Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Random sampling for the monomer-dimer model on a lattice

J. van den Berg, R. Brouwer

Probability, Networks and Algorithms (PNA)

PNA-R9917 December 31, 1999

Report PNA-R9917
ISSN 1386-3711

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Random Sampling for the Monomer-Dimer Model on a Lattice

J. van den Berg
CWI
P.O. Box 94079, 1090 GB Amsterdam
The Netherlands

R. Brouwer
Department of Mathematics, Informatics, Physics and Astronomy
University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam
The Netherlands

Abstract

In the monomer-dimer model on a graph, each matching (collection of non-overlapping edges) M has a probability proportional to $\lambda^{|M|}$, where $\lambda > 0$ is the model parameter, and $|M|$ denotes the number of edges in M. An approximate random sample from the monomer-dimer distribution can be obtained by running an appropriate Markov chain (each step of which involves an elementary local change in the configuration) sufficiently long. Jerrum and Sinclair have shown (roughly speaking) that for an arbitrary graph and fixed λ and ε (the maximal allowed variational distance from the desired distribution), $O(|\Lambda|^2|E|)$ steps suffice, where $|E|$ is the number of edges and $|\Lambda|$ the number of vertices of the graph. For sufficiently nice subgraphs (e.g. cubes) of the d-dimensional cubic lattice we give an explicit recipe to generate approximate random samples in (asymptotically) siginificantly fewer steps, namely (for fixed λ and ε) $O(|\Lambda|\,(\ln|\Lambda|)^2)$.

1991 Mathematics Subject Classification: 60K35 82B20 82B80 82C20.
Key words and Phrases: Random sampling, Markov chain Monte Carlo simulation, Monomer-dimer model, Absence of phase transition.

Note: Work by the first author carried out under project PNA3.1 "Probability". A slightly different version of this paper has been accepted for publication in the Journal of Mathematical Physics.

1 Introduction

The monomer-dimer model, described below, originates from Statistical Physics, where it has been used to study the absorption of oxygen molecules on a surface, and the properties of a binary mixture. See Heilmann and Lieb [8] for further background and references. More recently, the model has also drawn much attention in Operations Research, Combinatorics and Graph Theory (see [9] and [10]). Throughout this paper, the size (number of elements) of a finite set A will be denoted by $|A|$.

Consider a finite, undirected graph $G = (\Lambda, E)$, where Λ is the set of vertices of G and E is the set of edges. A *matching* on G is a subset $M \subset E$ such that no two edges in M have a common endpoint. Let $\lambda > 0$ (this is the model parameter). Now assign to each matching M a probability proportional to $\lambda^{|M|}$.

Alternatively, define the state space $\Omega = \{0,1\}^E$. Elements of Ω (called *configurations* on E) are typically denoted by $\omega = (\omega_e, e \in E)$. The monomer-dimer distribution for G (with parameter λ) is then defined as

$$\mu(\omega) = \frac{\lambda^{|\omega|} I(\omega \text{ is allowable})}{Z(\lambda)}, \tag{1.1}$$

where 'ω is allowable' means that the set $\{e : \omega_e = 1\}$ is a matching, $|\omega|$ denotes the size of that set, and $Z(\lambda)$ is the normalization factor (*partition function*). It is clear that the two descriptions (one with state space the set of all matchings, the other with state space $\{0,1\}^E$), are equivalent, and both descriptions will be used in this paper.

To continue, we need some more terminology and notation:
If two vertices i and j are adjacent, we write $i \sim j$. The *degree* $deg(v)$ of a vertex v is defined as the number of edges that have v as an endpoint. If two edges e_1 and e_2 share a common endpoint, we write $e_1 \sim e_2$. Let, in the rest of this subsection, Δ be a subset of E. We denote the set $\{0,1\}^\Delta$ by Ω_Δ. Similarly, if $\omega \in \Omega$, then ω_Δ denotes the 'restriction' of ω to Δ, i.e. the element $(\omega_e : e \in \Delta)$ of Ω_Δ. If ω, $\omega' \in \Omega_\Delta$ we call an edge $e \in \Delta$ an *edge of disagreement* (w.r.t. the pair (ω, ω')) if $\omega_e \neq \omega'_e$, and we denote the set of all such edges by $V(\omega, \omega')$. The *boundary* of Δ, denoted by $\partial\Delta$, consists of all elements $e \in E \setminus \Delta$ such that $e \sim e'$ for some $e' \in \Delta$.

Let $\alpha \in \Omega_{\partial\Delta}$, (so α is a configuration on the boundary of Δ). The monomer-dimer distribution for Δ *with boundary condition* α is defined as follows:

$$\mu^\alpha_\Delta(\omega) = \frac{\lambda^{|\omega|} I(\omega \text{ is allowable w.r.t } \alpha)}{Z_\Delta(\lambda)}, \tag{1.2}$$

where allowable with respect to α means that the set $\{e \in \Delta : \omega_e = 1\} \cup \{e \in \partial\Delta : \alpha_e = 1\}$ is a matching.

It is easy to check that the monomer-dimer model satisfies the following Markov property: Let σ denote a random configuration on E and let $\Delta \subset E$. Then the conditional distribution of σ_Δ, given $\sigma_{E\setminus\Delta}$, equals $\mu^{\sigma_{\partial\Delta}}_\Delta$ (and hence depends only on $\sigma_{\partial\Delta}$.

The paper by van den Berg (1999) shows that the monomer-dimer model on a lattice has certain very strong spatial mixing properties. In the present paper (see Section 3) we show explicitly how this can be used to improve, for 'nice' subgraphs of a lattice, earlier results in the literature concerning the generation of (approximate) random samples. Apart from a theorem by Jerrum and Sinclair, which is stated without proof in Section 2, and some easy to verify results on variational distance and coupling (see also Section 2), this paper is practically self-contained.

2 Preliminaries

In this section we give the background needed in Section 3. First we present some general and quite well-known results on coupling and variational distance. Then we will state the earlier mentioned result by Jerrum and Sinclair. Finally we will present and prove a result which is very similar (but more convenient for our purpose) to a result in Van den Berg ([5]).
Throughout this section Ω denotes an arbitrary finite set.

2.1 Coupling and variational distance

Suppose we have two probability distributions μ_1 and μ_2 on Ω. The *variational distance* $d_V(\mu_1, \mu_2)$ is defined by:

$$d_V(\mu_1, \mu_2) = \frac{1}{2} \cdot \sum_{\omega \in \Omega} |\mu_1(\omega) - \mu_2(\omega)|. \tag{2.1}$$

Another (but equivalent) definition of variational distance is the following:

$$d_V(\mu_1, \mu_2) = \max_{A \subset \Omega} |\mu_1(A) - \mu_2(A)|. \tag{2.2}$$

This equivalence is quite easy to check.

Suppose we have two probability distributions μ_1 and μ_2 on Ω. A *coupling* $\mathcal{P}$ of μ_1 and μ_2 is a distribution on $\Omega \times \Omega$ which has the following properties:

$$\sum_{\omega_1 \in \Omega} \mathcal{P}(\omega_1, \omega_2) = \mu_2(\omega_2) \quad \text{for all } \omega_2 \in \Omega, \tag{2.3}$$

and

$$\sum_{\omega_2 \in \Omega} \mathcal{P}(\omega_1, \omega_2) = \mu_1(\omega_1) \quad \text{for all } \omega_1 \in \Omega, \tag{2.4}$$

i.e. the marginal distributions of $\mathcal{P}$ are μ_1 and μ_2. Similarly, one can define couplings of more than two probability distributions. A trivial example of a coupling is the product coupling $\mu_1 \times \mu_2$.

Define the event "unequal" as the set $\{(\omega_1, \omega_2) \in \Omega \times \Omega : \omega_1 \neq \omega_2\}$. Likewise, we define the event "equal" as the set $\{(\omega_1, \omega_2) \in \Omega \times \Omega : \omega_1 = \omega_2\}$. The following results, Proposition 2.1, Lemma 2.2 and Proposition 2.3, are quite standard and not difficult to prove.

Proposition 2.1. *Let μ_1, μ_2 and μ_3 be probability distributions on Ω, and let $\mathcal{P}_{1,2}$ and $\mathcal{P}_{2,3}$ be couplings of μ_1 and μ_2, and of μ_2 and μ_3 respectively. Then there exists a coupling $\mathcal{P}_{1,3}$ of μ_1 and μ_3 with the following property:*

$$\mathcal{P}_{1,3}(\text{ "unequal"}) \leq \mathcal{P}_{1,2}(\text{ "unequal"}) + \mathcal{P}_{2,3}(\text{ "unequal"}). \tag{2.5}$$

Proof. Suppose we first sample an element x_0 from the distribution $\mu_1(x)$, then sample an element y_0 from the conditional distribution $\mathcal{P}_{1,2}(x_0, \cdot)/\mu_1(x_0)$, and finally an element z_0 from the conditional distribution $\mathcal{P}_{2,3}(y_0, \cdot)/\mu_2(y_0)$. Intuitively, the distribution of the pair (x_0, z_0) is a candidate for the coupling we are looking for. More formally, define the following "3-coupling" $\mathcal{P}_{1,2,3}$ of μ_1, μ_2 and μ_3.

$$\mathcal{P}_{1,2,3}(x, y, z) = \begin{cases} \frac{\mathcal{P}_{1,2}(x,y) \cdot \mathcal{P}_{2,3}(y,z)}{\mu_2(y)} & \text{if } \mu_2(y) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

It is easy to check that $\sum_{x,y} \mathcal{P}_{1,2,3}(x, y, z) = \mu_3(z)$, $\sum_{y,z} \mathcal{P}_{1,2,3}(x, y, z) = \mu_1(x)$, and $\sum_{x,z} \mathcal{P}_{1,2,3}(x, y, z) = \mu_2(y)$, so that $\mathcal{P}_{1,2,3}$ is indeed a coupling of μ_1, μ_2 and μ_3. From this it follows immediately that $\mathcal{P}_{1,3}$ defined by

$$\mathcal{P}_{1,3}(x, z) = \sum_y \mathcal{P}_{1,2,3}(x, y, z), \tag{2.7}$$

is a coupling of μ_1 and μ_3.

We also have $\sum_x \mathcal{P}_{1,2,3}(x, y, z) = \mathcal{P}_{2,3}(y, z)$, and $\sum_z \mathcal{P}_{1,2,3}(x, y, z) = \mathcal{P}_{1,2}(x, y)$. Property (2.5) now follows easily from the above, using

$$\{(x, y, z) : x \neq z\} \subset \{\{(x, y, z) : x \neq y\} \cup \{(x, y, z) : y \neq z\}\}.$$

□

We proceed with a lemma that states some basic properties of variational distance.

Lemma 2.2. *Let μ, μ' and ν be probability distributions on Ω. We have:*

1. $d_V(\mu,\nu) \geq 0$
2. $d_V(\mu,\nu) = d_V(\nu,\mu)$
3. $d_V(\mu,\nu) \leq d_V(\mu,\mu') + d_V(\mu',\nu)$
4. $d_V(\gamma\cdot\mu + (1-\gamma)\cdot\mu',\nu) \leq \gamma\cdot d_V(\mu,\nu) + (1-\gamma)\cdot d_V(\mu',\nu)$ *for all* $\gamma \in [0,1]$

Proof. The first two properties follow directly from the definition of variational distance, the third property follows from the triangle inequality, and the proof of the fourth property is an easy exercise. □

The following proposition relates the two notions of variational distance and couplings. Recall the notions "equal" and "unequal" defined earlier.

Proposition 2.3. *For all probability distributions μ and ν on Ω:*

$$d_V(\mu,\nu) = \min_{\mathcal{P}} \mathcal{P}(\text{ "unequal"}), \tag{2.8}$$

where the minimum is taken over all couplings $\mathcal{P}$ of μ and ν.

Proof. Let $\mathcal{P}$ be a coupling of μ and ν. It is a useful (for those not yet familiar with these notions) and fairly easy exercise to show that $d_V(\mu,\nu) \leq \mathcal{P}(\text{"unequal"})$. To complete the proof, we need to construct a coupling $\tilde{\mathcal{P}}$ for which

$$d_V(\mu,\nu) = \tilde{\mathcal{P}}(\text{"unequal"}). \tag{2.9}$$

To achieve this, define

$$\tilde{\mathcal{P}}(x,x) = \min\{\mu(x),\nu(x)\}. \tag{2.10}$$

With this definition, the probability $\tilde{\mathcal{P}}(\text{"equal"})$ (and hence the probability $\tilde{\mathcal{P}}(\text{"unequal"})$) is fixed. It is easy to show that if a coupling $\tilde{\mathcal{P}}$, satisfying (2.10) exists, it also satisfies (2.9). We still have to define $\tilde{\mathcal{P}}(x,y)$ for $x \neq y$ in such way that $\tilde{\mathcal{P}}$ is truly a coupling. In general, there are several ways to do this. For example, we can define:

$$\tilde{\mathcal{P}}(x,y) = \frac{(\mu(x) - \min\{\mu(x),\nu(x)\})\,(\nu(y) - \min\{\mu(y),\nu(y)\})}{\tilde{\mathcal{P}}(\text{ "unequal" })}, \tag{2.11}$$

for all $x \neq y$. Straightforward calculation will show that $\tilde{\mathcal{P}}$ is indeed a coupling. □

A coupling that reaches the minimum in Proposition (2.3) is called *optimal*.
For an extensive treatment of coupling methods, see [11].

2.2 Mixing times and the Jerrum-Sinclair result

Suppose we have an ergodic Markov chain on Ω. Let π be the stationary distribution of this chain and let $x \in \Omega$. Let $\mu^{x,t}$ denote the distribution of the Markov chain at time t , when it has started in initial state x. Let $\varepsilon > 0$. Define the *mixing time with respect to initial state* x of the Markov chain as follows:

$$\tau_x(\varepsilon) = \min_t\{d_V(\mu^{x,t'},\pi) \leq \varepsilon \text{ for all } t' \geq t\}. \tag{2.12}$$

The *(total) mixing time* of the Markov chain is defined by

$$\tau(\varepsilon) = \max_{x\in\Omega} \tau_x(\varepsilon). \tag{2.13}$$

Jerrum and Sinclair [9] have studied the mixing time of a suitable Markov chain for the monomer-dimer model. More precisely, they have proved the following: Let $G = (\Lambda, E)$ be a finite graph, and let $\Omega = \{$ all matchings on G $\}$. Consider the monomer-dimer distribution with parameter $\lambda > 0$ on Ω. Denote this distribution by π_λ. To sample from this distribution, they study a specific Markov chain mc_λ whose stationary distribution is π_λ. A transition $M \to M'$ in the Markov chain mc_λ is described as follows:

1. With probability $\frac{1}{2}$ let $M' = M$; otherwise
2. Choose uniformly at random an edge $e = (u, v) \in E$.
   Define M' as follows:

$$M' = \begin{cases} M + e & \text{if u,v unmatched in M,} \\ M - e & \text{if } e \in M, \\ M + e - e' & \text{if either u or v (but noth both) is matched and e' is the matching edge,} \\ M & \text{otherwise.} \end{cases}$$

3. Move to M' with probability $\min\{1, \frac{\pi_\lambda(M')}{\pi_\lambda(M)}\}$.

Note that mc_λ is aperiodic because $P(M, M) \geq \frac{1}{2} > 0$ for all matchings M. It is also clear that mc_λ is irreducible (because all matchings communicate via the empty matching), and easy to check that mc_λ satisfies the detailed balance condition

$$\pi_\lambda(M)P(M, M') = \pi_\lambda(M')P(M', M). \tag{2.14}$$

We conclude that mc_λ has stationary distribution π_λ and that, for any initial state, the distribution of the chain converges to π_λ. By a clever application of the so-called canonical path method, Jerrum and Sinclair [9] obtained the following bound for the mixing time of mc_λ.

Theorem 2.4. *The mixing time of mc_λ satisfies*[1] *:*

$$\tau(\varepsilon) \leq 4|E|n\lambda'\big(n \ln 4n + n \ln \lambda' + \ln(\varepsilon^{-1})\big), \tag{2.16}$$

where $\lambda' = \max\{1, \lambda\}$, and $n = \lceil |\Lambda|/2 \rceil$.

2.3 A result on the spatial dependencies of the monomer-dimer model.

The following Theorem is very similar to a result in Section 3 of [5] (the ideas in which go back to [2]- [4]), but slightly stronger and more convenient for our purpose. Therefore, and for completeness, we give a fairly detailed proof. Recall the definitions of $V(\omega, \omega')$ and $deg(v)$ in Section 1.

Theorem 2.5. *Let, for a given value of λ, μ be the monomer-dimer distribution on a graph $G = (V, E)$. Let $\Delta \subset E$ and let $\alpha, \beta \in \Omega_{\partial\Delta}$. Then a coupling $\mathcal{P}_{\Delta,\alpha,\beta}$ of μ^α_Δ and μ^β_Δ exists such that:*

$$\mathcal{E}_{\Delta,\alpha,\beta}\left(|\{e \in \Delta : e \text{ edge of disagreement}\}|\right) \leq 2c\lambda \cdot |V(\alpha, \beta)|, \tag{2.17}$$

where $\mathcal{E}_{\Delta,\alpha,\beta}$ denotes the expectation with respect to $\mathcal{P}_{\Delta,\alpha,\beta}$ and c equals $\max_{v\in\Delta}\{deg(v)\} - 1$.

Proof. Let Δ, α and β be as in the statement of the theorem. We construct the desired coupling $\mathcal{P}_{\Delta,\alpha,\beta}$ on $\Omega_\Delta \times \Omega_\Delta$ as follows:
Let x and y be independent configurations with distribution μ^α_Δ and μ^β_Δ respectively. Modify these configurations in the following way. For every $\tilde{e} \in V(\alpha, \beta)$, define the set

$$\begin{aligned} \tilde{e}_{diff}(x, y) = \{e \in V(x, y) : \exists \text{ a sequence } \tilde{e} \sim e_1 \sim e_2 \sim \cdots \sim e_n = e \\ \text{of distinct edges in } \Delta, \text{with } \forall i \in \{1 \cdots n\} : x_{e_i} \neq y_{e_i}\}. \end{aligned} \tag{2.18}$$

We call such a sequence a *path of disagreement* of length n from v to e, where v is the common endpoint of $\tilde{e}$ and e_1. Let the set $DIFF_{x,y,\alpha,\beta}$ be the union of paths of disagreement leaving from $V(\alpha, \beta)$:

$$DIFF_{x,y,\alpha,\beta} := \bigcup_{\tilde{e}\in V(\alpha,\beta)} \tilde{e}_{diff}(x, y). \tag{2.19}$$

[1] In fact, Proposition 12.4 of Jerrum and Sinclair [9] states

$$\tau(\varepsilon) \leq 4|E|n\lambda'\big(n(\ln n + \ln \lambda') + \ln(\varepsilon^{-1})\big). \tag{2.15}$$

However, we could only verify the proof when the factor $\ln n$ is replaced by $\ln 4n$, in 2.15.

The modified configurations $\tilde{x}$ and $\tilde{y}$ are defined by:

$$\begin{aligned} \tilde{x}_e &= \begin{cases} x_e & \text{if } e \in DIFF_{x,y,\alpha,\beta}, \\ y_e & \text{else.} \end{cases} \\ \tilde{y}_e &= \quad y_e \quad \text{for all } e \in \Delta. \end{aligned} \tag{2.20}$$

Note that the configurations $\tilde{x}$ and $\tilde{y}$ only differ from each other on $DIFF_{\tilde{x},\tilde{y},\alpha,\beta}$ and that this set equals $DIFF_{x,y,\alpha,\beta}$. We define $\mathcal{P}_{\Delta,\alpha,\beta}$ as the distribution of the pair $(\tilde{x},\tilde{y})$ constructed as above.

Lemma 2.6. *The distribution of $\mathcal{P}_{\Delta,\alpha,\beta}$ defined above is indeed a coupling of μ^{α}_{Δ} and μ^{β}_{Δ}.*

Proof. Since we have defined $\mathcal{P}_{\Delta,\alpha,\beta}$ as the distribution of $(\tilde{x},\tilde{y})$ it is sufficient to prove that $\tilde{x}$ has distribution μ^{α}_{Δ} and $\tilde{y}$ has distribution μ^{β}_{Δ}. Clearly, since $\tilde{y}$ equals y it has distribution μ^{β}_{Δ}. It remains to show that $\tilde{x}$ has distribution μ^{α}_{Δ}. To do this first introduce configurations $\hat{x}$ and $\hat{y}$ as follows: In words, $(\hat{x},\hat{y})$ is the pair of configurations obtained from (x,y) by exchanging x and y on the set of edges that do not have a path of disagreement to $V(\alpha,\beta)$. More precisely,

$$\begin{aligned} \hat{x}_e &= \begin{cases} x_e & \text{if } e \in DIFF_{x,y,\alpha,\beta}, \\ y_e & \text{else.} \end{cases} \\ \hat{y}_e &= \begin{cases} y_e & \text{if } e \in DIFF_{x,y,\alpha,\beta}, \\ x_e & \text{else.} \end{cases} \end{aligned} \tag{2.21}$$

By an appropriate use of the Markov property (see the proof of Lemma 1 in [5]), the pair $(\hat{x},\hat{y})$ has the same distribution as the pair (x,y). Finally, from the definitions of $\hat{x}$ and $\tilde{x}$, it follows that $\hat{x} = \tilde{x}$. Hence $\tilde{x}$ has distribution μ^{α}_{Δ}.

□

We now show that this coupling $\mathcal{P}_{\Delta,\alpha,\beta}$ has property (2.17). First recall (see the note before Lemma 2.6) that l.h.s. of (2.17) is equal to the expected size of $DIFF_{x,y,\alpha,\beta}$, where x and y are drawn independently from μ^{α}_{Δ} and μ^{β}_{Δ} respectively. Therefore we study the paths of disagreement for the pair (x,y). So consider an edge $\tilde{e} \in V(\alpha,\beta)$, say $\tilde{e} = (v_1, v_2)$. Observe that if a path of disagreement of length k from v_1 exists, then this path is unique. (Otherwise, as one can easily check, there would be three distinct edges, which share a common endpoint, and on each of which $x \neq y$. But then at least two of these edges have $x = 1$, or at least two of these edges have $y = 1$, which contradicts the fact that that x and y are allowable.) For v_2 a similar statement holds. Define $l_1(\tilde{e})$ ($l_2(\tilde{e})$) as the path of disagreement of maximal length, starting from v_1 (v_2 respectively). From the above observations we conclude that the l.h.s. of (2.17) is at most

$$\begin{aligned} &\sum_{\tilde{e} \in V(\alpha,\beta)} E\left[|l_1(\tilde{e})| + |l_2(\tilde{e})|\right] \\ &= \sum_{\tilde{e} \in V(\alpha,\beta)} \sum_{k=1}^{\infty} P(|l_1(\tilde{e})| \geq k) + P(|l_2(\tilde{e})| \geq k). \end{aligned} \tag{2.22}$$

To complete the proof of Theorem 2.5 we must, in view of equation (2.22), bound the probability

$$P(l_1(\tilde{e}) \text{ has length} \geq k), \tag{2.23}$$

and its analog for $l_2(\tilde{e})$. Before we do this, we first state a simple general lemma. Consider the monomer-dimer model on the very special 'star-shape' graph, which consists of n edges and $n+1$ vertices, one of which (the 'center of the star') has one edge to each of the other n vertices. It is clear that each allowable configuration has either 0 or 1 edge with value 1, and that the latter has probability $\lambda n/(1+\lambda n)$. Note that this is increasing in n. This observation, together with the Markov property mentioned in Section 1 (below ((1.2))) implies immediately the following:

Lemma 2.7. *Consider the monomer-dimer model with parameter λ on an arbitrary finite graph G. Let v be a vertex of G and let A be a subset of the edges of v. Then the conditional probability that there exists an edge in A with value 1, given the values of all edges outside A, is at most*

$$\frac{\lambda\,|A|}{1+\lambda\,|A|}.$$

We now proceed with the proof of Theorem 2.5. Suppose a path of disagreement of length k exists. What is the conditional probability that a path of disagreement of length $k+1$ exists? Let $e_1 \sim \cdots \sim e_k = e$ be the (unique) path of length k leaving from v_1, so that $\tilde{e} \sim e_1$. By the uniqueness property mentioned before, we have that the path of disagreement of length $k+1$ (if it exists) is an extension of the path of length k. Define

$$Adj(e) = \{b \in \Delta : b \sim e \text{ and } b \not\sim e_{k-1}\}. \tag{2.24}$$

Note that $|Adj(e)| \leq c$, with c as in the statement of the theorem.

By assumption $x_e = 0$ and $y_e = 1$ or vice versa. Without loss of generality we assume the former. Since y is a matching, we have $y_b = 0$ for every edge $b \in Adj(e)$. Hence we have a path of disagreement of length $k+1$, if and only if an edge $a \in Adj(e)$ exists with $x_a = 1$. By Lemma 2.7 above, the (conditional) probability of this event is at most $\lambda c/\lambda c + 1$. Iterating the above we get

$$\mathcal{P}_{\Delta,\alpha,\beta}(l_1(\tilde{e}) \text{ has length} \geq k) \leq \left(\frac{\lambda c}{\lambda c+1}\right)^k. \tag{2.25}$$

The same result holds for $l_2(\tilde{e})$.

Combining (2.25) with (2.22), it follows that the l.h.s. of (2.17) is at most

$$2 \sum_{\tilde{e} \in V(\alpha,\beta)} \sum_{k=1}^{\infty} \left(\frac{\lambda c}{\lambda c+1}\right)^k = 2\lambda c \cdot |V(\alpha,\beta)|. \tag{2.26}$$

This completes the proof of Theorem 2.5. □

Remark 2.8. In Sections 3.1 and 3.2, we will only work with d-dimensional hypercubes Δ. For such sets Δ, each edge on the boundary $\partial\Delta$ has exactly one vertex in common with an edge in the box Δ. For these special cases, the above result is improved by a factor 2, so that

$$\mathcal{E}_{\Delta,\alpha,\beta}\left(\#\{e \in \Delta : e \text{ edge of disagreement}\}\right) \leq \lambda(2d-1) \cdot |V(\alpha,\beta)|, \tag{2.27}$$

for every hypercube Δ.

3 Random sampling on subgraphs of the d-dimensional lattice

3.1 Description and motivation of the method

In subsection 2.2 we stated the Jerrum-Sinclair result. This result holds for general graphs. In the present section we study certain specifically 'nice' graphs, say a d-dimensional torus (described more precisely below). Suppose we want to sample (approximately) from the monomer-dimer model for such a graph. According to the Jerrum-Sinclair result (Theorem 2.4) we can do this by running the Markov chain mc_λ (described in subsection 2.2) a number of steps given by (2.16). For the torus this is, for fixed λ and ε, asymptotically of order $(\text{Volume})^3 \times \log(\text{Volume})$. Here Volume is the number of edges in the graph (or the number of vertices, which for these graphs differs a constant factor from the number of edges).

Can this, for these special graphs, be improved? There are several possibilities. One approach is to use logarithmic Sobolev inequalities: the results on spatial dependencies in Subsection 2.3 imply a mixing condition which, in turn, following a quite general theory developed by Stroock and Zegarlinski

(see [16]), could lead to a bound on the mixing time of order Volume $\times$ log(Volume). (We write *could* because there is an extra, quite subtle, condition which has to be checked to obtain such a bound from the Stroock-Zegarlinski theory: see Theorem 1 in the survey paper [7] by Frigessi, Martinelli and Stander). This result would be very interesting, but when one really wants to generate random samples, one not only wants to know the *asymptotic order* of the mixing time, but one needs an *explicit upper bound* to carry out the algorithm. To get (reasonable) explicit bounds from the Stroock-Zegarlinski theory is probably a lot of work which (in our opinion) is certainly worth the effort.

However, in the present paper we follow a somewhat different approach, which is based on a small modification of coupling and rescaling arguments which have become quite standard (see Aizenman and Holley ([1]), and Martinelli and Olivieri ([13]). This approach has the advantage that it gives, with relatively simple and few computations, an explicit bound whose asymptotic order is 'only a little worse' than the above mentioned Volume $\times$ log(Volume). (We get an extra factor of order log(Volume)).

Our approach is to combine (using rescaling and coupling arguments) the result of Jerrum and Sinclair (Theorem 2.4) with the result on spatial dependencies in Subsection 2.3. Although this approach applies to a larger class of graphs (see Remark 2 Subsection 3.3), we concentrate for simplicity on a graph Γ, which corresponds to a d-dimensional torus. More precisely, let N be a positive integer, and define Γ as the pair $(\Lambda_\Gamma, E_\Gamma)$ where the set of vertices Λ_Γ is defined as

$$\Lambda_\Gamma := \{0, \ldots, N\}^d, \tag{3.1}$$

and the set of edges E_Γ is

$$E_\Gamma := \{(v_1, v_2) : v_1, v_2 \in \Lambda_\Gamma \text{ and } |v_1 - v_2| = 1 \pmod{(N-1)}\}, \tag{3.2}$$

where $|\cdot|$ denotes the l_1 distance. We would like to sample from the monomer-dimer distribution π_Γ with parameter λ on this graph.

One way of approximate sampling from this distribution on Γ is the following: Let $\delta > 0$. Let Δ be a d-dimensional cube of length l. (Here l depends on λ and d; a suitable value will be determined later.) More precisely, Δ is the following set of edges.

$$\Delta := \{(v_1, v_2) : v_1, v_2 \in \{0, \ldots, l\}^d \text{ and } |v_1 - v_2| = 1\}. \tag{3.3}$$

Let $\hat{X}(t), t = 0, 1, \ldots$ be the Markov chain with state space $\{0,1\}^{E_\Gamma}$, which starts in some $x_0 \in \Gamma$, and of which the transitions are described as follows:. Suppose $\hat{X}(t) = x$. Choose u.a.r. a vertex $i \in \Lambda_\Gamma$. Let $\tilde{\Delta}$ be the box Δ shifted over i in the torus, i.e.

$$\tilde{\Delta} = \{((v_1 + i) \pmod N, (v_2 + i) \pmod N) : (v_1, v_2) \in \Delta\}. \tag{3.4}$$

Consider the monomer-dimer distribution on $\tilde{\Delta}$ with boundary condition $x_{\partial\tilde{\Delta}}$ (and parameter λ), denoted by $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$. Now sample a configuration $\tilde{x}$ from this distribution. At time $t+1$ the state becomes:

$$\hat{X}(t+1)_e = \begin{cases} \hat{X}(t)_e & \text{if } e \notin \tilde{\Delta}, \\ \tilde{x}_e & \text{if } e \in \tilde{\Delta}. \end{cases} \tag{3.5}$$

It can be proved, using the spatial mixing properties mentioned before, that for l sufficiently large, the mixing time of this Markov chain for fixed λ is of order $\mathcal{O}(|\Lambda_\Gamma| \cdot \log|\Lambda_\Gamma|)$ i.e. of the same order we mentioned above in connection with logarithmic Sobolev inequalities.

However, a problem arises when one actually tries to execute this algorithm. How to compute the above mentioned distribution $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$? Even for relatively small Δ, this is a huge problem. For example, if $d = 2$ and the length l of the hypercube is 10, the state space of $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$ has already more than 2^{100} elements. In practice, this algorithm cannot be used.

One way to proceed now would be to use certain comparison theorems to obtain a bound on the mixing time of the Markov chain mc_λ for this model from the bound on the mixing time of the above 'block dynamics' (see Diaconis and Saloff-Coste [6], Randall and Tetali [14] and Martinelli [12]). However, these comparison arguments do not involve the two mixing times directly but indirectly (via

the spectral gap or logarithmic Sobolev constant). Since the relation between the mixing time (2.12)-(2.13) and these quantities is not tight, this method would introduce a factor of order Volume, so the final result would be of order $(\text{Volume})^2 \log(\text{Volume})$.

Therefore we do the following: Instead of drawing a configuration *exactly* from the distribution $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$ mentioned before, we will sample *approximately* from this distribution. In other words, we replace each (*macro*) step in the Markov chain by a number of *micro* steps where each micro step corresponds with a transition of the Markov chain mc_λ (on $\tilde{\Delta}$, with boundary condition $x_{\partial\tilde{\Delta}}$) studied by Jerrum and Sinclair. It will turn out that (for fixed δ and λ), the total number of micro steps needed to obtain a 'δ-close' approximate sample from π_Γ is at most of order $\text{Volume}\,(\log(\text{Volume}))^2$ (see Corollary 3.3 at the end of this subsection).

More precisely, the modified Markov chain, which we denote by $X(t), t = 0, 1, \ldots$, has the same state space and initial state as $\hat{X}(t)$, but the transitions are now as follows: Suppose $X(t) = x$. As before, choose u.a.r. a vertex $i \in \Gamma$; determine the box $\tilde{\Delta}$, and consider the monomer-dimer distribution $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$. We will approximate this distribution. To do this, first define

$$\varepsilon = \left(\frac{|\Delta| - \lambda(2d-1)\,|\partial\Delta|}{|E_\Gamma||\Delta|} \right) \frac{\delta}{2}. \tag{3.6}$$

The choice of this value will become clear later. Now consider the (auxiliary) Markov chain mc_λ (w.r.t the monomer- dimer model on $\tilde{\Delta}$, with boundary condition $x_{\partial\tilde{\Delta}}$) described in Subsection 2.2. Although the initial state does not matter in the computations below, it is natural to take it equal to $x_{\tilde{\Delta}}$. Denote the distribution of this chain at time t by $\nu_{\tilde{\Delta}}^{t,x}$. Let Δ^* be the set of vertices which are endpoints of edges of Δ. From Theorem 2.4 it follows that $\nu_{\tilde{\Delta}}^{t,x}$ converges to $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$, and that, if the number of steps made by that Markov chain is at least T, given by

$$T := |\Delta||\Delta^*|\lambda' \left[|\Delta^*| \ln(2|\Delta^*|) + (|\Delta^*|) \ln \lambda' + 2\ln(\varepsilon^{-1}) \right], \tag{3.7}$$

then

$$d_V(\nu_{\tilde{\Delta}}^{T,x}, \mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}) \leq \varepsilon. \tag{3.8}$$

Let $\tilde{x}$ be the configuration on $\tilde{\Delta}$ after T transitions; this is a sample from $\nu_{\tilde{\Delta}}^{T,x}$. Now take

$$X(t+1)_e = \begin{cases} X(t)_e & \text{if } e \notin \tilde{\Delta}, \\ \tilde{x}_e & \text{if } e \in \tilde{\Delta}. \end{cases} \tag{3.9}$$

This completes our description of a (macro) step in the Markov chain X.

In the next section we will give an upper bound for the number of macro steps after which the variational distance between μ^t (the distribution of $X(t)$ at time t) and π_Γ becomes smaller than δ. The total number of micro steps needed then simply follows from multiplying this by the number T in (3.7).

3.2 A bound on the number of steps

In this section we will bound the number of steps of the Markov chain $X(t)$ to approximately reach the stationary distribution π_Γ. First, we define a suitably coupled system $(X(t), Y(t))$, $t = 0, 1, \ldots$ where $X(t)$ is the Markov chain introduced in the previous subsection, and $Y(t)$ is a Markov chain with the same transition probabilities as $\hat{X}$ (which was also introduced in the previous subsection), but which starts with the stationary distribution π_Γ (and hence keeps this distribution). Using the results in Section 2, we will obtain an upper bound for the variational distance between the distributions of $X(t)$ and $Y(t)$ for every time t. This is done by studying the number of edges of disagreement $|V(X(t), Y(t))|$.

More precisely, let $X(0) = x_0$ and let $Y(0)$ be drawn from the distribution π_Γ. Suppose at time t, $X(t) = x$ and $Y(t) = y$. Now we follow the description of a transition of $X(t)$ given in the previous subsection. However, instead of sampling a single configuration $\tilde{x}$ on $\tilde{\Delta}$, we now sample a pair $(\tilde{x}, \tilde{y})$ as follows. First consider the following three distributions on $\Omega_{\tilde{\Delta}}$: $\nu_{\tilde{\Delta}}^{T,x}$, $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$ and $\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}$. Let

$$\mathcal{P}_{opt,\, \nu_{\tilde{\Delta}}^{T,x},\, \mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}} \tag{3.10}$$

be an optimal coupling of $\nu_{\tilde{\Delta}}^{T,x}$ and $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$, and

$$\mathcal{P}_{\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}},\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}} \tag{3.11}$$

be a coupling of $\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}$ and $\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}$ which satisfies Theorem 2.5. Finally, let

$$\mathcal{P}_{\nu_{\tilde{\Delta}}^{T,x},\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}} \tag{3.12}$$

be a coupling of $\nu_{\tilde{\Delta}}^{T,x}$ and $\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}$ obtained from the two previous couplings as described in the proof of Proposition 2.1. The expectation with respect to the distribution (3.10) is denoted by $\mathcal{E}_{opt,\nu_{\tilde{\Delta}}^{T,x},\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}}$. The expectations for the other two couplings are denoted similarly. Now, sample a pair $(\tilde{x},\tilde{y})$ from this last coupling (3.12). Now take

$$X(t+1)_e = \begin{cases} X(t)_e & \text{if } e \notin \tilde{\Delta}, \\ \tilde{x}_e & \text{if } e \in \tilde{\Delta}. \end{cases} \qquad Y(t+1)_e = \begin{cases} Y(t)_e & \text{if } e \notin \tilde{\Delta}, \\ \tilde{y}_e & \text{if } e \in \tilde{\Delta}. \end{cases} \tag{3.13}$$

This completes the description of the transitions of the pair $(X(t),Y(t))$. Note that $\tilde{x}$ has been drawn from $\nu_{\tilde{\Delta}}^{T,x}$ so that the Markov chain $X(t)$ has indeed the same transition probabilities as in Subsection 3.1. Similarly, note that $Y(t)$ has indeed distribution π_Γ for each t.

Let μ^t denote the distribution of $X(t)$. Let $E(t)$ denote the expectation of $|V(X(t),Y(t))|$. Using Proposition 2.3 we have:

$$d_V(\mu^t,\pi_\Gamma) \le P[X(t)\neq Y(t)] \le E(t). \tag{3.14}$$

Therefore we will study $E(t)$. In particular, we study the change in this quantity after one (macro) step of the coupled Markov chain:
Using a property analogous to equation (2.5) in Proposition 2.1 we get

$$\begin{aligned} &\mathcal{E}_{\nu_{\tilde{\Delta}}^{T,x},\,\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|) \le \\ &\quad \mathcal{E}_{opt,\,\nu_{\tilde{\Delta}}^{T,x},\,\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|) + \\ &\quad \mathcal{E}_{\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}},\,\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|). \end{aligned} \tag{3.15}$$

So we need upper bounds for the expectations in the r.h.s. of (3.15). By Theorem 2.5 (and equation (2.27))

$$\mathcal{E}_{\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}},\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|) \le \lambda(2d-1)|V(x_{\partial\tilde{\Delta}},y_{\partial\tilde{\Delta}})|. \tag{3.16}$$

Because the coupling $\mathcal{P}_{opt,\,\nu_{\tilde{\Delta}}^{T,x},\,\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}}$ is optimal, we have

$$\begin{aligned} \mathcal{E}_{opt,\,\nu_{\tilde{\Delta}}^{T,x},\,\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|) &\le |\tilde{\Delta}|\cdot\mathcal{P}_{opt,\nu_{\tilde{\Delta}}^{T,x},\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}}(\text{“unequal”}) \\ &= |\tilde{\Delta}|\cdot d_V(\nu_{\tilde{\Delta}}^{T,x},\mu_{\tilde{\Delta}}^{x_{\partial\tilde{\Delta}}}) \\ &\le \varepsilon\cdot|\tilde{\Delta}|. \end{aligned} \tag{3.17}$$

The last inequality follows from (3.8). Together, equations (3.15) to (3.17) yield:

$$\begin{aligned} &\mathcal{E}_{\nu_{\tilde{\Delta}}^{T,x},\mu_{\tilde{\Delta}}^{y_{\partial\tilde{\Delta}}}}(|\{e\in\tilde{\Delta} : e \text{ edge of disagreement}\}|) \le \\ &\qquad \varepsilon\cdot|\tilde{\Delta}| + \lambda\cdot(2d-1)|V(x_{\partial\tilde{\Delta}},y_{\partial\tilde{\Delta}})|. \end{aligned} \tag{3.18}$$

We now state and prove the following Lemma.

Lemma 3.1.

$$E(t+1) \le b\cdot E(t) + \varepsilon|\Delta|, \tag{3.19}$$

where

$$b := 1 - \frac{|\Delta| - \lambda(2d-1)|\partial\Delta|}{|E_\Gamma|}. \tag{3.20}$$

Proof. Let $M(t) = |V(X(t), Y(t))|$. Note that the expectation of $M(t)$ is equal to $E(t)$. Suppose that $X(t)$, $Y(t)$, i and hence $\tilde{\Delta}$ are known. Consider the conditional expectation of the number of edges of disagreement that disappear during the transition $t \to t+1$:

$$\begin{aligned} &E[M(t) - M(t+1) \mid X(t) = x, Y(t) = y, \tilde{\Delta} = A] \\ &\quad = |V(x_A, y_A)| - \mathcal{E}_{\nu_A^{T,x}, \mu_A^{y_{\partial A}}}(|V(\tilde{x}, \tilde{y})|). \end{aligned} \tag{3.21}$$

By (3.18) this is larger than or equal to

$$|V(x_A, y_A)| - \varepsilon|A| - \lambda(2d-1)|V(x_{\partial A}, y_{\partial A})|. \tag{3.22}$$

Averaging over A we get

$$\begin{aligned} &E[M(t) - M(t+1) \mid X(t) = x, Y(t) = y] \\ &\quad \geq E(|V(x_{\tilde{\Delta}}, y_{\tilde{\Delta}})|) - \varepsilon|\Delta| - \lambda(2d-1)\, E(|V(x_{\partial\tilde{\Delta}}, y_{\partial\tilde{\Delta}})|), \end{aligned} \tag{3.23}$$

where the expectation in the r.h.s. refers to the distribution of $\tilde{\Delta}$. Recall that this is the uniform distribution, so that by symmetry each edge of E_Γ has the same probability, $|\Delta|/|E_\Gamma|$, to belong to $\tilde{\Delta}$. Similarly, the probability that a given edge belongs to $\partial\tilde{\Delta}$ equals $|\partial\Delta|/|E_\Gamma|$. Hence the r.h.s. of (3.23) equals

$$\frac{|\Delta|}{|E_\Gamma|} \cdot |V(x,y)| - \varepsilon|\Delta| - \lambda(2d-1)\frac{|\partial\Delta|}{|E_\Gamma|}|V(x,y)|.$$

So we have

$$E\left[M(t) - M(t+1) | X(t), Y(t)\right] \geq M(t)\left(\frac{|\Delta| - \lambda(2d-1)|\partial\Delta|}{|E_\Gamma|}\right) - \varepsilon|\Delta|. \tag{3.24}$$

Taking expectations in (3.24), we get

$$E(t) - E(t+1) \geq E(t)\left(\frac{|\Delta| - \lambda(2d-1)|\partial\Delta|}{|E_\Gamma|}\right) - \varepsilon|\Delta|, \tag{3.25}$$

from which the lemma follows immediately. □

For the moment we assume that the following inequalities,

$$0 < 1 - \frac{|\Delta| - \lambda(2d-1)|\partial\Delta|}{|E_\Gamma|} < 1 \tag{3.26}$$

hold, and will come back to this later.

Iterating equation (3.19) we get:

$$\begin{aligned} E(t+1) &\leq b^{t+1}|E_\Gamma| + \varepsilon|\Delta| \cdot \sum_{i=0}^{t} b^i \\ &= b^{t+1}|E_\Gamma| + \frac{(1-b^{t+1})}{1-b}\varepsilon|\Delta| \\ &= b^{t+1}|E_\Gamma| + (1-b^{t+1})\frac{\delta}{2}. \end{aligned} \tag{3.27}$$

Here we use the definition of ε and b (see equations (3.6) and (3.20)), and the fact that $E(0) \leq |E_\Gamma|$.

With (3.14) this gives

$$d_V(\mu^t, \pi_\Gamma) \leq b^t|E_\Gamma| + (1-b^t)\frac{\delta}{2}. \tag{3.28}$$

If we want to find t, such that the above mentioned variational distance is smaller than δ, it suffices to solve

$$b^t|E_\Gamma| \leq \frac{\delta}{2}. \tag{3.29}$$

Taking logarithms on both sides of (3.29) and using that $\ln(1-x) \leq -x$ for $0 < x < 1$, we find that (3.29) holds if

$$t \geq \frac{(\ln(2|E_\Gamma|) + \ln(\delta^{-1})) \cdot |E_\Gamma|}{|\Delta| - \lambda(2d-1)|\partial\Delta|}. \tag{3.30}$$

Recall that every step of the Markov chain $X(t)$ is in fact a macro step which corresponds with T micro steps in some box $\tilde{\Delta}$, where T is given by (3.7). Hence the total number of micro steps $\tau(\delta)$ after which the distribution of $X(t)$ has variational distance $\leq \delta$ from π_Γ is at most T times the r.h.s. of (3.30), i.e.

$$\begin{aligned} \tau(\delta) &\leq |\Delta||\Delta^*|\lambda' \cdot \left[|\Delta^*| \ln(2|\Delta^*|) + |\Delta^*| \ln(\lambda') + 2\ln(\varepsilon^{-1})\right] \times \\ &\quad \Big(\frac{(\ln(2|E_\Gamma|) + \ln(\delta^{-1}))|E_\Gamma|}{|\Delta| - \lambda(2d-1)|\partial\Delta|}\Big), \end{aligned} \tag{3.31}$$

where $\varepsilon = \varepsilon(\delta)$ is defined as in (3.6). Optimization considerations on a simplified modification of the r.h.s. of (3.31) lead to the following choice of the length l of Δ:

$$l := \lceil \lambda(4d+2) \rceil. \tag{3.32}$$

Note that $|\Delta^*| = (l+1)^d$, $|\Delta| = dl(l+1)^{d-1}$, and $|\partial\Delta| = 2d(l+1)^{d-1}$, so that, with l given by (3.32),

$$|\Delta^*| = (\lceil \lambda(4d+2) \rceil + 1)^d, \tag{3.33}$$

$$|\Delta| = d\lceil \lambda(4d+2) \rceil (\lceil \lambda(4d+2) \rceil + 1)^{d-1}, \tag{3.34}$$

$$|\partial\Delta| = 2d(\lceil \lambda(4d+2) \rceil + 1)^{d-1}. \tag{3.35}$$

Using (3.33)-(3.35), it is easy to check that, for every $\lambda > 0$ and every $d \geq 2$, the above choice of l implies the upper bound in (3.26). The lower bound in (3.26) is satisfied if $|E_\Gamma| > |\Delta|$, i.e. (in terms of λ and d) if $|E_\Gamma|$ is larger than the r.h.s. of (3.34). Using (3.33), (3.34), (3.35) and (3.6), we can now express the upper bound (3.31) on $\tau(\delta)$ completely in terms of δ, λ, d and $|E_\Gamma|$.

Summary of the algorithm and the main result
Concluding, we can state the following: Let $0 < \delta < 1$ and $\lambda > 0$. Consider the monomer-dimer distribution π_Γ with parameter λ on the d-dimensional torus Γ, as described in Subsection 3.1. Take $l = \lceil (4d+2)\lambda \rceil$ and let Δ be the hypercube of length l as described in Subsection 3.1. Compute ε from (3.6). Finally, compute T for the above choice of l, as in (3.7). Consider the Markov chain $X(t)$ with state space $\{0,1\}^{E_\Gamma}$, with transitions described as follows: Choose u.a.r. a vertex $i \in \Gamma$ and consider the box $\tilde{\Delta} = (i + \Delta)$. On this box (with the current $X(t)$ values on the boundary fixed) run the Markov chain mc_λ (described in Subsection 2.2) for T steps. These steps are called micro steps. This completes one transition (macro step) in the Markov chain $X(t)$.

Theorem 3.2. *In the algorithm described above, the number of micro steps $\tau(\delta)$ after which the distribution of $X(t)$ has variational distance smaller than or equal to δ from the stationary distribution π_Γ satisfies:*

$$\begin{aligned} \tau(\delta) &\leq T \cdot \frac{(\ln(2|E_\Gamma|) + \ln(\delta^{-1}))|E_\Gamma|}{|\Delta| - \lambda(2d-1)|\partial\Delta|} \\ &= |\Delta||\Delta^*|\lambda' \cdot \left[|\Delta^*| \ln(2|\Delta^*|) + |\Delta^*| \ln(\lambda') + 2\ln(\varepsilon^{-1})\right] \times \\ &\quad \Big(\frac{(\ln(2|E_\Gamma|) + \ln(\delta^{-1}))|E_\Gamma|}{|\Delta| - \lambda(2d-1)|\partial\Delta|}\Big), \end{aligned} \tag{3.36}$$

where $|\Delta^*|$, $|\Delta|$, $|\partial\Delta|$ and ε are given by (3.33), (3.34), (3.35) *and* (3.6) *respectively, and $\lambda' = \max(\lambda, 1)$.*

This result gives immediately (note the dependence of ε on $|E_\Gamma|$):
Corollary 3.3. For the algorithm above, if λ, d and δ are fixed, $\tau(\delta)$ satisfies

$$\tau(\delta) = \mathcal{O}\left(|E_\Gamma|(\ln(|E_\Gamma|))^2\right). \tag{3.37}$$

Remark 3.4. From (3.37) it follows that, for fixed λ and δ, on a large torus our bound is considerably better than the bound of Jerrum and Sinclair (Theorem 2.4). (Note that on a torus, the number of edges equals the dimension times the number of vertices, so $|E_\Gamma| = d|\Lambda_\Gamma|$.) However, our bound (3.31) involves a factor $\lambda^{2d}\lambda'$, while the bound of Jerrum and Sinclair is linear in λ', which is important for certain applications (see [9]). Hence if the size of the torus is relatively small with respect to λ, their bound is better than ours.

3.3 Remarks

1. Since the definition of the Markov chain $X(t)$ depends on δ, it is, strictly speaking, not correct to call $\tau(\delta)$ in (3.31) its mixing time.

2. The algorithm in the previous section, was described for a torus Γ. A similar result is still valid when the algorithm is applied to a sufficiently nice finite subset of $\mathbf{Z}^d$, for instance a hypercube $\Gamma = (\Lambda_\Gamma, E_\Gamma)$ where $\Lambda_\Gamma = \{0, \cdots, m\}^d$ and $E_\Gamma = \{(v_1, v_2) : v_1, v_2 \in \Lambda_\Gamma \text{ and } |v_1 - v_2| = 1\}$. Since Γ is not a torus, the box $\tilde{\Delta}$ must now be defined as $\tilde{\Delta} = (i + \Delta) \cap E_\Gamma$, where the vertex i is now the center of the box $\tilde{\Delta}$. The fact that in some cases $\tilde{\Delta}$ consists of roughly $|\Delta|/2^d$ elements leads to an increase of the size of a suitable Δ. This in turn leads to a number of micro steps needed in the procedure which is a constant (depending on the dimension d) larger than that for our torus.

3. One may think of several modifications of our computations to improve (decrease) the r.h.s. of (3.36). For instance it would be interesting and worth trying to improve Theorem 2.5. As to alternative methods, see the remark about logarithmic Sobolev inequalities in the beginning of this Section.

Acknowledgments

JvdB's interest in the mathematics of random sampling was triggered in the spring of 1997, when he participated, with financial support from Rutgers University, in the DIMACS program on Discrete Probability. He thanks M. Jerrum, P. Tetali, E. Vigoda and P. Winkler for interesting discussions during and after this program.

References

[1] Aizenman, M. and Holley, R. (1987). Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin-Shlosman regime. *Percolation Theory and Ergodic Theory of Infinite Particle Systems*, H. Kesten (ed.), IMS Volumes in Math. and Appl. **8**, 1-11.

[2] van den Berg, J. (1993). A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet. *Commun. Math. Phys.* **152**, 161-166.

[3] van den Berg, J. and Steif, J.E. (1994). Percolation and the hard-core lattice gas model. *Stoch. Proc. Appl.* **49**, 179-197.

[4] van den Berg, J. and Maes, C. (1994). Disagreement percolation in the study of Markov fields. *Ann. Probab.* **22**, 749-763.

[5] van den Berg, J. (1999). On the absence of phase transition in the monomer-dimer model, *Perplexing Problems in Probability* (Festschrift in honor of Harry Kesten), M. Bramson and R. Durrett eds., Birkhäuser, pp. 185-195.

[6] Diaconis, P. and Saloff-Coste, L., *Comparison Theorems for Reversible Markov Chains*, The Annals of Applied Probability, 1993, volume 3, no.3, pp. 696-730

[7] Frigessi, A., Martinelli, F. and Stander, J. (1997). Computational complexity of Markov Chain Monte Carlo methods for finite Markov random fields. *Biometrika* **84** (1), 1-18.

[8] Heilmann, O. and Lieb, E. (1972). Theory of monomer-dimer systems. *Commun. Math. Phys.* **25**, 190-232

[9] Jerrum, M. and Sinclair A. (1996). The Markov Chain Monte Carlo method: An approach to approximate counting and integration. in: Approximation Algorithms for NP-hard Problems, D. Hochbaum, ed., PWS Publishing, Boston

[10] Kahn, J. and Kayll, P.M., (1997) On the stochastic independence properties of hard-core distributions, *Combinatorica* **17**, 369-391.

[11] Lindvall, T. (1992). *Lectures on the Coupling Method*, Wiley series in Probability and Mathematical Statistics

[12] Martinelli, F., *Lecture notes on Glauber Dynamics for Discrete Spin Models*, Saint Flour Summer School in Probability, 1997

[13] Martinelli, F. and Olivieri, E. (1994). Approach to equilibrium in the one-phase region I: the attractive case. *Commun. Math. Phys.* **161**, 447-486.

[14] Randall, D. and Tetali, P. (1998). Analyzing Glauber dynamics using comparison of Markov chains. *Proc. of LATIN'98*, Campinas, Brazil; appeared in LNCS **48** (1998), 392-404.

[15] Sinclair, A., *Algorithms for random generation and counting, A Markov chain approach*, Progress in Theor.Comp.Science, Birkhäuser, 1993, pp. 42-62.

[16] Stroock, D.W. and Zegarlinski, B. (1992). The Logarithmic Sobolev inequality for discrete spin systems on a lattice. *Commun. Math. Phys.* **149**, 175-193.