Appendixes

Appendix A

In this appendix, we express the four-vertex model on a 4-regular graph $G = (V, E)$ as a Holant problem with the constraint function f^*.

The four-vertex model is naturally expressed as a Holant problem, which we define as follows. A function $f : \{0,1\}^k \to \mathbb{C}$ is called a constraint function of arity k. In this paper, we restrict f to take non-negative values in $\mathbb{Q}^+$. Fix a set of constraint functions $\mathcal{F}$. A function grid $\Gamma = (G, \xi)$ is a tuple, where $G = (V, E)$, ξ labels each $v \in V$ with a function $f_v \in \mathcal{F}$ of arity $\deg(v)$, and the incident edges $E(v)$ at v are identified as input variables to f_v, also labeled by ξ. Every assignment $\sigma : E \to \{0,1\}$ gives an evaluation $\prod_{v \in V} f_v(\sigma|_{E(v)})$, where $\sigma|_{E(v)}$ denotes the restriction of σ to $E(v)$. The problem Holant($\mathcal{F}$) on an instance Γ is to compute

$$\text{Holant}(\Gamma; \mathcal{F}) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\sigma|_{E(v)}).$$

We use Holant $(\mathcal{F}|\mathcal{G})$ for Holant problems over function grids with a bipartite graph (U, V, E), where each vertex in U (or V) is assigned a function in $\mathcal{F}$ (or $\mathcal{G}$, respectively).

To write the four-vertex model on a 4-regular graph $G = (V, E)$ as a Holant problem, we consider the edge-vertex incidence graph $G' = (U_E, U_V, E')$ of G. G' is a bipartite graph, where $(u_e, u_v) \in U_E \times U_V$ is an edge in E' iff $e \in E$ is incident to $v \in V$. We model the orientation of an edge in G by assigning the Disequality function $(\neq_2)$, which outputs 1 for inputs 01 and 10, and outputs 0 for inputs 00 and 11, at U_E in G'. We say that an orientation on edge $e = (v_i, v_j) \in E$ is directed from v_i to v_j in G if the edge (u_{v_i}, u_e) in G' takes value 1, and $(u_e, u_{v_j}) \in E'$ takes value 0. We write a 4-ary function $f(x_1, x_2, x_3, x_4)$ as a matrix

$$M(f) = M_{x_1 x_2, x_3 x_4}(f) = \begin{pmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{pmatrix}.$$

The constraint function for the four-vertex model can be expressed as a function with the constraint function

$$M(f) = \begin{pmatrix} 0 & 0 & 0 & \omega_1 \\ 0 & 0 & \omega_3 & 0 \\ 0 & \omega_4 & 0 & 0 \\ \omega_2 & 0 & 0 & 0 \end{pmatrix}.$$

If $\omega_1 = \omega_6 = a, \omega_3 = \omega_4 = b, \omega_2 = \omega_5 = c$, then we say that the four-vertex model has arrow reversal symmetry (ARS). In this paper, we consider the four-vertex model with the constraint function

$$M(f^*) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & b & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1}$$

Appendix B

In this appendix, we prove that f^* is unwindable.

We firstly give the definition of windable.

Definition 1. (Windable). For any finite set J and any configuration $x \in \{0,1\}^J$ define $\mathcal{M}_x$ to be the set of partitions of $\{i | x_i = 1\}$ into pairs and at most one singleton. A function $f : \{0,1\}^J \to \mathbb{Q}^+$ is windable if there exist values $B(x, y, M) \geq 0$ for all $x, y \in \{0,1\}^J$ and all $M \in \mathcal{M}_{x \oplus y}$ satisfying:

- $f(x)f(y) = \sum_{M \in \mathcal{M}_{x \oplus y}} B(x, y, M)$ for all $x, y \in \{0,1\}^J$.
- $B(x, y, M) = B(x \oplus S, y \oplus S, M)$ for all $x, y \in \{0,1\}^J$ and all $S \in M$.

Here, $x \oplus S$ denotes the vector obtained by changing x_i to $1 - x_i$ for the one or two elements i in S.

In the following, we will prove that the function defined in Eq. (1) is unwindable.

Proof. Suppose that the constraint function f^* is windable. Then for x=0101 and y=1010, we have

$$\mathcal{M}_{x\oplus y} = \{M_1 : \{(x_1, x_2), (x_3, x_4)\},$$
$$M_2 : \{(x_1, x_3), (x_2, x_4)\}, M_3 : \{(x_1, x_4), (x_2, x_3)\}\},$$

and there exists $B(x, y, M) \geq 0$ such that $f^*(0101) \cdot f^*(1010) = \sum_{M\in\mathcal{M}_{x\oplus y}} B(x, y, M) = b^2$ by Definition 1. Moreover, we have $B(x, y, M) = B(x \oplus S, y \oplus S, M)$ for $S = \{x_1, x_2\}$ by Definition 1, i.e.,

$$B(0101, 1010, M_1) = B(0101 \oplus \{x_1, x_2\},$$
$$1010 \oplus \{x_1, x_2\}, M_1) = B(1001, 0110, M_1).$$

Since $f^*(1001) \cdot f^*(0110) = \sum_{M\in\mathcal{M}_{x\oplus y}} B(x, y, M) = 0$, we have

$$B(0101, 1010, M_1) = B(1001, 0110, M_1) = 0. \quad (2)$$

Similarly, for $S = \{x_1, x_3\}$ and $S = \{x_1, x_4\}$, we have

$$B(0101, 1010, M_2) = B(1111, 0000, M_2) = 0 \quad (3)$$

and

$$B(0101, 1010, M_3) = B(1100, 0011, M_3) = 0. \quad (4)$$

The Eqs. (2), (3) and (4) implies that $f^*(0110) \cdot f^*(1001) = b^2 = 0$. This contradicts $b \neq 0$. □

Appendix C

In this appendix, we present some properties of Markov chain $\{Z_t\}$ for subsequent proofs.

A valid configuration of the four-vertex model is an assignment to all edges that results in a non-zero evaluation. By the transition probability matrix P, we have

$$P(\sigma'(C_i) = 1 | \sigma(\Gamma(C_i) = \mathbf{0}) = \frac{b^{\delta_{C_i}}}{1 + b^{\delta_{C_i}}}. \quad (5)$$

The Markov chain $\{Z_t\}$ has the following properties.

Proposition 1. For an arbitrary configuration σ, it must have a circuit C_i that satisfies $\sigma(\Gamma(C_i)) = \mathbf{0}$, i.e., all neighbors of C_i are assigned the value 0.

Proof. If σ is an all-zero string, then each circuit satisfies that its neighbors are assigned the value 0.

Otherwise, there exists a circuit C_i in σ assigned the value 1. Note that $\sigma \in \Omega$. Therefore, for two arbitrary adjacent circuits in σ, at most one of them can take the value 1 in a valid configuration. That implies that an arbitrary circuit C_i assigned the value 1 satisfies $\sigma(\Gamma(C_i)) = \mathbf{0}$. This completes the proof. □

Proposition 2. The transition probability matrix P of $\{Z_t\}$ has the following properties:

- aperiodicity: $\gcd\{t : P^t(i, i) > 0\} = 1$ for all $\sigma_i \in \Omega$;
- irreducibility: there exists t such that there is a positive probability of going from state σ_i to state σ_j after t steps, $P^t(i, j) > 0$, for all $\sigma_i, \sigma_j \in \Omega$.

Proof. For each state $\sigma_i \in \Omega$, there is a positive probability such that it stays the current state. Therefore, we have $gcd\{t : P^t(i.i) > 0\} = 1$ and aperiodicity is proved.

For each state $\sigma \in \Omega$, by taking the value 1 in σ to 0 step by step, there is a path to the state σ_0 in which all circuits take the value 0 and vice versa, i.e., for two arbitrary states, there is at least one path connected them by crossing σ_0. Irreducibility is proved. □

Based on the aperiodicity and irreducibility, we know that the chain $\{Z_t\}$ has a unique limiting distribution which is called the stationary distribution π, i.e.

$$\lim_{t\to\infty} P^t(i,j) = \pi_j, \text{ for all } \sigma_i, \sigma_j \in \Omega.$$

The distribution of the valid configuration $\mu(\sigma)$ is as follows:

$$\mu(\sigma) = \frac{\prod_{C_i:\sigma(C_i)=1} b^{\delta_{C_i}}}{\sum_{\sigma\in\Omega}\prod_{C_i:\sigma(C_i)=1} b^{\delta_{C_i}}}.$$

Note that $\sum_{\sigma\in\Omega}\prod_{C_i:\sigma(C_i)=1} b^{\delta_{C_i}}$ is the partition function of the four-vertex model with the constraint function f^* and the distribution μ is a Gibbs distribution. Moreover, we emphasize that the distribution μ is the unique stationary distribution of $\{Z_t\}$, because μ satisfies the detailed balance conditions:

$$\mu_i P^t(i,j) = \mu_j P^t(j,i),$$

for all $\sigma_i, \sigma_j \in \Omega$.

Our goal is to bound the time when the chain is close to the stationary distribution, i.e., we are interested in the mixing time $\tau(\varepsilon)$:

$$\tau(\varepsilon) = \max_i \min\{t : \Delta_i(t') \le \varepsilon \text{ for all } t' \ge t\},$$

where $\Delta_i(t)$ is the total variation distance between P_i^t and π, i.e.,

$$\Delta_i(t) = \frac{1}{2}\sum_{\sigma_i\in\Omega}\left|P^t(i,\cdot) - \pi_i\right|.$$

Appendix D

In this appendix, we introduce the applications of coupling to prove the rapid mixing of the Markov chain. And we give the mixing time for the four-vertex model on two-by-two-intersection free graph (Theorem 1). If there are no three circuits intersecting with each other, then we say that the underlying graph is two-by-two-intersection free.

Coupling:

We use coupling to prove the rapid mixing of $\{Z_t\}$. Coupling constructs a stochastic process (X_t, Y_t) such that both $\{X_t\}$ and $\{Y_t\}$ are copies of $\{Z_t\}$, and if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$. To reduce the difficulty of constructing the coupling, we use path coupling as a tool in the following analysis.

We begin by defining the notions of neighbors and paths in Ω. For a pair of states $\sigma, \sigma_{C_i} \in \Omega$, if $\sigma(C_i) = 0$, $\sigma_{C_i}(C_i) = 1$, and $\sigma(C_j) = \sigma_{C_i}(C_j)$ for all $j \ne i$, then σ and σ_{C_i} are said to be neighbors, denoted by $\sigma \sim \sigma_{C_i}$. We call $p = (p_0, p_1, \ldots, p_k)$ a simple path if all p_i are distinct and $p_0 \sim p_1 \sim \cdots \sim p_k$. For each $\sigma_i, \sigma_j \in \Omega$, we denote the set of simple paths connecting them as $p(\sigma_i, \sigma_j)$,

$$p(\sigma_i,\sigma_j) = \{p : \sigma_i = p_0, \sigma_j = p_k, \\ p \text{ is a simple path}\}.$$

We assume $X_t = \sigma$ and $Y_t = \sigma_{C_i}$, and denote the next states as $X_{t+1} = \sigma'$ and $Y_{t+1} = \sigma'_{C_i}$, respectively. Note that both Markov chains X_t and Y_t follow the same random walk in each step of the path coupling; if the move yields an invalid state, the chain remains in its current state. The following lemma on path coupling is adopted from [1].

Lemma 1. Let Φ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$ and for all $\sigma_i, \sigma_j \in \Omega$, there exists $p \in p(\sigma_i, \sigma_j)$

such that

$$\Phi(\sigma_i, \sigma_j) = \sum_k \Phi(p_k, p_{k+1}).$$

Suppose there exists $\beta < 1$ and a coupling of the Markov chain such that for all $\sigma, \sigma_{C_i} \in \Omega$:

$$E\left[\Phi(\sigma', \sigma'_{C_i})\right] \leq \beta \cdot \Phi(\sigma, \sigma_{C_i}).$$

Then the mixing time is

$$\tau(\varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{1-\beta}.$$

Proof of Theorem 1

Theorem 1. (Mixing time for two-by-two-intersection free graphs). For the four-vertex model with the constraint function f^*, whose underlying graph is two-by-two-intersection free under the circuit decomposition C, the Markov chain $\{Z_t\}$ mixes in time

$$O\left(\frac{n(1+b^\delta)}{b^\delta + 1 - b\delta} \log\left(\frac{n\delta}{\varepsilon}\right)\right),$$

for $\delta \geq 2$, $b \leq \frac{1}{\delta}$.

Now we analyze the mixing speed of $\{Z_t\}$ on two-by-two-intersection free graphs by path coupling.

Potential function

We will define the potential function as the distance of states associated with the path coupling at step t, such that if the potential function is zero, then the two chains will eventually reach the same state. According to path coupling, we consider a pair of states $\sigma \sim \sigma_{C_i}$. If the next move is on the "bad neighbors" of C_i, then the move will not work in one of chains and causes more differences between two chains. Therefore, we will define a potential function associated with the number of the "bad neighbors" of C_i.

First, we identify the good neighbors of C_i, i.e., the circuits that do not increase the difference between σ and σ_{C_i} even when moved. For a pair of adjacent states $\sigma \sim \sigma_{C_i}$ of $\{X_t\}$ and $\{Y_t\}$, the circuit C_j is blocked if C_j has at least one neighbor which takes the value 1 in both σ and σ_{C_i}. Denote the set of blocked neighbors of C_i by

$$B(\sigma, C_i) = \{C_j : C_j \in \Gamma(C_i), \text{ there exists } C_k \in \Gamma(C_j) \text{ s.t. } \sigma(C_k) = 1\}.$$

The set of unblocked circuits is $\Gamma(C_i) \backslash B(\sigma, C_i)$ for σ and σ_{C_i}. For the current state (σ, σ_{C_i}) of the coupling (X_t, Y_t), the move on an unblocked circuit only works in σ, and it makes more difference between $\{X_t\}$ and $\{Y_t\}$. The move on a blocked circuit will keep the difference between the states of $\{X_t\}$ and $\{Y_t\}$. Note that unlocked circuits are the "bad neighbors" we mentioned above. For some technical reasons, we define the potential function as the number of neighbors of C_i minus a positive constant $w < 1$ times the number of blocked neighbors of C_i, instead of the number of the unblocked neighbors of C_i. Specifically, for $w = \frac{b\delta}{b^\delta + 2}$ (note that $w < 1$ if $b \leq \frac{1}{\delta}$), the potential function Φ is as follows:

$$\Phi(\sigma, \sigma_{C_i}) = \delta_{C_i} - w|B(\sigma, C_i)|. \tag{6}$$

For two arbitrary states σ_i and σ_j, the potential function is

$$\Phi(\sigma_i, \sigma_j) = \min_{p \in p(\sigma_i, \sigma_j)} \sum_k \Phi(p_k, p_{k+1}),$$

where k is an integer and $i \leq k \leq j-1$. This potential function Φ clearly satisfies the following conditions for any $\sigma_i, \sigma_j \in \Omega$:

- $\Phi(\sigma_i, \sigma_j) \leq \Phi(\sigma_i, \sigma_k) + \Phi(\sigma_k, \sigma_j)$, for any $\sigma_k \in \Omega$;
- $\Phi(\sigma_i, \sigma_j) = \Phi(\sigma_j, \sigma_i)$;
- $\Phi(\sigma_i, \sigma_j) \geq 0$;
- $\Phi(\sigma_i, \sigma_j) = 0 \Leftrightarrow \sigma_i = \sigma_j$.

Analysis on the potential function

Recall that we set $X_t = \sigma$ and $Y_t = \sigma_{C_i}$, where σ and σ_{C_i} are neighbors, $\sigma_{C_i}(C_i) = 1$ and $\sigma(C_j) = \sigma_{C_i}(C_j)$ for $j \neq i$, and the next states are $X_{t+1} = \sigma'$ and $Y_{t+1} = \sigma'_{C_i}$. Let $\Phi = \Phi(\sigma, \sigma_{C_i})$ and $E[\Delta\Phi] = E[\Phi(\sigma', \sigma'_{C_i}) - \Phi(\sigma, \sigma_{C_i})]$. We now analyze $E[\Delta\Phi]$ to find the constant β in Lemma 1.

We classify the circuits in C into four types: C_i, C_j as a neighbor of C_i, C_k as a neighbor of the neighbors of C_i, and C_l as others. In the following analysis, we will only discuss the effect of moves on C_i, C_j and C_k, because the moves on C_l have no effect on Φ. For an arbitrary circuit C_i, let

$$E[\Delta^{+C_i}\Phi] = E[\Delta\Phi \mid \text{Markov chain attempts to move } C_i \text{ to 1}],$$

$$E[\Delta^{-C_i}\Phi] = E[\Delta\Phi \mid \text{Markov chain attempts to move } C_i \text{ to 0}],$$

$$E[\Delta^{C_i}\Phi] = E[\Delta\Phi \mid \text{Markov chain attempts to move } C_i],$$

then

$$E[\Delta^{C_i}\Phi] = \frac{b^{\delta_{C_i}}}{1 + b^{\delta_{C_i}}} E[\Delta^{+C_i}\Phi] + \frac{1}{1 + b^{\delta_{C_i}}} E[\Delta^{-C_i}\Phi].$$

By the above notations, we can rewrite $E[\Delta\Phi]$ as follows:

$$E[\Delta\Phi] = \frac{E[\Delta^{C_i}\Phi]}{n} + \frac{\sum_{C_j \in \Gamma(C_i)} E[\Delta^{C_j}\Phi]}{n} + \frac{\sum_{C_k \in \Gamma(\Gamma(C_i))} E[\Delta^{C_k}\Phi]}{n}, \tag{7}$$

where $\Gamma(\Gamma(C_i))$ is the set of the neighbors of the neighbors of C_i.

Now we consider the moves on C_i, C_j or C_k, separately.

- Move on C_i:
  In this case, a move on C_i will work in both $\{X_t\}$ and $\{Y_t\}$. After the move, $\{X_t\}$ and $\{Y_t\}$ will reach the same state which is σ or σ_{v_i}. Therefore, we have

  $$E[\Delta^{+C_i}\Phi] = E[\Delta^{-C_i}\Phi] = -\delta_{C_i} + w|B(\sigma, C_i)|,$$

  i.e.,

  $$E[\Delta^{C_i}\Phi] = -\delta_{C_i} + w|B(\sigma, C_i)|.$$

- Move on C_j:
  Since C_j is a neighbor of C_i, C_j must take the value 0 in σ_{C_i}. Note that σ and σ_{C_i} differ only at C_i, so C_j takes the value 0 in σ. Therefore, $E[\Delta^{-C_j}\Phi] = 0$.
  Consider the move which attempts to move C_j to 1. If C_j is unblocked, then this move will only work in $\{X_t\}$, and we have

  $$\begin{aligned}\Delta^{+C_j}\Phi &= \Phi(\sigma', \sigma'_{C_i}) - \Phi(\sigma, \sigma_{C_i}) \\ &= \Phi(\sigma_{C_j}, \sigma_{C_i}) - \Phi(\sigma, \sigma_{C_i}) \\ &\leq \Phi(\sigma_{C_j}, \sigma) + \Phi(\sigma, \sigma_{C_i}) - \Phi(\sigma, \sigma_{C_i}) \\ &= \Phi(\sigma_{C_j}, \sigma) \\ &= \delta_{C_j} - w|B(\sigma, C_j)|.\end{aligned}$$

  If C_j is blocked, then this move will work neither in $\{X_t\}$ nor in $\{Y_t\}$. And both chains will stay the current state. Therefore, $\Delta^{+C_j}\Phi = 0$. In conclusion, $E[\Delta^{C_j}\Phi]$ satisfies:
  - If $C_j \notin B(\sigma, C_i)$,
    $$E[\Delta^{C_j}\Phi] \leq \frac{b^{\delta_{C_j}}(\delta_{C_j} - w|B(\sigma, C_j)|)}{1 + b^{\delta_{C_j}}}. \tag{8}$$
  - Otherwise,
    $$E[\Delta^{C_j}\Phi] = 0. \tag{9}$$

- Move on C_k:
  Firstly, we assume that C_k takes the value 1 in σ, then C_k also takes the value 1 in σ_{C_i}.

The move which attempts to move C_k to 1 will work in both $\{X_t\}$ and $\{Y_t\}$. And both of them will stay the current state. Therefore,

$$\Delta^{+C_k}\Phi = 0. \tag{10}$$

Then we consider the move which attempts to move C_k to 0. This move will work in both $\{X_t\}$ and $\{Y_t\}$ and it might change some C_j from blocked to unblocked. The set of such C_j is defined as follows:

$$\alpha_{C_k} = \{C_j : C_j \in B(\sigma, C_i),\ \sigma(C'_k) = 0 \text{ for any } C'_k \in \Gamma(C_j) \setminus C_k\}.$$

We have

$$\begin{aligned}
&\Delta^{-C_k}\Phi \\
&= \Phi(\sigma', \sigma'_{C_i}) - \Phi(\sigma, \sigma_{C_i}) \\
&= \delta_{C_i} - w|B((\sigma_{C_i})^-_{C_k}, C_i)| - (\delta_{C_i} - w|B(\sigma, C_i)|) \\
&= w(|B(\sigma, C_i)| - |B((\sigma_{C_i})^-_{C_k}, C_i)|) \\
&= w|\alpha_{C_k}|,
\end{aligned}$$

where $(\sigma_{C_i})^-_{C_k}$ denotes the configuration in which all the circuits take the same value as σ_{C_i} except for $(\sigma_{C_i})^-_{C_k}(C_k) = 0$ and $\sigma_{C_i}(C_k) = 1$. Therefore,

$$\Delta^{-C_k}\Phi = \begin{cases} w|\alpha_{C_k}| & \text{if } \sigma(C_k) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Secondly, we assume that C_k takes the value 0 in σ. Note that σ and σ_{C_i} differ only at C_i, so C_k takes the value 0 in σ. The move which attempts to move C_k to 0 will work in both $\{X_t\}$ and $\{Y_t\}$. And both will stay the current state. Therefore,

$$\Delta^{-C_k}\Phi = 0. \tag{12}$$

Then consider the move which attempts to move C_k to 1. Recall that the graph is two-by-two-intersection free, so C_k and C_i are not adjacent. If there exists a neighbor of C_k which takes the value 1 in σ, then this move will not work in neither of two chains. Therefore, both chains stay the current states and $\Delta^{+C_k}\Phi = 0$; if all the neighbors of C_k take the value 0 in σ and denote the fact by $\sigma(\Gamma(C_k)) = 0$, then this move will work in both chains. Note that moving C_k to 1 which works in both chains effects the potential function by changing C_j from unblocked to blocked. The set of such C_j is defined as follows:

$$\beta_{C_k} = \{C_j : C_j \notin B(\sigma, C_i), C_j \in \Gamma(C_i) \cap \Gamma(C_k)\}.$$

We have

$$\begin{aligned}
&\Delta^{+C_k}\Phi \\
&= \Phi(\sigma', \sigma'_{C_i}) - \Phi(\sigma, \sigma_{C_i}) \\
&= \delta_{C_i} - w|B(\sigma_{C_k}, C_i)| - (\delta_{C_i} - w|B(\sigma, C_i)|) \\
&= w(|B(\sigma, C_i)| - |B(\sigma_{C_k}, C_i)|) \\
&= -w|\beta_{C_k}|.
\end{aligned}$$

Therefore, we obtain:

- If $\sigma(\Gamma(C_k)) = 0$ and $\sigma(C_k) = 0$,

$$\Delta^{+C_k}\Phi = -w|\beta_{C_k}|. \tag{13}$$

- Otherwise,

$$\Delta^{+C_k}\Phi = 0. \tag{14}$$

Combining Eqs. (11) and (12), the value of $\Delta^{-C_k}\Phi$ is as follows.

- If $\sigma(C_k) = 0$,

$$\Delta^{-C_k}\Phi = 0.$$

- If $\sigma(C_k) = 1$,

$$\Delta^{-C_k}\Phi = w|\alpha_{C_k}|.$$

Combining Eqs. (10), (13) and (14), the value of $\Delta^{+C_k}\Phi$ is as follows:

– If $\sigma(C_k) = 1$,

$$\Delta^{+C_k}\Phi = 0.$$

– If $\sigma(\Gamma(C_k)) = 0$ and $\sigma(C_k) = 0$,

$$\Delta^{+C_k}\Phi = -w|\beta_{C_k}|.$$

– Otherwise,

$$\Delta^{+C_k}\Phi = 0.$$

Combining the values of $\Delta^{-C_k}\Phi$ and $\Delta^{+C_k}\Phi$, for an arbitrary neighbor of neighbors of C_i, i.e., C_k, we obtain:

– If $\sigma(C_k) = 1$,

$$E[\Delta^{C_k}\Phi] = \frac{|\alpha_{C_k}|}{1 + b^{\delta_{C_k}}}w.$$

– If $\sigma(\Gamma(C_k)) = 0$ and $\sigma(C_k) = 0$,

$$E[\Delta^{C_k}\Phi] = -\frac{b^{\delta_{C_k}}|\beta_{C_k}|}{1 + b^{\delta_{C_k}}}w.$$

– Otherwise,

$$E[\Delta^{C_k}\Phi] = 0.$$

After analyzing the moves on C_i, C_j and C_k separately, we are ready to bound $E[\Delta\Phi]$. It seemed that we can compute $E[\Delta\Phi]$ by Eq. (7). But it can not work in practice, since we do not know the number of C_k. Therefore, we can not give a sharp bound for $E[\Delta\Phi]$ by Eq. (7). One way to handle this problem is "packing": Each C_j has a set of C_k which change C_j from unblocked to blocked (or from blocked to unblocked), and we pack such C_k into a set. This idea will give a "holistic" property which tends to give a sharper bound. In the following analysis, we will formalize this idea. For a blocked circuit $C_j \in \Gamma(C_i)$, we define the set

$$\Gamma'(C_j) = \{C_k : C_k \in \Gamma(C_j)\backslash\{C_i\}, C_j \in \alpha_{C_k}\},$$

and

$$E[\Delta^{*C_j}\Phi] = E[\Delta^{C_j}\Phi] + \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\alpha_{C_k}|}E[\Delta^{C_k}\Phi].$$

For an unblocked circuit $C_j \in \Gamma(C_i)$, we define

$$\Gamma'(C_j) = \{C_k : C_k \in \Gamma(C_j)\backslash\{C_i\}, C_j \in \beta_{C_k}\}$$

and

$$E[\Delta^{*C_j}\Phi] = E[\Delta^{C_j}\Phi] + \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\beta_{C_k}|}E[\Delta^{C_k}\Phi].$$

We can rewrite Eq. (7) as

$$E[\Delta\Phi] = \frac{1}{n}\left[E[\Delta^{C_i}\Phi] + \sum_{C_j\in\Gamma(C_i)} E[\Delta^{*C_j}\Phi]\right].$$

Firstly, we bound $E[\Delta^{*C_j}\Phi]$.

- If C_j is blocked.
  By Eqs. (9), it flows that $E[\Delta^{C_j}\Phi] = 0$. According to the definition of α_{C_k}, if $C_j \in \alpha_{C_k}$, then there exists exactly one C_k such that $C_k \in \Gamma(C_j)$ and $\sigma(C_k) = 1$. We have $|\Gamma'(C_j)| \le 1$. Therefore, we have

$$\begin{aligned}
&E[\Delta^{*C_j}\Phi]\\
&= E[\Delta^{C_j}\Phi] + \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\alpha_{C_k}|}E[\Delta^{C_k}\Phi]\\
&= 0 + \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\alpha_{C_k}|}\frac{|\alpha_{C_k}|}{1 + b^{\delta_{C_k}}}w\\
&= \sum_{C_k\in\Gamma'(C_j)} \frac{w}{1 + b^{\delta_{C_k}}}\\
&\le \sum_{C_k\in\Gamma'(C_j)} \frac{w}{1 + b^{\delta}}\\
&= |\Gamma'(C_j)|\frac{w}{1 + b^{\delta}}\\
&\le \frac{w}{1 + b^{\delta}}.
\end{aligned}$$

- If C_j is unblocked.

$$
\begin{aligned}
nE[\Delta\Phi] &= E[\Delta^{C_i}\Phi] + \sum_{C_j\in\Gamma(C_i)} E[\Delta^{*C_j}\Phi] \\
&= \left[-\delta_{C_i} + w|B(\sigma,C_i)| + \sum_{C_j\in\Gamma(C_i),C_j\in B(\sigma,C_i)} E[\Delta^{*C_j}\Phi] + \sum_{C_j\in\Gamma(C_i),C_j\notin B(\sigma,C_i)} E[\Delta^{*C_j}\Phi]\right] \\
&\le \left[-\delta_{C_i} + \sum_{C_j\in\Gamma(C_i),C_j\in B(\sigma,C_i)} \frac{2+b^{\delta}}{1+b^{\delta}}w + \sum_{C_j\in\Gamma(C_i),C_j\notin B(\sigma,C_i)} \frac{b\delta}{1+b^{\delta}}\right] \\
&= \frac{1}{(1+b^{\delta})}[-(1+b^{\delta})\delta_{C_i} + |B(\sigma,C_i)|(2+b^{\delta})w + (\delta_{C_i} - |B(\sigma,C_i)|)b\delta] \\
&= \frac{\delta_{C_i}}{(1+b^{\delta})}(b\delta - 1 - b^{\delta}).
\end{aligned}
\tag{16}
$$

By Eq. (8), we have $E[\Delta^{C_j}\Phi] \le \frac{b^{\delta_{C_j}}}{1+b^{\delta_{C_j}}}(\delta_{C_j} - w|B(\sigma,C_j)|)$. Moreover, we have

$$
\begin{aligned}
&E[\Delta^{*C_j}\Phi] \\
&= E[\Delta^{C_j}\Phi] + \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\beta_{C_k}|}E[\Delta^{C_k}\Phi] \\
&\le \frac{b^{\delta_{C_j}}}{1+b^{\delta_{C_j}}}(\delta_{C_j} - w|B(\sigma,C_j)|) \\
&- \sum_{C_k\in\Gamma'(C_j)} \frac{1}{|\beta_{C_k}|}\frac{b^{\delta_{C_k}}|\beta_{C_k}|}{1+b^{\delta_{C_k}}}w \\
&\le \frac{b^{\delta_{C_j}}}{1+b^{\delta_{C_j}}}(\delta_{C_j} - w|B(\sigma,C_j)|) \\
&\le \frac{b\delta}{1+b^{\delta}}.
\end{aligned}
\tag{15}
$$

Note that $w = \frac{b\delta}{b^{\delta}+2}$. Thus, $E[\Delta\Phi]$ is bounded in Eq. (16). Notice that $E[\Delta\Phi] \le 0$, when $b \le \frac{1}{\delta}$.

Now we are ready to bound $E[\Delta\Phi]$, which will bound the mixing time. We define $\beta_{\sigma,\sigma_{C_i}}$ such that

$$E[\Phi(\sigma',\sigma'_{C_i})] = \beta_{\sigma,\sigma_{C_i}}\Phi(\sigma,\sigma_{C_i}).$$

Therefore, we have

$$
\begin{aligned}
E[\Delta\Phi] &= E[\Phi(\sigma',\sigma'_{C_i})] - \Phi(\sigma,\sigma_{C_i}) \\
&= (\beta_{\sigma},\sigma_{C_i} - 1)\Phi(\sigma,\sigma_{C_i}),
\end{aligned}
$$

and

$$\beta_{\sigma,\sigma_{C_i}} = 1 + \frac{E[\Delta\Phi]}{\Phi(\sigma,\sigma_{C_i})}.$$

By Lemma 1, we set $\beta = \max_{\sigma,\sigma_{C_i}}\beta_{\sigma,\sigma_{C_i}}$. From Eq. (6), i.e., the definition of Φ, it is easy to know that $\Phi(\sigma,\sigma_{C_i}) \le \delta_{C_i}$. By the bound of $E[\Delta\Phi]$ in Eq. (16), we have

$$\beta \le 1 + \frac{1}{n(1+b^{\delta})}(b\delta - 1 - b^{\delta}).$$

The Lemma (1) produces a bound of the mixing time

$$\tau(\varepsilon) \le \frac{n(1+b^{\delta})}{b^{\delta}+1-b\delta}\log\left(\frac{n\delta}{\varepsilon}\right).$$

Appendix E

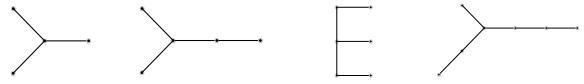
In this appendix, we first introduce the subdivided claw. Then we introduce the modified log-Sobolev (MLSI) and its related concepts. Finally, we give the mixing time of the four-vertex model on claw-free graphs and extend it to $S_{1,1,t}$-free graphs.

Subdivided claw:

A claw is the complete bipartite graph $K_{1,3}$, i.e., there is one central vertex which is connected to three other non-adjacent vertices.

A subdivided claw is constructed by inserting an arbitrary number of intermediate vertices along each of three edges of a claw. For integers $1 \le i \le j \le k$, we denote by $S_{i,j,k}$ the subdivided claw

whose three leaves are at distances i, j and k from the central vertex of degree three. Smaller subdivided claws have specific names: $S_{1,1,1}$ is called the claw, $S_{1,1,2}$ is the fork, $S_{1,2,2}$ is known as the E, and $S_{1,2,3}$ is the skew star (see Fig. 1).



(a) Claw (b) Fork (c) E (d) Skew star

Fig. 1 The claw, the fork, the E and the skew star.

Under the circuit decomposition C of the underlying graph G of the four-vertex model, if there are no circuits in G such that the vertices corresponding to them in the black face graph $H(G)$ form a subdivided claw, then we say that the underlying graph is subdivided claw-free.

Modified Log-Sobolev Constant:

Glauber dynamics is a canonical Markov chain for sampling from the Gibbs distribution. Thus, the Markov chain $\{Z_t\}$ constitutes a Glauber dynamics.

For the Glauber dynamics, MLS constant plays a crucial role to analyzing Markov chains, since the mixing time of Markov chains can be obtained by MLS [2].

For the function $f : \sigma \to \mathbb{R}^+$, the Dirichlet form is defined by:

$$\begin{aligned}&\varepsilon_p(f, \log(f))\\ &= \langle f, (I - P)\log(f)\rangle_\mu\\ &= \sum_{\sigma_i \in \Omega} f(\sigma_i)\mu(\sigma_i)\\ &\cdot \sum_{\sigma_j \in \Omega} [(I - P)_{ij} \log(f(\sigma_j)) - \log(f(\sigma_i))],\end{aligned}$$

where I is the $|\Omega| \times |\Omega|$ identity matrix, P is the transition probability matrix of Z_t. The entropy is defined as:

$$\mathrm{Ent}_\mu[f] = E_\mu[f \log(f)] - E_\mu[f] \log(E_\mu[f]),$$

where $E_\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$. The MLS constant of the chain $\{Z_t\}$ is defined as

$$\rho^{GD}(\mu) = \inf\left\{\frac{\varepsilon_p(f, \log f)}{\mathrm{Ent}_\mu[f]} | f : \sigma \to R^+, \mathrm{Ent}_\mu[f] \neq 0\right\}.$$

To obtain the MLS constant, researchers establish the spectral independence and the marginal boundedness for the stationary distribution of the Markov chain [2]. We now present the definitions of spectral independence and the marginal boundedness.

Marginal boundedness:

Definition 2. (Marginal boundedness) We say μ is $\mathcal{B}$-marginally bounded if for all pinnings τ on $\Lambda \subseteq C$ and all $C_i \in C \setminus \Lambda$, we have $\mathcal{B} \leq \mu^\tau(\sigma(C_i) = 1) \leq 1 - \mathcal{B}$.

In this appendix, we use the monotonic coupling to establish the marginal boundedness. Thus, we proceed to introduce the monotonic coupling in detail.

We define the monotone coupling of two Markov chains $\{Z_t\}$ and $\{\hat{Z}_t\}$, both defined on the state space $\Omega \subseteq \{0, 1\}^C$.

Each state in Ω is a binary string representing an assignment on the circuits. For an arbitrary configuration σ, we denote the set of all circuits C except C_i as $C \setminus C_i$, and we denote the assignment of $(C \setminus C_i)$ in σ as $\sigma(C \setminus C_i)$. We say that $\sigma \leq \hat{\sigma}$ in the product order on Ω if $\sigma(C_i) \leq \hat{\sigma}(C_i)$ for every circuit C_i. We assume that μ and $\hat{\mu}$ are the stationary distributions of the Markov chains $\{Z_t\}$ and $\{\hat{Z}_t\}$, respectively.

If $\mu(\sigma(C_i) = 1 | \sigma(C \setminus C_i)) \leq \hat{\mu}(\hat{\sigma}(C_i) = 1 | \hat{\sigma}(C \setminus C_i))$ for all σ and $\hat{\sigma}$ such that $\sigma \leq \hat{\sigma}$ in the product order on Ω, we can construct a monotonic coupling using the following walk rules:

- At each step of the process, we set the same random number $h \in [0, 1]$ and the same circuit C_i for both chains.
- In $\{Z_t\}$, we set $\sigma'(C_i) = 1$ if $h \leq \mu(\sigma(C_i) = 1|\sigma(C \setminus C_i))$; otherwise, we set it to 0. The same update rule is applied to $\{\hat{Z}_t\}$.

The important observation is that

$$\hat{Z}_t \geq Z_t \Rightarrow \hat{Z}_{t+1} \geq Z_{t+1},$$

where $(\Rightarrow)$ represents that the left part is a sufficient condition for the right part.

Spectral independence:

Definition 3. (Pinning) A pinning is a partial configuration $\tau \in \{0, 1\}^\Lambda$ on $\Lambda \subseteq C$ such that $\mu_\Lambda(\tau) > 0$, where μ_Λ is the marginal distribution on Λ.

To distinguish from μ_Λ, we denote the conditional distribution on $C \setminus \Lambda$ as μ^τ. And let $\mu^{(C_i,0)}$ and $\mu^{(C_i,1)}$ denote the conditional distributions for the circuit C_i fixed at 0 and 1, respectively.

Definition 4. (Influence matrix) Let μ be a distribution over $\{0, 1\}^C$. Suppose that for an arbitrary pinning τ on $\Lambda \subseteq C$, the pairwise influence matrix $J^\tau_\mu \in \mathbb{R}^{(C\setminus\Lambda)\times(C\setminus\Lambda)}$ is defined as following: for all $C_i, C_j \in C \setminus \Lambda$, let

$$\begin{aligned} J^\tau_\mu(C_i, C_j) &= \mu^\tau(\sigma(C_i) = 1 \mid \sigma(C_j) = 1) \\ &- \mu^\tau(\sigma(C_i) = 1 \mid \sigma(C_j) = 0), \end{aligned}$$

when $i \neq j$ and $\min\{\mu^\tau(\sigma(C_i) = 1), \mu^\tau(\sigma(C_i) = 0)\} > 0$, and let $J^\tau_\mu(C_i, C_j) = 0$ otherwise. All eigenvalues of the influence matrix J^τ_μ are real.

Definition 5. (Spectral independence) We say μ is η-spectrally independent if we have $\lambda_{\max(J^\tau_\mu)} \leq \eta$ for all pinnings τ, where $\lambda_{\max}(J^\tau_\mu)$ denotes the largest eigenvalue of the influence matrix J^τ_μ.

Determining an upper bound on the eigenvalues of the influence matrix is challenging due to the constraints imposed by its dimensionality. To overcome this difficulty, [3] proposed using the Wasserstein distance to bound the parameter η.

Let μ and $\hat{\mu}$ be the probability distributions over $\{0, 1\}^C$. The Wasserstein distance between μ and $\hat{\mu}$ is defined as

$$W_1(\mu, \hat{\mu}) = \inf_{\sigma\sim\mu,\, \hat{\sigma}\sim\hat{\mu}} d_{\mathrm{H}}(\sigma, \hat{\sigma}) P(\sigma, \hat{\sigma}), \tag{17}$$

where d_{H} denotes the Hamming distance, and the infimum is taken over all couplings of random variables σ and $\hat{\sigma}$ with the respective marginal distributions.

Theorem 2. [3] Let μ be the distribution of a Holant problem on graph $G = (V, E)$. Suppose that for a constant $\eta > 0$, the following is true: For two arbitrary pinnings $\tau, \hat{\tau}$ on the same subset $\Lambda \subseteq E$ which differ on exactly one edge, we have

$$W_1(\mu^\tau, \mu^{\hat{\tau}}) \leq \eta.$$

Then μ is η-spectrally independent.

Totally-connected distribution:

By establishing the spectral independence and the marginal boundedness, Chen and Liu [2] give the mixing time for the Markov chain on totally-connected distribution. Next, we present the definition of totally-connected distribution and other related concepts.

For a subset $S \subseteq C$, the Hamming graph $\mathcal{H}_S$ is defined to be the graph with the circuit set $\{0, 1\}^S$. Note that each circuit in $\mathcal{H}_S$ has a unique valid configuration in $\{0, 1\}^S$ corresponding to it one-to-one. In $\mathcal{H}_S$, there is an edge connecting the two vertices iff their corresponding valid configurations in $\{0, 1\}^S$ differ at exactly one circuit. A state space

$\Omega_S \subseteq \{0,1\}^S$ is said to be connected if the induced subgraph $\mathcal{H}_S[\Omega_S]$ is connected, where $\mathcal{H}_S[\Omega_S]$ is formed by the set of vertex corresponding to Ω_S.

Definition 6. (Totally-connected distribution) A distribution μ on $\{0,1\}^C$ is said to be totally-connected if for an arbitrary non-empty subset $S \subseteq C$ and an arbitrary pinning $\tau \in \{0,1\}^\Lambda$ for $\Lambda \subseteq C \setminus S$, the set $\Omega_S^\tau = \{\sigma \in \{0,1\}^S \mid \mu^\tau(\sigma) > 0\}$ is connected.

The following theorem provides the mixing time of the Glauber dynamics under the conditions of η-spectral independence and $\mathcal{B}$-marginal boundedness.

Theorem 3. [2] Let $\Delta \geq 3$ be an integer and $\mathcal{B}, \eta > 0$ be reals. Suppose that $G = (V, E)$ is an n-vertex graph of maximum degree at most Δ and μ is the totally-connected Gibbs distribution of some spin system on G. If μ is both $\mathcal{B}$-marginally bounded and η-spectrally independent, then the Glauber dynamics for sampling from μ satisfies the modified log-Sobolev inequality with constant $\frac{C_1}{n}$, where

$$C_1 = \left(\frac{\Delta}{\mathcal{B}}\right)^{O\left(\frac{\eta}{\mathcal{B}^2}+1\right)}.$$

Furthermore, the mixing time of the Glauber dynamics satisfies

$$\tau(\varepsilon) = \left(\frac{\Delta}{\mathcal{B}}\right)^{O\left(\frac{\eta}{\mathcal{B}^2}+1\right)} \times O\left(n \log\left(\frac{n}{\varepsilon}\right)\right).$$

According to Theorem 2 and Definition 2, proving μ is η-spectrally independent and $\mathcal{B}$-marginally bounded requires considering all pinnings. To simplify the subsequent proof, we "removing" some circuits from the function grid of the four-vertex model with f^*, so that during the proof, we only need to consider pinning being a single circuit.

The operation "removing the circuit set C_r from the function grid $\Gamma = (G, f^*)$" is divided into two sequential steps, the details are as follows:

Step 1: For all circuits in C_r, we apply the same processing. Without loss of generality, we take an arbitrary circuit $C_i \in C_r$ as an example for illustration. For each edge in C_i, we remove it. For each vertex on C_i, we remove the vertex while preserving two edges that pass through it but do not belong to C_i. Then we merge these two edges into a single edge.

Step 2: For all circuits that are adjacent to those in C_r and not in C_r. Without loss of generality, we take an arbitrary circuit C_j adjacent to C_i as an example for illustration. Note that the circuit C_j loses some vertices after "Step 1" which causes it to become a new circuit. Thus, we denote this new circuit as C_j^λ, where λ is the maximum weight among the circuits in C_λ — herein, the weight of a circuit is the contribution to the evaluations of valid configurations when this circuit takes 1. Note that we set C_j^λ as an isolated circuit with only one vertex after "Step 1" if all neighbors of C_j are in C_r. The constraint function of each vertex on C_j^λ is as follows:

$$M(f_{C_j^\lambda}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & b^{\delta_{C_j}/\delta_{C_j^\lambda}} & 0 \\ 0 & b^{\delta_{C_j}/\delta_{C_j^\lambda}} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Interestingly, the weights of C_j and C_j^λ are both $b^{\delta_{C_j}}$. This implies that removing some circuits from the function gird Γ leaves the weight of the others unchanged. Thus, by the function gird Γ, we have

$$\lambda \leq \begin{cases} b & 0 < b < 1; \\ b^\delta & b \geq 1. \end{cases} \tag{18}$$

To make the statement concise, we denote the set of all circuits after removing some circuits from Γ as C_λ. And we denote the planar graph corresponding to the circuit decomposition C_λ as G_λ. Note that all vertices of an arbitrary circuit in C_λ are labeled with the same constraint function. Thus, for circuit C_i^λ, we denote the same constraint function labeled on all its vertices as $f_{C_i^\lambda}$. And we denote the set of constraint functions obtained by removing some circuits from Γ as $\mathcal{F} = \{f_{C_i^\lambda} \mid C_i^\lambda \in C_\lambda\}$. We say that σ_λ is a valid configuration for the four-vertex model on the planar graph G_λ with the constraint function set $\mathcal{F}$. We denote the set of all valid configurations as Ω_λ and denote the Gibbs distribution on Ω_λ as μ_λ .

Note that G is a planar graph with the circuit decomposition C. We suppose that μ is the Gibbs distribution of the four-vertex model on the planar graph G with the constraint function f^*, it exhibits a convenient property: For an arbitrary pinning $\tau \in \{0,1\}^\Lambda$ and an arbitrary circuit $C_i \in \Lambda$, we have that μ^τ is equivalent to $\mu_\lambda^{C_i^\lambda}$. The distribution $\mu_\lambda^{C_i^\lambda}$ is the Gibbs distribution of the four-vertex model on G_λ with $\mathcal{F}$, where G_λ and $\mathcal{F}$ are obtained by removing the circuit set Λ' from Γ — herein, Λ' is the set of circuits, which is partitioned into two parts: one part consists of circuits in τ that take the value 0 (excluding C_i), one part consists of circuits in τ that take the value 1 (excluding C_i) and their neighbors.

Thus, for two arbitrary pinnings $\tau, \hat{\tau}$ on Λ which differ on circuit C_i, there always exists a circuit set Λ' such that "removing Λ' from Γ" constructs two distributions $\mu_\lambda^{(C_i^\lambda,0)}$ and $\mu_\lambda^{(C_i^\lambda,1)}$ corresponding to τ and $\hat{\tau}$, satisfying the following conditions:

$$\mu^\tau(\sigma(C \setminus \Lambda)) = \mu_\lambda^{(C_i^\lambda,0)}(\sigma_\lambda(C_\lambda \setminus C_i^\lambda))$$

and

$$\mu^{\hat{\tau}}(\sigma(C \setminus \Lambda)) = \mu_\lambda^{(C_i^\lambda,1)}(\sigma_\lambda(C_\lambda \setminus C_i^\lambda)).$$

This implies that

$$W_1\left(\mu^\tau, \mu^{\hat{\tau}}\right) = W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right). \tag{19}$$

Similarly, for an arbitrary pinnings τ on Λ, there always exists a circuit set Λ' such that "removing Λ' from Γ" constructs a distributions μ_λ for an arbitrary circuit $C_i \in C \setminus \Lambda'$ satisfying the following condition:

$$\mu^\tau(C_i) = \mu_\lambda(C_i^\lambda), \tag{20}$$

where Λ' is the set of circuits, which is partitioned into two parts: one part consists of circuits in τ that take the value 0, one part consists of circuits in τ that take the value 1 and their neighbors.

By Eqs. (19) and (20), there is no need to consider all pinnings when proving that μ is η-spectrally independent and $\mathcal{B}$-marginally bounded. Next, by analyzing the distribution μ_λ, we obtain the following corollary.

Corollary 1. Let $n > 0$ be integers, and $\mathcal{B}$, $\eta > 0$ be reals. And let $\mathcal{H}$ be a hereditary class. Suppose that G is a planar graph with n circuits under the circuit decomposition C and its black face graph $H(G) \in \mathcal{H}$. And suppose that μ is the Gibbs distribution of the four-vertex model with the function grid $\Gamma = (G, f^*)$ and μ_λ is the Gibbs distribution of the four-vertex model on the graph G_λ with the constraint function set $\mathcal{F} = \{f_{C_i^\lambda} \mid C_i^\lambda \in C_\lambda\}$, where G_λ and $\mathcal{F}$ are obtained by removing some circuits from Γ and C_λ is the circuit decomposition for G_λ. If there exists η and $\mathcal{B}$ such that for all G_λ satisfying $H(G_\lambda) \in \mathcal{H}$ and all $C_i^\lambda \in C_\lambda$, the following two conditions hold:

$$W_1(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}) \le \eta$$

and

$$\mathcal{B} \le \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \le 1 - \mathcal{B},$$

where $\mu_\lambda^{(C_i^\lambda,0)}$ and $\mu_\lambda^{(C_i^\lambda,1)}$ are the conditional distributions on the graph G_λ when the circuit $C_i^\lambda = 0$ and $C_i^\lambda = 1$ are fixed, respectively. Then the Glauber dynamics for sampling from μ satisfies the modified log-Sobolev inequality with constant $\frac{C_1}{n}$ where

$$C_1 = \left(\frac{\delta}{\mathcal{B}}\right)^{O\left(\frac{\eta}{\mathcal{B}^2}+1\right)}.$$

Furthermore, the mixing time of the Glauber dynamics satisfies

$$\tau(\varepsilon) = \left(\frac{\delta}{\mathcal{B}}\right)^{O\left(\frac{\eta}{\mathcal{B}^2}+1\right)} \times O\left(n \log\left(\frac{n}{\varepsilon}\right)\right).$$

Proof. By Theorem 3, we know that the mixing time of μ can be established iff μ is totally-connected Gibbs distribution of the spin system and it exhibits the marginally boundedness and the spectral independence.

We begin with proving that μ is the Gibbs distribution of some spin system on the black face graph of G. By the circuit decomposition C, the planar four-vertex model on the graph G can be regarded as the hard-core model on the black face graph of G. Note that the hard-core model is a specific instance of a spin system. Thus, μ is the Gibbs distribution of the spin system.

Then, we prove that μ is the totally-connected. We assume that τ is an arbitrary pinning on $\{0,1\}^{C\setminus S}$, where S is an arbitrary non-empty subset. And we assume that $\sigma_1, \sigma_2 \in \{0,1\}^S$ are two arbitrary valid state, i.e. we assume that $\mu^\tau(\sigma_1) > 0$ and $\mu^\tau(\sigma_2) > 0$. By taking the value 1 in σ_1 to 0 step by step, there is a path from σ_1 to the valid state σ_0 in which all circuits in subset S take the value 0 and vice versa. Similarly, there is a path from σ_2 to the valid state σ_0 and vice versa. Note that $\mu^\tau(\sigma_0) > 0$. Thus, for two arbitrary states, there is at least one path connected them by crossing σ_0, i.e., $\Omega_S^\tau = \{\sigma \in \{0,1\}^S \mid \mu^\tau(\sigma) > 0\}$ is connected for an arbitrary non-empty subset S and an arbitrary pinning τ. By Definition 6, μ is totally-connected.

Next, we prove that μ is $\mathcal{B}$-marginally bounded and η-spectrally independent.

We first prove that μ is $\mathcal{B}$-marginally bounded. Given an arbitrary pinning τ on $\Lambda \subseteq C$, we construct the graph $\dot{G}_\lambda$, specifically as follows:

- The circuits in τ assigned the value 0 are removed from Γ.
- The circuits in τ assigned the value 1, along with their neighbors, are removed from Γ.

Obviously, $\dot{G}_\lambda$ can be decomposed into some circuits. We denote the set of these circuits as $\dot{C}_\lambda = \{\dot{C}_1^\lambda, \dot{C}_2^\lambda, \ldots, \dot{C}_{|\dot{C}_\lambda|}^\lambda\}$. Let $\dot{\mu}_\lambda$ denote the Gibbs distribution for the four-vertex model on the graph $\dot{G}_\lambda$ with the constraint function set $\dot{\mathcal{F}} = \{f_{\dot{C}_i^\lambda} \mid \dot{C}_i^\lambda \in \dot{G}_\lambda\}$ obtained by removing $C \setminus \dot{C}_\lambda$ from Γ. For an arbitrary state $\sigma \in \{0,1\}^C$, there exists a state $\dot{\sigma}_\lambda \in \{0,1\}^{\dot{C}_\lambda}$ such that

$$\mu^\tau(\sigma) = \dot{\mu}_\lambda(\dot{\sigma}_\lambda).$$

This implies that

$$\mu^\tau(\sigma(C_i) = 1) = \dot{\mu}_\lambda(\dot{\sigma}_\lambda(\dot{C}_i^\lambda) = 1), \tag{21}$$

for an an arbitrary state σ and an arbitrary circuit $C_i \in C \setminus \Lambda$.

Since $\mathcal{H}$ is a hereditary class and $H(G) \in \mathcal{G}$, it follows that $H(\dot{G}_\lambda) \in \mathcal{H}$. Notably, for an arbitrary graph G_λ satisfying $H(G_\lambda) \in \mathcal{H}$ and an arbitrary circuit $C_i^\lambda \in C_\lambda$, there exists a constant $\mathcal{B}$ such that $\mathcal{B} \le \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \le 1 - \mathcal{B}$. Therefore, we have

$$\mathcal{B} \le \dot{\mu}_\lambda(\dot{\sigma}_\lambda(\dot{C}_i^\lambda) = 1) \le 1 - \mathcal{B}. \tag{22}$$

By Eqs. (21) and (22), it follows that

$$\mathcal{B} \le \mu^\tau(\sigma(C_i) = 1) \le 1 - \mathcal{B},$$

for an arbitrary pinning τ on $\Lambda \subseteq C$ and an arbitrary circuit $C_i \in C \setminus \Lambda$. This establishes that μ is $\mathcal{B}$-marginally bounded.

Next, we show that μ is η-spectrally independent. By Theorem 2, to bound spectral independence, we need to pin some edges. However, in fact, pinning an edge for the Holant problem is equivalent to pinning a circuit for the four-vertex model. Therefore, we assume that τ and $\hat{\tau}$ are two arbitrary pinnings on $\Lambda \subseteq C$, which differ only at circuit C_i.

Given an arbitrary pinning τ on $\Lambda \subseteq C$, we construct the graph $\ddot{G}_\lambda$, specifically as follows:

- The circuits in $\tau \setminus C_i$ assigned the value 0 are removed from Γ.
- The circuits in $\tau \setminus C_i$ assigned the value 1, along with their neighbors, are removed from Γ.

Obviously, $\ddot{G}_\lambda$ can be decomposed into some circuits. We denote the set of these circuits as $\ddot{\mathcal{C}}_\lambda = \{\ddot{C}_1, \ddot{C}_2, \ldots, \ddot{C}_{|\ddot{\mathcal{C}}_\lambda|}\}$. Let $\ddot{\mu}_\lambda$ denote the Gibbs distribution for the four-vertex model with the constraint function set $\ddot{\mathcal{F}} = \{f_{\ddot{C}_i} \mid \ddot{C}_i \in \ddot{\mathcal{C}}_\lambda\}$ obtained by removing $C \setminus \ddot{\mathcal{C}}_\lambda$. For an arbitrary pair of pinnings τ and $\hat{\tau}$, we have

$$W_1(\mu^\tau, \mu^{\hat{\tau}}) = W_1\left(\ddot{\mu}_\lambda^{(\ddot{C}_i,0)}, \ddot{\mu}_\lambda^{(\ddot{C}_i,1)}\right). \tag{23}$$

Since $\mathcal{H}$ is a hereditary class and $H(G) \in \mathcal{H}$, it follows that $H(\ddot{G}_\lambda) \in \mathcal{H}$. Notably, for an arbitrary graph G_λ satisfying $H(G_\lambda) \in \mathcal{H}$ and an arbitrary circuit $C_i^\lambda \in \mathcal{C}_\lambda$, there exists a constant η such that $W_1(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}) \leq \eta$. Therefore, for two arbitrary pinnings τ and $\hat{\tau}$, we have

$$W_1(\ddot{\mu}_\lambda^{(\ddot{C}_i,0)}, \ddot{\mu}_\lambda^{(\ddot{C}_i,1)}) \leq \eta. \tag{24}$$

By Eqs. (23) and (24), it follows that

$$W_1(\mu^\tau, \mu^{\hat{\tau}}) \leq \eta,$$

for two arbitrary pinnings τ and $\hat{\tau}$. This implies that μ is η-spectrally independent. This completes the proof. □

Claw-free graphs and upwards:

In this appendix, we consider the case where the underlying graph G is claw-free and $S_{1,1,t}$-free under the circuit decomposition C.

Our proof of the marginal boundedness and spectral independence of μ is inspired by [4]. However, unlike their approach, our result removes the requirement that the weight of each circuit be the same in the underlying graph.

When proving that distribution μ is $\mathcal{B}$-marginally bounded and η-spectrally independent, the distribution we consider instead is μ_λ. By Eqs. (19) and (20), we transform $W_1(\mu^\tau, \mu^{\hat{\tau}})$ into $W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right)$ and transform $\mu^\tau(C_i)$ into $\mu_\lambda(C_i^\lambda)$, respectively. The advantage of doing this is that we may focus solely on a single circuit, rather than all the pinnings.

We first define the Markov chain $\{Z_t^\lambda\}$ on the state space $\Omega_\lambda \subseteq \{0,1\}^{\mathcal{C}_\lambda}$. To make the statement concise, we denote the weight of C_i^λ as $W(C_i^\lambda)$ and set $\lambda = \max_{C_i^\lambda \in \mathcal{C}_\lambda} W(C_i^\lambda)$. For an arbitrary configuration $\sigma_\lambda \in \Omega_\lambda$, the transition probability matrix P_λ of $\{Z_t^\lambda\}$ is as follows:

- Choose an arbitrary circuit $C_i^\lambda \in \mathcal{C}_\lambda$ at random.
- If $\sigma_\lambda(\Gamma(C_i^\lambda)) = \mathbf{0}$, where $\mathbf{0}$ denotes the all-zero string of length $\delta_{C_i^\lambda}$, then:

$$\sigma'_\lambda = \begin{cases} \text{only take } C_i^\lambda \text{ to 1 in } \sigma_\lambda \\ \text{with the probability } \frac{W(C_i^\lambda)}{1+W(C_i^\lambda)}; \\ \\ \text{only take } C_i^\lambda \text{ to 0 in } \sigma_\lambda \\ \text{with the probability } \frac{1}{1+W(C_i^\lambda)}. \end{cases}$$

- else : set $\sigma'_\lambda \leftarrow \sigma_\lambda$.

- The new state is σ'_λ.

Similar to the Markov chain Z_t, the Markov chain Z_t^λ has the aperiodicity and irreducibility. Therefore, the Markov chain Z_t^λ admits a unique stationary distribution, which is explicitly given by:

$$\mu_\lambda(\sigma_\lambda) = \frac{\prod\limits_{C_i^\lambda:\sigma_\lambda(C_i^\lambda)=1} W(C_i^\lambda)}{\sum\limits_{\sigma_\lambda\in\Omega_\lambda}\prod\limits_{C_i^\lambda:\sigma_\lambda(C_i^\lambda)=1} W(C_i^\lambda)}.$$

For a short while, we now change the perspective very slightly. We define the Markov chain $\{\hat{Z}_t^\lambda\}$ on the state space $\{0,1\}^{C_\lambda}$. For an arbitrary configuration $\hat{\sigma}_\lambda \in \{0,1\}^{C_\lambda}$, the transition probability matrix $\hat{P}_\lambda$ of $\{\hat{Z}_t^\lambda\}$ is defined as follows:

- Choose an arbitrary circuit $C_i^\lambda \in C_\lambda$ at random.
- Let

$$\hat{\sigma}'_\lambda = \begin{cases} \text{only move } C_i \text{ to 1 in } \hat{\sigma}_\lambda \\ \text{with the probability } \frac{\lambda}{1+\lambda}; \\ \\ \text{only move } C_i \text{ to 0 in } \hat{\sigma}_\lambda \\ \text{with the probability } \frac{1}{1+\lambda}. \end{cases}$$

- The new state is $\hat{\sigma}'_\lambda$.

The transition probability matrix $\hat{P}_\lambda$ of $\{\hat{Z}_t^\lambda\}$ has the aperiodicity and irreducibility.

Proposition 3. The transition probability matrix $\hat{P}_\lambda$ of $\{\hat{Z}_t^\lambda\}$ has the following properties:

- aperiodicity: $\gcd\left\{t : \hat{p}_\lambda^t(i,i) > 0\right\} = 1$ for all $\hat{\sigma}_{\lambda,i} \in \{0,1\}^{C_\lambda}$;
- irreducibility: there exists t such that there is a positive probability of going from state $\hat{\sigma}_{\lambda,i}$ to state $\hat{\sigma}_{\lambda,j}$ after t steps, $\hat{P}_\lambda^t(i,j) > 0$, for all $\hat{\sigma}_{\lambda,i}, \hat{\sigma}_{\lambda,j} \in \{0,1\}^{C_\lambda}$.

Proof. For each state $\hat{\sigma}_{\lambda,i} \in \Omega_\lambda$, there is a positive probability such that it stays the current state. Thus, we have $gcd\left\{t : \hat{P}_\lambda^t(i,i) > 0\right\} = 1$ and aperiodicity is proved.

For each state $\hat{\sigma}_\lambda \in \Omega_\lambda$, by moving the circuits which take the value 1 in $\hat{\sigma}_\lambda$ to 0 step by step, there is a path to the state $\hat{\sigma}_{\lambda,0}$ in which all circuits take the value 0 and vice versa, i.e., for two arbitrary states, there is at least one path connected them by crossing $\hat{\sigma}_{\lambda,0}$. Irreducibility is proved. □

Based on the aperiodicity and irreducibility, the Markov chain $\{\hat{Z}_t^\lambda\}$ admits a unique stationary distribution, which is explicitly given by:

$$\hat{\mu}_\lambda(\hat{\sigma}_\lambda) = \prod_{C_i^\lambda:\hat{\sigma}_\lambda(C_i^\lambda)=1}\left(\frac{\lambda}{1+\lambda}\right)^{|C_i^\lambda|}\prod_{C_j^\lambda:\hat{\sigma}_\lambda(C_j^\lambda)=0}\left(\frac{1}{1+\lambda}\right)^{|C_j^\lambda|}.$$

The distribution $\hat{\mu}_\lambda$ can be viewed as the product of independent Bernoulli random variables, each defined on a circuit in G_λ and having a success probability $\frac{\lambda}{1+\lambda}$. Note that the distribution μ_λ is defined on $\Omega_\lambda \in \{0,1\}^{C_\lambda}$. To establish a connection between μ_λ and $\hat{\mu}_\lambda$, we need to unify their domains to $\{0,1\}^{C_\lambda}$. Therefore, we rewrite μ_λ in the following form:

$$\mu_\lambda(\sigma_\lambda) = \begin{cases} \dfrac{\prod\limits_{C_i^\lambda:\sigma_\lambda(C_i^\lambda)=1} W(C_i^\lambda)}{\sum\limits_{\sigma_\lambda\in\Omega_\lambda}\prod\limits_{C_i^\lambda:\sigma_\lambda(C_i^\lambda)=1} W(C_i^\lambda)} & \sigma_\lambda \in \Omega_\lambda; \\ 0 & \sigma_\lambda \in \{0,1\}^{C_\lambda} \setminus \Omega_\lambda. \end{cases}$$

The following elementary fact will play an important role in the subsequent analysis.

Lemma 2. For the two distributions μ_λ and $\hat{\mu}_\lambda$ defined above, $\hat{\mu}_\lambda$ stochastically dominates μ_λ. That is, there is a coupling of random variables σ_λ and $\hat{\sigma}_\lambda$, distributed as μ_λ and $\hat{\mu}_\lambda$, respectively, such that $\sigma_\lambda \le \hat{\sigma}_\lambda$ in the product order on $\{0,1\}^{C_\lambda}$.

Proof. A simple way to see this is by monotonic coupling of Markov chains. Recall that μ_λ and $\hat{\mu}_\lambda$

are the stationary distributions of the Markov chains $\{Z_t^\lambda\}$ and $\{\hat{Z}_t^\lambda\}$, respectively. By the definition of monotone coupling, the couplings $\{Z_t^\lambda\}$ and $\{\hat{Z}_t^\lambda\}$ can form a monotone coupling iff the following condition holds:

$$\begin{aligned}&\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1 \mid \sigma_\lambda(C_\lambda \setminus C_i^\lambda))\\ &\leq \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1 \mid \hat{\sigma}_\lambda(C_\lambda \setminus C_i^\lambda)),\end{aligned} \tag{25}$$

for all $\sigma_\lambda, \hat{\sigma}_\lambda$ such that $\sigma_\lambda \preceq \hat{\sigma}_\lambda$. We first show that Eq. (25) holds. We consider the following cases separately.

1. $\sigma_\lambda(\Gamma(C_i^\lambda)) = \hat{\sigma}_\lambda(\Gamma(C_i^\lambda)) = \mathbf{0}$. By the transition probability matrix P_λ, we have

$$\begin{aligned}&\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1 \mid \sigma_\lambda(C_\lambda \setminus C_i^\lambda))\\ &= \frac{W(C_i^\lambda)}{1 + W(C_i^\lambda)}.\end{aligned} \tag{26}$$

   By the transition probability matrix $\hat{P}_\lambda$, we have

$$\begin{aligned}&\hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1 \mid \hat{\sigma}_\lambda(C_\lambda \setminus C_i^\lambda))\\ &= \frac{\lambda}{1 + \lambda}.\end{aligned} \tag{27}$$

   Note that $\lambda = \max_{C_i^\lambda \in \mathcal{C}_\lambda} W(C_i^\lambda)$. Thus, Eq. (25) holds.
2. $\sigma_\lambda(\Gamma(C_i^\lambda)) \neq \mathbf{0}$ and $\hat{\sigma}_\lambda(\Gamma(C_i^\lambda)) \neq \mathbf{0}$. By the transition probability matrix P_λ, we have

$$\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1 \mid \sigma_\lambda(C_\lambda \setminus C_i^\lambda)) = 0.$$

   Thus, Eq. (25) holds.
3. $\sigma_\lambda(\Gamma(C_i^\lambda)) = \mathbf{0}$ and $\hat{\sigma}_\lambda(\Gamma(C_i^\lambda)) \neq \mathbf{0}$. By the transition probability matrix P_λ and $\hat{P}_\lambda$, we have Eqs. (26) and (27). Thus, Eq. (25) holds.
4. $\sigma_\lambda(\Gamma(C_i^\lambda)) \neq \mathbf{0}$ and $\hat{\sigma}_\lambda(\Gamma(C_i^\lambda)) = \mathbf{0}$. This implies that there exists a $C_j^\lambda \in \Gamma(C_i^\lambda)$ such that $\sigma_\lambda(C_j^\lambda) = 1$ and $\hat{\sigma}_\lambda(C_j^\lambda) = 0$. This contradicts $\sigma_\lambda \preceq \hat{\sigma}_\lambda$.

By Eq. (25), we can construct a coupling between the Markov chains $\{Z_t^\lambda\}$ and $\{\hat{Z}_t^\lambda\}$. We initialize Z_0^λ with the all-zero configuration and $\hat{Z}_0^\lambda$ with the all-one configuration.

According to the properties of monotonic coupling, we have $Z_t^\lambda \preceq \hat{Z}_t^\lambda$ in the product order on $\{0,1\}^{\mathcal{C}_\lambda}$ for $t \geq 0$. This completes the proof. □

Next, we prove that there exists a constant $\mathcal{B}$ such that for an arbitrary claw-free graph G and all pinnings $C_i \in \mathcal{C}$, it holds that $\mathcal{B} \leq \mu(\sigma(C_i)) \leq 1-\mathcal{B}$.

Corollary 2. Let $\delta \geq 3$ and $n > 0$ be integers, and b be a real. And let G_λ be a planar graph that can be decomposed into n_λ circuits under the decomposition $\mathcal{C}_\lambda$. For the four-vertex model on the graph G_λ with the constraint function set $\mathcal{F} = \{f_{C_i^\lambda} \mid C_i^\lambda \in \mathcal{C}_\lambda\}$, the distribution of its valid configurations μ_λ for all $C_i^\lambda \in \mathcal{C}_\lambda$ satisfies:

- When $0 < b \leq 1$,

$$\frac{b^\delta}{(1+b^\delta)^\delta(1+b)} \leq \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \frac{b}{1+b}.$$

- When $b > 1$,

$$\frac{b}{(1+b)(1+b^\delta)^\delta} \leq \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \frac{b^\delta}{1+b^\delta}.$$

That is, μ is $\frac{b^\delta}{(1+b^\delta)^\delta(1+b)}$-marginally bounded for $0 < b \leq 1$ and $\frac{b}{(1+b)(1+b^\delta)^\delta}$-marginally bounded for $b > 1$.

Proof. Consider the upper bound of $\mu_\lambda(\sigma_\lambda(C_i) = 1)$. By Lemma 2, we have that $\hat{\mu}$ stochastically dominates the distribution μ. In [5], we have that if this dominance holds, then for an arbitrary non-decreasing function $f : \{0,1\}^{\mathcal{C}_\lambda} \to \mathbb{R}$, μ_λ and $\hat{\mu}_\lambda$ satisfy:

$$\sum_{\sigma_\lambda \in \Omega_\lambda} f(\sigma_\lambda)\mu_\lambda(\sigma_\lambda) \leq \sum_{\hat{\sigma}_\lambda \in \Omega_\lambda} f(\hat{\sigma}_\lambda)\hat{\mu}_\lambda(\hat{\sigma}_\lambda).$$

For an arbitrary circuit $C_i^\lambda \in \mathcal{C}_\lambda$, we construct a specific non-decreasing functions $f_i(x)$:

$$f_i(x) = \begin{cases} 1 & \text{if } C_i^\lambda \text{ takes the value 1 in } x, \\ 0 & \text{otherwise.} \end{cases}$$

Further, we have

$$\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) = \sum_{\sigma_\lambda \in \Omega_\lambda} f_i(\sigma_\lambda)\mu_\lambda(\sigma_\lambda)$$

and

$$\hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1) = \sum_{\hat{\sigma}_\lambda \in \Omega_\lambda} f_{\hat{\sigma}_\lambda^*}(\hat{\sigma}_\lambda)\hat{\mu}_\lambda(\hat{\sigma}_\lambda).$$

Since $\hat{\mu}_\lambda$ stochastically dominates μ_λ, we have

$$\sum_{\hat{\sigma}_\lambda \in \Omega_\lambda} f_{\sigma_x}(\hat{\sigma}_\lambda)\hat{\mu}_\lambda(\hat{\sigma}_\lambda) \geq \sum_{\sigma_\lambda \in \Omega_\lambda} f_{\sigma_x}(\sigma_\lambda)\mu_\lambda(\sigma_\lambda).$$

This implies that for an arbitrary circuit $C_i^\lambda \in \mathcal{C}_\lambda$ we have

$$\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1). \tag{28}$$

Note that $\hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1) = \frac{\lambda}{1+\lambda}$. Thus, we have

$$\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \begin{cases} \frac{b}{1+b} & 0 < b \leq 1, \\ \frac{b^\delta}{1+b^\delta} & b > 1. \end{cases} \tag{29}$$

Next, we consider the lower of $\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1)$. Note that C_j^λ is the neighbor of C_i^λ and $\sigma_\lambda(j)$ is the assignment of all the circuits neighboring C_i^λ. By the equation of conditional probability, we have

$$\begin{aligned} &\mu_\lambda(\sigma_\lambda(C_i = 1)) \\ &= \mu_\lambda(\sigma_\lambda(C_i = 1) \mid \sigma_\lambda(j) = \mathbf{0})\mu_\lambda(\sigma_\lambda(j) = \mathbf{0}). \end{aligned} \tag{30}$$

For two arbitrary circuits C_u^λ and C_v^λ in the graph G_λ, $\sigma_\lambda(C_u^\lambda) = 0$ and $\sigma_\lambda(C_v^\lambda) = 0$ are independent of each other. Therefore, we have

$$\mu_\lambda(\sigma_\lambda(j) = \mathbf{0}) = \prod_{C_j^\lambda \in \Gamma(C_i^\lambda)} \mu_\lambda(\sigma_\lambda(C_j^\lambda) = 0). \tag{31}$$

Note that $\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 1)$ and $\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 0) = 1 - \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1)$. Thus, we have $\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 0) \geq \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 0)$. Further, we have

$$\prod_{C_j^\lambda \in \Gamma(C_i^\lambda)} \mu_\lambda(\sigma_\lambda(C_j^\lambda) = 0) \geq \prod_{C_j^\lambda \in \Gamma(C_i^\lambda)} \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_j^\lambda) = 0). \tag{32}$$

By Eqs. (30), (31) and (32), it follows that

$$\begin{aligned} &\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \\ &\geq \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1 \mid \sigma_\lambda(j) = \mathbf{0}) \\ &\cdot \prod_{C_j^\lambda \in \Gamma(C_i^\lambda)} \hat{\mu}_\lambda(\sigma_\lambda(C_j^\lambda) = 0). \end{aligned} \tag{33}$$

By the transition probability matrix P_λ, we have

$$\begin{aligned} &\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1 | \sigma_\lambda(j) = \mathbf{0}) \\ &\geq \begin{cases} \frac{b^\delta}{1+b^\delta} & 0 < b \leq 1, \\ \frac{b}{1+b} & b > 1. \end{cases} \end{aligned} \tag{34}$$

Note that $|\Gamma(C_i)| \leq \delta$. By the transition probability matrix $\hat{P}_\lambda$, we have

$$\begin{aligned} &\prod_{C_j^\lambda \in \Gamma(C_i^\lambda)} \hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_j^\lambda) = 0) \\ &\geq \begin{cases} \left(\frac{1}{1+b}\right)^\delta & 0 < b \leq 1, \\ \left(\frac{1}{1+b^\delta}\right)^\delta & b > 1. \end{cases} \end{aligned} \tag{35}$$

By Eqs. (33), (34) and (35), it follows that

$$\begin{aligned} &\mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \\ &\geq \begin{cases} \frac{b^\delta}{(1+b^\delta)(1+b)^\delta}, & 0 < b \leq 1, \\ \frac{b}{(1+b^\delta)^\delta(1+b)}, & b > 1. \end{cases} \end{aligned} \tag{36}$$

By Eqs (29) and (36), it follows that

- if $0 < b \leq 1$, we have

$$\frac{b^\delta}{(1+b^\delta)^\delta(1+b)} \leq \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \leq \frac{b}{1+b}$$

and

$$\frac{b^\delta}{(1+b^\delta)^\delta(1+b)} \leq 1 - \frac{b}{1+b}.$$

- if $b > 1$,

$$\frac{b}{(1+b)(1+b^\delta)^\delta} \le \mu_\lambda(\sigma_\lambda(C_i^\lambda) = 1) \le \frac{b^\delta}{1+b^\delta}$$

and

$$\frac{b}{(1+b)(1+b^\delta)^\delta} \le 1 - \frac{b^\delta}{1+b^\delta}.$$

Thus, if $0 < b \le 1$, then μ is $\frac{b^\delta}{(1+b^\delta)^\delta(1+b)}$-marginally bounded. Otherwise, μ is $\frac{b}{(1+b)(1+b^\delta)^\delta}$-marginally bounded. This completes the proof. □

Next, we prove that there exists η such that for all $C_i^\lambda \in C_\lambda$, we have $W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \le \eta$.

Without loss of generality, we assume that $\mu_\lambda^{(C_i^\lambda,0)}$ and $\mu_\lambda^{(C_i^\lambda,1)}$ are the stationary distributions of $\{Z_t^0\}$ and $\{Z_t^1\}$, respectively. Let σ_λ^0 and σ_λ^1 denote samples drawn from $\mu_\lambda^{(C_i^\lambda,0)}$ and $\mu_\lambda^{(C_i^\lambda,1)}$, respectively. If we construct a coupling of $\{Z_t^0\}$ and $\{Z_t^1\}$ such that

$$\sum_{\substack{\sigma_\lambda^0 \sim \mu_\lambda^{(C_i^\lambda,0)},\\ \sigma_\lambda^1 \sim \mu_\lambda^{(C_i^\lambda,1)}}} d_H(\sigma_\lambda^0, \sigma_\lambda^1) P(\sigma_\lambda^0, \sigma_\lambda^1) \le \eta,$$

then, by Eq. (17), we have

$$\begin{aligned} & W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \\ & \le \sum_{\substack{\sigma_\lambda^0 \sim \mu_\lambda^{(C_i^\lambda,0)},\\ \sigma_\lambda^1 \sim \mu_\lambda^{(C_i^\lambda,1)}}} d_H(\sigma_\lambda^0, \sigma_\lambda^1) P(\sigma_\lambda^0, \sigma_\lambda^1) \\ & \le \eta, \end{aligned}$$

which completes the proof. Therefore, the key to the proof is to construct a coupling of which the expectation in terms of the Hamming distance between σ_λ^0 and σ_λ^1 is bounded.

Before presenting the proof, we first introduce the relevant notations: For an arbitrary state $\sigma_\lambda^{(C_i^\lambda,0)}$, let B denote the random 'blue' independent set consisting of circuits with value 1 in $\sigma_\lambda^{(C_i^\lambda,0)}$; For an arbitrary state $\sigma_\lambda^{(C_i^\lambda,1)}$, let R denote the random 'red' independent set consisting of circuits assigned value 1 in $\sigma_\lambda^{(C_i^\lambda,1)}$. Note that both B and R are sets composed of circuits. We then consider the subgraph $G[R \cup B]$ of G induced by vertices of the ciruits in the union of the two independent sets. Let $C \subseteq C_\lambda$ be the set of connected circuits in $G[R \cup B]$ that contains C_i^λ. We refer to this set C as a cluster.

Next, we construct the desired coupling of $\{Z_t^0\}$ and $\{Z_t^1\}$, which performs a random walk at each step as follows:

- We select the same circuits C_i^λ for both chains. $\{Z_t^0\}$ and $\{Z_t^1\}$ perform the same random update at C_i, even if the resulting state is invalid (in which case the chain remains in its current state).
- Update the blue independent set by setting $B \leftarrow (B \cap C) \cup (R \setminus C)$. Note that B agrees with the original B inside C and with R outside; the red independent set R remains unchanged.
- Z_{t+1}^0 is set to the state $\sigma_\lambda^{(C_\lambda^i,0)}$ corresponding to the independent set B, while Z_{t+1}^1 is set to the state $\hat{\sigma}$ corresponding to R.

By Lemma 7 of [4], the constructed coupling of $\{Z_t^0\}$ and $\{Z_t^1\}$ is correct: σ is a sample from $\mu_\lambda^{(C_i^\lambda,0)}$ and σ_λ^1 is a sample from $\mu_\lambda^{(C_i^\lambda,1)}$. Therefore, by analyzing this coupling via Algorithm 1, we can derive an upper bound on $W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right)$.

Lemma 3. Suppose that $\delta \ge 3$ and $b > 0$. For the four-vertex model with the constraint function set $\mathcal{F}$, whose underlying graph under the circuit decomposition C_λ is a claw-free, the distribution of its partition function μ_λ for all $C_i^\lambda \in C_\lambda$ satisfies:

$$W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \le \begin{cases} 2(1+b)^{2\delta} & 0 < b \le 1, \\ 2(1+b^\delta)^{2\delta} & b > 1, \end{cases}$$

Algorithm 1 Growing a red-blue cluster by Breadth-First Search (BFS)

function CLUSTER BFS(C_i)
  $L_0 \leftarrow \{C_i\}$
  $U \leftarrow C \setminus C_i$
  $d \leftarrow 0$
  while $L_d \neq \emptyset$ **do**
    $N \leftarrow U \cap \partial L_d$
    if d is even **then**
      set $L_{d+1} \leftarrow N \cap B$ (blue vertices)
    else
      set $L_{d+1} \leftarrow N \cap R$ (red vertices)
    end if
    $U \leftarrow U \setminus N$
    $d \leftarrow d + 1$
  end while
  $d \leftarrow d - 1$
end function
Remark: U is the set of unprocessed circuits and $\partial L_d = \{C_u \in C \setminus L_d : C_u$ is a neighbor of C_w for some $C_w \in L_d\}$. The red-blue cluster containing C_i is $C = L_0 \cup L_1 \cup \cdots \cup L_d$,

where $\delta = \max_{C_i^\lambda \in C_\lambda} \delta_{C_i^\lambda}$ and $\delta_{C_i^\lambda}$ is the number of neighbors of the circuit C_i^λ.

Proof. The claimed bound on $W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right)$ is justified by the coupling just analyzed.

We interpret σ_λ^0, sampled from $\mu_\lambda^{(C_i^\lambda,0)}$, as the corresponding random "blue" independent circuit set B. Similarly, we interpret σ_λ^1, sampled from $\mu_\lambda^{(C_i^\lambda,1)}$, as the corresponding random "red" independent circuit set R. Therefore, we have $d_H(\sigma_\lambda^0, \sigma_\lambda^1) = |B \oplus R|$.

Since B agrees with the original B inside C and with R outside, it follows that $C = B \oplus R$. By Eq. (17), we have

$$E[|C|] \geq W_1(\mu_\lambda^{(C_i,0)}, \mu_\lambda^{(C_i,1)}).$$

Let $G_\lambda[C]$ is the subgraph of G_λ induced by vertices of the circuits in the cluster C. We first describe the properties of an arbitrary cluster $G_\lambda[C]$, as follows:

1. An arbitrary pair of adjacent circuits in C belong to B and R respectively, i.e. The circuits of C are bipartite. This holds because both B and R are independent circuit sets.
2. $G_\lambda[C]$ is a connected claw-free graph which can be decomposed into some circuits. This follows from the fact that $G_\lambda[C]$ is a subgraph of G_λ and G_λ is a connected claw-free graph under the circuit decomposition C_λ.
3. $|C| = |L_0| + \cdots + |L_d|$, where d is the number of iterations of the algorithm and L_d is the set of vertices added to the cluster during the d-th iteration. This holds because C is constructed in d stages by Algorithm 1.

By property 1, two arbitrary circuits adjacent to the same circuit are not adjacent to each other in C. By property 2, the number of neighbors of each circuit is at most 2. Therefore, $G_\lambda[C]$ is either a path or a cycle under the circuit decomposition C_λ. By property 3, we have $|C| = \sum_{j=1}^{d} |L_j|$. Note that the vertices added to the current cluster at each step of the algorithm are essentially those adjacent to the current cluster. Moreover, $G_\lambda[C]$ is either a path or a cycle under the circuit decomposition C_λ. Therefore, $|L_j| \leq 2$. If we assume that the number of iterations of the algorithm is a random variable, then we have $E[|C|] \leq E[2d]$.

Next, we consider the upper bound of $E[|2d|]$. By the process of Algorithm 1, we know that the termination condition of the algorithm is $|L_d| = 0$. If there exists the lower bound for $P(|L_d| = 0)$, then the search process of the algorithm for C can be regarded as the d -Bernoulli trials with this lower bound as the probability of success, where the results of the first d - 1 trials are all failures and only the result of the d-th trial is a success. Therefore, we can use the expected value of the geometric dis-

tribution as the upper bound of $E(|d|)$.

We now consider the lower bound of $P(|L_d| = 0)$. To make the statement concise, we set

$$\partial' L_{d-1} = \{C_i^\lambda : C_i^\lambda \in \partial L_{d-1} \cap U \text{ and } \sigma_\lambda^{(C_i^\lambda,1)}(C_i^\lambda) = 0\}$$

and

$$\partial'' L_{d-1} = \{C_i^\lambda : C_i^\lambda \in \partial L_{d-1} \cap U \text{ and } \sigma_\lambda^{(C_i^\lambda,0)}(C_i^\lambda) = 0\}.$$

By Algorithm 1, we have:

- if d is even, then

$$P(|L_d| = 0) = \prod_{C_i^\lambda \in \partial' L_{d-1}} \mu_\lambda^{(C_i^\lambda,1)}\left(\sigma_\lambda^{(C_i^\lambda,1)}(C_i^\lambda) = 0\right).$$

- if d is odd, then

$$P(|L_d| = 0) = \prod_{C_i^\lambda \in \partial'' L_{d-1}} \mu_\lambda^{(C_i^\lambda,1)}\left(\sigma_\lambda^{(C_i^\lambda,0)}(C_i^\lambda) = 0\right).$$

By Eq. (28), we have

$$P(|L_d| = 0) \geq \prod_{C_i^\lambda \in \partial L_{d-1}} \hat{\mu}_\lambda(\sigma_\lambda(C_i^\lambda) = 0), \tag{37}$$

where $\hat{\mu}_\lambda(\hat{\sigma}_\lambda(C_i^\lambda) = 0) = \frac{1}{1+\lambda}$. Since $|L_j| \leq 2$, we have $|\partial L_{d-1}| \leq 2\delta$. By Eq. (37), we have

$$P(|L_d = 0|) \geq \frac{1}{(1+\lambda)^{2\delta}}.$$

So $E[|d|]$ is bounded above by the expectation of a random variable having a geometric distribution with success probability $\frac{1}{(1+\lambda)^{2\delta}}$, which is $(1+\lambda)^{2\delta}$.

Since $E[|C|] \leq E[2d]$, we have $E[|C|] \leq 2(1+\lambda)^{2\delta}$. By Eq. (18), we have

$$E[|C|] \leq \begin{cases} 2(1+b)^{2\delta} & 0 < b \leq 1, \\ 2(1+b^\delta)^{2\delta} & b > 1, \end{cases}$$

This completes the proof. □

By Corollary 1, Corollary 2 and Lemma 3, we obtain:

Theorem 4. (Mixing time for claw-free graphs). For the four-vertex model with the constraint function f^*, whose underlying graph is claw-free,

- if $\delta \geq 3$ and $0 < b \leq 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b^\delta)^\delta(1+b)}{b^\delta}\right)^{O\left(\frac{2(1+b^\delta)^{2\delta}(1+b)^{2\delta+2}}{b^{2\delta}}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right);$$

- if $\delta \geq 3$ and $b > 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b)(1+b^\delta)^\delta}{b}\right)^{O\left(\frac{2(1+b)^2(1+b^\delta)^{4\delta}}{b^2}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right).$$

Next, we consider G as a connected $S_{1,1,t}$-free graph, which can be specifically divided into two parts:

- If the black face graph of G has no claws, then G is a claw-free graph. By Theorem 4, we can obtain the mixing time of Z_t.
- If the black face graph of G has the claw, then the number of circuits in G satisfies $|n| \leq 1 + \delta\frac{(\delta-1)^{t+1}-1}{\delta-2}$. This implies that the state space of Z_t is a constant independent of n. Thus, the mixing time of Z_t is $O(1)$.

Combining the above two parts, we complete the proof of Theorem 5.

Theorem 5. (Mixing time for $S_{1,1,t}$-free graphs). For the four-vertex model with the constraint function f^*, whose underlying graph is $S_{1,1,t}$-free ($t \geq 2$) under the circuit decomposition C,

- if $\delta \geq 3$ and $0 < b \leq 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b^\delta)^\delta(1+b)}{b^\delta}\right)^{O\left(\frac{2(1+b^\delta)^{2\delta}(1+b)^{2\delta+2}}{b^{2\delta}}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right);$$

- if $\delta \geq 3$ and $b > 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b)(1+b^\delta)^\delta}{b}\right)^{O\left(\frac{2(1+b)^2(1+b^\delta)^{4\delta}}{b^2}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right).$$

Appendix F

In this appendix, we consider the case where the underlying graph G is composed of blow-ups of paths and cycles under the circuit decomposition C.

We first give the definition of the "blow-up", in detail as follows: A graph is called a blow-up of a graph if it is constructed by replacing each vertex of graph with an independent set, and each edge of graph with a complete bipartite graph between the corresponding sets.

If the black face graph $H(G)$ is composed of blow-ups of paths and cycles, then we say that the underlying graph G is composed of blow-ups of paths and cycles under the circuit decomposition C.

Skew-star-free graphs:

We first define $G_B = (V_1, V_2, E)$ as a bipartite graph, where V_1 and V_2 are independent sets and $E \subseteq V_1 \times V_2$ is the set of edges on graph G_B. We call G_B a complete bipartite graph if $E = V_1 \times V_2$. For a bipartite graph $G_B = (V_1, V_2, E)$, we denote its bipartite complement by $\overline{G}_B$, i.e., $\overline{G}_B = (V_1, V_2, (V_1 \times V_2) \setminus E)$. A bipartite graph is said to be prime if every pair of distinct vertices has different neighborhoods. Moreover, a graph is called a blow-up of a graph G if it is constructed by replacing each vertex of G with an independent set, and each edge of G with a complete bipartite graph between the corresponding sets. The following lemma on the structural properties of $S_{1,2,t}$-free (also called skew-star-free) graphs is from [6].

Lemma 4. Let G_B be a prime bipartite graph without a skew star. Then

- either G_B is disconnected,
- or G_B is the bipartite complement to a disconnected graph,
- or G_B can be partitioned into a independent set and a complete bipartite graph,
- or G_B is claw-free,
- or G_B is composed of blow-ups of paths and cycles.

Thus, we have that G is skew-star-free by Lemma 4 under the circuit decomposition C.

Proof of Theorem 6:

Theorem 6. (Mixing time for the composition of blow-ups of paths and cycles). For the four-vertex model with the constraint function f^*, whose underlying graph is composed of blow-ups of paths and cycles under the circuit decomposition C,

- if $\delta \geq 3$ and $0 < b \leq 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b^\delta)^\delta(1+b)}{b^\delta}\right)^{O\left(\frac{2(\delta-1)(1+b^\delta)^{2\delta}(1+b)^{2\delta(\delta-1)+2}}{b^{2\delta}}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right);$$

- if $\delta \geq 3$ and $b > 1$, the Markov chain $\{Z_t\}$ mixes in time:

$$\left(\frac{\delta(1+b)(1+b^\delta)^\delta}{b}\right)^{O\left(\frac{2(\delta-1)(1+b)^2(1+b^\delta)^{2\delta^2}}{b^2}+1\right)} \times O\left(n\log\left(\frac{n}{\varepsilon}\right)\right).$$

As before, we still derive the mixing time of the Markov chain on G by finding the appropriate parameters $\mathcal{B}$ and η. By Corollary 2, we have

$$\mathcal{B} = \begin{cases} \frac{b^\delta}{(1+b^\delta)^\delta(1+b)} & 0 < b \leq 1; \\ \frac{b}{(1+b)(1+b^\delta)^\delta} & b > 1. \end{cases}$$

So, we only need to re-find the parameter η.

Lemma 5. Suppose that $\delta = \max_{C_i^\lambda \in C_\lambda} \delta_{C_i^\lambda} \geq 3$ and $\delta_{C_i^\lambda}$ is the number of neighbors of the circuit C_i^λ. For the four-vertex model with the constraint function set $\mathcal{F}$, whose underlying graph under the circuit decomposition C_λ is composed of blow-ups of paths and cycles, the distribution of its partition function μ_λ for all $C_i^\lambda \in C_\lambda$ satisfies:

- when $0 < b \leq 1$,

$$W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \leq 2(\delta-1)(1+b)^{2(\delta-1)};$$

- when $b > 1$,

$$W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \leq 2(\delta-1)(1+b^\delta)^{2(\delta-1)}.$$

Proof. The proof is similar to Lemma 3, both of which use the BFS of Algorithm 1 to analyze the coupling of $\{Z_t^0\}$ and $\{Z_t^1\}$, where $\{Z_t^0\}$ and $\{Z_t^1\}$ are sampled from the distributions $\mu_\lambda^{(C_i^\lambda,0)}$ and $\mu_\lambda^{(C_i^\lambda,1)}$ respectively. But the difference is that the layers L_j created by the BFS of Algorithm 1 no longer satisfy $|L_j| \leq 2$. Before reconsidering the bounds of $|L_j|$, we consider the properties of the cluster C obtained from the BFS. Since G_λ is composed of blow-ups of paths and cycles, G is a prime bipartite graph under the circuit decomposition C. Since $G_\lambda[C]$ is a connected subgraph of G_λ, $G_\lambda[C]$ is connected and is composed of blow-ups of paths and cycles under the circuit decomposition C_λ. This implies that $|L_j| \leq 2(\delta-1)$. Thus, $|\partial L_{d-1}| \leq 2\delta(\delta-1)$. By Eq. (37), we have

$$P(|L_d| = 0|) \geq \frac{1}{(1+\lambda)^{2\delta(\delta-1)}}.$$

So $E|d|$ is bounded above by the expectation of a random variable having a geometric distribution with success probability $\frac{1}{(1+\lambda)^{2\delta(\delta-1)}}$, which is $(1+\lambda)^{2\delta(\delta-1)}$. Since $|C| \leq 2(\delta-1)d$, we have $E[|C|] \leq E[2(\delta-1)d]$ and $E[|C|] \leq 2(\delta-1)(1+\lambda)^{2\delta(\delta-1)}$. Note that $E(|C|) \geq W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right)$. Therefore, we have $W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \leq 2(\delta-1)(1+\lambda)^{2\delta(\delta-1)}$. By Eq. (18), we have

$$W_1\left(\mu_\lambda^{(C_i^\lambda,0)}, \mu_\lambda^{(C_i^\lambda,1)}\right) \leq \begin{cases} 2(\delta-1)(1+b)^{2\delta(\delta-1)} & 0 < b \leq 1; \\ 2(\delta-1)(1+b^\delta)^{2\delta(\delta-1)} & b > 1. \end{cases}$$

This completes the proof. □

By Corollary 1, Corollary 2 and Lemma 5, we complete the proof of Theorem 6.

References

1. Dyer M E, Greenhill C S. On markov chains for independent sets. Journal of Algorithms, 2000, 35(1): 17–49
2. Chen Z C, Liu K K, Vigoda E. Optimal mixing of glauber dynamics: entropy factorization via high-dimensional expansion. In: Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC. 2021, 1537–1550
3. Anari N, Kui L K, Gharan S O. Spectral independence in high dimensional expanders and applications to the hardcore model. In: Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS). 2020, 1319–1330
4. Jerrum M. Glauber dynamics for the hard-core model on bounded-degree-free graphs. Combinatorics, Probability and Computing, 2025, 34(6): 803–814
5. Friedli S, Velenik Y. Statistical mechanics of lattice systems: A concrete mathematical introduction. Cambridge: Cambridge University Press, 2018
6. Lozin V V. Bipartite graphs without a skew star. Discrete Mathematics, 2002, 257(1): 83–100