

OPTIMAL MIXING FOR THE HARD-CORE MODEL ON ARBITRARY TREES AT SMALL FUGACITY

ABSTRACT. We prove that, for every fixed $0 < \lambda \leq 2^{-10}$, single-site heat-bath Glauber dynamics on every n -vertex tree mixes in $O_\lambda(n \log(n+1))$ discrete steps. More generally, its mixing time to accuracy ε is at most $C_\lambda n(\log(n+1) + \log(1/\varepsilon))$. The main finite-time regularization estimate is that two units of continuous-time single-site dynamics on a star contract relative entropy by a fixed factor independent of the number of leaves. The estimate tensorizes over forests of stars and holds under every feasible boundary condition. We combine it with an entropy factorization over two consecutive depth levels. The resulting auxiliary chain uses exact fixed-time single-site kernels, rather than block heat-bath updates that must be approximated to high accuracy. Commutation parallelizes its updates into logarithmically many constant-duration rounds. Monotone censoring then transfers the estimate to the original dynamics.

Notes. This work was fully AI generated using GPT 6-Astra and the Codex harness, prompted by Zhangsong Li without any mathematical input.

1. STATEMENT AND THE FINITE-TIME REGULARIZATION PRINCIPLE

For a finite tree $T = (V, E)$, let Ω_T be its independent sets and write

$$\mu(\sigma) = \frac{\lambda^{|\sigma|}}{Z_{\lambda,T}}, \quad \sigma \in \Omega_T.$$

A single-site heat-bath update chooses a vertex, proposes occupation with probability $p = \lambda/(1+\lambda)$ and vacancy with probability $1-p$, and rejects an occupation proposal if a neighbor is occupied. We use rate-one clocks at the vertices in continuous time. For $A \subseteq V$, define

$$P_A f = \mathbb{E}_\mu[f \mid \sigma_{V \setminus A}], \quad L_A = \sum_{v \in A} (P_{\{v\}} - I), \quad L = L_V.$$

Thus P_A is an exact block heat-bath kernel, whereas e^{tL_A} runs the actual single-site dynamics in A for time t , with the exterior frozen. The discrete-time transition matrix in the problem is

$$M = \frac{1}{n} \sum_{v \in V} P_{\{v\}}, \quad L = n(M - I).$$

All logarithms are natural.

Theorem 1.1. *Set $\lambda_0 = 2^{-10}$. For every fixed $0 < \lambda \leq \lambda_0$, there is $C_\lambda < \infty$ such that, for every $n \geq 1$, every n -vertex tree and every $0 < \varepsilon \leq 1/4$,*

$$t_{\text{mix}}^{\text{disc}}(\varepsilon) \leq C_\lambda n(\log(n+1) + \log(1/\varepsilon)).$$

In particular, the requested $C_\lambda n \log n$ bound holds for every $n \geq 2$ at accuracy $1/4$. The continuous-time mixing time with rate one per vertex satisfies the same bound without the factor n .

The finite-time viewpoint is essential. The supplied monotone-censoring argument separates regularization from a dimension-free infinitesimal logarithmic Sobolev inequality. We use the same separation, but replace global self-return cores by a local event on which a star's center is vacant and its leaves evolve by product noise. This gives a *multiplicative finite-time entropy contraction*, with no additive error.

A direct fixed-density reduction to the product cube would not work. Indeed, under independent Bernoulli(p) occupations on a path, the probability of being an independent set is at most $(1-p^2)^{\lfloor n/2 \rfloor}$, by considering disjoint edges. Instead, we use blocks consisting of two depth levels, each of which is a forest of stars.

There are three substantive estimates. First, two units of dynamics on any star forest lose a fixed fraction of the conditional entropy. Second, the total entropy is at most seven times the sum of the conditional entropies on depth bands. Third, the auxiliary band-update process can be implemented exactly in logarithmically many constant-duration rounds. In particular, there is no approximate simulation of block heat-bath updates and no per-round error that must be made inverse-polynomially small.

2. ENTROPY CONTRACTION FOR CHANNELS

For a probability measure ρ and $f \geq 0$, write

$$\text{Ent}_\rho(f) = \rho(f \log f) - \rho(f) \log \rho(f).$$

For probabilities $\nu \ll \rho$, write $D(\nu \parallel \rho) = \text{Ent}_\rho(d\nu/d\rho)$. Kernels act on row measures from the right. The following elementary channel facts are included to keep the argument self-contained; see also [3] for the general strong data-processing framework.

Lemma 2.1 (Tensorization). *Let K_i be a channel with reference input probability ρ_i , and suppose that, for every probability ν_i ,*

$$D(\nu_i K_i \parallel \rho_i K_i) \leq \eta D(\nu_i \parallel \rho_i), \quad 0 \leq \eta \leq 1.$$

Then the same inequality holds for $K = \bigotimes_i K_i$ and $\rho = \bigotimes_i \rho_i$, including arbitrary correlated input probabilities ν .

Proof. Let $X = (X_1, \dots, X_m) \sim \nu$, and generate Y_i independently conditional on X through K_i . The reference output is the product of $\rho_i K_i$. The relative-entropy chain rule and the one-coordinate hypothesis give

$$\begin{aligned} D(\mathcal{L}(Y) \parallel \rho K) &= \sum_i \mathbb{E} D(\mathcal{L}(Y_i \mid Y_{<i}) \parallel \rho_i K_i) \\ &\leq \eta \sum_i \mathbb{E} D(\mathcal{L}(X_i \mid Y_{<i}) \parallel \rho_i) \\ &\leq \eta \sum_i \mathbb{E} D(\mathcal{L}(X_i \mid X_{<i}) \parallel \rho_i) = \eta D(\nu \parallel \rho). \end{aligned}$$

For the second inequality, $Y_{<i}$ is obtained from $X_{<i}$ by a channel independent of X_i conditional on $X_{<i}$. Thus the conditional law given $Y_{<i}$ is a mixture of the conditional laws given $X_{<i}$, and convexity of relative entropy applies. Zero-probability conditioning events can be omitted. $\square$

Lemma 2.2 (Bounded change of reference). *Suppose a channel K satisfies*

$$D(\nu K \parallel \beta K) \leq \eta D(\nu \parallel \beta)$$

for every probability ν , and let ρ have the same support as β , with

$$0 < m \leq \frac{d\rho}{d\beta} \leq M < \infty.$$

Then

$$D(\nu K \parallel \rho K) \leq \left(1 - \frac{m}{M}(1 - \eta)\right) D(\nu \parallel \rho) \tag{2.1}$$

for every probability ν on that support.

Proof. Put $\Psi(u, v) = u \log(u/v) - u + v \geq 0$, for $u \geq 0$ and $v > 0$. For an input probability σ , consider the joint probability $\sigma(x)K(x, y)$, and let

$$g_\sigma(y) = \mathbb{E}_{\sigma K}[f(X) \mid Y = y], \quad \Delta_{\sigma, K}(f) = \text{Ent}_\sigma(f) - \text{Ent}_{\sigma K}(g_\sigma),$$

where in $\text{Ent}_{\sigma K}(g_\sigma)$ the notation σK denotes the output marginal. The variational identities

$$\begin{aligned} \text{Ent}_\sigma(f) &= \inf_{a>0} \mathbb{E}_\sigma \Psi(f, a), \\ \Delta_{\sigma, K}(f) &= \inf_{h>0} \mathbb{E}_{\sigma(x)K(x, y)} \Psi(f(x), h(y)) \end{aligned}$$

follow by minimizing over the constant a , or separately over $h(y)$. Values zero at the minimizer are handled by limits. Comparing the two input measures gives

$$\text{Ent}_\rho(f) \leq M \text{Ent}_\beta(f), \quad \Delta_{\rho, K}(f) \geq m \Delta_{\beta, K}(f).$$

The assumed contraction, extended to unnormalized f by homogeneity, says that $\Delta_{\beta, K}(f) \geq (1 - \eta) \text{Ent}_\beta(f)$. Consequently

$$\Delta_{\rho, K}(f) \geq \frac{m}{M} (1 - \eta) \text{Ent}_\rho(f).$$

Apply this to $f = d\nu/d\rho$ to obtain (2.1). $\square$

Corollary 2.3 (Product refresh). *Let β be any product probability and U_t its rate-one coordinate-refresh semigroup. Then*

$$D(\nu U_t \parallel \beta) \leq e^{-t} D(\nu \parallel \beta).$$

If $m \leq d\rho/d\beta \leq M$, then

$$D(\nu U_t \parallel \rho U_t) \leq \left(1 - \frac{m}{M} (1 - e^{-t})\right) D(\nu \parallel \rho).$$

Proof. In one coordinate, $\nu U_t = e^{-t} \nu + (1 - e^{-t}) \beta$, so the first claim follows from convexity of relative entropy. Lemma 2.1 extends it to a product. The second claim is Lemma 2.2. $\square$

3. FINITE-TIME ENTROPY CONTRACTION ON STARS

For the remainder of the proof set

$$p = \frac{\lambda}{1 + \lambda}, \quad a_\lambda = e^{-2p} (1 - e^{-(1-p)}), \quad \alpha_\lambda = a_\lambda \frac{1 - e^{-1}}{1 + \lambda} > 0. \quad (3.1)$$

No smallness assumption on λ is needed in this section.

Proposition 3.1 (A degree-independent fixed-time contraction). *Let $\mu_\star$ be the hard-core measure with fugacity $\lambda > 0$ on a star with any number of leaves, and let S_t be its rate-one-per-vertex single-site semigroup. Then, for every probability ν ,*

$$D(\nu S_t \parallel \mu_\star) \leq (1 - \alpha_\lambda) D(\nu \parallel \mu_\star). \quad (3.2)$$

The same estimate holds for every forest of stars, with any feasible fixed boundary condition, using the same α_λ .

Proof. Write C for the center occupation and $Y \in \{0, 1\}^d$ for the leaf vector. Let

$$\beta = \text{Bern}(p)^{\otimes d}, \quad b = \beta(0^d) = (1 + \lambda)^{-d}, \quad q = \mu_\star(C = 1) = \frac{\lambda b}{1 + \lambda b}.$$

The stationary measure has the exact decomposition

$$\mu_\star = (1 - q)(\delta_0 \otimes \beta) + q\delta_{(1, 0^d)}. \quad (3.3)$$

Construct the center proposals as independent occupation and vacancy Poisson processes, of rates p and $1 - p$. Let $\mathcal{E}$ be the event that there is no center occupation proposal during $[0, 2]$, and

at least one center vacancy proposal during $[0, 1]$. It is independent of the initial state and of all leaf randomness, and $\mathbb{P}(\mathcal{E}) = a_\lambda$.

Conditional on $\mathcal{E}$, let $s \leq 1$ be the first center vacancy-proposal time. Its conditional distribution is also independent of the initial state. The center is certainly vacant during $[1, 2]$. Let H_s be the channel from the initial star state to the leaves at time 1, conditional on this center event and this first vacancy time. If the initial center is vacant, the leaves evolve freely throughout $[0, 1]$. If it is occupied, all leaves stay vacant until s and then evolve freely. Hence the reference law at time 1 is

$$\rho_s := \mu_\star H_s = (1 - q)\beta + q U_{1-s}(0^d, \cdot).$$

Stationarity of β for U_t implies, pointwise,

$$b U_t(0^d, y) \leq \beta(y).$$

Using $q/b = (1 - q)\lambda$, we obtain the dimension-free bounds

$$1 - q \leq \frac{\rho_s(y)}{\beta(y)} \leq (1 - q)(1 + \lambda). \quad (3.4)$$

In particular, the ratio of the upper and lower bounds is at most $1 + \lambda$, even when q is exponentially small in d .

During the last unit interval, the center remains vacant and the leaves undergo independent product refresh. For every input law ν , Corollary 2.3, followed by data processing through H_s , gives

$$\begin{aligned} D(\nu H_s U_1 \| \mu_\star H_s U_1) &\leq \left(1 - \frac{1 - e^{-1}}{1 + \lambda}\right) D(\nu H_s \| \rho_s) \\ &\leq \left(1 - \frac{1 - e^{-1}}{1 + \lambda}\right) D(\nu \| \mu_\star). \end{aligned}$$

Adjoining the deterministic final center state does not change relative entropy. Averaging over s by joint convexity proves the same contraction for the conditional transition kernel $K_{\mathcal{E}}$.

Let $K_{\mathcal{E}^c}$ denote the transition kernel conditional on the complementary center event. This is a channel independent of the input distribution, so ordinary data processing gives

$$D(\nu K_{\mathcal{E}^c} \| \mu_\star K_{\mathcal{E}^c}) \leq D(\nu \| \mu_\star).$$

Since

$$S_2 = a_\lambda K_{\mathcal{E}} + (1 - a_\lambda) K_{\mathcal{E}^c}, \quad \mu_\star S_2 = \mu_\star,$$

joint convexity now yields

$$\begin{aligned} D(\nu S_2 \| \mu_\star) &\leq a_\lambda D(\nu K_{\mathcal{E}} \| \mu_\star K_{\mathcal{E}}) + (1 - a_\lambda) D(\nu K_{\mathcal{E}^c} \| \mu_\star K_{\mathcal{E}^c}) \\ &\leq (1 - \alpha_\lambda) D(\nu \| \mu_\star). \end{aligned}$$

The reference measures on the two branches need not be stationary; they are kept separately until the last joint-convexity step.

The measure and the dynamics on a forest of stars are products, so Lemma 2.1 proves the forest assertion. A feasible occupied boundary vertex only forces its adjacent interior vertices to be vacant. Deleting those forced-vacant vertices leaves smaller stars and isolated vertices. An isolated vertex contracts entropy by e^{-2} in two time units, which is at most $1 - \alpha_\lambda$ because $\alpha_\lambda \leq 1 - e^{-1}$. Deterministic components have zero entropy. These observations prove the assertion with boundary conditions as well. $\square$

Remark 3.2 (What the estimate does not assert). Proposition 3.1 is an estimate at the fixed time 2, not an infinitesimal entropy inequality. For example, on a large star the center-occupied state has stationary mass $q = \lambda / ((1 + \lambda)^d + \lambda)$ and leaves that state at rate $1/(1 + \lambda)$. For $g = \mathbf{1}_{\{C=1\}} / \sqrt{q}$,

$$\text{Ent}_{\mu_\star}(g^2) = \log(1/q), \quad -\mu_\star(g L g) = \frac{1}{1 + \lambda}.$$

Thus a dimension-free classical logarithmic Sobolev inequality is unavailable. Also, the proof never waits for every leaf to update. It only uses a fixed fraction of product-noise entropy contraction during the vacant-center window.

4. A TWO-LEVEL TREE ESTIMATE

We next prove an entropy factorization adapted to depth bands. This is a direct, level-grouped version of the tree martingale method; the broader connection between spatial mixing and functional inequalities on trees is developed in [1].

Lemma 4.1. *Consider a finite rooted tree with vertex fugacities in $(0, \lambda_0]$. Let X be the root occupation, let Y collect the children, and let Z collect the vertices at distance at least two from the root. For every real g ,*

$$\|\mathbb{E}[g(X) \mid Z]\|_8 \leq \|g(X)\|_2. \quad (4.1)$$

The assertion also holds when the root is forced vacant, and it tensorizes over a product of such rooted-tree measures.

Proof. The forced-vacant case is immediate. Otherwise let $a \leq \lambda_0$ be the root fugacity. Conditional on $X = 0$, the child subtrees are independent. For the i th child, write Y_i for its occupation and Z_i for its strict descendants, and put

$$q_i = \mathbb{P}(Y_i = 1 \mid X = 0), \quad Q_i(Z_i) = \mathbb{P}(Y_i = 1 \mid Z_i, X = 0).$$

With $p_0 = \lambda_0/(1 + \lambda_0)$, the heat-bath conditional law gives $0 \leq Q_i \leq p_0$. Let ν_i be the law of Z_i conditional on $X = 0$, and define

$$\nu = \bigotimes_i \nu_i, \quad U = \prod_i (1 - Q_i), \quad k = \prod_i (1 - q_i).$$

Then $\nu_i(Q_i) = q_i$ and $\nu(U) = k$. Bayes' rule gives

$$q := \mathbb{P}(X = 1) = \frac{ak}{1 + ak}, \quad R(Z) := \mathbb{P}(X = 1 \mid Z) = \frac{aU}{1 + aU}, \quad \frac{d\mu_Z}{d\nu} = (1 - q)(1 + aU). \quad (4.2)$$

To verify these identities, the likelihood ratio of Z_i under $Y_i = 0$ relative to its law under $X = 0$ is $(1 - Q_i)/(1 - q_i)$, and $X = 1$ forces every Y_i to be zero.

For $0 \leq x \leq p_0$, convexity on $[0, p_0]$ and the second-order Taylor bound imply

$$(1 - x)^8 \leq 1 - \frac{1 - (1 - p_0)^8}{p_0} x \leq 1 - 7x,$$

because $p_0 \leq 1/28$ and $1 - (1 - p_0)^8 \geq 8p_0 - 28p_0^2 \geq 7p_0$. Consequently

$$\nu_i[(1 - Q_i)^8] \leq 1 - 7q_i \leq (1 - q_i)^7, \quad \nu(U^8) \leq k^7.$$

By (4.2),

$$\mathbb{E}_{\mu_Z} R^8 = (1 - q)a^8 \mathbb{E}_\nu \frac{U^8}{(1 + aU)^7} \leq a^8 k^7.$$

Set $\phi(X) = (X - q)/\sqrt{q(1 - q)}$ and $h(Z) = \mathbb{E}[\phi(X) \mid Z]$. Then $\mathbb{E}h = 0$ and

$$\begin{aligned} \|h\|_8 &\leq \frac{ak^{7/8} + q}{\sqrt{q(1 - q)}} \\ &= \sqrt{a}[(1 + ak)k^{3/8} + \sqrt{k}] \leq (2 + a)\sqrt{a} \leq 3\sqrt{\lambda_0} = \frac{3}{32} < \frac{1}{8}. \end{aligned}$$

Every function of X is $g = A + B\phi$. By homogeneity suppose $A^2 + B^2 = 1$. Expanding the eighth power and bounding all terms of degree at least two by their absolute values gives

$$\begin{aligned}\mathbb{E}(A + B\phi)^8 &\leq A^8 + B^2 \sum_{j=2}^8 \binom{8}{j} 8^{-j} \\ &= A^8 + B^2[(9/8)^8 - 2] \leq A^8 + B^2 \leq 1.\end{aligned}$$

Here $|A|^{8-j}|B|^j \leq B^2$ for $j \geq 2$, and $(9/8)^8 - 2 < 1$. This proves (4.1).

For tensorization, let K_1, K_2 be two of the conditional-expectation operators. The one-component bounds and the mixed-norm form of Minkowski's inequality yield

$$\begin{aligned}\|(K_1 \otimes K_2)g\|_{L^8(Z_1, Z_2)} &\leq \|K_1 g\|_{L^8(Z_1; L^2(X_2))} \\ &\leq \|K_1 g\|_{L^2(X_2; L^8(Z_1))} \leq \|g\|_{L^2(X_1, X_2)}.\end{aligned}$$

Iteration proves the result for every finite product. $\square$

For $A \subseteq V$, use the notation

$$H_A(f) = \text{Ent}_\mu(f) - \text{Ent}_\mu(P_A f), \quad f \geq 0.$$

This is the mean conditional entropy in A given its complement.

Lemma 4.2 (Local entropy estimate). *In the setting of Lemma 4.1, including a product of rooted trees, let X denote the root vector, let W be its set of vertices, and let B consist of those roots and their children. Then*

$$\text{Ent}_\mu(\mathbb{E}[f \mid X]) \leq \frac{7}{3}H_B(f) + \frac{1}{3}H_{V \setminus W}(f), \quad f \geq 0. \quad (4.3)$$

Proof. Normalize $\mu(f) = 1$. Write $D = \text{Ent}_\mu(f)$, and let D_X, D_Z be the relative entropies of the X and Z marginals of $f\mu$ relative to the corresponding marginals of μ . Lemma 4.1 and Hölder's inequality give

$$\mathbb{E}_\mu[a(X)b(Z)] \leq \|a\|_2 \|b\|_{8/7}.$$

For the marginal density ratios r_X, r_Z , substitute $a = r_X^{1/2}$ and $b = r_Z^{7/8}$. Their indicated norms equal one, so

$$\mathbb{E}_\mu \exp \left\{ \frac{1}{2} \log r_X + \frac{7}{8} \log r_Z \right\} \leq 1.$$

The variational formula for relative entropy implies

$$\frac{1}{2}D_X + \frac{7}{8}D_Z \leq D.$$

Using $D_Z = D - H_B(f)$ and $D = D_X + H_{V \setminus W}(f)$ gives (4.3). Zeros in marginal density ratios are handled by approximation, and homogeneity removes the normalization. $\square$

5. ENTROPY FACTORIZATION OVER DEPTH BANDS

Root T arbitrarily, let V_j be the vertices at depth j , and let h be the maximum depth. Set $V_j = \emptyset$ for $j > h$, and define

$$B_j = V_j \cup V_{j+1}, \quad F_j = \bigcup_{i \geq j} V_i, \quad M_j f = P_{F_j} f.$$

Thus $M_0 f = \mu(f)$ and $M_j f = f$ for $j \geq h+1$. Every B_j induces a forest of stars.

Proposition 5.1 (Band entropy factorization). *For $0 < \lambda \leq \lambda_0$ and every $f \geq 0$,*

$$\text{Ent}_\mu(f) \leq 7 \sum_{j=0}^h H_{B_j}(f). \quad (5.1)$$

Proof. Put $b_j(f) = \text{Ent}_\mu(M_{j+1}f) - \text{Ent}_\mu(M_jf)$ for $0 \leq j \leq h$ and $b_j(f) = 0$ for $j > h$. Conditional Jensen and telescoping give

$$b_j(f) \geq 0, \quad \sum_{j=0}^h b_j(f) = \text{Ent}_\mu(f).$$

Condition on all levels above V_j . The remaining law is a product of rooted-tree measures, whose roots are the vertices of V_j . A root with an occupied parent is forced vacant. Apply Lemma 4.2 to $g = M_{j+3}f$ in this conditional forest and then average over the conditioning. The tower property identifies the left side as $b_j(f)$, giving

$$b_j(f) \leq \frac{7}{3}H_{B_j}(M_{j+3}f) + \frac{1}{3}H_{F_{j+1}}(M_{j+3}f).$$

The sets B_j and F_{j+3} are disjoint and have no edge between them. Conditional on the remaining vertices, their configurations are independent. Convexity of conditional entropy therefore gives

$$H_{B_j}(M_{j+3}f) \leq H_{B_j}(f).$$

The nested conditional expectations give the exact identity

$$\begin{aligned} H_{F_{j+1}}(M_{j+3}f) &= \text{Ent}_\mu(M_{j+3}f) - \text{Ent}_\mu(M_{j+1}f) \\ &= b_{j+1}(f) + b_{j+2}(f). \end{aligned}$$

Consequently

$$b_j(f) \leq \frac{7}{3}H_{B_j}(f) + \frac{1}{3}(b_{j+1}(f) + b_{j+2}(f)).$$

Summing in j and moving the resulting $(2/3)\text{Ent}_\mu(f)$ to the left proves (5.1). $\square$

Remark 5.2. The proof works directly with the entire depth bands. It does not replace a sum of individual-star conditional entropies by a band entropy, an inequality that would have the wrong direction for general densities.

6. EXACT FINITE-TIME BAND KERNELS

For every depth band define

$$K_j = e^{2L_{B_j}}. \tag{6.1}$$

This is two units of actual single-site evolution in B_j , not P_{B_j} . It is reversible with respect to μ and preserves the exterior configuration.

Lemma 6.1. *For every $f \geq 0$,*

$$\text{Ent}_\mu(f) - \text{Ent}_\mu(K_j f) \geq \alpha_\lambda H_{B_j}(f). \tag{6.2}$$

Proof. Condition on the exterior of B_j . The conditional measure is the hard-core measure on a forest of stars with feasible boundary condition. Proposition 3.1, applied to the normalized conditional density and then extended by homogeneity, contracts its entropy by $1 - \alpha_\lambda$. Average over the exterior. The conditional mean of f is unchanged by K_j , so the entropy chain rule gives precisely (6.2). $\square$

Let the auxiliary continuous-time chain have generator

$$\mathcal{L} = \sum_{j=0}^h (K_j - I).$$

Its graphical construction has a rate-one Poisson clock for each band and applies K_j at a ring of that band. These clocks describe *virtual* time. Every K_j preserves μ and is self-adjoint on $L^2(\mu)$, so the same is true of the auxiliary semigroup.

Proposition 6.2 (Virtual-time entropy decay). *If f_t is the density of this auxiliary chain relative to μ , then*

$$\text{Ent}_\mu(f_t) \leq e^{-\alpha_\lambda t/7} \text{Ent}_\mu(f_0). \quad (6.3)$$

Proof. For a stationary Markov kernel K and positive f of mean one,

$$\mu[(f - Kf) \log f] \geq \text{Ent}_\mu(f) - \text{Ent}_\mu(Kf),$$

because the difference between the left and right sides is $D((Kf)\mu \| f\mu) \geq 0$. Therefore Lemma 6.1 and Proposition 5.1 imply

$$\begin{aligned} -\frac{d}{dt} \text{Ent}_\mu(f_t) &= \sum_j \mu[(f_t - K_j f_t) \log f_t] \\ &\geq \sum_j (\text{Ent}_\mu(f_t) - \text{Ent}_\mu(K_j f_t)) \\ &\geq \alpha_\lambda \sum_j H_{B_j}(f_t) \geq \frac{\alpha_\lambda}{7} \text{Ent}_\mu(f_t). \end{aligned}$$

Integrating proves the result. Approximation by positive densities handles arbitrary initial laws. $\square$

For $\lambda < 1$, put $c_\lambda = \log((1 + \lambda)/\lambda)$. Since $Z_{\lambda,T} \leq (1 + \lambda)^n$ and $\lambda^{|\sigma|} \geq \lambda^n$,

$$\mu(\sigma) \geq \left(\frac{\lambda}{1 + \lambda}\right)^n, \quad \text{Ent}_\mu\left(\frac{\mathbf{1}_{\{\sigma\}}}{\mu(\sigma)}\right) \leq c_\lambda n. \quad (6.4)$$

Thus the auxiliary chain mixes to any accuracy δ after virtual time

$$t_* = \frac{7}{\alpha_\lambda} \log \frac{c_\lambda n}{2\delta^2}, \quad (6.5)$$

provided the displayed logarithm is positive. Indeed, (6.3), (6.4) and Pinsker's inequality bound its total-variation distance by δ . The parameters chosen below always make this logarithm positive.

7. PARALLEL IMPLEMENTATION WITHOUT BLOCK APPROXIMATION

Lemma 7.1 (Depth of a band-update word). *Run the independent rate-one band clocks for virtual time t . Their update word can be rearranged, without changing its transition kernel, into D rounds. Every round consists of bands whose indices have pairwise distance at least three, and its kernel equals e^{2L_A} , where A is the union of those bands. For every integer $r \geq 1$,*

$$\mathbb{P}(D \geq r) \leq n \left(\frac{5et}{r}\right)^r. \quad (7.1)$$

The round schedule depends only on the virtual Poisson clocks, not on any spin values or any single-site proposal randomness.

Proof. If $|i - j| \geq 3$, then B_i, B_j are disjoint and nonadjacent. All single-site kernels in the two sets commute, hence so do L_{B_i}, L_{B_j} and K_i, K_j . Moreover,

$$K_i K_j = e^{2(L_{B_i} + L_{B_j})} = e^{2L_{B_i \cup B_j}}.$$

Construct a directed acyclic graph on the virtual clock events: draw an edge from each earlier event to each later event whose label differs by at most two. Assign to an event the maximum number of vertices in a directed path ending there. Events with the same assigned number form one round. Any two labels in a round differ by at least three. Sorting by round number is a topological ordering and can be achieved by interchanging only commuting kernels. This proves the kernel identity and the exact implementation assertion.

If $D \geq r$, there is a sequence of r chronologically ordered clock events with labels $j_1, \dots, j_r$ satisfying $|j_{k+1} - j_k| \leq 2$. There are at most $(h + 1)5^{r-1} \leq n5^{r-1}$ possible label sequences. For any

fixed sequence, the expected number of ordered event tuples is $t^r/r!$, also when labels repeat, by the factorial moment formula for Poisson processes. A union bound and $r! \geq (r/e)^r$ prove (7.1). $\square$

Fix $0 < \varepsilon \leq 1/4$, and choose

$$\delta = \frac{\varepsilon}{8n}, \quad t_* = \frac{7}{\alpha_\lambda} \log \frac{c_\lambda n}{2\delta^2}, \quad R = \left\lceil 10et_* + 10 \log \frac{8n^2}{\varepsilon} + 10 \right\rceil, \quad S = 2R. \quad (7.2)$$

By (7.1) and $R \geq 10et_*$,

$$\mathbb{P}(D > R) \leq \mathbb{P}(D \geq R) \leq n2^{-R} \leq \delta.$$

Here $R \geq 10 \log(n/\delta)$, and $10 \log 2 > 1$.

Sample the virtual clocks on $[0, t_*]$ independently of all randomness of the actual single-site chain. Retain only the first R rounds of the resulting schedule, appending empty rounds if there are fewer. Denote the round unions by $A_1, \dots, A_R$. During the r th real interval of length 2, censor all updates outside A_r and perform the ordinary single-site updates inside A_r .

By Lemma 7.1, on $\{D \leq R\}$ this censored chain implements *exactly* the conditional transition kernel of the auxiliary chain at virtual time t_* . On the complementary event both constructions are valid probability kernels. Therefore their unconditional laws differ by at most $\mathbb{P}(D > R) \leq \delta$, uniformly over the initial state. The auxiliary-chain estimate gives

$$\sup_{\sigma} \|\mathcal{L}_{\sigma}(X_S^{\text{cens}}) - \mu\|_{\text{TV}} \leq 2\delta. \quad (7.3)$$

There is no other approximation error. In particular,

$$S = O_\lambda(\log(n+1) + \log(1/\varepsilon)). \quad (7.4)$$

8. MONOTONE CENSORING AND THE ORIGINAL CHAIN

We include the required censoring argument, rather than appealing to a comparison of spectral gaps. It is the finite-state binary version of the Peres–Winkler censoring principle [2].

Let $V = V_e \sqcup V_o$ be the bipartition and transform the occupations by

$$\xi_v = \sigma_v \quad (v \in V_e), \quad \xi_v = 1 - \sigma_v \quad (v \in V_o).$$

On an even–odd edge the hard-core constraint becomes $\xi_u \leq \xi_v$. The transformed state space has a bottom $\mathbf{0}$ and a top $\mathbf{1}$, corresponding respectively to occupation of all odd or all even vertices. Every single-site kernel is monotone in this order. For an even vertex its conditional probability of $\xi_v = 1$ is p if all neighboring transformed spins equal one and zero otherwise. For an odd vertex it is 1 if any neighboring transformed spin equals one and $1 - p$ otherwise. Both probabilities are increasing functions of the neighboring configuration.

Lemma 8.1 (Censoring from the extrema). *Starting from the top or from the bottom, deleting updates according to a random schedule independent of the single-site proposal randomness cannot improve total-variation convergence to μ .*

Proof. It suffices first to fix a finite coordinate-update word and a subsequence. Suppose the current density f relative to μ is increasing. Reversibility implies that its density after updating v is $P_{\{v\}}f$, which remains increasing by monotonicity of the kernel. For every increasing g ,

$$\mu(fg) - \mu[(P_{\{v\}}f)g] = \mu[\text{Cov}(f, g \mid \sigma_{V \setminus \{v\}})] \geq 0.$$

The covariance is nonnegative because each conditional fiber has at most two ordered states. Thus an update decreases this law in stochastic order; applying subsequent monotone kernels preserves the order comparison.

The top point mass has increasing density. Repeatedly inserting the missing updates shows that the censored law β stochastically dominates the uncensored law α , and both have increasing densities. The set $A = \{d\alpha/d\mu \geq 1\}$ is increasing, so

$$\|\alpha - \mu\|_{\text{TV}} = \alpha(A) - \mu(A) \leq \beta(A) - \mu(A) \leq \|\beta - \mu\|_{\text{TV}}.$$

For continuous-time and random schedules, condition on the independent schedule and the coordinate clocks, apply the finite-word comparison, and average. Stochastic order and increasing densities are preserved under mixtures, so the same total-variation argument applies to the averaged laws. Reversing the order gives the result from the bottom. $\square$

Proof of Theorem 1.1. Use the independent censoring schedule constructed in the preceding section. By (7.3) and Lemma 8.1, the uncensored chain at time S from each extremal state is within 2δ of μ .

Couple the uncensored processes X_t^+, X_t^- from the two extrema by common rate-one clocks and monotone proposal uniforms in transformed coordinates. Then $X_t^+ \geq X_t^-$ at all times. At a fixed vertex, the disagreement probability equals the difference of the two one-site means, hence is at most the sum of the two total-variation distances to μ , or 4δ . Therefore

$$\mathbb{P}(X_S^+ \neq X_S^-) \leq 4n\delta = \varepsilon/2. \quad (8.1)$$

Every other initial configuration, including an initially stationary one, can be coupled between these extrema. Their coalescence is permanent. This proves the continuous-time mixing bound, with error at most $\varepsilon/2$, using (7.4).

For the discrete-time statement, superpose all vertex clocks. The total count N_S has law $\text{Pois}(nS)$, and the successive coordinate marks are uniform vertices. The embedded chain is exactly the discrete-time heat-bath chain in the problem. Set $m = \lceil 2nS \rceil$. If the extremal coupling has coalesced by time S and $N_S \leq m$, the embedded chains have coalesced by step m . Thus

$$\sup_{\sigma} \|M^m(\sigma, \cdot) - \mu\|_{\text{TV}} \leq \varepsilon/2 + \mathbb{P}(N_S > m).$$

Using $\mathbb{E}2^{N_S} = e^{nS}$ gives

$$\mathbb{P}(N_S > m) \leq \exp(-(2\log 2 - 1)nS) \leq \varepsilon/2.$$

For the last inequality, (7.2) gives $nS \geq 20 \log(8n^2/\varepsilon)$, and $20(2\log 2 - 1) > 1$. This is a pathwise union bound: no independence between coalescence and N_S is assumed. Finally, $m \leq 2nS + 1$ and (7.4) give the asserted order. The argument also covers $n = 1$; the expression $\log(n + 1)$ avoids the zero in $\log n$ at that size. $\square$

9. HOW THIS REMOVES THE EXTRA LOGARITHMS

The transferable part of the monotone-censoring draft is the decision to prove finite-time regularization directly, instead of demanding a uniform infinitesimal logarithmic Sobolev inequality. Here the local vacant-center window replaces the global self-return-core construction. Its occurrence has fixed positive probability, and the reference law of the leaves after the first unit interval is a bounded perturbation of a product measure, by (3.4). Relative-entropy contraction then supplies a multiplicative estimate without discarding any mass.

The crucial implementation choice is (6.1): every virtual band update is already two units of single-site dynamics. There is no need to run a star forest to inverse-polynomial total-variation accuracy before moving to the next round. Although the virtual process makes $O(n \log n)$ individual band calls, the band interaction graph has bounded range. Its dependency depth is only $O(\log n)$ with the required high probability. Each parallel round costs exactly two real time units.

Four distinctions keep the argument valid. The stationary reference is not retained after conditioning on a center-clock event; each branch has its own reference image, and stationarity is used only after the branches are recombined. Channel contraction tensorizes for arbitrary correlated input distributions. The band entropy inequality is proved for whole bands, not by grouping individual-star

entropy terms. Finally, the actual censoring schedule depends only on auxiliary clocks, never on the spin configuration or proposal outcomes. These points remove, respectively, a conditioning error, an independence assumption, a wrong-way entropy inequality, and a state-dependent-censoring issue.

REFERENCES

- [1] F. Martinelli, A. Sinclair, and D. Weitz, *Glauber dynamics on trees: Boundary conditions and mixing time*, Communications in Mathematical Physics **250** (2004), 301–334. arXiv:math/0307336.
- [2] Y. Peres and P. Winkler, *Can extra updates delay mixing?*, Communications in Mathematical Physics **323** (2013), 1007–1016. arXiv:1112.0603.
- [3] M. Raginsky, *Strong data processing inequalities and Φ -Sobolev inequalities for discrete channels*, arXiv:1411.3575, version 4 (2016).