time $\mathrm{poly}(m)$.

4.1 Sampling From the Gibbs Distribution

We discuss under what assumptions one can approximately sample from the Gibbs distribution of a computationally feasible polymer model in time polynomial in the size of the clique cover. Our main result is the following theorem.

▶ **Theorem 3.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model with polymer clique cover Λ of size m, assume $\mathcal{P}$ and Λ are computationally feasible as stated in Definition 12, and let $Z_{\max} = \max_{i \in [m]} \{Z_{|\Lambda_i}\}$. Further assume that

(a) $\mathcal{P}$ satisfies the clique dynamics condition for a function f such that, for all $\gamma \in C$, it holds that $\mathrm{e}^{-\mathrm{poly}(m)} \leq f(\gamma) \leq \mathrm{e}^{\mathrm{poly}(m)}$,

(b) $Z_{\max} \in \text{poly}(m)$, and that,

(c) for all $i \in [m]$, we can sample from $\mu_{|\Lambda_i}$ in time $\text{poly}(m)$.

Then, for all $\varepsilon \in (0, 1]$, we can ε-approximately sample from μ in time $\text{poly}(m/\varepsilon)$. ◀

Proof. In order to sample from μ, we utilize the polymer Markov chain $\mathcal{M}(\mathcal{P})$ based on Λ. By Theorem 2, it holds that

$$\tau_{\mathcal{M}(\mathcal{P})}(\varepsilon) \in \mathrm{O}\left(m^3 Z_{\max} \ln\left(m^2 Z_{\max}^2 \frac{f_{\max}}{f_{\min}}\right)^2 \ln\left(\frac{1}{\varepsilon}\right)\right).$$

Due to assumptions (a) and (b), it holds that $\tau_{\mathcal{M}(\mathcal{P})}(\varepsilon) \in \text{poly}(m/\varepsilon)$. It remains to show that each step of $\mathcal{M}(\mathcal{P})$, as laid out in Definition 7, can be computed in time $\text{poly}(m)$. To this end, let X_t denote the current state of $\text{poly}(m)$.

We first sample $i \in [m]$ uniformly at random. Because of assumption (c), we can then sample $\Gamma \in \mathcal{F}_{|\Lambda_i}$ according to $\mu_{|\Lambda_i}$ in time $\text{poly}(m)$. This covers lines 1 and 2.

Regarding Line 3, note that we can check whether $\Gamma = \emptyset$ in time $\text{poly}(m)$. Assume that $\Gamma = \emptyset$, and note that $|X_t| \leq m \in \text{poly}(m)$, as X_t contains at most one polymer per polymer clique. In order to compute $X_t \setminus \Lambda_i$, it suffices to iterate over every $\gamma \in \Gamma$ and check if $\gamma \in \Lambda_i$, which can be done in time $\text{poly}(m)$, by assumption (1). Once we found $\gamma \in \Lambda_i$, we remove it in time $\text{poly}(m)$.

Regarding Line 4, assume now that $\Gamma = \{\gamma\}$ for some $\gamma \in \Lambda_i$. In order to decide if $X_t \cup \Gamma$ is a valid polymer family, it is sufficient to iterate over all $\gamma' \in X_t$ and check whether any of them is incompatible to γ. By assumption (2) this can be done in time $\text{poly}(m)$, which concludes the proof. ■

By making a slightly stronger assumption about the polymer model, assumptions (a) and (b) of Theorem 3 are easily satisfied.

▶ **Observation 13.** If $\mathcal{P}$ satisfies, for all $\gamma \in C$,

$$\sum_{\gamma' \in C \,:\, \gamma' \nsim \gamma} f(\gamma') w_{\gamma'} \leq f(\gamma), \tag{8}$$

then the clique dynamics condition is satisfied for the same function f. Thus, if equation (8) holds for a function f with $\mathrm{e}^{-\text{poly}(m)} \leq f(\gamma) \leq \mathrm{e}^{\text{poly}(m)}$ for all $\gamma \in C$, assumption (a) also holds. Further, by setting γ to be the polymer in Λ_i that minimizes f, equation (8) implies that $Z_{|\Lambda_i} = 1 + \sum_{\gamma' \in \Lambda_i} w_{\gamma'} \leq 2$, meaning that assumption (b) is trivially satisfied. ◀

Although the condition above is slightly more restrictive than the clique dynamics condition, it is more convenient to use for algorithmic applications. It can be seen as a weaker and more general version of the mixing condition by Chen et al. [8].

4.2 Approximation of the Partition Function

By now, we mainly discussed conditions for approximately sampling from the Gibbs distribution. We now discuss how to turn this into a randomized approximation for the partition function. To this end, we apply self-reducibility [25]. However, note that the obvious way for applying self-reducibility, namely based on single polymers, might take $|C|$ reduction steps. This is not feasible in many algorithmic applications of polymer models.

To circumvent this problem, we propose a self-reducibility argument based on polymer cliques. By doing so, the number of reductions is bounded by the size of the clique cover that is used, thus adding no major overhead to the runtime of our proposed approximate sampling scheme. Besides this idea of applying self-reducibility based on cliques, most of our arguments are analogous to known applications, like in [24, Chapter 3].

We proceed by formalizing clique-based self-reducibility. Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, and let Λ be a polymer clique cover of $\mathcal{P}$ with size m. We define a sequence of subsets of polymers $(K_i)_{0 \leq i \leq m}$ with $K_0 = \emptyset$ and, for $i \in [m]$, with $K_i = K_{i-1} \cup \Lambda_i$.

Further, for all $i \in [m]$, let $\sigma_i = Z_{|K_{i-1}} / Z_{|K_i}$. Note that $Z_{|K_0} = 1$ and $Z_{|K_m} = Z$. Thus,

$$Z = \prod_{i \in [m]} \frac{Z_{|K_i}}{Z_{|K_{i-1}}} = \Big(\prod_{i \in [m]} \sigma_i \Big)^{-1}.$$

Hence, when approximating Z, it is sufficient to focus, for all $i \in [m]$, on approximating σ_i.

For all $i \in [m]$, a similar relation holds with respect to the probability that a random $\Gamma \in \mathcal{F}_{|K_i}$ is already in $\mathcal{F}_{|K_{i-1}}$. More formally, let $i \in [m]$, and let $\Gamma \sim \mu_{|K_i}$. Note that

$$\mathrm{E}\big[\mathbb{1}\{\Gamma \in \mathcal{F}_{|K_{i-1}}\}\big] = \sum_{\Gamma \in \mathcal{F}_{|K_i}} \mu_{|K_i}(\Gamma) \cdot \mathbb{1}\{\Gamma \in \mathcal{F}_{|K_{i-1}}\} = \sum_{\Gamma \in \mathcal{F}_{|K_{i-1}}} \mu_{|K_i}(\Gamma) = \frac{Z_{|K_{i-1}}}{Z_{|K_i}} = \sigma_i. \tag{9}$$

We use these observations in order to obtain a randomized approximation of Z, as given in Algorithm 1, by iteratively approximating σ_i via sampling from $\mu_{|K_i}$ for all $i \in [m]$.

Algorithm 1: Randomized approximation of the partition function of a polymer model

Input: polymer model $\mathcal{P} = (C, w, \nsim)$, polymer clique cover of $\mathcal{P}$ with size m, number of samples $s \in \mathbb{N}_{>0}$, sampling error $\varepsilon_s \in (0, 1]$

Output: ε-approximation of $Z(\mathcal{P})$ according to Lemma 14

for $i \in [m]$ **do**
  for $j \in [s]$ **do**
    $\Gamma^{(j)} \leftarrow \varepsilon_s$-approximate sample from $\mu_{|K_i}$;
  $\widehat{\sigma}_i \leftarrow \frac{1}{s} \sum_{j \in [s]} \mathbb{1}\{\Gamma^{(j)} \in \mathcal{F}_{|K_{i-1}}\}$;
$\widehat{\sigma} \leftarrow \prod_{i \in [m]} \widehat{\sigma}_i$;
return $1/\widehat{\sigma}$;

The following result bounds, for all $\varepsilon \in (0,1]$, the number of samples s and the sampling error ε_s that are required by Algorithm 1 to obtain an ε-approximation of Z. Our main approximation result, Theorem 4, is then implied by this bound.

▶ **Lemma 14.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, let $Z_{\max} = \max_{i\in[m]}\{Z_{|\Lambda_i}\}$, and let $\varepsilon \in (0,1]$. Consider Algorithm 1 for $\mathcal{P}$ with $s = 1 + 125 Z_{\max} m/\varepsilon^2$ and $\varepsilon_s = \varepsilon/(5 Z_{\max} m)$.

Then Algorithm 1 returns a randomized ε-approximation of Z. ◀

Proof. Let $i \in [m]$. We start by bounding σ_i. Note that $Z_{|K_i} \geq Z_{|K_{i-1}}$ and $Z_{|K_i} \leq Z_{|K_{i-1}} Z_{|\Lambda_i}$. Thus, $1/Z_{\max} \leq \sigma_i \leq 1$.

The remaining proof is split into two parts. First, we bound $\mathrm{E}[\widehat{\sigma}]$ with respect to $1/Z$. Second, we bound the absolute difference of $\widehat{\sigma}$ and $\mathrm{E}[\widehat{\sigma}]$. Combining both errors concludes the proof.

Bounding $\mathrm{E}[\widehat{\sigma}]$. Note that, for all $i \in [m]$, it holds that $\sigma_i - \varepsilon_s \leq \mathrm{E}[\widehat{\sigma}_i] \leq \sigma_i + \varepsilon_s$, since $\widehat{\sigma}_i$ is the mean of ε_s-approximate samples. By the bounds on σ_i and our choice of ε_s, we get

$$\left(1 - \frac{\varepsilon}{5m}\right)\sigma_i \leq \mathrm{E}[\widehat{\sigma}_i] \leq \left(1 + \frac{\varepsilon}{5m}\right)\sigma_i.$$

Recall that $1/Z = \prod_{i\in[m]} \sigma_i$. Further, since $\{\widehat{\sigma}_i\}_{i\in[m]}$ are mutually independent, we have $\mathrm{E}[\widehat{\sigma}] = \prod_{i\in[m]} \mathrm{E}[\widehat{\sigma}_i]$. Consequently, since, for all $x \in [0,1]$ and all $k \in \mathbb{N}_{>0}$, it holds that $\mathrm{e}^{-x/k} \leq 1 - x/(k+1)$ [24, Chapter 3], we obtain

$$\mathrm{e}^{-\varepsilon/4}\frac{1}{Z} \leq \left(1 - \frac{\varepsilon}{5m}\right)^m \frac{1}{Z} \leq \mathrm{E}[\widehat{\sigma}] \leq \left(1 + \frac{\varepsilon}{5m}\right)^m \frac{1}{Z} \leq \mathrm{e}^{\varepsilon/5}\frac{1}{Z} \leq \mathrm{e}^{\varepsilon/4}\frac{1}{Z}. \tag{10}$$

Bounding the absolute difference of $\widehat{\sigma}$ *and* $\mathrm{E}[\widehat{\sigma}]$. By Chebyshev's inequality, we get

$$\Pr\left[|\widehat{\sigma} - \mathrm{E}[\widehat{\sigma}]| \geq \frac{\varepsilon}{5}\mathrm{E}[\widehat{\sigma}]\right] \leq \frac{25}{\varepsilon^2}\frac{\mathrm{Var}[\widehat{\sigma}]}{\mathrm{E}[\widehat{\sigma}]^2} = \frac{25}{\varepsilon^2}\left(\frac{\mathrm{E}\left[\widehat{\sigma}^2\right]}{\mathrm{E}[\widehat{\sigma}]^2} - 1\right).$$

Again, by the mutual independence of $\{\widehat{\sigma}_i\}_{i\in[m]}$, we have $\mathrm{E}[\widehat{\sigma}]^2 = \prod_{i\in[m]} \mathrm{E}[\widehat{\sigma}_i]^2$ and $\mathrm{E}\left[\widehat{\sigma}^2\right] = \prod_{i\in[m]} \mathrm{E}\left[\widehat{\sigma}_i^2\right]$. Thus,

$$\Pr\left[|\widehat{\sigma} - \mathrm{E}[\widehat{\sigma}]| \geq \frac{\varepsilon}{5}\mathrm{E}[\widehat{\sigma}]\right] \leq \frac{25}{\varepsilon^2}\left(\prod_{i\in[m]} \frac{\mathrm{E}\left[\widehat{\sigma}_i^2\right]}{\mathrm{E}[\widehat{\sigma}_i]^2} - 1\right) = \frac{25}{\varepsilon^2}\left(\prod_{i\in[m]}\left(1 + \frac{\mathrm{Var}[\widehat{\sigma}_i]}{\mathrm{E}[\widehat{\sigma}_i]^2}\right) - 1\right).$$

For bounding the variance of $\widehat{\sigma}_i$, recall that $\widehat{\sigma}_i = \frac{1}{s}\sum_{j\in[s]} \mathbb{1}\left\{\Gamma^{(j)} \in \mathcal{F}_{|K_{i-1}}\right\}$, where $\{\Gamma^{(j)}\}_{j\in[s]}$ are independently drawn from an ε_s-approximation of $\mu_{|K_i}$. By equation (9), we have

$$\mathrm{Var}[\widehat{\sigma}_i] = \frac{1}{s^2}\sum_{j\in[s]} \mathrm{Var}\left[\mathbb{1}\left\{\Gamma^{(j)} \in \mathcal{F}_{|K_{i-1}}\right\}\right] = \frac{1}{s}\mathrm{E}[\widehat{\sigma}_i](1 - \mathrm{E}[\widehat{\sigma}_i]).$$

Noting that $\mathrm{E}[\widehat{\sigma_i}] \geq \big(1 - \varepsilon/(5m)\big)\sigma_i \geq 4/(5Z_{\max})$, we bound

$$\frac{\mathrm{Var}[\widehat{\sigma_i}]}{\mathrm{E}[\widehat{\sigma_i}]^2} = \frac{1}{s\mathrm{E}[\widehat{\sigma_i}]} - \frac{1}{s} \leq \frac{5Z_{\max}}{4s}.$$

Hence, using that, for all $x \in [0,1]$ and all $k \in \mathbb{N}_{>0}$, it holds that $\mathrm{e}^{x/(k+1)} \leq 1 + x/k$, we obtain

$$\begin{aligned}\Pr\Big[|\widehat{\sigma} - \mathrm{E}[\widehat{\sigma}]| \geq \frac{\varepsilon}{5}\mathrm{E}[\widehat{\sigma}]\Big] &\leq \frac{25}{\varepsilon^2}\left(\left(1 + \frac{5Z_{\max}}{4s}\right)^m - 1\right) \leq \frac{25}{\varepsilon^2}\left(\mathrm{e}^{5Z_{\max}m/(4s)} - 1\right) \\ &\leq \frac{25}{\varepsilon^2}\frac{5Z_{\max}m}{4s-1}.\end{aligned}$$

Due to our choice of s, and using the same approach as in bounding equation (10), with probability at least $3/4$, it holds that

$$\mathrm{e}^{-\varepsilon/4}\mathrm{E}[\widehat{\sigma}] \leq \Big(1 - \frac{\varepsilon}{5}\Big)\mathrm{E}[\widehat{\sigma}] \leq \widehat{\sigma} \leq \Big(1 + \frac{\varepsilon}{5}\Big)\mathrm{E}[\widehat{\sigma}] \leq \mathrm{e}^{\varepsilon/4}\mathrm{E}[\widehat{\sigma}]. \tag{11}$$

Combining the results. Combining equations (10) and (11) yields that

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

with probability at least $3/4$, which concludes the proof. ■

We can now prove our main approximation result.

▶ **Theorem 4.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model with polymer clique cover Λ of size m such that $\mathcal{P}$ and Λ are computationally feasible as stated in Definition 12. Assume that $\mathcal{P}$ satisfies the conditions of Theorem 3. For all $\varepsilon \in (0,1]$, there is a randomized ε-approximation of Z computable in time $\mathrm{poly}(m/\varepsilon)$. ◀

Proof. The statement follows from Lemma 14, choosing the parameters of Algorithm 1 accordingly. Note that Theorem 3 assumes that $Z_{\max} \in \mathrm{poly}(m)$. This implies that $s \in \mathrm{poly}(m/\varepsilon)$ and that we can sample ε_s-approximately from μ in time $\mathrm{poly}(m/\varepsilon)$. Note that, for all $i \in [m]$, the same holds for $\mu_{|K_i}$, as this only requires the Markov chain to ignore some of the polymer cliques in each step. ■

5 Truncation of Polymer Cliques

In Section 4, we discuss under which assumptions the partition function of a polymer model $\mathcal{P}$ with polymer clique cover Λ of size m can be approximated in time polynomial in m (Theorem 4). One of the assumptions requires to be able to sample, for all $i \in [m]$, from $\mu_{|\Lambda_i}$ in time $\mathrm{poly}(m)$. Unfortunately, for many algorithmic problems, the number of polymer families of each polymer

clique is large, and efficient sampling from $\mu_{|\Lambda_i}$ is non-trivial. However, as we only require to approximately sample from $\mu_{|\Lambda_i}$, it is sufficient to ignore polymer families with low probabilities, that is, with low weight.

To this end, we define a size function for polymers. We aim to remove polymers of large size (low weight), which still yields a sufficient approximation of $\mu_{|\Lambda_i}$ (Lemma 17). As a consequence, we can still approximate $Z(\mathcal{P})$ in time polynomial in m (Theorem 20) by *truncating* the polymer model to polymers of small size.

▶ **Definition 15 (Size Function).** Given a polymer model $(C, w, \nsim)$, a size function is a function $|\cdot| : C \to \mathbb{R}_{>0}$. Given a polymer $\gamma \in C$, we call $|\gamma|$ the *size* of γ. ◀

▶ **Definition 16 (Truncation).** Let $(C, w, \nsim)$ be a polymer model equipped with a size function $|\cdot|$, and let $\mathcal{B} \subseteq C$. For all $k \in \mathbb{R}$, we call $\mathcal{B}^{\leq k} = \{\gamma \in \mathcal{B} \mid |\gamma| \leq k\}$ the *truncation* of $\mathcal{B}$ to size k. Further, we write $\mathcal{B}^{>k} = \mathcal{B} \setminus \mathcal{B}^{\leq k}$. ◀

Note that $\mathcal{B} \subseteq C$ and that $\mathcal{B}^{\leq k}, \mathcal{B}^{>k}$ is a partition of $\mathcal{B}$, implying $\mathcal{B}^{\leq k}, \mathcal{B}^{>k} \subseteq C$. Thus, we can apply our notions of restricted polymer families, partition function, and Gibbs distribution as in Section 2.1 to $\mathcal{B}^{\leq k}$ and $\mathcal{B}^{>k}$ as well. The case $\mathcal{B} = C$ (i.e., we truncate the entire polymer model) plays a special role, which is why we use the shorter notation $\mathcal{F}_{\leq k} = \mathcal{F}_{C^{\leq k}}$, $Z_{\leq k} = Z_{C^{\leq k}}$, and $\mu_{\leq k} = \mu_{C^{\leq k}}$. Analogously, we define $\mathcal{F}_{>k}$, $Z_{>k}$, and $\mu_{>k}$.

▶ **Lemma 17.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, and let $|\cdot|$ be a size function for $\mathcal{P}$. Assume that there is a $k \in \mathbb{R}$ and an $\varepsilon \in (0, 1)$ such that, for all $i \in [m]$, it holds that

$$\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \frac{\varepsilon}{m}. \tag{12}$$

Then $\mathrm{e}^{-\varepsilon} \leq Z_{\leq k}/Z \leq 1$ and $d_{\mathrm{TV}}(\mu, \mu_{\leq k}) \leq \varepsilon$. ◀

Proof. We start by proving $\mathrm{e}^{-\varepsilon} \leq Z_{\leq k}/Z \leq 1$. Since $Z_{\leq k} \leq Z$, as removing polymers does not increase the partition function, it remains to show that $Z \leq \mathrm{e}^{\varepsilon} Z_{\leq k}$.

We observe that $Z \leq Z_{\leq k} Z_{>k}$ with equality if and only if, for all $\Gamma \in \mathcal{F}_{\leq k}$ and all $\Gamma' \in \mathcal{F}_{>k}$, it holds that $\Gamma \cup \Gamma' \in \mathcal{F}$. We proceed by showing that $Z_{>k} \leq \mathrm{e}^{\varepsilon}$.

Note that $C^{>k} = \bigcup_{i \in [m]} \Lambda_i^{>k}$ and that each polymer family in $\mathcal{F}_{>k}$ contains at most one polymer from each $\Lambda_i^{>k}$. Thus, we obtain

$$Z_{>k} \leq \prod_{i \in [m]} Z_{|\Lambda_i^{>k}} = \prod_{i \in [m]} \left(1 + \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma\right).$$

Due to equation (12), we get $Z_{>k} \leq (1 + \varepsilon/m)^m \leq \mathrm{e}^{\varepsilon}$, which proves the first claim.

Observing that

$$d_{\mathrm{TV}}(\mu, \mu_{\leq k}) = \frac{Z - Z_{\leq k}}{Z} \leq 1 - \mathrm{e}^{-\varepsilon} \leq \varepsilon$$

proves the second claim. ■

Recall the clique truncation condition (Condition 5). If the clique truncation condition is satisfied, by choosing a reasonable k for truncating a polymer model, little overall weight is removed. That is, the truncated model represents a good approximation of the original.

▶ **Lemma 18.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, let $|\cdot|$ be a size function for $\mathcal{P}$, and let $i \in [m]$. Assume that Λ_i satisfies Condition 5 for a function g and a bound B.

Then, for all $\varepsilon' \in (0,1)$ and all $k \geq g^{-1}(B/\varepsilon')$, it holds that $\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \varepsilon'$. ◀

Proof. Let $\varepsilon' \in (0,1)$ and $k \geq g^{-1}(B/\varepsilon')$. Due to the clique truncation condition and the monotonicity of g, we observe that

$$g(k) \sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \sum_{\gamma \in \Lambda_i^{>k}} g(|\gamma|) w_\gamma \leq \sum_{\gamma \in \Lambda_i} g(|\gamma|) w_\gamma \leq B.$$

As g is positive, dividing by $g(k)$ yields $\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq B/g(k)$. Substituting our bound for k and noting that g is invertible, we conclude that

$$\sum_{\gamma \in \Lambda_i^{>k}} w_\gamma \leq \frac{B}{g\left(g^{-1}\left(\frac{B}{\varepsilon'}\right)\right)} = \varepsilon',$$

which proves the claim. ■

As a direct consequence of Lemma 18, we get that the partition function of the truncated model is a useful approximation of the original partition function. Combining Lemmas 17 and 18 and choosing $\varepsilon' = \varepsilon/m$ directly implies the following result.

▶ **Corollary 19.** Let $\mathcal{P} = (C, w, \nsim)$ be a polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, and let $|\cdot|$ be a size function for $\mathcal{P}$. Assume that there is a $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ and a $B \in \mathbb{R}_{>0}$ such that, for $i \in [m]$, the polymer clique Λ_i satisfies Condition 5 for g and B.

Then, for all $\varepsilon \in (0,1)$ and $k \geq g^{-1}(Bm/\varepsilon)$, it holds that $\mathrm{e}^{-\varepsilon} \leq Z_{\leq k}/Z \leq 1$ and $d_{\mathrm{TV}}(\mu, \mu_{\leq k}) \leq \varepsilon$. ◀

Using the truncated polymer model, we achieve an ε-approximation result of the partition function of the original model that is computable in time $\mathrm{poly}(m/\varepsilon)$, similar to Theorem 4. Note that Observation 13 applies to Theorem 20 as well.

▶ **Theorem 20.** Let $\mathcal{P} = (C, w, \nsim)$ be a computationally feasible polymer model, let Λ be a polymer clique cover of $\mathcal{P}$ with size m, and let $|\cdot|$ be a size function for $\mathcal{P}$. Further, let $Z_{\max} = \max_{i \in [m]}\{Z_{|\Lambda_i}\}$, and let $t(k)$ denote an upper bound, for all $i \in [m]$, on the time to enumerate $\Lambda_i^{\leq k}$. Last, assume that

(a) $Z_{\max} \in \mathrm{poly}(m)$,

(b) $\mathcal{P}$ satisfies the clique dynamics condition for a function f such that, for all $\gamma \in C$, it holds that $\mathrm{e}^{-\mathrm{poly}(m)} \leq f(\gamma) \leq \mathrm{e}^{\mathrm{poly}(m)}$, and that

(c) there are $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ and $B \in \mathbb{R}_{>0}$ with $B \in \mathrm{poly}(m)$ and $t(g^{-1}(x)) \in \mathrm{poly}(x)$ (for all $x \in \mathbb{R}_{>0}$) such that, for all $i \in [m]$, it holds that Λ_i satisfies Condition 5.

Then, for all $\varepsilon \in (0,1]$, we can ε-approximately sample from μ in time $\mathrm{poly}(m/\varepsilon)$, and there is a randomized ε-approximation of Z computable in time $\mathrm{poly}(m/\varepsilon)$. ◀

Proof. As in the proof of Theorem 3, we consider the polymer Markov chain $\mathcal{M}(\mathcal{P})$. Further, let $k = g^{-1}(2Bm/\varepsilon)$, let $\mathcal{M}_k$ denote the polymer Markov chain on $(C_{\leq k}, w, \nsim)$, and let P_k denote its transitions. We aim to run $\mathcal{M}_k$ for at least $t^* = \tau_{\mathcal{M}(\mathcal{P})}(\varepsilon/2)$ iterations, starting from $\emptyset \in \mathcal{F}_{\leq k}$.

We prove that $d_{\mathrm{TV}}\big(\mu, P_k^{t^*}(\emptyset,\cdot)\big) \leq \varepsilon$. By the triangle inequality, we obtain

$$d_{\mathrm{TV}}\big(\mu, P_k^{t^*}(\emptyset,\cdot)\big) \leq d_{\mathrm{TV}}\big(\mu, \mu_{\leq k}\big) + d_{\mathrm{TV}}\big(\mu_{\leq k}, P_k^{t^*}(\emptyset,\cdot)\big).$$

By our choice of k and by Corollary 19 together with assumption (c), we get that $d_{\mathrm{TV}}\big(\mu, \mu_{\leq k}\big) \leq \varepsilon/2$. Further, note that truncation preserves the clique dynamics condition for the same function f. Thus, Theorem 2 applies to $\mathcal{M}_k$, and we obtain $d_{\mathrm{TV}}\big(\mu_{\leq k}, P_k^{t^*}(\emptyset,\cdot)\big) \leq \varepsilon/2$ for our choice of t^*.

It remains to show that the runtime is bounded by $\mathrm{poly}(m/\varepsilon)$. Analogously to the proof of Theorem 3, due to assumptions (a) and (b), we know that $\tau_{\mathcal{M}(\mathcal{P})}(\varepsilon/2) \in \mathrm{poly}(m/\varepsilon)$, which implies $\tau_{\mathcal{M}_k}(\varepsilon/2) \in \mathrm{poly}(m/\varepsilon)$. Also analogously, it holds that each step can be done in $\mathrm{poly}(m)$, except for sampling, for all $i \in [m]$, from $\mu_{|\Lambda_i}$. However, note that, for all $i \in [m]$, we only need to sample from $\mu_{|\Lambda_i^{\leq k}}$. We do so by enumerating $\Lambda_i^{\leq k}$ in time $t(k)$. By our choice of k and by assumption (c), this takes time at most

$$t(k) = t\left(g^{-1}\left(\frac{2Bm}{\varepsilon}\right)\right) \in \mathrm{poly}\left(\frac{m}{\varepsilon}\right),$$

which proves that we can ε-approximately sample from μ in the desired runtime.

Showing that we can ε-approximate Z in time $\mathrm{poly}(m/\varepsilon)$ is done analogously. By Corollary 19 and assumption (c), we know that for $k = g^{-1}(2Bm/\varepsilon)$ it holds that $\mathrm{e}^{-\varepsilon/2} \leq Z_{\leq k}/Z \leq 1$, which implies $\mathrm{e}^{-\varepsilon/2} Z \leq Z_{\leq k} \leq Z$. As argued above, the truncation of the polymer model $\mathcal{P}$ to this size k satisfies the conditions of Theorem 3, where the sampling from each clique is done by ignoring polymers larger than k. Thus, by Theorem 4, we obtain an $\varepsilon/2$-approximation for $Z_{\leq k}$ in time $\mathrm{poly}(2m/\varepsilon) = \mathrm{poly}(m/\varepsilon)$. Noting that, for $\varepsilon \leq 1$, it holds that

$$1 - \varepsilon \leq \left(1 - \frac{\varepsilon}{2}\right)\mathrm{e}^{-\varepsilon/2} \text{ and } \left(1 + \frac{\varepsilon}{2}\right) \leq 1 + \varepsilon,$$

which concludes the proof. ■

We apply Theorem 20 together with Observation 13 to obtain the results from Table 1. In Section 6, we calculate the bound for the hard-core model on bipartite expanders, and we discuss

how to obtain the other results and how to choose the function f for Condition 1.

5.1 Necessity of Clique Truncation

Before we showcase how our results are applied to improve algorithmic bounds for specific problems, we briefly discuss the interplay between the clique dynamics condition and the clique truncation condition. More specifically, we argue that, unless NP = RP, no polynomial running time guarantees in terms of the size of a clique cover can be given for approximating the partition function of an abstract polymer model if it only satisfies Definition 12 (i.e., being computationally feasible) and the clique dynamics condition. To this end, we encode the Unambiguous Sat problem in such a polymer model. Note that this is not a polynomial time reduction in the usual sense. Similarly to other applications of polymer models, the resulting model has size exponential in the original problem instance.

Unambiguous Sat is defined as the problem of deciding if a Boolean formula in conjunctive normal form on n variables has a satisfying assignment, restricted to instances with at most one such satisfying assignment. Due to Valiant and Vazirani [33], it is known that no (randomized) polynomial time algorithm for Unambiguous Sat exists unless NP = RP. By a standard fact from complexity theory, this even holds when considering randomized algorithms with two-sided errors (see [30, Problem 11.5.18]).

We encode an instance of Unambiguous Sat by considering a polymer model $\mathcal{P} = (C, w, \nsim)$ that contains one polymer $\gamma \in C$ for each assignment. For a polymer $\gamma \in C$, define the weight $w_\gamma = \frac{1}{8}$ if the corresponding assignment is satisfying and $w_\gamma = \frac{1}{2^{n+4}}$ otherwise. Moreover, we set all polymers to be pairwise incompatible.

Note that the resulting polymer model consists of one giant clique. However, to ensure that the weight w_γ can be computed in time polynomial in the size of the clique cover, we have to choose a clique cover that is at most polynomial in n. One option to do so is to assume some ordering of the variables $v_1, \dots, v_n$ and assign each polymer to the clique Λ_i if $i \in [n]$ is the first variable set to *true*. If all variables are *false*, we may assign it to an arbitrary clique (say Λ_n). This way, we get a clique cover Λ of size n and the resulting polymer model together with the clique cover is computationally feasible. We proceed by arguing that $\mathcal{P}$ satisfies the clique dynamics condition.[4] To this end, set $f(\gamma) = 1$ for all polymers $\gamma \in C$ and observe that

$$\sum_{\gamma' \nsim \gamma} f(\gamma') w_{\gamma'} \leq \frac{2^n}{2^{n+4}} + \frac{1}{8} < f(\gamma).$$

If there is no satisfying assignment, then it holds that

$$Z(\mathcal{P}) = 1 + \sum_{\gamma \in C} w_\gamma = 1 + \frac{2^n}{2^{n+4}} = 1 + \frac{1}{16}.$$

[4] In fact, it even satisfies the widely applied and much more restrictive Kotecký–Preiss condition [26].

On the other hand, if there is such a satisfying assignment, we have $Z(\mathcal{P}) \geq 1 + \frac{1}{8}$. Thus, assume we can compute $\widetilde{Z}$ such that $\left(1 - \frac{1}{128}\right)Z(\mathcal{P}) \leq \widetilde{Z} \leq \left(1 + \frac{1}{128}\right)Z(\mathcal{P})$ with probability at least $\frac{3}{4}$ in time polynomial in the size of the clique cover Λ then this gives us a randomized algorithm to decide if there is a satisfying assignment in time poly(n) with two-sided error probability of at most $\frac{1}{4}$. Assuming NP $\neq$ RP, this should not be possible.

In a sense, the above construction shows that, even if the clique dynamics of a polymer model is rapidly mixing, the clique dynamics condition and the fact that the model is computationally feasible might not be sufficient to argue that each step of the Markov chain runs efficiently. More precisely, the sheer size of each polymer clique and a lack of structural knowledge makes it hard to sample from each cliques distribution. The clique truncation condition is one way to circumvent this problem by giving a sufficient condition to ignore most of the polymers in each clique. In general, this shows that some additional condition or detailed structural knowledge is required.

6 Example Application: Hard-Core Model on Bipartite Expanders

In order to demonstrate how Theorem 20 improves known bounds for the algorithmic use of polymer models, we investigate the hard-core model for high fugacity $\lambda \in \mathbb{R}_{>0}$ on bipartite α-expanders with bounded maximum degree Δ. For a graph (V, E) and an $S \subseteq V$, let $N_G(S)$ denote the set of all vertices that are adjacent to a vertex in S.

▶ **Definition 21 (bipartite α-expander).** Let $G = (V, E)$ be a bipartite graph with partition $V = V_{\mathrm{L}} \cup V_{\mathrm{R}}$. For all $i \in \{\mathrm{L}, \mathrm{R}\}$, we call $S \subseteq V_i$ *small* if and only if $|S| \leq |V_i|/2$. For all $\alpha \in (0, 1)$, graph G is a *bipartite α-expander* if and only if, for all small sets of vertices S, it holds that $|N_G(S)| \geq (1 + \alpha)|S|$. ◀

For any graph G, the hard-core partition function is a graph polynomial of some parameter $\lambda \in \mathbb{R}_{>0}$, called fugacity. Let $\mathcal{I}_G$ be the set of all independent sets in G. The hard-core partition function for fugacity λ is now formally defined as

$$Z(G, \lambda) = \sum_{I \in \mathcal{I}_G} \lambda^{|I|}.$$

We approximate $Z(G, \lambda)$ in terms of the partition function of two polymer models, constructed as proposed by Jenssen et al. [22]. For a bipartite α-expander G with bounded 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. For all $i \in \{\mathrm{L}, \mathrm{R}\}$, we define a polymer model $\mathcal{P}^{(i)} = (C^{(i)}, w^{(i)}, \nsim)$ as follows:

- each polymer $\gamma \in C^{(i)}$ is defined by a non-empty set of vertices $\overline{\gamma} \subseteq V_i$ such that $\overline{\gamma}$ is small and induces a connected subgraph in G^2,
- for $\gamma \in C^{(i)}$, let $w_\gamma^{(i)} = \lambda^{|\overline{\gamma}|} / \left((1 + \lambda)^{|N_G(\overline{\gamma})|}\right)$, and

- two polymers $\gamma, \gamma' \in C^{(i)}$ are incompatible if and only if there are vertices $v \in \overline{\gamma}, w \in \overline{\gamma'}$ with graph distance at most 1 in G^2.

To ease notation, for all $i \in \{\mathrm{L}, \mathrm{R}\}$, we write $\mu^{(i)}$ and $Z^{(i)}$ instead of $\mu^{(\mathcal{P}^{(i)})}$ and $Z(\mathcal{P}^{(i)})$, respectively.

We use $\mathcal{P}^{(\mathrm{L})}$ and $\mathcal{P}^{(\mathrm{R})}$ for approximating the hard-core partition function of bipartite α-expanders in the following sense.

▶ **Lemma 22 ([22, Lemma 19]).** Given a bipartite α-expander $G = (V_\mathrm{L} \cup V_\mathrm{R}, E)$ with $|V_\mathrm{L} \cup V_\mathrm{R}| = n$, let $Z(G, \lambda)$ denote its hard-core partition function with fugacity $\lambda \in \mathbb{R}_{>0}$, and let the polymer models $\mathcal{P}^{(\mathrm{L})}, \mathcal{P}^{(\mathrm{R})}$ be defined as above. For all $\lambda \geq \mathrm{e}^{11/\alpha}$, it holds that

$$(1 - \mathrm{e}^{-n})Z(G, \lambda) \leq (1+\lambda)^{|V_\mathrm{R}|}Z^{(\mathrm{L})} + (1+\lambda)^{|V_\mathrm{L}|}Z^{(\mathrm{R})} \leq (1 + \mathrm{e}^{-n})Z(G, \lambda). \quad ◀$$

To apply Theorem 20, we have to fix a polymer clique cover Λ for each polymer model $\mathcal{P}^{(i)}$ with $i \in \{\mathrm{L}, \mathrm{R}\}$. Based on the incompatibility relation, a natural choice is to define, for each $v \in V_i$, a clique Λ_v such that $\gamma \in \Lambda_v$ if and only if $v \in \overline{\gamma}$. As we need to verify the clique dynamics condition, it is useful to have a bound on the number of incompatible polymers, which the following lemma provides.

▶ **Lemma 23 ([3, Lemma 2.1]).** For an undirected graph $G = (V, E)$ with maximum degree Δ and for all $v \in V$, the number of vertex-induced connected subgraphs that contain v and have at most $k \in \mathbb{N}_{>0}$ vertices is bounded from above by $\mathrm{e}^k \Delta^{k-1} / \left(k^{3/2}\sqrt{2\pi}\right)$. ◀

Commonly, the bound $(\mathrm{e}\Delta)^{k-1}/2$ is applied, as it is more convenient to work with. However, this bound actually only holds for $k \geq 2$. Further, note that Jenssen et al. [22] used a weaker bound, namely $(\mathrm{e}\Delta)^k$. Although this bound holds for all $k \in \mathbb{N}_{>0}$, it yields a much worse dependency on Δ. For a fair comparison, we added the result of refined calculations for the approach by Jenssen et al. [22] to Table 1.

In order to apply truncation, we further need a notion of size for polymers. An obvious choice is to set $|\gamma| = |\overline{\gamma}|$. The following lemma then bounds the time for enumerating polymers in a clique up to some size $k \in \mathbb{N}_{>0}$.

▶ **Lemma 24 ([31, Lemma 3.7]).** Let $G = (V, E)$ be an undirected graph with maximum degree Δ, and let $v \in V$. 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 $\mathrm{e}^{\mathrm{O}(k \log(\Delta))}$. ◀

We now prove our bound on λ for an efficient approximation of the hard-core partition function on bipartite α-expanders. Most of the calculations are similar to those of Jenssen et al. [22], except that we use our newly obtained conditions.

▶ **Proposition 25.** Let $G(V_\mathrm{L} \cup V_\mathrm{R}, E)$ be a bipartite α-expander with $|V_\mathrm{L} \cup V_\mathrm{R}| = n$ and with maximum degree $\Delta \in \mathbb{N}_{>0}$. For $\lambda \geq \max\{(\mathrm{e}\Delta^2/0.8)^{1/\alpha}, \mathrm{e}^{11/\alpha}\}$ and for all $\varepsilon \in (0, 1]$, there is an FPRAS for $Z(G, \lambda)$ with runtime $(n/\varepsilon)^{\mathrm{O}(\ln(\Delta))}$. ◀

Proof. If $\varepsilon \in \mathrm{O}(\mathrm{e}^{-n})$, we compute $Z(G,\lambda)$ by enumerating all independent sets. Since there are at most 2^n independent sets, which is polynomial in $1/\mathrm{e}^{-n}$, the statement then follows. It remains to analyze the case $\varepsilon \in \Omega(\mathrm{e}^{-n})$. To this end, assume that $\varepsilon \geq 4\mathrm{e}^{-n}$.

By Lemma 22, $Z(G,\lambda)$ can be e^{-n}-approximated using $Z^{(\mathrm{L})}$ and $Z^{(\mathrm{R})}$. We aim for an $\varepsilon/4$-approximation of $Z^{(\mathrm{L})}$ and $Z^{(\mathrm{R})}$, each with failure probability at most $1-\sqrt{3}/2$. Note that

$$1-\varepsilon \leq (1-\mathrm{e}^{-n})\Big(1-\frac{\varepsilon}{4}\Big) \text{ and } (1+\mathrm{e}^{-n})\Big(1+\frac{\varepsilon}{4}\Big) \leq 1+\varepsilon.$$

Thus, with probability at least $(\sqrt{3}/2)^2 = 3/4$ the result is an ε-approximation of $Z(G,\lambda)$. We can obtain the desired error probability of at most $1-\sqrt{3}/2$ for the approximations of $Z^{(\mathrm{L})}$ and $Z^{(\mathrm{R})}$ by taking the median of $\mathrm{O}\big(\ln(2/(2-\sqrt{3}))\big) = \mathrm{O}(1)$ independent approximations with failure probability at most $1/4$.

Let $i \in \{\mathrm{L},\mathrm{R}\}$. In order to approximate $Z^{(i)}$, we aim to apply Theorem 20. To this end, for all $v \in V_i$, we define a polymer clique Λ_v containing all polymers $\gamma \in C^{(i)}$ with $v \in \overline{\gamma}$. This results in a polymer clique cover of size n.

We proceed by proving that the polymer model satisfies the clique dynamics condition for $f(\gamma) = |\overline{\gamma}|$. We use Observation 13 to simplify this step. This also implies that assumption (a) of Theorem 20 is satisfied. For any $\gamma \in C^{(i)}$ we start by bounding the set of polymers $\gamma' \nsim \gamma$ by

$$\sum_{\gamma' \in C^{(i)}\colon \gamma' \nsim \gamma} f(\gamma') w_{\gamma'}^{(i)} \leq \sum_{v \in N_{G^2}(\overline{\gamma})} \sum_{\gamma' \in \Lambda_v} f(\gamma') w_{\gamma'}^{(i)} = \sum_{v \in N_{G^2}(\overline{\gamma})} \sum_{k \in \mathbb{N}_{>0}} \sum_{\substack{\gamma' \in \Lambda_v \\ |\overline{\gamma'}|=k}} f(\gamma') w_{\gamma'}^{(i)}.$$

Because G is a bipartite α-expander, for all $\gamma \in C^{(i)}$, we have $w_\gamma^{(i)} \leq 1/\lambda^{\alpha|\overline{\gamma}|}$. Further, note that the degree of G^2 is bounded by Δ^2. By Lemma 23 and our definition of f, we obtain

$$\sum_{v \in N_{G^2}(\overline{\gamma})} \sum_{k \in \mathbb{N}_{>0}} \sum_{\substack{\gamma' \in \Lambda_v \\ |\overline{\gamma'}|=k}} f(\gamma') w_{\gamma'}^{(i)} \leq \Delta^2 |\overline{\gamma}| \sum_{k \in \mathbb{N}_{>0}} \frac{\mathrm{e}^k (\Delta^2)^{k-1}}{k^{3/2}\sqrt{2\pi}} \cdot k \cdot \frac{1}{\lambda^{\alpha k}} = \frac{|\overline{\gamma}|}{\sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \Big(\frac{\mathrm{e}\Delta^2}{\lambda^\alpha}\Big)^k \frac{1}{\sqrt{k}}.$$

For $\lambda \geq \big(\mathrm{e}\Delta^2/0.8\big)^{1/\alpha}$, we get

$$\frac{|\overline{\gamma}|}{\sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \Big(\frac{\mathrm{e}\Delta^2}{\lambda^\alpha}\Big)^k \frac{1}{\sqrt{k}} \leq \frac{|\overline{\gamma}|}{\sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} (0.8)^k \frac{1}{\sqrt{k}} \leq \frac{|\overline{\gamma}|}{\sqrt{2\pi}} \sqrt{2\pi} = f(\gamma).$$

It remains to show, for all $v \in V_i$, that Λ_v satisfies the clique truncation condition for a $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ and a $B \in \mathbb{R}_{>0}$. To this end, for all $\gamma \in C^{(i)}$, let $|\gamma| = |\overline{\gamma}|$, let $g(|\gamma|) = \mathrm{e}^{0.2|\gamma|}$, and let $B = 1$.

Analogously to our verification of the clique dynamics condition, we see, for all $v \in V_i$, that

$$\sum_{\gamma \in \Lambda_v} g(|\gamma|) w_\gamma^{(i)} \leq \frac{1}{\Delta^2 \sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \left(\frac{e\Delta^2}{\lambda^\alpha} \right)^k \frac{1}{k^{3/2}} e^{0.2k} = \frac{1}{\Delta^2 \sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \left(\frac{e^{1.2}\Delta^2}{\lambda^\alpha} \right)^k \frac{1}{k^{3/2}}.$$

For $\lambda \geq \left(e\Delta^2/0.8\right)^{1/\alpha}$, we get

$$\frac{1}{\Delta^2 \sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \left(\frac{e^{1.2}\Delta^2}{\lambda^\alpha} \right)^k \frac{1}{k^{3/2}} \leq \frac{1}{\Delta^2 \sqrt{2\pi}} \sum_{k \in \mathbb{N}_{>0}} \left(0.8e^{0.2}\right)^k \frac{1}{k^{3/2}} < \frac{1}{\Delta^2 \sqrt{2\pi}} 2.2 \leq B.$$

Last, we bound the runtime of the FPRAS. By Lemma 24, we can enumerate each polymer clique up to size k in time $t(k) \in e^{O(k \log(\Delta))}$. As $g^{-1} \colon x \mapsto 5\ln(x)$, we have $t \circ g^{-1} \colon x \mapsto x^{O(\ln(\Delta))}$, which is polynomial for $\Delta \in \Theta(1)$. For the runtime bound, note that we truncate to size $k = g^{-1}(n/\varepsilon)$. Thus, the time for computing each step of the polymer Markov chain is bounded by $t(k) = (n/\varepsilon)^{O(\ln(\Delta))}$, which dominates the runtime. ■

Choice of f and Calculation of Table 1 Note that the choice of the function f used in the clique dynamics condition is very sensitive to the bound on the number of subgraphs. For the bound stated in Lemma 23, it turns out that using $f(\gamma) = |\overline{\gamma}|$ yields the best bounds on λ (see the proof of Proposition 25 for details). With this choice of f, the condition that we identified in Observation 13 is similar to the mixing condition of [8, Definition 1], except that we do not require a strict inequality. Further, note that such a choice of f is not possible for the Kotecký–Preiss condition [26]. If purely exponential bounds on the number of subgraphs are used, the best results are usually obtained by setting f to take an exponential form. A detailed understanding of how to choose f might be of interest for applications to specific graph classes and other combinatorial structures.

The results for the remaining applications in Table 1 are derived via similar calculations. For the Potts model on expander graphs and the hard-core model on unbalanced bipartite graphs, we use Lemmas 23 and 24 together with the same function f for the clique dynamics condition as in the proof of Proposition 25. For the perfect matching polynomial, we use the bounds for the number of polymers and for polymer enumeration that are stated by Casel et al. [7], and we choose $f(\gamma) = e^{a|\gamma|}$ for $a \approx 0.2$.

Acknowledgments

Andreas Göbel was funded by the project PAGES (project No. 467516565) of the German Research Foundation (DFG). This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 945298-ParisRegionFP. Marcus Pappik was funded by the HPI Research School on Data Science and Engineering.

References

[1] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *Proc. of FOCS'20*. 2020, pp. 1319–1330. DOI: 10.1109/FOCS46700.2020.00125 (see page 7).

[2] Antonio Blanca, Sarah Cannon, and Will Perkins. "Fast and perfect sampling of subgraphs and polymer systems". In: *arXiv preprint arXiv:2202.05907* (2022) (see page 8).

[3] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. "Left and right convergence of graphs with bounded degree". In: *Random Structures & Algorithms* 42.1 (2013), pp. 1–28. DOI: 10.1002/rsa.20414 (see page 25).

[4] Christian Borgs, Jennifer T. Chayes, Tyler Helmuth, Will Perkins, and Prasad Tetali. "Efficient sampling and counting algorithms for the Potts model on $\mathbb{Z}^d$ at all temperatures". In: *Proc. of STOC'20*. 2020, pp. 738–751. DOI: 10.1145/3357713.3384271 (see page 2).

[5] Russ Bubley and Martin E. Dyer. "Path Coupling: A Technique for Proving Rapid Mixing in Markov Chains". In: *Proc. of FOCS'97*. IEEE Computer Society, 1997, pp. 223–231. DOI: 10.1109/SFCS.1997.646111. URL: https://doi.org/10.1109/SFCS.1997.646111 (see page 7).

[6] Sarah Cannon and Will Perkins. "Counting independent sets in unbalanced bipartite graphs". In: *Proc. of SODA'20*. 2020, pp. 1456–1466. DOI: 10.1137/1.9781611975994.88 (see pages 2, 6).

[7] Katrin Casel, Philipp Fischbeck, Tobias Friedrich, Andreas Göbel, and J. A. Gregor Lagodzinski. "Zeros and approximations of holant polynomials on the complex plane". In: *CoRR* abs/1905.03194 (2019). URL: http://arxiv.org/abs/1905.03194 (see pages 2, 4, 6, 27).

[8] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. "Fast algorithms at low temperatures via Markov chains". In: *Random Struct. Algorithms* 58.2 (2021), pp. 294–321. DOI: 10.1002/rsa.20968. URL: https://doi.org/10.1002/rsa.20968 (see pages 3, 4, 6, 16, 27).

[9] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Rapid mixing of Glauber dynamics up to uniqueness via contraction". In: *CoRR* abs/2004.09083 (2020). URL: http://arxiv.org/abs/2004.09083 (see page 7).

[10] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. "The statistical physics of real-world networks". In: *Nature Reviews Physics* 1 (2019), pp. 58–71. DOI: 10.1038/s42254-018-0002-6 (see page 1).

[11] Roland L. Dobrushin. "Estimates of semi-invariants for the Ising model at low temperatures". In: *Translations of the American Mathematical Society-Series 2* 177 (1996), pp. 59–82 (see pages 3, 14).

[12] Rick Durrett. *Probability: theory and examples*. Available at https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf. Cambridge University Press, 2019. ISBN: 978-1-108-47368-2 (see page 32).

[13] Martin Dyer and Catherine Greenhill. "On Markov chains for independent sets". In: *Journal of Algorithms* 35.1 (2000), pp. 17–49. DOI: 10.1006/jagm.1999.1071 (see page 4).

[14] Roberto Fernández, Pablo A. Ferrari, and Nancy L. Garcia. "Loss network representation of Peierls contours". In: *Annals of Probability* 29.2 (2001), pp. 902–937. DOI: 10.1214/aop/1008956697 (see page 3).

[15] Roberto Fernández and Aldo Procacci. "Cluster expansion for abstract polymer models. New Bounds from an old approach". In: *Communications in Mathematical Physics* 274.1 (2007), pp. 123–140. DOI: 10.1007/s00220-007-0279-2 (see pages 3, 5, 14).

[16] Sacha Friedli and Yvan Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, 2017. ISBN: 978-1-107-18482-4. DOI: 10.1017/9781316882603 (see page 1).

[17] Tobias Friedrich, Andreas Göbel, Martin S. Krejca, and Marcus Pappik. "A spectral independence view on hard spheres via block dynamics". In: *Proc. of ICALP'21 (to appear)*. 2021 (see page 7).

[18] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. "Fast algorithms for general spin systems on bipartite expanders". In: *Proc. of MFCS'20*. To appear. 2020. URL: https://arxiv.org/abs/2004.13442 (see page 2).

[19] Sam Greenberg, Dana Randall, and Amanda Pascoe Streib. "Sampling biased monotonic surfaces using exponential metrics". In: *Combinatorics, Probability and Computing* 29.5 (2020), pp. 672–697. DOI: 10.1017/S0963548320000188 (see pages 9, 31).

[20] Tyler Helmuth, Will Perkins, and Guus Regts. "Algorithmic Pirogov–Sinai theory". In: *Proc. of STOC'19*. 2019, pp. 1009–1020. DOI: 10.1145/3313276.3316305 (see pages 1–3).

[21] Ernst Ising. "Contribution to the theory of ferromagnetism". In: *Zeitschrift für Physik* 31 (1925), pp. 253–258. DOI: 10.1007/BF02980577 (see page 1).

[22] Matthew Jenssen, Peter Keevash, and Will Perkins. "Algorithms for #BIS-hard problems on expander graphs". In: *Proc. of SODA'19*. 2019, pp. 2235–2247. DOI: 10.1137/1.9781611975482.135 (see pages 2, 6, 24, 25).

[23] Matthew Jenssen, Will Perkins, and Aditya Potukuchi. "Approximately counting independent sets in bipartite graphs via graph containers". In: *arXiv preprint arXiv:2109.03744* (2021) (see page 2).

[24] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Springer Science & Business Media, 2003. DOI: 10.1007/978-3-0348-8005-3 (see pages 17, 18).

[25] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical Computer Science* 43 (1986), pp. 169–188. DOI: 10.1016/0304-3975(86)90174-X (see page 17).

[26] Roman Kotecký and David Preiss. "Cluster expansion for abstract polymer models". In: *Communications in Mathematical Physics* 103.3 (1986), pp. 491–498. DOI: 10.1007/BF01211762 (see pages 2–5, 14, 23, 27).

[27] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. "Counting independent sets and colorings on random regular bipartite graphs". In: *CoRR* abs/1903.07531 (2019). URL: http://arxiv.org/abs/1903.07531 (see page 2).

[28] Michael Mitzenmacher and Eli Upfal. *Probability and computing: randomization and probabilistic techniques in algorithms and data analysis*. 2nd ed. Cambridge university press, 2017. ISBN: 110715488X (see page 34).

[29] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. "On the hardness of sampling independent sets beyond the tree threshold". In: *Probability Theory and Related Fields* 143.3-4 (2009), pp. 401–439. DOI: 10.1007/s00440-007-0131-9 (see page 7).

[30] Christos H Papadimitriou. *Computational complexity*. John Wiley and Sons Ltd., 2003 (see page 23).

[31] Viresh Patel and Guus Regts. "Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials". In: *Electronic Notes in Discrete Mathematics* 61 (2017), pp. 971–977. DOI: 10.1016/j.endm.2017.07.061 (see pages 3, 6, 25).

[32] Han Peters and Guus Regts. "On a conjecture of Sokal concerning roots of the independence polynomial". In: *Michigan Mathematical Journal* 68.1 (2019), pp. 33–55. DOI: 10.1307/mmj/1541667626 (see pages 2, 5).

[33] L.G. Valiant and V.V. Vazirani. "NP is as easy as detecting unique solutions". In: *Theoretical Computer Science* 47 (1986), pp. 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 23).

[34] Dror Weitz. "Counting independent sets up to the tree threshold". In: *Proc. of STOC'06*. 2006, pp. 140–149. DOI: 10.1145/1132516.1132538 (see page 2).

[35] H. Peyton Young. *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press, 1998. DOI: 10.2307/j.ctv10h9d35 (see page 1).

A Appendix: Coupling Theorem

We closely follow the proof of Greenberg et al. [19]. Central to this is the following theorem, which we present in a slightly different fashion than Greenberg et al. [19, Lemma 3.5].

▶ **Theorem 26.** Let $d, D \in \mathbb{R}$ with $d \leq D$, let $(S_t)_{t\in\mathbb{N}}$ be a stochastic process adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$, and let T be a stopping time with respect to $\mathcal{F}$. Assume that, for all $t \in \mathbb{N}$, it holds that $S_{\min\{t,T\}} \in [d, D]$, that

$$\mathrm{E}[S_{t+1} \mid \mathcal{F}_t] \cdot \mathbb{1}\{t < T\} \leq S_t \cdot \mathbb{1}\{t < T\}, \tag{13}$$

and that there is a $Q \in \mathbb{R}_{>0}$ such that

$$\mathrm{E}\left[(S_{t+1} - S_t)^2 \,\middle|\, \mathcal{F}_t\right] \cdot \mathbb{1}\{t < T\} \geq Q \cdot \mathbb{1}\{t < T\}. \tag{14}$$

Then

$$\mathrm{E}[T \mid \mathcal{F}_0] \leq \frac{\mathrm{E}\left[(D - S_T)^2 \,\middle|\, \mathcal{F}_0\right] - (D - S_0)^2}{Q}.$$ ◀

Before we prove the theorem, we discuss the changes we made in the phrasing of the theorem. Different to the original theorem by Greenberg et al. [19, Lemma 3.5], we include a filtration and we use indicator functions. Our reasons are as follows. The proof of Theorem 26 aims to apply the optional-stopping theorem for submartingales. A submartingale is, by definition, a stochastic process $(Z_t)_{t\in\mathbb{N}}$ adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that, for all $t \in \mathbb{N}$, the expectation of Z_t is finite and $\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t] \geq Z_t$. It is important to note that the expectation $\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t]$ is itself a random variable and that the inequality $\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t] \geq Z_t$ is stronger than $\mathrm{E}[Z_{t+1}] \geq \mathrm{E}[Z_t]$ (which follows by the law of total expectation). Hence, we require a filtration.

Second, the indicator functions make sure that equations (13) and (14) (and the boundedness of S) only have to hold as long as S did not stop. Afterward, they are trivially satisfied. This is important, as S is bounded from below by d and its conditional expectation does not increase. Assume that we did not use indicator functions. If there is a $t \in \mathbb{N}$ such that $S_t = d$, then $S_{t+1} = d$ holds as well, as otherwise the inequality $\mathrm{E}[S_{t+1} \mid \mathcal{F}_t] \leq S_t$ does not hold. However, this implies that $\mathrm{E}\left[(S_{t+1} - S_t)^2 \,\middle|\, \mathcal{F}_t\right] = 0$ (since the process is now almost surely deterministic), which violates equation (14) if not for the indicator functions.

Note that our additional assumptions in Theorem 26 only fix issues in the proof of Greenberg et al. [19, Lemma 3.5]. The proof itself remains mostly unchanged.

Proof of Theorem 26. Let $M \in \mathbb{R}_{\geq 0}$ such that, for all $t \in \mathbb{N}$, it holds that $M \geq (D - S_t)^2 \cdot \mathbb{1}\{t \leq T\}$. Note that such an M exists, as, for all $t \in \mathbb{N}$, by assumption, $S_{\min\{t,T\}}$ is bounded. For all $t \in \mathbb{N}$, let $Z_t = \left((D - S_t)^2 - Qt - M\right) \cdot \mathbb{1}\{t \leq T\}$. Note that, for all $t \in \mathbb{N}$, due to the definition of M and due to $\mathbb{1}\{t < T\} \leq \mathbb{1}\{t \leq T\}$, it holds that

$$\left((D - S_t)^2 - Qt - M\right) \cdot \mathbb{1}\{t < T\} \geq \left((D - S_t)^2 - Qt - M\right) \cdot \mathbb{1}\{t \leq T\} = Z_t. \tag{15}$$

We show that Z is a submartingale with respect to $\mathcal{F}$. Let $t \in \mathbb{N}$. By noting that $\mathbb{1}\{t+1 \leq T\} = \mathbb{1}\{t < T\}$, since T is integer, and by applying equations (13) to (15), we get

$$\begin{aligned}
\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t] &= \mathrm{E}\left[\left((D - S_{t+1})^2 - Q(t+1) - M\right) \cdot \mathbb{1}\{t+1 \leq T\} \,\middle|\, \mathcal{F}_t\right] \\
&= \mathrm{E}\left[\left((D - S_{t+1})^2 - Q(t+1) - M\right) \cdot \mathbb{1}\{t < T\} \,\middle|\, \mathcal{F}_t\right] \\
&= \left(D^2 - 2D\mathrm{E}[S_{t+1} \mid \mathcal{F}_t] + \mathrm{E}\left[S_{t+1}^2 \,\middle|\, \mathcal{F}_t\right] - Q - Qt - M\right) \cdot \mathbb{1}\{t < T\} \\
&\overset{(14)}{\geq} \left(D^2 - 2D\mathrm{E}[S_{t+1} \mid \mathcal{F}_t] + 2S_t\mathrm{E}[S_{t+1} \mid \mathcal{F}_t] - S_t^2 - Qt - M\right) \cdot \mathbb{1}\{t < T\} \\
&= \left(D^2 - 2\mathrm{E}[S_{t+1} \mid \mathcal{F}_t](D - S_t) - S_t^2 - Qt - M\right) \cdot \mathbb{1}\{t < T\} \\
&\overset{(13)}{\geq} \left(D^2 - 2S_t(D - S_t) - S_t^2 - Qt - M\right) \cdot \mathbb{1}\{t < T\} \\
&= \left((D - S_t)^2 - Qt - M\right) \cdot \mathbb{1}\{t < T\} \\
&\overset{(15)}{\geq} Z_t.
\end{aligned}$$

Since Z is bounded, it is uniformly integrable [12, Theorems 4.2.11 and 4.6.4]. By applying the optional-stopping theorem for uniformly integrable submartingales [12, Theorem 4.8.3], we get that $\mathrm{E}[Z_T \mid \mathcal{F}_0] \geq \mathrm{E}[Z_0 \mid \mathcal{F}_0]$. Solving this inequality for $\mathrm{E}[T \mid \mathcal{F}_0]$ and noting that Z_0 is $\mathcal{F}_0$-measurable concludes the proof. ■

While we state Theorem 26 in an elaborate fashion, we use it in a different way in the following proof of Theorem 6. Instead of considering expected values conditional on a σ-algebra, we consider expected values conditional on each of the outcomes of the random process before it stops. We formalize this statement in the following remark.

▶ **Remark 27.** Let $(X_t)_{t \in \mathbb{N}}$ denote a random process over $\mathbb{R}$, defined over a probability space with sample space Ω, let $(\mathcal{F}_t)_{t \in \mathbb{N}}$ denote the natural filtration of X, and let T be a stopping time with respect to $\mathcal{F}$. If X is discrete and Markovian, then, for all $t \in \mathbb{N}$ and all $\delta \in \mathbb{R}$, the following equivalence holds:

$$\forall x \in \{y \in \mathbb{R} \mid \exists \omega \in \Omega \colon X_t(\omega) = y \wedge t < T(\omega)\} \colon \mathrm{E}[X_t - X_{t+1} \mid X_t = x] \leq \delta \tag{16}$$

$$\leftrightarrow$$

$$\mathrm{E}[(X_t - X_{t+1}) \cdot \mathbb{1}\{t < T\} \mid \mathcal{F}_t] \leq \delta \cdot \mathbb{1}\{t < T\}. \tag{17}$$

◀

Informally, on the one hand, proposition (17) considers all instantiations $\omega \in \Omega$ of X where X did not stop up to time t. On the other hand, proposition (16) partitions Ω with respect to X_t and shows the relevant inequality for the sets $X_t^{-1}(x)$ (for the values of x as specified in the proposition). Since for all $\omega \in X_t^{-1}(x)$ the value $\mathrm{E}[X_t - X_{t+1} \mid X_t](\omega)$ is the same (since X is Markovian), it suffices to consider their average, which is $\mathrm{E}[X_t - X_{t+1} \mid X_t = x]$.

More formally, first note that the indicator functions in proposition (17) relate to the predicate "$t < T(\omega)$" in the set of values for x in proposition (16), which makes sure to consider only such

values x for which the process did not stop yet, as these are the only instantiations of X for which proposition (17) is not trivially satisfied.

By definition of the conditional expectation for discrete random variables, it holds that

$$\begin{aligned}\mathrm{E}[X_t - X_{t+1} \mid X_t = x] &= \sum_{\omega \in \Omega} \mathrm{E}[X_t - X_{t+1} \mid X_t](\omega) \cdot \Pr\left[\{\omega\} \,\middle|\, X_t^{-1}(x)\right] \\ &= \sum_{\substack{\omega \in \Omega: \\ X_t(\omega) = x}} \mathrm{E}[X_t - X_{t+1} \mid X_t](\omega) \cdot \frac{\Pr[\{\omega\}]}{\Pr\left[X_t^{-1}(x)\right]}.\end{aligned}$$

Since X is Markovian, for all $\omega \in \Omega$ where $X_t(\omega) = x$, the random variable $\mathrm{E}[X_t - X_{t+1} \mid X_t]$ has the same value $v \in \mathbb{R}$, as the distribution of X_{t+1} only depends on $X_t = x$. That is, $\mathrm{E}[X_t - X_{t+1} \mid X_t = x] = \mathrm{E}[X_t - X_{t+1} \mid X_t](\omega) = v$. Since X is Markovian and thus conditioning on X_t is the same as conditioning on $\mathcal{F}_t$, this shows that $\mathrm{E}[X_t - X_{t+1} \mid X_t = x] = \mathrm{E}[X_t - X_{t+1} \mid \mathcal{F}_t](\omega)$.

Proof of Theorem 6 We now start with the stating the proof of the coupling theorem, which we restate here for convenience.

▶ **Theorem 6.** Let $\mathcal{M}$ be an ergodic Markov chain with state space Ω and with transition matrix P such that, for all $x \in \Omega$, it holds that $P(x, x) > 0$. For $d, D \in \mathbb{R}_{>0}$, $d \leq D$, let $\delta \colon \Omega^2 \to \{0\} \cup [d, D]$ be such that $\delta(x, y) = 0$ if and only if $x = y$. Assume that there is a coupling between the transitions of two copies $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$ of $\mathcal{M}$ such that, for all $t \in \mathbb{N}$ and all $x, y \in \Omega$, it holds that

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \leq \delta(x, y). \tag{3}$$

Furthermore, assume that there are $\kappa, \eta \in (0, 1)$ such that, for the same coupling and all $t \in \mathbb{N}$ and all $x, y \in \Omega$ with $x \neq y$, it holds that

$$\Pr[|\delta(X_{t+1}, Y_{t+1}) - \delta(x, y)| \geq \eta\delta(x, y) \mid X_t = x, Y_t = y] \geq \kappa. \tag{4}$$

Then, for all $\varepsilon \in (0, 1]$, it holds that

$$\tau_{\mathcal{M}}(\varepsilon) \leq \frac{\left(\ln(D/d) + 2\ln(2)\right)^2}{\ln(1+\eta)^2 \kappa} \ln\left(\frac{1}{\varepsilon}\right).$$

If $\ln(D/d) \in \Omega(1)$, then this bound simplifies to

$$\tau_{\mathcal{M}}(\varepsilon) \in \mathrm{O}\left(\frac{\ln(D/d)^2}{\ln(1+\eta)^2 \kappa} \ln\left(\frac{1}{\varepsilon}\right)\right). \qquad \blacktriangleleft$$

Proof. We aim to bound the expected time until δ hits 0 for the coupled copies $(X_t)_{t \in \mathbb{N}}$ and $(Y_t)_{t \in \mathbb{N}}$ of $\mathcal{M}$ and for all pairs of starting states $x, y \in \Omega$. This results in a bound on the expected coupling

time, and, because $\mathcal{M}$ is ergodic, also bounds $\tau_{\mathcal{M}}$ (see, for example, Chapter 11 by Mitzenmacher and Upfal [28] for a detailed discussion).

We start by defining a scaled potential δ' such that, for all $x, y \in \Omega$, it holds that $\delta'(x, y) = \delta(x, y)/d$. Note that δ' takes values in $\{0\} \cup [1, D/d]$, and, for all $t \in \mathbb{N}$, it holds that

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

Further, for all $x, y \in \Omega$, by the linearity of expectation and by equation (3), it holds that

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

and, by equation (4), that

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

We define the stochastic processes whose expected hitting time we bound by $\varphi = (\varphi_t)_{t\in\mathbb{N}}$, where $\varphi_t = \delta'(X_t, Y_t)$. Further, for all $x \in [0, D/d]$, let

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

and let $\psi = (\psi_t)_{t\in\mathbb{N}}$ with $\psi_t = \overline{\ln}(\varphi_t)$ for all $t \in \mathbb{N}$. Note that

$$\psi_t = -2\ln(2) \leftrightarrow \varphi_t = 0 \leftrightarrow X_t = Y_t.$$

Let $T = \inf_{t\in\mathbb{N}}\{\psi_t \leq -2\ln(2)\}$. To obtain the desired bound on the mixing time, we bound $\mathrm{E}[T \mid X_0 = x, Y_0 = y]$ for every pair $x, y \in \Omega$.

To this end, we aim to apply Theorem 26 to ψ (with the natural filtration of $(X_t, Y_t)_{t\in\mathbb{N}}$) together with Remark 27, which requires showing, for all $t \in \mathbb{N}$, all $x \in \mathrm{rng}(X_t)$, and all $y \in \mathrm{rng}(Y_t)$ such that there is an instantiation ω from the sample space with $X_t(\omega) = x$, $Y_t(\omega) = y$, and $t < T(\omega)$, that $\mathrm{E}[\psi_{t+1} \mid X_t = x, Y_t = y] \leq \psi_t$ (equation (13)) and obtaining a lower bound on $\mathrm{E}\left[(\psi_{t+1} - \psi_t)^2 \,\middle|\, X_t = x, Y_t = y\right]$ (equation (14)).

Let $t \in \mathbb{N}$, $x \in \mathrm{rng}(X_t)$, and $y \in \mathrm{rng}(Y_t)$, all as described above. Note that this implies $\psi_t \in [0, \ln(D/d)]$ and $X_t \neq Y_t$. Further note that $\overline{\ln}$ is a concave function and that $\varphi_t \geq 1$. Applying Jensen's inequality to the conditional expectation, we obtain

$$\mathrm{E}[\psi_{t+1} \mid X_t = x, Y_t = y] \leq \overline{\ln}(\mathrm{E}[\varphi_{t+1} \mid X_t = x, Y_t = y]) \leq \overline{\ln}(\varphi_t) = \psi_t,$$

which shows equation (13).

We proceed by bounding $\mathrm{E}\big[(\psi_{t+1}-\psi_t)^2 \,\big|\, X_t = x, Y_t = y\big]$ from below. Let A be the event that ψ jumps from $\psi_t \geq 0$ directly to $\psi_{t+1} = -2\ln(2)$ (i.e., $\varphi_t \geq 1$ and $\varphi_{t+1} = 0$). The positive self-loop probability of $\mathcal{M}$ implies that $\Pr[A \mid X_t = x, Y_t = y] < 1$ and $\Pr\big[\overline{A} \,\big|\, X_t = x, Y_t = y\big] > 0$. By the law of total expectation, we obtain

$$\begin{aligned}
&\mathrm{E}\big[(\psi_{t+1}-\psi_t)^2 \,\big|\, X_t = x, Y_t = y\big] \\
&\quad = \mathrm{E}\big[(\psi_{t+1}-\psi_t)^2 \,\big|\, X_t = x, Y_t = y, A\big]\Pr[A \mid X_t = x, Y_t = y] \\
&\qquad + \mathrm{E}\Big[(\psi_{t+1}-\psi_t)^2 \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big](1 - \Pr[A \mid X_t = x, Y_t = y]).
\end{aligned}$$

We lower-bound each term in the sum separately. Because of $\psi_t \geq 0$, we have

$$\begin{aligned}
&\mathrm{E}\big[(\psi_{t+1}-\psi_t)^2 \,\big|\, X_t = x, Y_t = y, A\big]\Pr[A \mid X_t = x, Y_t = y] \\
&\qquad = \big(-2\ln(2) - \psi_t\big)^2 \cdot \Pr[A \mid X_t = x, Y_t = y] \\
&\qquad \geq 4\ln(2)^2 \cdot \Pr[A \mid X_t = x, Y_t = y]. \qquad (18)
\end{aligned}$$

Furthermore, because $\eta > 0$, by conditional Markov's inequality, we get

$$\begin{aligned}
&\mathrm{E}\Big[(\psi_{t+1}-\psi_t)^2 \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad \geq \ln(1+\eta)^2 \Pr\Big[(\psi_{t+1}-\psi_t)^2 \geq \ln(1+\eta)^2 \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad = \ln(1+\eta)^2 \Pr\Big[|\psi_{t+1}-\psi_t| \geq \ln(1+\eta) \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big].
\end{aligned}$$

We decompose the probability as

$$\begin{aligned}
&\Pr\Big[|\psi_{t+1}-\psi_t| \geq \ln(1+\eta) \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad = \Pr\Big[\psi_{t+1}-\psi_t \geq \ln(1+\eta) \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad\quad + \Pr\Big[\psi_{t+1}-\psi_t \leq -\ln(1+\eta) \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big].
\end{aligned}$$

We rewrite the first of these probabilities as

$$\begin{aligned}
&\Pr\Big[\psi_{t+1}-\psi_t \geq \ln(1+\eta) \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad = \Pr\left[\ln\left(\frac{\varphi_{t+1}}{\varphi_t}\right) \geq \ln(1+\eta) \,\middle|\, X_t = x, Y_t = y, \overline{A}\right] \\
&\qquad = \Pr\left[\frac{\varphi_{t+1}}{\varphi_t} \geq 1+\eta \,\middle|\, X_t = x, Y_t = y, \overline{A}\right] \\
&\qquad = \Pr\Big[\varphi_{t+1}-\varphi_t \geq \eta\varphi_t \,\Big|\, X_t = x, Y_t = y, \overline{A}\Big]. \qquad (19)
\end{aligned}$$

Since, for all $z \in (0,1)$, it holds that $-\ln(1+z) \geq \ln(1-z)$, we bound the other probability by

$$\begin{aligned}
&\Pr\left[\psi_{t+1} - \psi_t \leq -\ln(1+\eta) \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad \geq \Pr\left[\psi_{t+1} - \psi_t \leq \ln(1-\eta) \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad = \Pr\left[\ln\left(\frac{\varphi_{t+1}}{\varphi_t}\right) \leq \ln(1-\eta) \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad = \Pr\left[\frac{\varphi_{t+1}}{\varphi_t} \leq 1-\eta \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad = \Pr\left[\varphi_{t+1} - \varphi_t \leq -\eta\varphi_t \;\middle|\; X_t = x, Y_t = y, \overline{A}\right]. \qquad (20)
\end{aligned}$$

Combining equations (19) and (20), we obtain

$$\begin{aligned}
&\Pr\left[|\psi_{t+1} - \psi_t| \geq \ln(1+\eta) \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad \geq \Pr\left[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \;\middle|\; X_t = x, Y_t = y, \overline{A}\right]. \qquad (21)
\end{aligned}$$

For bounding the right-hand side of equation (21), recall that we consider an instantiation of the process such that $\varphi_t \geq 1$. Consider the probability that φ takes steps of at least size $\eta\varphi_t$. By the law of total probability,

$$\begin{aligned}
&\Pr[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \mid X_t = x, Y_t = y] \\
&\quad = \Pr[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \mid X_t = x, Y_t = y, A]\Pr[A \mid X_t = x, Y_t = y] \\
&\qquad + \Pr\left[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \;\middle|\; X_t = x, Y_t = y, \overline{A}\right](1 - \Pr[A \mid X_t = x, Y_t = y]).
\end{aligned}$$

Since A is the event to go from $\varphi_t \geq 1$ to $\varphi_{t+1} = 0$, for all $\eta \in (0,1)$, it holds that

$$\Pr[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \mid X_t = x, Y_t = y, A] = 1.$$

Thus, and by equation (4), we obtain

$$\begin{aligned}
&\Pr\left[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \;\middle|\; X_t = x, Y_t = y, \overline{A}\right] \\
&\quad = \frac{\Pr[|\varphi_{t+1} - \varphi_t| \geq \eta\varphi_t \mid X_t = x, Y_t = y] - \Pr[A \mid X_t = x, Y_t = y]}{1 - \Pr[A \mid X_t = x, Y_t = y]} \\
&\quad \geq \frac{\kappa - \Pr[A \mid X_t = x, Y_t = y]}{1 - \Pr[A \mid X_t = x, Y_t = y]}. \qquad (22)
\end{aligned}$$

By combining equations (21) and (22) and observing that for $\varphi_t \geq 1$ it holds that $\varphi_t = e^{\psi_t}$, we get

$$\begin{aligned}
&\mathrm{E}\Big[(\psi_{t+1} - \psi_t)^2 \;\Big|\; X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad \geq \ln(1+\eta)^2 \Pr\Big[|\psi_{t+1} - \psi_t| \geq \ln(1+\eta) \;\Big|\; X_t = x, Y_t = y, \overline{A}\Big] \\
&\qquad \geq \ln(1+\eta)^2 \frac{\kappa - \Pr[A \mid X_t = x, Y_t = y]}{1 - \Pr[A \mid X_t = x, Y_t = y]}.
\end{aligned} \tag{23}$$

Last, we use equations (18) and (23) and that $\eta < 1$ implies $\ln(1+\eta) \leq \ln(2)$ to obtain

$$\begin{aligned}
&\mathrm{E}\Big[(\psi_{t+1} - \psi_t)^2 \;\Big|\; X_t = x, Y_t = y\Big] \\
&\qquad \geq 4\ln(2)^2 \Pr[A \mid X_t = x, Y_t = y] \\
&\qquad\quad + (1 - \Pr[A \mid X_t = x, Y_t = y]) \ln(1+\eta)^2 \frac{\kappa - \Pr[A \mid X_t = x, Y_t = y]}{1 - \Pr[A \mid X_t = x, Y_t = y]} \\
&\qquad = 4\ln(2)^2 \Pr[A \mid X_t = x, Y_t = y] + \ln(1+\eta)^2\kappa - \ln(1+\eta)^2 \Pr[A \mid X_t = x, Y_t = y] \\
&\qquad \geq 3\ln(2)^2 \Pr[A \mid X_t = x, Y_t = y] + \ln(1+\eta)^2\kappa \\
&\qquad \geq \ln(1+\eta)^2\kappa,
\end{aligned}$$

which shows equation (14).

By Theorem 26 and $\psi_0 \leq \ln(D/d)$, we get for all $x, y \in \Omega$ that

$$\mathrm{E}[T \mid X_0 = x, Y_0 = y] \leq \frac{\big(\ln(D/d) + 2\ln(2)\big)^2}{\ln(1+\eta)^2\kappa}.$$

This results in the desired mixing time bound of

$$\tau_{\mathcal{M}}(\varepsilon) \leq \frac{\big(\ln(D/d) + 2\ln(2)\big)^2}{\ln(1+\eta)^2\kappa} \ln\left(\frac{1}{\varepsilon}\right). \qquad \blacksquare$$