Mixing Time for the Solid-on-Solid Model

Fabio Martinelli
Department of Mathematics
University of Roma Tre
Largo San Murialdo 1
00146 Roma, Italy.
martin@mat.uniroma3.it.

Alistair Sinclair*
Computer Science Division
University of California
Berkeley CA 94720-1776, U.S.A.
sinclair@cs.berkeley.edu.

ABSTRACT

We analyze the mixing time of a natural local Markov chain (the Glauber dynamics) on configurations of the solid-on-solid model of statistical physics. This model has been proposed, among other things, as an idealization of the behavior of contours in the Ising model at low temperatures. Our main result is an upper bound on the mixing time of $\tilde{O}(n^{3.5})$, which is tight within a factor of $\tilde{O}(\sqrt{n})$. The proof, which in addition gives insight into the actual evolution of the contours, requires the introduction of several novel analytical techniques that we conjecture will have other applications.

Categories and Subject Descriptors: G.3 [Probability and Statistics]: Markov processes; J.2 [Physical Sciences and Engineering]: Physics

General Terms: Algorithms, Theory

Keywords: Markov chain Monte Carlo (MCMC), mixing time, statistical physics, Ising model, Glauber dynamics

1. INTRODUCTION

In the $n \times n$ *solid-on-solid (SOS)* model [18, 19], a configuration is an assignment of an integer *height* $\eta(i) \in [0, n]$[†] to each of n positions $i \in [1, n]$, with fixed boundary conditions $\eta(0) = \eta(n+1) = 0$. The probability of a configuration $\eta = \{\eta(i)\}$ is given by the *Gibbs distribution*

$$\mu(\eta) = Z_\beta^{-1} \exp\left\{-\beta \sum_{i=1}^{n+1} |\eta(i-1) - \eta(i)|\right\}. \qquad (1)$$

Here β is a parameter called the "inverse temperature", and Z_β is a normalizing factor (the "partition function"). Thus a configuration of the SOS model may be pictured as a "contour" with fixed endpoints $(0, 0)$ and $(n+1, 0)$ (see Fig. 1(a)). The Gibbs distribution favors contours that are "smooth" (i.e., have no large jumps in height), this bias being more pronounced for larger values of β.

*Supported in part by NSF Grant CCF-0635153.

[†]Throughout the paper, unless otherwise stated $[a, b]$ will denote the integer points in the interval $[a, b]$.

STOC'09, May 31–June 2, 2009, Bethesda, Maryland, USA.

In this paper we analyze the *Glauber dynamics* for the SOS model. This is a natural local Markov chain on configurations which in each step updates the height at a randomly chosen position i from $\eta(i)$ to $\eta(i) \pm 1$; the transition probabilities are chosen so that the dynamics is reversible wrt the Gibbs distribution μ and thus converges to it from any initial configuration. Our goal is to determine the *mixing time*, i.e., the number of steps until the dynamics is close to its equilibrium distribution μ.

Although dynamics for the SOS and related models have been studied extensively in many contexts connected with the behavior of random surfaces (see, e.g., [6, 7, 8, 17]), to the best of our knowledge the mixing time has not been rigorously analyzed. Our specific motivations for studying this question are twofold. The first comes from a connection with the more familiar Ising model, whose Glauber dynamics has been the focus of much attention in both statistical physics and computer science (see, e.g., [1, 2, 12, 14, 23]). In the Ising model in an $n \times n$ box $\Lambda_n \subseteq \mathbb{Z}^2$, the configurations are assignments σ of spin values $\{+, -\}$ to the vertices of Λ_n. The Gibbs distribution is $\mu(\sigma) = Z_\beta^{-1} \exp(-\beta D(\sigma))$, where $D(\sigma)$ is the number of neighboring pairs of vertices in Λ_n whose spins differ and β is inverse temperature. One of the most important open problems concerning the dynamics of the Ising model is to determine the mixing time at low temperatures (large β), when the boundary conditions around the edges of Λ_n are fixed to be $+$. (In this case the Gibbs distribution puts most of the weight on configurations that are almost entirely $+$.) This can essentially be reduced to the following question: if the box is initially filled with $-$ spins, how long does it take until this large region of $-$ is destroyed under the influence of the boundary conditions? This in turn is equivalent to the question of how the outer contour

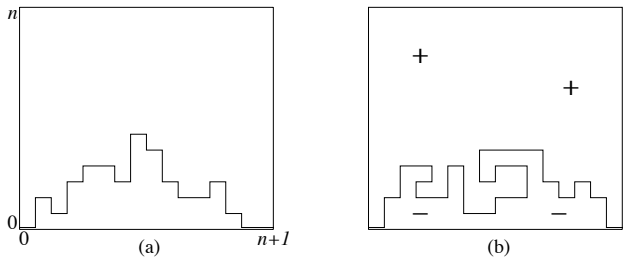

Figure 1: (a) A contour in the SOS model. (b) A contour in the Ising model.

of the − region contracts towards the center of the box. For large β, it is conjectured [9] that this happens in polynomial time (independent of β); however, only very weak upper bounds of the form $\exp(n^{1/2+\epsilon})$ are known [11] (except in the qualitatively different case of zero temperature, which is analyzed in [3]). The SOS model has been proposed [19] as an idealized model of this Ising contour, in which we think of the sites above and below the SOS contour as being + and − respectively. (Note that the sum $\sum_i |\eta(i-1) - \eta(i)|$ in the Gibbs distribution (1) is, up to an additive constant, exactly $D(\sigma)$ under this interpretation.) The mixing time is essentially the number of steps until the maximal contour (i.e., with $\eta(i) = n$ for $1 \le i \le n$) drops down close to the bottom of the box under the influence of the boundary conditions of height 0. The principal simplification here is that, unlike the Ising model, the SOS contour has no "overhangs" (see Fig. 1(b)). However, for large β one may hope that overhangs are rare, so the approximation should give useful insight into the behavior of the true Ising contour (see [5] for much more on this point). More concretely, we conjecture that the proof techniques we develop in this paper may be useful in analyzing the Ising model. (For recent progress in this direction, see [15].)

Our second motivation comes from the challenge that the SOS model poses to standard techniques. The two most natural approaches seem to be the following:

1. *Coupling.* It is not hard to check that, under the natural monotone coupling of the SOS model (see Section 2), the expected Hamming distance between two coupled copies of the dynamics does not increase. This leads to a mixing time bound of $\tilde{O}(n^5)$[‡] which, as we shall see, is very weak. It also gives little insight into the actual evolution of the contour.

2. *Comparison.* Another standard approach is to first analyze a "non-local" dynamics in which transitions are allowed to update the height $\eta(i)$ to any value in $[0, n]$. Typically, non-local dynamics are easier to analyze (see, e.g., [10, 24]). One can then use the machinery of Diaconis and Saloff-Coste [4] to relate the mixing time of the local dynamics to that of the non-local one, as was done, for example, by Randall and Tetali [21] for the related "lozenge tilings" model. However, since such comparisons proceed via the spectral gap, they are usually quite wasteful; in particular, for the SOS model this approach leads to a mixing time of $\tilde{O}(n^8)$.

In this paper we aim for a more refined analysis that, in addition to an almost tight bound, gives greater insight into the actual evolution of the contour in the SOS model. Our main result is the following:

THEOREM 1.1. *For the $n \times n$ SOS model at any inverse temperature $\beta > \frac{1}{2}\ln 3$, the mixing time is $\tilde{O}(n^{3.5})$.*

We note that the lower bound on β is imposed only to simplify our analysis, and is in practice not really a restriction as the regime of interest is low temperatures (large β). The bound on mixing time is tight up to a factor of $\sqrt{n}$ (and logarithmic factors), as a lower bound of $\Omega(n^3)$ follows from a straightforward argument (see Theorem 3.5 below).

Our analysis rests on the following four key ingredients:

(i) First, we give a tight analysis of the non-local dynamics mentioned above, showing that its mixing time is $O(n^3 \log n)$ (see Theorem 3.1). This analysis, which we believe to be of independent interest, follows an idea of Wilson, developed in the context of the lozenge tilings model [24], in using an eigenvector of the discrete Laplacian to obtain a contraction in distance. However, to get this approach to work in our setting we need to bound a certain "entropy repulsion" effect due to the height barriers at 0 and n (see Lemma 3.2).

(ii) We then relate the local to the non-local dynamics using a recent "censoring inequality" of Peres and Winkler [16], which says that *censoring* (i.e., not applying) some subset of updates in a monotone dynamics can only increase the distance from stationarity. This allows one to simulate a single move of the non-local dynamics, at position i, by censoring all local moves except those that update $\eta(i)$; by the censoring inequality, this can only increase the mixing time. As a result, the mixing time of the local dynamics is bounded above by that of the non-local dynamics times a factor related to the mixing time of the one-dimensional local process within the ith "column".

(iii) A naïve application of the censoring inequality would entail a substantial overhead of $O(n^2)$ from the mixing time within a column, which is essentially the square of the maximum height difference between the two neighboring columns. To overcome this, we need to control the height differences, or "gradients," along the contour. We do this by introducing a sequence of "bounding dynamics" with gradually decreasing boundary conditions; since the boundary conditions are—intuitively at least—the source of large gradients, this gives us control of the gradients. As a result, we are able to cut the simulation overhead between the local and non-local dynamics to $\tilde{O}(\sqrt{n})$. We note that this sequence of bounding dynamics captures some of the intuition about the actual evolution of the contour.

(iv) Making rigorous the above bound on gradients requires detailed information about the non-equilibrium shape of the contour, which is notoriously difficult to obtain. We get around this difficulty by starting the bounding dynamics *in equilibrium*, but *conditioned on a certain rare event A.* (The conditioning is necessary to ensure that the bounding property holds.) By choosing A such that its probability, though tiny, is nonetheless larger than the probability of large gradients in equilibrium, we can argue that large gradients do not occur during the evolution.

2. PRELIMINARIES

Gibbs distribution. We denote by $\Omega_n = [0, n]^n$ the set of all configurations $\eta = \{\eta(i)\}_{i=1}^n$ of the $n \times n$ solid-on-solid model, as defined in the Introduction. The probability of a configuration η is given by the Gibbs distribution defined in equation (1). This distribution induces a conditional distribution on the height $\eta(i)$ at position i, given the heights $\eta(i \pm 1)$ at its neighbors, as follows. Let $a = \min\{\eta(i-1), \eta(i+1)\}$, $b = \max\{\eta(i-1), \eta(i+1)\}$. Then $\mu_{ab}(j) := \Pr[\eta(i) = j \mid a, b]$ is given by

$$\mu_{ab}(j) = \begin{cases} e^{-\beta(b-a)-2\beta(a-j)}/Z & \text{if } 0 \le j < a; \\ e^{-\beta(b-a)}/Z & \text{if } a \le j \le b; \\ e^{-\beta(b-a)-2\beta(j-b)}/Z & \text{if } b < j \le n, \end{cases} \tag{2}$$

where Z is a normalizing factor. Note that μ_{ab} is uniform on the interval $[a, b]$ and decays exponentially (at a rate depending on β) outside it.

[‡]Throughout, the notation $\tilde{O}(\cdot)$ hides factors of polylog(n).

Single-site dynamics. Our goal is to analyze the *single-site Glauber dynamics*, which is a reversible Markov chain $\mathcal{M}_n^{\rm ss}$ on Ω_n with transitions defined as follows, where $\eta = \eta_t$ denotes the current configuration at time t:

1. Pick $i \in [1,n]$ u.a.r.
2. Replace $\eta(i)$ by $\max\{\eta(i)-1, 0\}$ or $\min\{\eta(i)+1, n\}$ with probabilities p^-, p^+ respectively, determined as follows (where a, b are the minimum and maximum heights of the neighbors, as above): if $\eta(i) \le a$ then $p^- = \frac{1}{4}e^{-2\beta}$, else $p^- = \frac{1}{4}$; if $\eta(i) \ge b$ then $p^+ = \frac{1}{4}e^{-2\beta}$, else $p^+ = \frac{1}{4}$. With the remaining probability $1-(p^-+p^+)$, leave $\eta(i)$ unchanged.

It is standard that $\mathcal{M}_n^{\rm ss}$ is an ergodic, reversible Markov chain that converges to the stationary distribution μ on Ω_n. Our goal is to estimate its *mixing time*, i.e., the number of steps required for the distribution to get close to μ from an arbitrary initial configuration.

Column dynamics. We will analyze $\mathcal{M}_n^{\rm ss}$ by first analyzing a related Glauber dynamics $\mathcal{M}_n^{\rm col}$, called the "column dynamics," that makes *non-local* moves. (The term "column" refers to the set $[0,n]$ of possible heights at i.) If the configuration at time t is $\eta_t = \eta$, $\mathcal{M}_n^{\rm col}$ makes a transition as follows:

1. Pick $i \in [1,n]$ u.a.r.
2. Replace $\eta(i)$ by a random height j chosen from the conditional distribution (2).

$\mathcal{M}_n^{\rm col}$ is again ergodic and reversible with stationary distribution μ. Note that both $\mathcal{M}_n^{\rm ss}$ and $\mathcal{M}_n^{\rm col}$ update the height at a randomly chosen position i in a manner that is reversible wrt the conditional distribution (2). The difference is that $\mathcal{M}_n^{\rm ss}$ considers only local moves (changing the height by ± 1), while $\mathcal{M}_n^{\rm col}$ allows the height at i to be set to any value.

Mixing time. Let $\mathcal{M}$ by any reversible Markov chain on Ω_n with stationary distribution μ. Following standard practice, we measure the convergence rate of $\mathcal{M}$ via the quantity

$$\tau_{\mathcal{M}}(\varepsilon) = \min\{t : \|\nu_t^\xi - \mu\| \le \varepsilon,\ \forall \xi \in \Omega_n\},$$

where ν_t^ξ denotes the distribution of the configuration at time t starting from configuration ξ at time 0, and $\|\cdot\|$ denotes variation distance. Thus $\tau_{\mathcal{M}}(\varepsilon)$ is the number of steps until the variation distance from μ drops to ε, for an arbitrary initial configuration. For concreteness we define the *mixing time* as $\tau_{\mathcal{M}}^{mix} = \tau_{\mathcal{M}}(1/2e)$; it is well known that $\tau_{\mathcal{M}}(\varepsilon) \le \lceil \ln \varepsilon^{-1} \rceil \times \tau_{\mathcal{M}}^{mix}$ for all $\varepsilon > 0$.

Monotonicity and coupling. We define a natural partial order on Ω_n as follows: for configurations $\eta, \xi \in \Omega_n$, we say that $\eta \preceq \xi$ iff $\eta(i) \le \xi(i)$ for all $i \in [1,n]$. Note that $\preceq$ has unique maximal and minimal elements $\eta^{\max}$ and $\eta^{\min}$ given by $\eta^{\max}(i) = n$ and $\eta^{\min}(i) = 0$ for $1 \le i \le n$.

A key fact we shall exploit throughout is the existence of a *complete coupling* of the Glauber dynamics (single-site or column) that is monotone w.r.t. $\preceq$. A complete coupling of a Markov chain $\mathcal{M}$ on Ω_n is a random function $f : \Omega_n \to \Omega_n$ that preserves the transition probabilities of $\mathcal{M}$, i.e., $\Pr_f[f(\eta) = \eta'] = \Pr_{\mathcal{M}}[\eta \to \eta']$ for all $\eta, \eta' \in \Omega_n$. Note that f simultaneously couples the evolution of the Markov chain at all configurations. For the column dynamics we define f as follows, where η denotes the current configuration:

1. Pick $i \in [1,n]$ and a real number $r \in [0,1]$ independently and u.a.r.
2. Let $g(k) = \sum_{j=0}^{k} \mu_{ab}(j)$ be the cumulative distribution function of the height at position i, given neighboring heights a, b. Set $\eta'(i) = \min\{k : g(k) \le r\}$.

An analogous definition holds for the single-site dynamics. It is simple to check that these couplings are *monotone* w.r.t. the partial order $\preceq$, in the sense that if $\eta_t \preceq \xi_t$, and η_{t+1}, ξ_{t+1} are the corresponding configurations at the next time step under the coupling, then $\eta_{t+1} \preceq \xi_{t+1}$.

A further standard fact we will need is that the mixing time of the Glauber dynamics is bounded above by the time until the coupled evolutions started in the two extremal configurations, $\eta^{\max}$ and $\eta^{\min}$, coincide with constant probability. More precisely:

PROPOSITION 2.1. **[20]** *Let $(\eta_t^{\max})$, $(\eta_t^{\min})$ denote the coupled evolutions of two copies of a monotone Glauber dynamics $\mathcal{M}$ on Ω_n started in configurations $\eta^{\max}$, $\eta^{\min}$ respectively. Then $\tau_{\mathcal{M}}(\varepsilon) \le \min\{t : \Pr[\eta_t^{\max} \ne \eta_t^{\min}] \le \varepsilon\}$.*

Censoring. In our analysis of the single-site dynamics, we shall also need a useful tool from recent work of Peres and Winkler, which says that *censoring* (i.e., not applying) any subset of updates in the dynamics can only increase the distance from stationarity. This so-called "censoring inequality" applies to any monotone single-site dynamics.

LEMMA 2.2. **[16]** *Suppose a monotone single-site dynamics is started in a random initial configuration with distribution ν_0 s.t. ν_0/μ is increasing[§] w.r.t. $\preceq$. Let ν denote the distribution after updates at positions $i_1, i_2, \ldots, i_m$, and ν' the distribution after updates at a subsequence of these positions $i_{j_1}, i_{j_2}, \ldots, i_{j_{m'}}$ (chosen a priori). Then $\|\nu - \mu\| \le \|\nu' - \mu\|$.*

Remark: [16, Thm 16.5] states this result for the special case in which ν_0 is concentrated on the maximal state $\eta^{\max}$. However, it is easy to see that the proof requires only the weaker assumption that ν_0/μ is increasing.

The censoring inequality can be used to relate the single-site and column dynamics via the following observation. If we censor all moves of the single-site dynamics except for those that update a certain position i, then after some fixed number of steps T (which depends on the mixing time of the single-site dynamics just within the ith column, with its neighbors fixed) we will, up to small error, have simulated one move of the column dynamics. By Lemma 2.2 the censoring can only slow down convergence of the single-site dynamics, so the mixing time of $\mathcal{M}_n^{\rm ss}$ is bounded above by roughly T times that of $\mathcal{M}_n^{\rm col}$. We shall use a more sophisticated version of this argument in Section 4.

3. THE COLUMN DYNAMICS

Our goal in this section is to provide a tight analysis of the column dynamics $\mathcal{M}_n^{\rm col}$. Specifically, we will prove:

THEOREM 3.1. *For any $\beta > \frac{1}{2}\ln 3$, the mixing time of the column dynamics $\mathcal{M}_n^{\rm col}$ is $O(n^3 \log n)$.*

[§]A real-valued function f on Ω_n is *increasing* w.r.t. $\preceq$ if $\eta \preceq \xi$ implies $f(\eta) \le f(\xi)$.

We believe this result, which we show is tight up to the $\log n$ factor (see Theorem 3.5 below) is interesting in its own right. It will also be a key ingredient in our analysis of the single-site dynamics later.

Remark: The lower bound on β is required only for convenience in the proof of Lemma 3.2 below, and is in any case not really a restriction as the interesting case is when β is large (low temperature). We believe that the proof of the lemma can be adapted to any $\beta > 0$ at the cost of further technical complication.

Recall that, if the current configuration of $\mathcal{M}_n^{\mathrm{col}}$ is η_t and we choose position $i \in [1,n]$ at the next step, then the new height $\eta_{t+1}(i)$ is drawn from the conditional distribution (2), where a, b are the minimum and maximum heights of the neighbors $\eta_t(i \pm 1)$. A key observation is that, under such a move, the expected value of the new height $\eta_{t+1}(i)$ is close to the average $\frac{a+b}{2}$ of its two neighbors; moreover, the error term satisfies a natural ordering property w.r.t. a, b.

LEMMA 3.2. *Let $\beta > \frac{1}{2}\ln 3$. In the above situation, and assuming $a + b \le n$, the expected value of the new height $\eta_{t+1}(i)$ satisfies*

$$\mathrm{E}[\eta_{t+1}(i) \mid a, b] = \frac{a+b}{2} + \varepsilon(a,b), \tag{3}$$

where $\varepsilon(a,b) \ge 0$. Moreover, $\varepsilon(a,b) \le \varepsilon(c,d)$ for any pair c, d with $c \le \min\{a,d\} \le \max\{a,d\} \le b$.

We defer the proof of the lemma, which is a technical calculation, to the full version of the paper [13]. However, the intuition is as follows. Note that the distribution of $\eta_{t+1}(i)$ is uniform on the interval $[a,b]$, and decays symmetrically on either side except for the effects of the barriers at heights 0 and n. Thus we would expect its mean to be close to $\frac{a+b}{2}$. The term $\varepsilon(a,b)$ captures the "entropy repulsion" effect of the barriers. This effect is more pronounced for pairs that are closer to 0, as is the case for the pair (c,d) in the lemma.

We can derive from Lemma 3.2 the following more symmetrical form that allows us to compare the heights of two ordered configurations under the monotone coupling. (The straightforward proof is omitted; see the full version [13].)

COROLLARY 3.3. *Let $\beta > \frac{1}{2}\ln 3$. Suppose η_t and ξ_t are two configurations satisfying $\eta_t \preceq \xi_t$, and let $a = \min\{\xi_t(i-1), \xi_t(i+1)\}$, $b = \max\{\xi_t(i-1), \xi_t(i+1)\}$, $c = \min\{\eta_t(i-1), \eta_t(i+1)\}$, $d = \max\{\eta_t(i-1), \eta_t(i+1)\}$. Then*

$$0 \le \mathrm{E}[\xi_{t+1}(i) \mid a, b] - \mathrm{E}[\eta_{t+1}(i) \mid c, d] \le \frac{a+b}{2} - \frac{c+d}{2}. \tag{4}$$

Armed with Corollary 3.3, we can now proceed to our analysis of $\mathcal{M}_n^{\mathrm{col}}$.

PROOF OF THEOREM 3.1. Following Proposition 2.1, it suffices to show that two coupled copies of $\mathcal{M}_n^{\mathrm{col}}$, started in configurations $\eta^{\max}$ and $\eta^{\min}$, will coincide with constant probability after $O(n^3 \log n)$ steps. Call these two copies $(\eta_t^{\max})$, $(\eta_t^{\min})$ respectively.

We will measure the distance between $\eta_t^{\max}$ and $\eta_t^{\min}$ using the quantity

$$D(t) = \sum_{i=1}^{n} w(i)(\eta_t^{\max}(i) - \eta_t^{\min}(i)), \tag{5}$$

where $w(i) \ge 0$ is a suitably chosen weight function. Note that $\eta_t^{\max}(i) \ge \eta_t^{\min}(i)$ for all i, t by monotonicity, so all terms in the sum are non-negative; and $D(t) = 0$ iff $\eta_t^{\max} = \eta_t^{\min}$. Following an idea of Wilson [24], we choose w as the second eigenvector of the discrete Laplacian operator Δ on $[1,n]$ with zero boundary conditions, defined by $\Delta g(i) = -\frac{1}{2}(g(i+1) + g(i-1)) + g(i)$, $g(0) = g(n+1) = 0$. It is well known (and easy to verify) that $w(i) = \cos(-\frac{\pi}{2} + \frac{\pi i}{n+1})$ with corresponding eigenvalue $\lambda = 1 - \cos(\frac{\pi}{n+1}) = \Theta(\frac{1}{n^2})$.

The reason for this choice is that, by Corollary 3.3, one step of the dynamics behaves very like the Laplacian, so choosing w as an eigenvector of Δ should give us a contraction of $(1 - \frac{\lambda}{n})$ in D at every step. The argument proceeds as follows:

$$\begin{aligned}
&\mathrm{E}[D(t+1) - D(t) \mid \eta_t^{\max}, \eta_t^{\min}] \\
&= \frac{1}{n}\sum_{i=1}^{n} w(i)\big\{\mathrm{E}[\eta_{t+1}^{\max}(i) \mid \eta_t^{\max}(i-1), \eta_t^{\max}(i+1)] \\
&\qquad - \mathrm{E}[\eta_{t+1}^{\min}(i) \mid \eta_t^{\min}(i-1), \eta_t^{\min}(i+1) \\
&\qquad - (\eta_t^{\max}(i) - \eta_t^{\min}(i))\big\} \\
&\le -\frac{1}{n}\sum_{i} w(i)(\Delta\eta_t^{\max}(i) - \Delta\eta_t^{\min}(i)) \\
&= -\frac{1}{n}\sum_{i} \Delta w(i)(\eta_t^{\max}(i) - \eta_t^{\min}(i)) = -\frac{\lambda}{n} D(t),
\end{aligned}$$

where in the inequality we have used Corollary 3.3.

Thus after t steps of the dynamics we have $\mathrm{E}[D(t)] \le (1 - \frac{\lambda}{n})^t D(0) \le (1 - \frac{c}{n^3})^t n^2$ for a constant $c > 0$. Taking $t = t^* = c' n^3 \log(\frac{n}{\varepsilon})$ for a sufficiently large constant c' ensures that $\mathrm{E}[D(t^*)] \ll \frac{\varepsilon}{n^2}$. Finally, we may bound the coupling probability at time t^* as follows:

$$\begin{aligned}
\Pr[\eta_{t^*}^{\max} \ne \eta_{t^*}^{\min}] &\le \sum_i \Pr[\eta_{t^*}^{\max}(i) - \eta_{t^*}^{\min}(i) \ge 1] \\
&\le w_{\min}^{-1} \sum_i w(i)\mathrm{E}[\eta_{t^*}^{\max}(i) - \eta_{t^*}^{\min}(i)] \\
&= w_{\min}^{-1}\mathrm{E}[D(t^*)] \le \varepsilon,
\end{aligned}$$

where in the second line we used Markov's inequality, and in the last line the fact that $w_{\min} := \min_i w(i) = \cos(-\frac{\pi}{2} + \frac{\pi}{n+1}) = \Theta(\frac{1}{n^2})$. Thus by Proposition 2.1 we have $\tau_{\mathcal{M}_n^{\mathrm{col}}}(\varepsilon) \le t^* = O(n^3 \log(n/\varepsilon))$, so $\tau_{\mathcal{M}_n^{\mathrm{col}}}^{mix} = O(n^3 \log n)$. □

For our analysis of the single-site dynamics, it will be convenient to introduce a "parallel" version $\mathcal{M}_n^{\mathrm{par}}$ of the column dynamics in which all odd-numbered (or all even-numbered) positions are updated simultaneously at each step. Moreover, since repeated updates of odd or even positions have no effect, we may as well assume that odd and even updates alternate. This leads to the following definition of $\mathcal{M}_n^{\mathrm{par}}$, in which O, E denote updates of all odd and even positions respectively, and the update at any given position is performed as in the column dynamics:

1. Flip a single fair coin.
2. If heads, perform t pairs of odd-even updates (i.e., $(OE)^t$), else if tails perform t pairs of even-odd updates (i.e., $(EO)^t$).

Note that $\mathcal{M}_n^{\mathrm{par}}$ is a convex combination of two reversible Markov chains, one performing the update sequence $(OE)^t$ and the other $(EO)^t$. We will call these chains $\mathcal{M}_n^{\mathrm{OE}}$ and $\mathcal{M}_n^{\mathrm{EO}}$ respectively.

Following our analysis of $\mathcal{M}_n^{\rm col}$, it is straightforward to see that $\mathcal{M}_n^{\rm par}$ inherits a similar bound on the mixing time, with a factor n speedup coming from the parallelization of the updates. The proof is similar to that of Theorem 3.1 and is omitted (see [13]).

THEOREM 3.4. *For any $\beta > \frac{1}{2}\ln 3$, the mixing time of $\mathcal{M}_n^{\rm par}$ is $O(n^2 \log n)$.*

Remark: The proofs of Theorems 3.1 and 3.4 show the stronger results that $\tau_{\mathcal{M}_n^{\rm col}}(\varepsilon) = O(n^3 \log(n/\varepsilon))$ and $\tau_{\mathcal{M}_n^{\rm par}}(\varepsilon) = O(n^2 \log(n/\varepsilon))$. We shall use this result for $\tau_{\mathcal{M}_n^{\rm par}}(\varepsilon)$ in the next section.

We close this section with a lower bound which shows that the above bound on the mixing time of the column dynamics is tight up to the $\log n$ factor. This lower bound also applies to the single-site dynamics, which will imply that our upper bound on its mixing time derived in the next section is tight within a factor of $\tilde{O}(\sqrt{n})$, as claimed in the Introduction.

THEOREM 3.5. *The mixing times of both $\mathcal{M}_n^{\rm col}$ and $\mathcal{M}_n^{\rm ss}$ are at least $\Omega(n^3)$.*

PROOF (SKETCH). Recall that the *spectral gap* of a reversible dynamics $\mathcal{M}$ is given by

$$\text{gap} = \frac{1}{2} \inf_f \frac{\sum_{\eta,\eta'} \mu(\eta) \Pr_{\mathcal{M}}[\eta \to \eta'](f(\eta) - f(\eta'))^2}{\text{Var}_\mu(f)} \qquad (6)$$

where the infimum is over all non-constant functions $f : \Omega_n \to \mathbb{R}$. Since the mixing time is bounded below by gap^{-1}, it suffices to show that $\text{gap} \le n^{-3}$. Now take the test function $f(\eta) = \sum_i w(i)(\eta(i+1) - \eta(i-1))$, where w is as in the proof of Theorem 3.1. Then straightforward calculations show that, for both $\mathcal{M}_n^{\rm col}$ and $\mathcal{M}_n^{\rm ss}$, the numerator of (6) is at most c_1/n^2 and the denominator is at least $c_2 n$, for constants $c_1, c_2 > 0$. The details are in the full version [13]. □

4. THE SINGLE-SITE DYNAMICS

In this section we prove our main result, Theorem 1.1 of the Introduction, which we restate here for convenience.

THEOREM 4.1. *The mixing time of the single-site dynamics $\mathcal{M}_n^{\rm ss}$ at any inverse temperature $\beta > \frac{1}{2}\ln 3$ is $\tilde{O}(n^{3.5})$.*

Remark: We have confined our analysis to the SOS model in an $n \times n$ region, where the maximum height is equal to the number of sites, n. This is the most natural setting, especially in view of the connection with the Ising model. If instead we allow heights in the range $[0, h]$ (so that the set of configurations is $[0, h]^n$) then, as will be apparent from our analysis below, we obtain a bound on the mixing time of $\tilde{O}(n^{2.5} \max\{n, h\})$.

Before embarking on the proof, which is quite involved, we give a brief informal sketch of our strategy. Recall from Proposition 2.1 that it suffices to bound the coupling time for two extremal copies of the dynamics, $(\eta_t^{\max})$ and $(\eta_t^{\min})$. We do this in three phases as follows.

Phase 1: We show that, after time $\tilde{O}(n^{3.5})$, $\eta_t^{\max}$ is (w.h.p.) below height $\sqrt{n}\log^2 n$. We do this in about $\sqrt{n}$ stages, each of length $\tilde{O}(n^3)$; in each stage, the height of $\eta_t^{\max}$ is reduced by $\sqrt{n}$. To bring the height down from $h_{j-1} = n - (j-1)\sqrt{n}$ to $h_j = n - j\sqrt{n}$, we introduce a "bounding" dynamics $\mathcal{M}_j$ on the expanded region $[-n, 2n]$ of width $3n$ with boundary conditions $k_j = h_j - \sqrt{n}\log^2 n$. To do this, we must start $\mathcal{M}_j$ above $\eta_t^{\max}$, i.e., above height h_{j-1} in the region $[1, n]$; by monotonicity, the time for $\eta_t^{\max}$ to reach height h_j is then bounded by the time for $\mathcal{M}_j$ to reach h_j. But since the boundary conditions of $\mathcal{M}_j$, and hence its equilibrium distribution, are far below h_j, this in turn is bounded by the mixing time of $\mathcal{M}_j$. This mixing time can be bounded by relating it to the parallel column dynamics using the censoring inequality (Lemma 2.2). A key point here is that the overhead in this comparison is determined by the maximum gradient in $\mathcal{M}_j$, so we need to keep this small; but by starting $\mathcal{M}_j$ in equilibrium conditioned on being above height h_{j-1} on $[1, n]$, we can relate the probability of developing a large ($>$ polylog(n)) gradient to the *equilibrium* probability of this event, which is very small.

Phase 2: After Phase 1, $\eta_t^{\max}$ is below a bounding dynamics with boundary conditions at 0, but on the wider region $[-n, 2n]$. Such a dynamics will have height $\tilde{O}(\sqrt{n})$ at positions 1 and n, so is still far from equilibrium on the desired region $[1, n]$. The role of Phase 2 is to shrink the region to $[1, n]$ by moving in the boundary conditions. We do this in three steps, motivated by the need to keep the gradients small so that comparison with the column dynamics has low overhead. First we show, again using a bounding dynamics, that after a further time $\tilde{O}(n^{3.5})$, $\eta_t^{\max}$ is below the equilibrium contour on a region $[-n^{3/4}, n + n^{3/4}]$. We then repeat this to shrink the region to $[-n^{1/2}, n + n^{1/2}]$, and then again to get it to $[1, n]$.

Phase 3: Once $\eta_t^{\max}$ is below the desired equilibrium contour on $[1, n]$, we show that after a further $\tilde{O}(n^{3.5})$ steps it is likely to couple with $\eta_t^{\min}$. This is done by comparing $\eta_t^{\min}$ with the parallel column dynamics, in similar fashion to the above, but a somewhat delicate *ad hoc* argument is needed to show that $\eta_t^{\min}$ itself does not develop large gradients.

We now proceed with the detailed proof. Let us first fix some notation. Let $\Lambda_n = [1, n]$ denote the usual SOS region. For a positive integer ℓ, let $\Lambda_{\ell,n} = [-\ell + 1, n + \ell]$ denote the enlargement of Λ_n by 2ℓ additional positions. We say that a configuration η on $\Lambda_{\ell,n}$ has *k-b.c.* if its boundary conditions are $\eta(-\ell) = \eta(n + \ell + 1) = k$.

It will be convenient to allow configurations of our dynamics to have arbitrarily large heights. Thus let $H_n = [0, n]$ denote the usual set of heights for the SOS model, and $H_\infty = \mathbb{N}$ the unrestricted set of heights. The corresponding sets of configurations are $\Omega_{\ell,n} = H_n^{\Lambda_{\ell,n}}$ and $\Omega_{\ell,n}^\infty = H_\infty^{\Lambda_{\ell,n}}$ respectively. The Gibbs distribution on $\Omega_{\ell,n}^\infty$ with k-b.c. is denoted $\mu_{\ell,n}^{(k)}$. (This is defined as for the original SOS model in (1); note that, despite the unbounded heights, the partition function Z_β is bounded for all $\beta > 0$. Note also that the column dynamics on $\Omega_{\ell,n}^\infty$ inherits the mixing time bounds from Section 3 provided its initial configuration is below height n.)

The single-site dynamics on $\Omega_n^\infty := \Omega_{0,n}^\infty$ with 0-b.c. starting in the constant configuration $\eta(i) = k$ for $i \in \Lambda_n$ will be denoted $(\eta_t^{(k)})$. We will in fact assume that this dynamics is implemented by choosing a position $i \in \Lambda_{n,n}$ (rather than in Λ_n) for updating, and doing nothing if i falls outside Λ_n. This just slows down the dynamics by a factor of 3. We will do the same for all the single-site dynamics we use. Since the largest region we will use is $\Lambda_{n,n}$, this device allows us to extend the monotone complete coupling of Section 2 to all our dynamics, even when they live on different regions $\Lambda_{\ell,n}$.

4.1 Phase 1: From height n to height $\sqrt{n}\log^2 n$

Following the sketch above, our goal is to bound the time for $\eta_t^{\max}$ to drop from height n to height $\sqrt{n}\log^2 n$ in stages of size $\sqrt{n}$. To this end, for $0 \le j \le N := \sqrt{n} - \log^2 n$,[¶] define

$$h_j := n - j\sqrt{n}; \quad k_j := h_j - \sqrt{n}\log^2 n.$$

Also, set $t_n = n^3 \log^{12} n$ and $\epsilon_n = 1/n^3$. The following lemma summarizes the outcome of Phase 1, and says that after each subsequence of $t_n = \tilde{O}(n^3)$ steps the single-site dynamics started at height n decreases its height by $\sqrt{n}$ whp. The total length of this phase is therefore $\tilde{O}(n^{3.5})$.

LEMMA 4.2. *With the above notation, we have* $\sum_{j=1}^{N} \Pr[\exists i \in \Lambda_n : \eta^{\max}_{jt_n}(i) \ge h_j] \le \epsilon_n$.

PROOF. Clearly, by monotonicity, it is enough to prove the lemma with $\eta_t^{\max}$ replaced by $\eta_t^{(n)}$ (which differs only in that its height set is H_∞ rather than H_n). Let B_j denote the event $\{\exists\, i \in \Lambda_n :\ \eta^{(n)}_{jt_n}(i) \ge h_j\}$ and let $q_j = \Pr[B_j \cap \overline{B}_{j-1}]$. Then

$$\sum_{j=1}^{N} \Pr[B_j] \le \sum_{j=1}^{N} \sum_{k=1}^{j} q_k \le N^2 \max_j q_j. \tag{7}$$

We now bound q_j. By definition of $\overline{B}_{j-1}$ and monotonicity,

$$q_j \le \Pr[\exists i \in \Lambda_n : \eta^{(h_{j-1})}_{t_n}(i) \ge h_j]. \tag{8}$$

In order to bound the r.h.s. of (8) we compare the dynamics to another "bounding" single-site dynamics on the enlarged interval $\Lambda_{n,n}$ with k_j-b.c. and with initial distribution ν_0 given by the equilibrium distribution $\mu^{(k_j)} := \mu^{(k_j)}_{n,n}$ conditioned on the increasing event

$$A := \{\eta(i) \ge h_{j-1} \text{ for every } i \in \Lambda_n\}. \tag{9}$$

Because of the choice of the initial distribution ν_0 and of the boundary condition, if (σ_t^ξ) denotes the bounding dynamics starting from the configuration ξ, it is clear that

$$\begin{aligned}\Pr[\exists i \in \Lambda_n : \eta^{(h_{j-1})}_{t_n}(i) \ge h_j] \\ \le \textstyle\sum_\xi \nu_0(\xi) \Pr[\exists i \in \Lambda_n : \sigma^\xi_{t_n}(i) \ge h_j].\end{aligned} \tag{10}$$

A key observation at this stage, which largely explains the choice of the scales h_j, k_j is the following. Since $h_j - k_j = \sqrt{n}\log^2 n$, the event on the r.h.s. of (10) is very unlikely in the equilibrium distribution $\mu^{(k_j)}$. More precisely, the following bound is proved in the appendix:[‖]

CLAIM 4.3. $\mu^{(k_j)}(\exists i \in \Lambda_n : \eta(i) \ge h_j) \le e^{-c\log^4 n}$.

Thus, in order to show that the r.h.s. of (10) is small, we only need to show that the distribution of the bounding dynamics at time t_n, ν_{t_n}, is close to its equilibrium distribution, or specifically

$$\|\nu_{t_n} - \mu^{(k_j)}\| = o(1/n^4). \tag{11}$$

This will ensure that $q_j = o(1/n^4)$, which by (7) and the fact that $N \sim \sqrt{n}$ gives the lemma.

[¶]Throughout, for clarity, we ignore rounding issues; clearly these do not affect our asymptotic results.

[‖]Throughout we shall use c to denote a generic positive constant.

To bound the mixing time of the bounding dynamics, we relate it to the corresponding parallel column dynamics using the censoring inequality (Lemma 2.2). Note that this is valid because the initial distribution ν_0 satisfies the requirement that $\nu_0/\mu^{(k_j)}$ is increasing wrt $\preceq$.

To do this, we split the time t_n into $M := n^2 \log^2 n$ *epochs* each of length $m := n\log^{10} n$. Given t_n random positions $\mathbf{i} = (i_1, i_2, \ldots, i_{t_n})$ in $\Lambda_{n,n}$, the measure ν_{t_n} can be written as the average over $\mathbf{i}$ of the measure $\nu_{\mathbf{i}}$ obtained by applying, in the given order, t_n single-site updates at positions $i_1, i_2, \ldots, i_{t_n}$. Next we write $\mathbf{w}(\mathbf{i}) = (\mathbf{w}_1, \ldots, \mathbf{w}_M)$ by grouping together positions in the same epoch. Finally, we define two censored versions of the dynamics as follows. In the first version, we delete all even positions from the odd epochs and all odd positions from the even epochs; denote the resulting censored vector $\mathbf{OE}(\mathbf{i}) = (\mathbf{OE}_1, \ldots, \mathbf{OE}_M)$ and the associated measure $\nu_{\mathbf{OE}(\mathbf{i})}$. In the second version, we reverse the roles of odd and even and denote the resulting censored vector $\mathbf{EO}(\mathbf{i}) = (\mathbf{EO}_1, \ldots, \mathbf{EO}_M)$ and the associated measure $\nu_{\mathbf{EO}(\mathbf{i})}$.

This construction gives us

$$\begin{aligned}\|\nu_{t_n} - \mu^{(k_j)}\| &= \|\mathrm{Av}_{\mathbf{i}}\, \nu_{\mathbf{i}} - \mu^{(k_j)}\| \le \mathrm{Av}_{\mathbf{i}} \|\nu_{\mathbf{i}} - \mu^{(k_j)}\| \\ &\le \mathrm{Av}_{\mathbf{i}} \|\tfrac{1}{2}(\nu_{\mathbf{OE}(\mathbf{i})} + \nu_{\mathbf{EO}(\mathbf{i})}) - \mu^{(k_j)}\|,\end{aligned} \tag{12}$$

where the last step relies on the censoring inequality. Note that the expected number of times any position i appears in $\mathbf{i}$ is $m/n = \log^{10} n$. Hence a standard Chernoff bound guarantees that, apart from an error that is exponentially small in $\log^{10} n$, the r.h.s. of (12) is bounded above by

$$\max_{\mathbf{i} \in \Sigma} \|\tfrac{1}{2}(\nu_{\mathbf{OE}(\mathbf{i})} + \nu_{\mathbf{EO}(\mathbf{i})}) - \mu^{(k_j)}\|, \tag{13}$$

where Σ consists of all $\mathbf{i}$ such that the censored vectors $\mathbf{OE}(\mathbf{i})$ and $\mathbf{EO}(\mathbf{i})$ each contain at least $\frac{1}{2}\log^{10} n$ updates of every position $i \in \Lambda_{n,n}$ in every epoch $k \in \{1, \ldots, M\}$.

Now we claim that the distribution $\frac{1}{2}(\nu_{\mathbf{OE}(\mathbf{i})} + \nu_{\mathbf{EO}(\mathbf{i})})$ is very close to the distribution at time $M = n^2 \log^2 n$ of the parallel column dynamics $\mathcal{M}_n^{\mathrm{par}}$, with the same initial distribution ν_0 and boundary conditions k_j. To establish this, we need to show that $\frac{1}{2}\log^{10} n$ single-site updates at position i, with its neighboring heights fixed, are enough to simulate (with small error) one column update at i. This relies crucially on the fact that $\mathcal{M}_n^{\mathrm{par}}$ is unlikely to produce configurations with large gradients, which we define to be at least $\log^{4.5} n$. Accordingly, define the set of "bad" configurations

$$B = \{\eta : |\eta(i+1) - \eta(i)| \ge \log^{4.5} n \text{ for some } i \in \Lambda_{n,n}\}. \tag{14}$$

CLAIM 4.4. *Let* ν_s^{OE} *and* ν_s^{EO} *denote the distributions of* $\mathcal{M}_n^{\mathrm{OE}}$ *and* $\mathcal{M}_n^{\mathrm{EO}}$ *respectively after* s *steps, starting from* ν_0. *Then* $\|\frac{1}{2}(\nu_{\mathbf{OE}(\mathbf{i})} + \nu_{\mathbf{EO}(\mathbf{i})}) - \nu_M^{\mathrm{par}}\| \le M\big(\max_s\{\nu_s^{\mathrm{OE}}(B) + \nu_s^{\mathrm{EO}}(B)\} + e^{-c\log^9 n}\big)$.

The intuition for this Claim, whose detailed proof is deferred to the full version [13], is the following. The first term on the r.h.s. bounds the probability of seeing a bad configuration, so we may assume that $\eta \notin B$. A sequence of single-site updates at position i (with its neighboring heights a, b fixed) can be viewed as a lazy nearest-neighbor random walk on column i with stationary distribution equal to the distribution of a column update. This distribution (see (2)) is uniform on the interval $[a, b]$ and decays exponentially outside

it. Hence its mixing time is essentially $O((b-a)^2 \log(b-a))$, which is $O(\log^9 n \log\log n)$ assuming $\eta \notin B$. Thus $\frac{1}{2}\log^{10} n$ steps suffice to simulate a column update with very small error $e^{-c\log^9 n}$, which is the second term in the bound. The factor M comes from a union bound over steps.

In order to use Claim 4.4, we need to bound $\nu_s^{\mathrm{OE}}(B)$ (and, symmetrically, $\nu_s^{\mathrm{EO}}(B)$), the probability of the dynamics creating a large gradient. This is in general a highly non-trivial task because it requires detailed non-equilibrium information about the contours. However, it is here that our choice of the initial distribution $\nu_0 = \mu^{(k_j)}(\,\cdot \mid A)$, where A is defined in (9), is crucial. Since $\mu^{(k_j)}$ remains invariant under any number of steps of $\mathcal{M}_n^{\mathrm{OE}}$ (and of $\mathcal{M}_n^{\mathrm{EO}}$), we can write, for any s,

$$\nu_s^{\mathrm{OE}}(B) \le \mu^{(k_j)}(B)/\mu^{(k_j)}(A),$$

and an identical bound for $\nu_s^{\mathrm{EO}}(B)$. But this is easy to evaluate as it is the ratio of the probabilities of two events *in equilibrium*! The following straightforward bound is proved in the appendix:

CLAIM 4.5. *There exists a constant $c > 0$ such that $\mu^{(k_j)}(B)/\mu^{(k_j)}(A) \le e^{-c\log^{4.5} n}$.*

We can now put everything together. For each $\mathbf{i} \in \Sigma$, the quantity in (13) is bounded by

$$\|\tfrac{1}{2}(\nu_{\mathbf{OE(i)}} + \nu_{\mathbf{EO(i)}}) - \nu_M^{\mathrm{par}}\| + \|\nu_M^{\mathrm{par}} - \mu^{(k_j)}\|, \qquad (15)$$

where ν_s^{par} denotes the distribution obtained from ν_0 after s steps of the parallel column dynamics. By Claims 4.4 and 4.5, the first term in (15) is bounded by $M(2e^{-c\log^{4.5} n} + e^{-c\log^9 n})$, which is certainly $o(1/n^4)$, while the second term is $o(1/n^4)$ by Theorem 3.4 and the fact that $M \gg n^2 \log n$ (the mixing time of $\mathcal{M}_n^{\mathrm{par}}$). Hence the variation distance of the dynamics is $o(1/n^4)$, as required in (11). This concludes the proof of the lemma and the analysis of Phase 1. □

4.2 Phase 2: From height $\sqrt{n}\log^2 n$ to equilibrium height

Phase 1 guarantees that the contour $\eta_t^{\max}$ has maximum height $\sqrt{n}\log^2 n$ with high probability after $\tilde{O}(n^{3.5})$ steps. In Phase 2 we show that a contour starting at this height is, after a further $\tilde{O}(n^{3.5})$ steps, "below" the equilibrium distribution μ_n on Λ_n, in a sense made precise by the following lemma. Recall that $\eta^{(k)}$ denotes the single-site dynamics in Ω_n^∞ with 0-b.c. starting at height $\eta(i) = k$ for all i.

LEMMA 4.6. *Let $s_n = n^{3.5}\log^{12} n$ and $\epsilon_n = 1/n^3$. Then for any increasing event E,*

$$\Pr[\eta_{3s_n}^{(\sqrt{n}\log^2 n)} \in E] \le \mu_n(E) + 3\epsilon_n.$$

PROOF. The proof proceeds via three "smoothing" steps, which bring the contour close to the equilibrium distribution on smaller intervals $\Lambda_{\ell,n}$ with 0-b.c. The first step uses $\ell = n^{3/4}$ and the second $\ell = n^{1/2}$; the third step then reduces the interval to the desired Λ_n. All steps use the same technology as Phase 1 by relating the single-site dynamics to the parallel column dynamics; as before, the key is to control the gradients along the contours—and it is this that dictates our choices of ℓ.

We proceed now with the first smoothing step. Let $h_n = \sqrt{n}\log^2 n$, $\ell = n^{3/4}$, and $\mu_{\ell,n}$ be the equilibrium distribution in the enlarged interval $\Lambda_{\ell,n}$ with 0-b.c. Let $A = \{\eta \in \Omega_{\ell,n}^\infty;\ \eta(i) \ge h_n\ \forall i \in \Lambda_n\}$ and let ν_{s_n} be the distribution at time $s_n = n^{3.5}\log^{12} n$ of the single-site dynamics in $\Lambda_{\ell,n}$ with 0-b.c. starting from the distribution $\nu_0 := \mu_{\ell,n}(\,\cdot \mid A)$.

LEMMA 4.7 (FIRST SMOOTHING). *With the above notation, and with $\epsilon_n = 1/n^3$, we have $\|\nu_{s_n} - \mu_{\ell,n}\| \le \epsilon_n$.*

Remark: Let $P_{s_n}^{(h_n)}$ denote the distribution at time s_n of the single-site dynamics in Λ_n with height set H_n and 0-b.c. starting from height h_n. Clearly $P_{s_n}^{(h_n)} \le \nu_{s_n}$ and therefore Lemma 4.7 implies that, for any increasing event $E \subseteq \Omega_n$, $P_{s_n}^{(h_n)}(E) \le \mu_{\ell,n}(E) + \epsilon_n$.

PROOF. The proof follows the same pattern as that of each stage in Lemma 4.2, with one major difference. Since the measure $\mu_{\ell,n}$ has 0-b.c. at distance $\ell = n^{3/4}$ from the interval Λ_n, the corresponding conditioning event $A = \{\eta(i) \ge h_n \text{ for every } i \in \Lambda_n\}$ is now much more unlikely than it was before (cf. equation (9)). Whereas previously we had $\mu_(A) \ge e^{-c\log^4 n}$ (see the proof of Claim 4.5), we must now make do with the following weaker lower bound (which is almost sharp):

CLAIM 4.8. $\mu_{\ell,n}(A) \ge e^{-ch_n^2/\ell} = e^{-cn^{1/4}\log^4 n}$ *for a suitable constant c.*

The proof is in the appendix. Recall that the proof of Lemma 4.2 hinged on the fact that the ratio $\mu(B)/\mu(A) \ll \epsilon_n$, where B is the set of contours with large (i.e., polylogarithmic) gradient, as defined in (14). Since now $\mu(A)$ is much smaller, we must weaken our definition of large gradient and redefine B as follows:

$$B = \{\eta : |\eta(i+1) - \eta(i)| \ge n^{1/4}\log^{4.5} n \text{ for some } i \in \Lambda_{\ell,n}\}. \qquad (16)$$

With this definition of B, a calculation analogous to that in the proof of Claim 4.5 gives $\mu_{\ell,n}(B) \le e^{-c' n^{1/4}\log^{4.5} n}$, and hence $\mu_{\ell,n}(B)/\mu_{\ell,n}(A)$ is of the same order.

We then divide the time s_n into $M = n^2\log^2 n$ epochs, each of length $n^{3/2}\log^{10} n$, and use the censoring inequality as before to simulate M steps of the parallel column dynamics (which is sufficient for it to mix within variation distance ϵ_n). Whp, in each epoch each position receives at least $\frac{1}{2}n^{1/2}\log^{10} n$ single-site updates, and by a random walk argument analogous to that in Claim 4.4 the number of updates needed to faithfully simulate one column update is essentially the square of the gradient, which is $(n^{1/4}\log^{4.5} n)^2 = n^{1/2}\log^9 n$ and thus much smaller than the number of updates actually performed. This concludes the proof of the first smoothing step. □

Thus after $n^{3.5}\log^{12} n$ steps, with high probability, our single-site dynamics is below the equilibrium distribution in $\Lambda_{\ell,n}$. However, the typical configurations for the latter at $i = 1, n$ have height $\tilde{O}(\sqrt{\ell}) = \tilde{O}(n^{3/8})$, which yields a large gradient if we put zero boundary conditions at 0 and n. (To keep the overhead down to $\tilde{O}(n^{1/2})$ we can only afford a gradient of $\tilde{O}(n^{1/4})$.) To fix this we need a second smoothing step, which we now describe.

Let $\ell = n^{3/4}$ as before, let $\ell' = n^{1/2}$ and let $\mu_{\ell',n}$ be the equilibrium distribution in the interval $\Lambda_{\ell',n}$ with 0-b.c. Let ν'_{s_n} be the distribution at time $s_n = n^{3.5}\log^{12} n$ of the single-site dynamics in $\Lambda_{\ell',n}$ with 0-b.c. starting from the distribution ν'_0 defined as follows. For any configuration $\eta \in$

$\Omega^{\infty}_{\ell',n}$, write $\eta = (\eta^{out}, \eta^{in})$ where $\eta^{out} = \{\eta(i)\}_{i\in\Lambda_{\ell',n}\setminus\Lambda_n}$ and $\eta^{in} = \{\eta(i)\}_{i\in\Lambda_n}$. Then

$$\begin{aligned}\nu'_0(\eta) &:= \mu_{\ell',n}(\eta^{out})\mu_{\ell,n}(\eta^{in}) \\ &= \mu_{\ell',n}(\eta^{out})\mu_{\ell',n}(\eta^{in} \mid \eta(1),\eta(n))\mu_{\ell,n}(\eta(1),\eta(n)) \\ &= \mu_{\ell',n}(\eta) f_{\ell,\ell',n}(\eta)\,,\end{aligned}$$

where $f_{\ell,\ell',n}(\eta) := \frac{\mu_{\ell,n}(\eta(1),\eta(n))}{\mu_{\ell',n}(\eta(1),\eta(n))}$. Here we used the fact that $\mu_{\ell,n}(\eta^{in} \mid \eta(1),\eta(n)) = \mu_{\ell',n}(\eta^{in} \mid \eta(1),\eta(n))$. Notice that the marginal of ν'_0 on η^{in} coincides with that of $\mu_{\ell,n}$.

LEMMA 4.9 (SECOND SMOOTHING). *With the above notation, and with $\epsilon_n = 1/n^3$, we have $\|\nu'_{s_n} - \mu_{\ell',n}\| \le \epsilon_n$.*

Remark: Let $P^{(\nu'_0)}_{s_n}$ denote the distribution at time s_n of the single-site dynamics in Λ_n with height set H_n and 0-b.c. starting from the distribution ν'_0. Clearly $P^{(\nu'_0)}_{s_n} \le \nu'_{s_n}$ and therefore Lemma 4.9 implies that, for any increasing event $E \subseteq \Omega_n$, $P^{(\nu'_0)}_{s_n}(E) \le \mu_{\ell',n}(E) + \epsilon_n$.

PROOF. The proof is identical to that of Lemma 4.7 provided that the distribution $\nu^{\rm par}_s$ induced by the parallel column dynamics starting from ν'_0 still satisfies $\nu^{\rm par}_s(B) \ll \epsilon_n$ at all times s, where B is as in (16). To this end, define $G = \{\eta :\ \eta(1) + \eta(n) \le n^{3/8}\log^2 n\}$. Then we can write

$$\begin{aligned}\nu^{\rm par}_s(B) &\le \nu'_0(\overline{G}) + \sum_{\eta\in G}\nu'_0(\eta)\Pr[\eta^{\rm par}_s \in B] \\ &\le \mu_{\ell,n}\big(\overline{G}\big) + \max_{\eta\in G} f_{\ell,\ell',n}(\eta)\mu_{\ell',n}(B) \\ &\le 2e^{-c\log^4 n} + e^{cn^{1/4}\log^4 n}e^{-c'n^{1/4}\log^{4.5} n} \\ &\le 3e^{-c\log^4 n} \qquad (17)\end{aligned}$$

for some constants $c, c' > 0$ and sufficiently large n. □

After the second smoothing step, the typical height at the points 1 and n is only $\tilde{O}(n^{1/4})$, giving a gradient small enough to allow us to relate the contour to the true equilibrium contour μ_n on Λ_n. This will be the last step in the proof of Lemma 4.6.

Using the previous notation, let ν''_0 denote the marginal on η^{in} of the equilibrium distribution $\mu_{\ell',n}$ and let $P^{(\nu''_0)}_{s_n}$ be the distribution of the single-site dynamics in Λ_n with 0-b.c. starting from ν''_0.

LEMMA 4.10. *With the above notation, and with $\epsilon_n = 1/n^3$, we have $\|P^{(\nu''_0)}_{s_n} - \mu_n\| \le \epsilon_n$.*

PROOF. The proof is identical to that of Lemma 4.9, provided that the distribution $\nu^{\rm par}_s$ induced by the parallel column dynamics starting from ν''_0 still satisfies $\nu^{\rm par}_s(B) \ll \epsilon_n$ at all times s, where B is again as in (16). The necessary computation is as in (17), with ν'_0 replaced by ν''_0 and the set G replaced by $G' = \{\eta : \eta(1) + \eta(n) \le n^{3/16}\log^2 n\}$. □

The proof of Lemma 4.6 now follows immediately from the above three lemmas. For any increasing event E on Ω_n, we have

$$\begin{aligned}\Pr[\eta^{(\sqrt{n}\log^2 n)}_{3s_n} \in E] &\le \epsilon_n + \Pr[\eta^{(\nu'_0)}_{2s_n} \in E] \\ &\le 2\epsilon_n + \Pr[\eta^{(\nu''_0)}_{s_n} \in E] \\ &\le 3\epsilon_n + \mu_n(E),\end{aligned}$$

as required. In the first inequality we used Lemma 4.7, in the second Lemma 4.9, and in the last Lemma 4.10. □

4.3 Phase 3: Coupling with $\eta^{\min}_t$

After Phases 1 and 2, which take total time $4s_n = \tilde{O}(n^{3.5})$, we may assume that the contour $\eta^{\max}_t$ is "below" the equilibrium distribution μ_n. To complete the proof of Theorem 4.1 it remains to couple this contour with the minimum contour $\eta^{\min}_t$. For this, we need the complementary fact that, after a similar time, $\eta^{\min}_t$ is "above" μ_n. This is provided by the following lemma.

LEMMA 4.11. *Let $s_n = n^{3.5}\log^{12} n$ and $\epsilon_n = 1/n^3$. Then for any increasing event E, $\Pr[\eta^{\min}_{2s_n} \in E] \ge \mu_n(E) - \epsilon_n$.*

PROOF. We begin with a high-level sketch of the argument, which proceeds in two stages. First we prove that, after time s_n and apart from an error term $\epsilon_n/2$, the distribution of $\eta^{\min}_{s_n}$ is above μ_n *conditioned on* the unlikely event $A := \{\eta(j\lfloor n^{3/4}\rfloor) = 0,\ j = 1,2,\dots\}$. Then we show, using the same comparison with the column dynamics as we used for $\eta^{\max}_t$, that after a further s_n steps the distribution of $\eta^{\min}_t$, starting from $\mu_n(\cdot \mid A)$, moves above μ_n, apart from another error term $\epsilon_n/2$. The role of the first stage here is to ensure that the probability of the conditioning event A is larger than the probability of developing large gradients, just as in the proof of Lemma 4.2. (The conditioning event of being entirely at zero, which is the initial configuration of $\eta^{\min}_t$, is far too rare to satisfy this condition.)

We now spell out the details. Consider the special positions in Λ_n of the form $i_j = j\lfloor n^{3/4}\rfloor$, $j = 1,2,\dots$, and consider an auxiliary single-site dynamics in Λ_n, starting from $\eta^{\min}$ and with the usual updating rules except that we do not update the heights at these positions, i.e., $\eta_t(i_j)$ is fixed to be 0. The distribution of this dynamics at time t is denoted $\tilde{P}^{\min}_t$. Clearly this dynamics is the product of the standard single-site dynamics in each interval $I_j := [i_j, i_{j+1}]$ with 0-b.c., and its equilibrium distribution $\tilde{\mu}_n$ is just a product measure over these intervals and is equal to $\mu_n(\cdot \mid A)$.

CLAIM 4.12. *The above single-site dynamics satisfies*

$$\|\tilde{P}^{\min}_{s_n} - \tilde{\mu}_n\| \le \epsilon_n/2.$$

PROOF. If $\tilde{\eta}^{\min}_t$ denotes the new dynamics at time t, then monotonicity implies that

$$\begin{aligned}&\Pr[\exists i \in \Lambda_n \text{ and } t \in [0,s_n] : \tilde{\eta}^{\min}_t(i) \ge \sqrt{n}\log^2 n\,] \\ &\qquad \le \sum\nolimits_i s_n\,\mu_n(\eta(i) \ge \sqrt{n}\log^2 n) \le e^{-c\log^4 n} \ll \epsilon_n\end{aligned}$$

for some constant $c > 0$. Thus, with probability greater than $1 - e^{-c\log^4 n}$, in the time interval $[0, s_n]$ the dynamics $\tilde{\eta}^{\min}_t$ never gets detached from the similar dynamics with height set $[0, \sqrt{n}\log^2 n]$ and it is enough to prove that the distribution of the latter satisfies the inequality in the Claim. Notice that with the above restriction on the height set the maximum gradient of the contour is bounded by $\sqrt{n}\log^2 n$, and therefore the penalty in the comparison argument between the single-site and (parallel) column dynamics is only $O(n\log^4 n)$. By Theorem 3.4, the time until the parallel column dynamics in each interval I_j (of width $O(n^{3/4})$) is within variation distance $\epsilon_n/4$ of equilibrium is $O(n^{3/2}\log n)$, and therefore we achieve variation distance $\epsilon_n/2$ for the single-site dynamics with height set $[0, \sqrt{n}\log^2 n]$ and 0-b.c. at the positions $\{i_j\}$ in time at most

$$O(n^{3/2}\log n) \times n\log^4 n \times n = O(n^{3.5}\log^5 n) \ll s_n\,. \quad \square$$

By Claim 4.12, for any increasing event E we can write

$$\Pr[\eta^{\min}_{2s_n} \in E] \geq \sum\nolimits_{\xi} \tilde{P}^{\min}_{s_n}(\xi) \Pr[\eta^{\xi}_{s_n} \in E] \geq P^{(\tilde{\mu}_n)}_{s_n}(E) - \epsilon_n/2.$$

We now bound $P^{(\tilde{\mu}_n)}_{s_n}(E)$ from below. This follows using the same comparison with the parallel column dynamics that we used for $\eta^{\max}_t$. The only thing we need to check is that large gradients do not appear in the column dynamics, and as we have seen this will follow if we can upper-bound the ratio $\mu_n(B)/\mu_n(A)$, where the large gradient set B is defined as

$$B = \{\eta : |\eta(i+1) - \eta(i)| \geq n^{1/4} \log^{4.5} n \text{ for some } i \in \Lambda_{n,n}\}.$$

By a calculation analogous to that in the proof of Claim 4.5, we have

$$\mu_n(B) \leq e^{-cn^{1/4} \log^{4.5} n}.$$

For the denominator, we can show the following (see the appendix for a proof):

CLAIM 4.13. *$\mu_n(A) \geq e^{-cn^{1/4} \log n}$ for a constant $c > 0$.*

Thus we have shown that $\mu_n(B)/\mu_n(A) \leq e^{-c' n^{1/4} \log^{4.5} n}$, which by comparison with the column dynamics is sufficient to guarantee that $P^{(\tilde{\mu}_n)}_{s_n}(E) \geq \mu_n(E) - \epsilon_n/2$. This completes the proof of the lemma. □

4.4 Putting it all together

Finally, we combine the above three phases to prove Theorem 4.1. Let $T_n = 4n^{3.5} \log^{12} n = 4s_n$, and let $\widehat{T}$ denote the first time t such that $\eta^{\max}_t = \eta^{\min}_t$. Then we have

$$\begin{aligned}
\Pr[\widehat{T} \geq T_n] &\leq \sum_{i \in \Lambda_n} \sum_{k=1}^{n} \Pr[\eta^{\max}_{T_n}(i) \geq k > \eta^{\min}_{T_n}] \\
&\leq \sum_{i \in \Lambda_n} \sum_{k=1}^{n} \big(\Pr[\eta^{\max}_{T_n}(i) \geq k] - \Pr[\eta^{\min}_{T_n}(i) \geq k]\big) \\
&\leq 5n^2\epsilon_n + \sum_{i \in \Lambda_n} \sum_{k=1}^{n} \big(\mu_n(\eta(i) \geq k) - \mu_n(\eta(i) \geq k)\big) \\
&= 5n^2\epsilon_n = o(1).
\end{aligned}$$

In the third line we have used Lemmas 4.2 and 4.6 to relate $\eta^{\max}_{T_n}$ to μ_n, and Lemma 4.11 to relate $\eta^{\min}_{T_n}$ to μ_n. Hence by Proposition 2.1 the mixing time of $\mathcal{M}^{\mathrm{ss}}_n$ is at most $T_n = \tilde{O}(n^{3.5})$. This completes the proof of Theorem 4.1. □

5. CONCLUDING REMARKS

1. We conjecture that the true mixing time of the single-site dynamics is $\tilde{O}(n^3)$, matching our lower bound within logarithmic factors. This would mean that our upper bound is off by $\sqrt{n}$. We should also note that we made no attempt to optimize the powers of $\log n$ in our bounds.

2. There are similarities between the SOS model and the lozenge tilings model studied in, e.g., [10, 21, 24]. For that model, Wilson [24] obtained a tight bound on the mixing time of a non-local dynamics (similar to our column dynamics), but only very weak upper bounds for the local dynamics are known. We conjecture that our results can be adapted to that model also.

3. We believe that some of our techniques may be useful for analyzing other Markov chains. In particular we envisage (i) further applications of the censoring inequality to relate local and non-local dynamics; (ii) the use of sequences of bounding dynamics, with carefully tuned boundary conditions, to capture geometric information about the evolution of a Markov chain; (iii) the use of an initial distribution that is in equilibrium conditioned on a rare event in order to obtain non-equilibrium information about the Markov chain.

References

[1] N. BERGER, C. KENYON, E. MOSSEL and Y. PERES. Glauber dynamics on trees and hyperbolic graphs. *Prob. Th. Rel. Fields* **131** (2005), pp. 311–340.

[2] F. CESI. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Prob. Th. Rel. Fields* **120** (2001), pp. 569–584.

[3] L. CHAYES, R. SCHONMANN and G. SWINDLE. Lifshitz law for the volume of a 2-dimensional droplet at zero temperature. *J. Stat. Phys.* **79** (1995), pp. 821–831.

[4] P. DIACONIS and L. SALOFF-COSTE. Comparison theorems for reversible Markov chains. *Annals of Applied Probability* **3** (1993), pp. 696–730.

[5] R.L. DOBRUSHIN, R. KOTECKY and S. SHLOSMAN. *The Wulff construction: a global shape from local interactions*. AMS Translations Of Mathematical Monographs **104**, Providence, 1992.

[6] F. DUNLOP, P. FERRARI and L. FONTES. A dynamic one-dimensional interface interacting with a wall. *J. Stat. Phys.* **107** (2002), pp. 705–727.

[7] T. FUNAKI. Stochastic interface models. *Lectures on Probability Theory and Statistics*, Springer Lecture Notes in Mathematics **1869**, 2005, pp. 103–294.

[8] G. GIACOMIN. *Random polymer models*. Imperial College Press, World Scientific, London, 2007.

[9] D. HUSE and D. FISHER. Dynamics of droplet fluctuations in pure and random Ising systems. *Physics Review B* **35** (1987), pp. 6841–6846.

[10] M. LUBY, D. RANDALL and A. SINCLAIR. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing* **31** (2001), pp. 167–192.

[11] F. MARTINELLI. Lectures on Glauber dynamics for discrete spin models. *Lectures on Probability Theory and Statistics*, Springer Lecture Notes in Mathematics **1717**, 1998, pp. 93–191.

[12] F. MARTINELLI and E. OLIVIERI. Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case. *Comm. Math. Phys.* **161** (1994), pp. 447–486.

[13] F. MARTINELLI and A. SINCLAIR. Mixing time for the solid-on-solid model. Full version, arxiv.org, 2009.

[14] F. MARTINELLI, A. SINCLAIR and D. WEITZ. The Ising model on trees: Boundary conditions and mixing time. *Proc. 44th IEEE FOCS*, 2003, pp. 628–639.

[15] F. MARTINELLI and F. TONINELLI. On the mixing time of the 2D Ising model with plus boundary conditions at low temperature. Manuscript, 2009.

[16] Y. PERES. *Mixing for Markov chains and spin systems.* Lecture Notes for PIMS Summer School at UBC, August 2005. www.stat.berkeley.edu/∼peres/ubc.pdf

[17] G. POSTA. Spectral gap for an unrestricted Kawasaki type dynamics. *ESAIM: Probability & Statistics* **1** (1995), pp. 145–181.

[18] V. Privman and N. Švrakić. Difference equations in statistical mechanics II: Solid-on-solid models in two dimensions. *J. Stat. Phys.* **51** (1988), pp. 1111–1126.

[19] V. Privman and N. Švrakić. Line interfaces in two dimensions: Solid-on-solid models. *Lecture Notes in Physics* **338**, Springer, 1989, pp. 32–60.

[20] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms* **9** (1996), pp. 223–252.

[21] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics* **41** (2000), pp. 1598–1615.

[22] F. Spitzer. *Principles of Random Walks.* Van Nostrand, Princeton, 1964.

[23] D.W. Stroock and B. Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.* **149** (1992), pp. 175–194.

[24] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Annals of Applied Probability* **14** (2004), pp. 274–325.

Appendix: Proofs of technical claims in Section 4

Here we collect together various bounds that we used on the probabilities of certain events in the equilibrium distribution of the SOS model. These bounds all follow easily from a single lemma, which we state first.

The key idea is that an equilibrium SOS contour can be viewed as a random walk with independent geometric increments, conditioned to start and end at 0 and to remain always above 0. More precisely, let $\{X(i)\}_{i\geq 0}$ be a random walk on $\mathbb{Z}$ with independent geometric increments, i.e.,

$$\Pr[|X(i+1)-X(i)|=\ell] \propto e^{-\beta\ell}.$$

Then the equilibrium distribution $\mu_n^{(k)}$ of the SOS model on Λ_n with boundary conditions at height $k\geq 0$ and height set H, where either $H=H_n$ or $H=H_\infty$, can be written as

$$\mu_n^{(k)}(\eta)=\Pr[X(i)=\eta(i)\mid C_{k,H,n}],$$

where the conditioning event is $C_{k,H,n}:=\{X(0)=X(n+1)=k;\, X(i)\in H\ \forall i\in[1,n]\}$. The following lemma bounds the probability of $C_{k,H,n}$.

Lemma A.1. *Assume $0\leq k\leq\sqrt{n}\log^2 n$. Then*

$$\Pr[C_{k,H,n}\mid X(0)=k]\geq e^{-c\log n}$$

for some constant $c>0$.

Proof. For any $\ell>0$, standard large deviation bounds based on the exponential Chebyshev inequality imply that

$$\Pr[\max_{1\leq i\leq n} X(i)\geq\ell\mid X(0)=k]\leq ne^{-c\ell^2/n}$$

for some constant $c>0$. Thus, using the assumption $0\leq k\leq\sqrt{n}\log^2 n$, we have

$$\Pr[C_{k,H,n}|X(0)=k]\geq \tag{18}$$
$$\Pr[X(n+1)=k\,;\min_i X(i)\geq k\mid X(0)=k]-ne^{-c\log^2 n}.$$

A standard random walk calculation [22] shows that the probability on the r.h.s. here is $\Omega(e^{-c'\log n})$ for some $c'>0$, which completes the proof. □

We now proceed to derive from the above lemma the various bounds we used in the proofs of Section 4.

Proof of Claim 4.3. Recall that $h_j=n-j\sqrt{n}$ and $k_j=h_j-\sqrt{n}\log^2 n$, so that $h_j-k_j=\sqrt{n}\log^2 n$. We write

$$\begin{aligned}\mu^{(k_j)}(\eta(i)\geq h_j) &\leq \frac{\Pr[X(i)\geq h_j\mid X(0)=k_j]}{\Pr[C_{k_j,H,3n}\mid X(0)=k_j]}\\ &\leq e^{-c(h_j-k_j)^2/n}e^{c\log n}\leq e^{-c'\log^4 n}.\end{aligned}$$

A union bound over i completes the proof. □

Proof of Claim 4.5. Recall that $h_j=n-j\sqrt{n}$ and $k_j=h_j-\sqrt{n}\log^2 n$, and that $B=\{\eta:\ |\eta(i+1)-\eta(i)|\geq\log^{4.5} n$ for some $i\in\Lambda_{n,n}\}$. Note that $h_{j-1}-k_j\sim\sqrt{n}\log^2 n$. Then, again using the above random walk representation, we can upper bound $\mu^{(k_j)}(B)$ by

$$\sum_{i=0}^{3n}\frac{\Pr[|X(i+1)-X(i)|\geq\log^{4.5} n\mid X(0)=X(3n)=k_j]}{\Pr[C_{k_j,H,3n}\mid X(0)=k_j]}$$
$$\leq 3ne^{-c\log^{4.5} n}e^{c\log n}$$

for some constant $c>0$.

On the other hand, a lower bound for $\mu^{(k_j)}(A)$ of the form

$$\mu^{(k_j)}(A)\geq e^{-c(h_{j-1}-k_j)^2/n}=e^{-c'\log^4 n}$$

can be computed following the same path as in the proof of Claim 4.8 below. □

Proof of Claim 4.8. Let $\ell=n^{3/4}$, $h_n=\sqrt{n}\log^2 n$, and $A=\{\eta\in\Omega^\infty_{n+2\ell}:\ \eta(i)\geq h_n\,\forall i\in\Lambda_n\}$. Using the above representation for the measure $\mu_{\ell,n}$ together with an obvious shift by ℓ, the Markov property, the FKG inequality and (18) we can bound $\mu_{\ell,n}(A)$ below by

$$\begin{aligned}&\Pr\big[X(i)\geq 0\,\forall i\in[0,\ell]\,;\,X(\ell+1)\geq h_n\mid X(0)=0\big]^2\\ &\qquad\times\Pr\big[X(i)\geq 0\,\forall i\in\Lambda_n\mid X(0)=X(n)=0\big]\\ &\geq e^{-3c\log n}\Pr[X(\ell+1)\geq h_n\mid X(0)=0]^2\\ &\geq e^{-3c\log n}e^{-c'h_n^2/\ell}\\ &\geq e^{-cn^{1/4}\log^4 n},\end{aligned}$$

where in the penultimate step we used standard results on large deviations. □

Proof of Claim 4.13. Using Lemma A.1, we can lower bound $\mu_n(A)$ by

$$\Pr\big[X(i_j)=0\ \forall j\,;\,C_{0,H_n,n}\mid X(0)=0\big]=$$
$$\Pr\big[C_{0,H_n,n^{3/4}}\mid X(0)=0\big]^{n^{1/4}}\geq e^{-cn^{1/4}\log n}.\quad\square$$