hexagon:2611.00580v1

Uniform Spectral Gap for the AKLT Model on the Square Lattice

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note, the entire document is machine-written. We have carefully supervised its preparation to ensure that the proofs are readable and that previous literature is appropriately cited. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

This paper is accompanied by a Lean formalization, which can be found at

`https://github.com/pmaronow/aklt-lean`.

Compared to the formalization, the proof here has been slightly streamlined for human consumption. The exact proof that was formalized can be found in a manuscript in the linked repository.

P. L. was partially supported by NSF grant DMS-2450004.

ABSTRACT

We prove that the spin-2 AKLT Hamiltonian on the square-lattice torus has a unique ground state and a spectral gap bounded below uniformly when both side lengths are sufficiently large, as expected since the work of Affleck, Kennedy, Lieb, and Tasaki. We also prove uniqueness of the infinite-volume frustration-free state on the square lattice. We first bound the quantum gap from below by the spectral gap of a classical block dynamics for the measure that the valence-bond-solid state defines in spin coherent states. We bound the latter through the coupling argument of Martinelli, Olivieri, and Schonmann, whose input we derive from the exponential decay of responses of pair correlations to boundary conditions. We prove this decay by an expansion into bond-self-avoiding walks, comparing the expansions with and without boundary conditions walk by walk, and closing the resulting inequality as a contraction. The proof is computer-assisted. Its local and combinatorial inputs are verified in exact integer arithmetic and in floating-point arithmetic with directed rounding.

Contents

1	Introduction	1
2	Overview of the proof	6
3	Coherent states and a classical block dynamics	14
4	From a window estimate to the gap of the block dynamics	23
5	The cut lemma and the window estimate	35
6	Walk expansions	53
7	Contraction of the boundary response	69
8	The computer-assisted inputs	80
A	Connectivity of the cut	87
B	Certificates for the local inequalities	91
C	The tail of the weighted trail sums	111
D	Explicit constants	123
E	The ancillary files	130
F	Uniqueness of the infinite-volume frustration-free state	136
References		140

SECTION 1

Introduction

1.1 Background

Haldane predicted that the antiferromagnetic Heisenberg chain has a spectral gap above its ground state when the spin per site is an integer [15, 16]. The first rigorous example of an isotropic model with this behavior was found by Affleck, Kennedy, Lieb, and Tasaki [2, 3]. They constructed a spin-1 chain whose ground state is built from valence bonds, and they proved that it has a unique ground state, a spectral gap, and exponentially decaying correlations. The same construction gives a rotation-invariant, frustration-free Hamiltonian on any lattice, with spin equal to half the number of neighbors at each site, and an explicit ground state of this Hamiltonian, the *valence-bond-solid* (VBS) state. These Hamiltonians are now called AKLT models. In [3], models in two and higher dimensions were presented that "are expected to have the same properties" as the chain. In particular, the AKLT models on the hexagonal lattice, with spin $3/2$, and on the square lattice, with spin 2, are expected to have a unique ground state in the thermodynamic limit and a spectral gap (see [20, Section 1.1] for this reading of [3]). Beyond their role in the study of Haldane's prediction, two-dimensional AKLT states have been studied as resources for measurement-based quantum computation [47].

Kennedy, Lieb, and Tasaki [23] proved that the spin-$3/2$ model on the hexagonal lattice has a unique ground state in the infinite-volume limit, in which all truncated correlation functions decay exponentially. They also characterized the finite-volume ground states of the VBS models in every dimension, and they proved that the two-point function of the spin-2 model on the square lattice decays exponentially on bipartite tori. Their bound on the staggered magnetization with Néel boundary conditions decays exponentially in the distance to the boundary, with a prefactor proportional to the boundary size. These results use the representation of the VBS state, due to Arovas, Auerbach, and Haldane [5], as the Boltzmann weight of a classical model of unit vectors with a factor $1 - s_x \cdot s_y$ for each bond. On the square lattice, the proof in [23] expands correlations into walks whose weights are nonnegative on bipartite tori and for Néel boundary conditions (in which the boundary spins are fixed in a staggered pattern), and it bounds them by a weighted count of bond-self-avoiding walks. Kennedy, Lieb, and Tasaki pointed out that the exponential decay for other boundary conditions and the uniqueness of the infinite-volume ground state on the square lattice were open questions. Jackson, Nachtergaele, and Young [20,

Section 1.2] left open the question of uniqueness among frustration-free states.

The spectral gap of two-dimensional AKLT models has been studied mostly through *finite-size criteria*, which bound the gap of a frustration-free Hamiltonian from below, uniformly in the volume, once the gap of a fixed finite system exceeds an explicit threshold [13, 26–28]. For the hexagonal lattice, Lemm, Sandvik, and Wang [29] and Pomata and Wei [41] combined such criteria with numerical computations for finite systems, by the density matrix renormalization group and by exact and Lanczos diagonalization, respectively, and obtained a gap. However, the numerical steps of these works are not certified, and Lemm, Sandvik, and Wang state that their result "may not be considered a rigorous mathematical proof as a matter of principle" [29]. Recently, Kastoryano [21] gave an analytic proof of a uniform gap for the hexagonal model on a family of rectangular regions, through the boundary-state criterion of Kastoryano, Lucia, and Pérez-García [22], with a scalar estimate verified in the proof assistant Lean. For decorated lattices, in which the bonds of a lattice are replaced by chains of spin-1 sites, gaps were proved in [1, 32, 40], and gaps for two further planar lattices with spin 2 at every site were obtained in [14]. Jackson, Nachtergaele, and Young [20] proved the uniqueness of the frustration-free infinite-volume state, local indistinguishability of finite-volume ground states, and the stability of the gap for the hexagonal lattice and for the decorated square lattices (see also [33]).

For the spin-2 model on the square lattice, the existence of a spectral gap has remained open [14, 37]. Numerical studies indicate a gap of order 10^{-2} [12], and Martyn, Kato, and Lucia [37] gave numerical evidence for the boundary-state criterion of [22]. In [14], the gap on the square lattice is reduced to a finite-size problem for a weighted Hamiltonian, which was left open in that work. The goal of this paper is to prove that the spin-2 AKLT model on the square lattice has a spectral gap bounded below uniformly when both side lengths of the torus are sufficiently large. The proof is computer-assisted.

1.2 Main result

We now define the model. For integers $L, N \geq 3$, let $\Lambda = \Lambda_{L,N} = \mathbb{Z}_L \times \mathbb{Z}_N$ denote the discrete torus, and let $\mathcal{E}$ denote its set of nearest-neighbor bonds. At every site $x \in \Lambda$ we place a spin 2, that is, a copy V_x of the five-dimensional irreducible representation of $\mathrm{SU}(2)$, and we let $\mathcal{H}_\Lambda = \bigotimes_{x\in\Lambda} V_x$. For a bond $e = \{x, y\} \in \mathcal{E}$, let P_e denote the orthogonal projection of $V_x \otimes V_y$ onto its subspace of total spin 4, extended by the identity to $\mathcal{H}_\Lambda$. We call

$$H_{L,N} = \sum_{e\in\mathcal{E}} P_e \tag{1.1}$$

the *spin*-2 *AKLT Hamiltonian* on $\Lambda_{L,N}$. For a nonnegative self-adjoint operator H on a finite-dimensional space that is not zero, we let $\mathrm{gap}(H)$ denote its smallest nonzero

eigenvalue. Our main result is the following theorem.

Theorem 1.1. *Set*

$$L_0 = 2\lceil \exp(10^{40}) \rceil + 5, \qquad c = \exp\Big[-\exp\big(\exp(10^{41})\big)\Big]. \tag{1.2}$$

For all integers $L, N \geq L_0$, the kernel of $H_{L,N}$ is one-dimensional and

$$\operatorname{gap}(H_{L,N}) \geq c.$$

The kernel of $H_{L,N}$ is spanned by the VBS state for all $L, N \geq 3$. This follows from the divisibility argument of [25, Sections 1–2] (see also [23] and [20, Theorem 2.1]), and we include the short proof in Section 3, since our comparison argument rests on the same divisibility property. The content of Theorem 1.1 is therefore the lower bound on the gap, which is uniform in the size of the torus and holds for both parities of L and N. (For odd L or N, the torus is not bipartite, and the classical measure that we use is frustrated; our arguments use the bipartite structure only locally.) We have made no attempt to optimize the constants in Theorem 1.1. They are far from the expected size of the gap, and they reflect the very large blocks that our argument requires.

The proof uses three statements that we verify with the help of a computer: a family of 7,962 polynomial inequalities, which we verify by testing the positive definiteness of integer matrices in exact arithmetic; an exponential bound on a weighted number of walks, which we verify with an integer vector in exact arithmetic; and a bound below one on a weighted sum over returning walks, which we verify in floating-point arithmetic with directed rounding. The programs, the certificates, and the outputs of the computations are included as ancillary files and described in Appendix E. The rest of the proof is written out in full.

The same coupling and coherent-state representation also give uniqueness of the infinite-volume frustration-free state on the square lattice. We prove this in Appendix F.

1.3 Methods and relation to previous work

We give an overview of the proof in Section 2. We now describe how its ingredients relate to the literature.

Our starting point, proved in Section 3, is a comparison of the gap of $H_{L,N}$ with the gap of a classical Markov process. The VBS state, written in spin coherent states [31, 39, 43], defines a probability measure on configurations of unit vectors, and this is the classical model of [5, 23]. We show that the gap of $H_{L,N}$ is bounded below by the spectral gap of a dynamics that resamples large square blocks from their conditional laws under this measure, times $bg_b/(b+1)$, where b is the side length of the block and g_b is the smallest positive eigenvalue of the sum of the bond projections over the bonds

that meet one block. Correspondences between classical dynamics satisfying detailed balance and frustration-free quantum Hamiltonians are known for Rokhsar–Kivelson-type Hamiltonians [7, 17, 46]. In these works, the quantum Hamiltonian is constructed from the classical dynamics on a discrete state space, and its excitation spectrum is that of the classical dynamics. In our setting, however, the Hamiltonian is given, the coherent states are not orthogonal, and the comparison is an inequality, which rests on the divisibility of the local ground states by the bond singlets. We are not aware of an earlier comparison of this kind for AKLT models. In contrast with finite-size criteria, the smallest positive eigenvalue of this part of $H_{L,N}$ enters only through its positivity, so no threshold has to be met by a finite quantum system.

The spectral gap of Glauber dynamics for classical spin systems is closely tied to spatial mixing; see [10, 11, 34, 36]. In two dimensions, Martinelli, Olivieri, and Schonmann [35] proved, for finite-range systems with a finite single-spin space, that weak mixing implies strong mixing on squares. The heart of their proof is a coupling that, after a change of the boundary condition at one site, makes the two configurations agree on successive shells around that site, with a probability bounded below at each shell. In Section 4, we follow the proof of [35, Proposition 2.1] for our continuous spins. In their proof, the finiteness of the spin space and the positivity of the weights give a lower bound for the probability that the two configurations agree near the boundary of the box. We replace this bound by an elementary overlap estimate for densities that may vanish. The step from a contracting coupling to a spectral gap is the standard argument of Chen [8] (see [30, Theorem 13.1] for finite state spaces, and [38] for general ones), which we write out for our block dynamics.

The input of the coupling argument is a statement about the laws of shell segments far from the boundary of the box, uniformly in all boundary conditions. Uniformity in the boundary condition is the difficulty that was left open in [23] for the square lattice. We obtain it in Section 5 through a cut lemma: once we remove sites that separate the segment from the boundary, the law of the segment no longer depends on the boundary condition, and the cost of the removal is controlled by the changes that the boundary condition causes in certain pair correlations, which we call *responses*. In Sections 6 and 7, we bound the responses through an expansion of correlations into bond-self-avoiding walks (trails), of the type used in [23]. With arbitrary boundary conditions, the terms of this expansion have no sign. Instead of positivity, we compare, for each walk, its contributions with and without boundary conditions, and the differences are again responses. The resulting linear inequality for the responses closes as a contraction in a weighted supremum norm, provided a weighted sum over returning trails is smaller than one.

The three computer-assisted statements are proved in Section 8 and Appendices B and C. Each polynomial inequality is verified through an explicit Hermitian operator whose expectations in products of spin-$\frac{1}{2}$ coherent states reproduce the polynomial;

its positivity reduces to the positive definiteness of integer matrices, which we test by fraction-free elimination [6]. The bound on the weighted number of trails uses a transfer matrix over the last six steps of a trail, in the spirit of the finite-memory bounds for self-avoiding walks [4, 42], with an exact integer certificate. The bound on returning trails combines an exact enumeration of short trails with a function that satisfies a one-step inequality for a walk that dominates the longer ones, and this inequality is verified in floating-point arithmetic with directed rounding [19] (see [45] for rigorous computation with directed rounding). In each case we state precisely what the programs compute and what a reader must trust.

1.4 Notation

We write x, y, z, u, v for sites of $\mathbb{Z}^2$ and $d(x,y) = |x_1 - y_1| + |x_2 - y_2|$ for the graph distance between them, and we let $|x-y|_\infty = \max(|x_1 - y_1|, |x_2 - y_2|)$ denote their ℓ^∞ distance. We write $x \sim y$ if $d(x,y) = 1$, and we call a set $\{x, y\}$ with $x \sim y$ a *bond*. For $A \subset \mathbb{Z}^2$ and a site x, we let $\operatorname{dist}(x, A)$ and $\operatorname{dist}_\infty(x, A)$ denote the minimum of $d(x,a)$ and of $|x-a|_\infty$ over $a \in A$, with both distances equal to $+\infty$ when A is empty. We use the same notation on a torus $\mathbb{Z}_L \times \mathbb{Z}_N$, with the distances taken in the torus. We let $|A|$ denote the cardinality of a finite set A, and we set $(t)_+ = \max(t, 0)$ for $t \in \mathbb{R}$.

For $a, b \in \mathbb{R}^3$, we let $a \cdot b$ denote the Euclidean inner product and $|a|$ the Euclidean norm. We let $\mathbb{S}^2$ denote the unit sphere of $\mathbb{R}^3$, $\mathrm{d}n$ the uniform probability measure on $\mathbb{S}^2$, and $\mathbb{B}$ the closed unit ball of $\mathbb{R}^3$. We call the elements of $\mathbb{S}^2$ *spins*, and we write $n = (n_x)_x$ for a configuration of spins. For a probability measure μ and an integrable function f, we write $\mu(f) = \int f \, \mathrm{d}\mu$.

Let μ be a probability measure on a measurable space, and let $\mathcal{L}$ be a bounded, self-adjoint, nonpositive operator on $L^2(\mu)$ that annihilates the constant functions, such as the generator of a Markov dynamics reversible with respect to μ with bounded jump rates. We call

$$\operatorname{gap}(\mathcal{L}) = \inf\left\{ \frac{\langle F, -\mathcal{L}F \rangle_{L^2(\mu)}}{\operatorname{Var}_\mu(F)} : F \in L^2(\mu) \text{ not constant} \right\}$$

the *spectral gap* of $\mathcal{L}$, where $\operatorname{Var}_\mu(F) = \mu(|F - \mu(F)|^2)$.

All constants in this paper are explicit.

SECTION 2

Overview of the proof

This section explains the proof of Theorem 1.1. It is written to be read on its own, with the notation of Section 1.4. The precise statements and proofs occupy Sections 3 to 8 and the appendices; a roadmap is given in Section 2.7.

2.1 The strategy

The proof is a chain of four reductions. First, we bound the gap of $H_{L,N}$ from below by the gap of a classical Markov process, whose stationary measure is the squared modulus of the VBS state, written as a function of spin coherent states, on configurations of unit vectors. Second, we bound the gap of this process through the coupling argument of Martinelli, Olivieri, and Schonmann [35]. Its only input is a weak form of spatial mixing, which asks that, for every boundary condition of a large box, the law of a segment of spins far from the boundary dominate a fixed positive multiple of one probability measure. Third, we deduce this property from the exponential decay of certain differences between correlations with and without boundary conditions, by removing a set of sites between the segment and the boundary. Fourth, we prove the required decay by an expansion of correlations into walks, in the spirit of Kennedy, Lieb, and Tasaki [23], combined with a contraction argument whose numerical inputs we verify by computer.

The first, third, and fourth steps are new. The comparison with a classical dynamics replaces the finite-size criteria used for the hexagonal lattice, which require a lower bound for the gap of a finite quantum system above an explicit threshold; in our argument, a finite quantum system enters only through the positivity of its gap. The removal of a separating set replaces the weak mixing hypothesis of [35] by a bound on the cost of removing sites, and this cost needs to be bounded, not small. The contraction argument replaces the positivity behind the expansion of [23], which holds on bipartite tori and for Néel boundary conditions, by a comparison of the expansions with and without boundary conditions, walk by walk; the local inequalities that this comparison requires are new as well. The second step follows the proof of [35, Proposition 2.1], with an elementary overlap bound in place of the finiteness of the spin space.

2.2 From the quantum gap to a classical dynamics

We use the polynomial representation of spin 2 [3, 5], in which we identify the states at one site with the homogeneous polynomials of degree four in a variable $z \in \mathbb{C}^2$, and the states of the torus with the polynomials in variables $z_x \in \mathbb{C}^2$, $x \in \Lambda$, that are homogeneous of degree four in each z_x. For a bond $\{x, y\}$, we call $z_{x,1}z_{y,2} - z_{x,2}z_{y,1}$ its *singlet polynomial*. The kernel of the projection onto total spin 4 on the bond consists of the polynomials divisible by this singlet polynomial. We let Ψ_{VBS} denote the product of the singlet polynomials over all bonds of the torus; this is the VBS state.

Let $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ denote the vector of Pauli matrices. For $a \in \mathbb{S}^2$, we let $\zeta(a) \in \mathbb{C}^2$ denote a unit vector with $\zeta(a)\zeta(a)^* = (1 + a \cdot \sigma)/2$, which is a spin-$\frac{1}{2}$ coherent state. We choose ζ measurably as in Section 3.1. Evaluation at the coherent states $\zeta(s_x)$ identifies the space of states isometrically with a subspace of $L^2((\mathbb{S}^2)^\Lambda, \prod_{x\in\Lambda} 5\,\mathrm{d}s_x)$. Let $\mathcal{E}$ denote the set of bonds of the torus, and define $\mu_{L,N}$ to be the probability measure on $(\mathbb{S}^2)^\Lambda$ given by

$$\mu_{L,N}(\,\mathrm{d}s) = \frac{1}{Z_{L,N}} \prod_{\{x,y\}\in\mathcal{E}} (1 - s_x \cdot s_y) \prod_{x\in\Lambda} \mathrm{d}s_x, \tag{2.1}$$

where $Z_{L,N}$ is the normalizing constant. The singlet polynomial of a bond $\{x, y\}$, evaluated at coherent states, has squared modulus $(1 - s_x \cdot s_y)/2$. Therefore the squared modulus of Ψ_{VBS}, as a function of s, is proportional to the density of $\mu_{L,N}$. This classical representation of the VBS state goes back to [5, 23].

We now explain the main observation of this step. Let φ be a state orthogonal to Ψ_{VBS}, and let F denote the ratio of φ and Ψ_{VBS}, as functions of s. Then $\mu_{L,N}(F) = 0$ and $\mathrm{Var}_{\mu_{L,N}}(F) = \|\varphi\|^2/\|\Psi_{\mathrm{VBS}}\|^2$. Fix a square $R \subset \Lambda$ of side b. Let H_R be the sum of the bond projections over the bonds that meet R, and let ψ be the orthogonal projection of φ onto the kernel of H_R. Every polynomial in this kernel is divisible by the singlet polynomials of all bonds that meet R, and the same holds for Ψ_{VBS}. These singlet polynomials exhaust the degree of every variable z_x with $x \in R$. Hence the ratio of ψ and Ψ_{VBS} does not depend on the spins in R. The conditional expectation given the spins outside R is the best mean-square approximation of F by such functions. Let $\mathrm{Var}_R F$ denote the conditional variance of F given the spins outside R, and let g_b denote the smallest positive eigenvalue of H_R. Then

$$\mu_{L,N}\big(\mathrm{Var}_R F\big) \le \frac{\|\varphi - \psi\|^2}{\|\Psi_{\mathrm{VBS}}\|^2} \le \frac{\langle \varphi, H_R\varphi\rangle}{g_b\|\Psi_{\mathrm{VBS}}\|^2}. \tag{2.2}$$

We call the Markov dynamics that resamples each translate of R, at rate b^{-2}, from its conditional law under $\mu_{L,N}$ the *block dynamics*. Up to the factor b^{-2}, the left side of (2.2) is the contribution of R to the Dirichlet form of the block dynamics. Summing (2.2) over the translates of R, we find that for $L, N \ge b + 4$, the gap of $H_{L,N}$ is at least $bg_b/(b + 1)$ times the spectral gap of the block dynamics. This is Theorem 3.5.

In (2.2), the finite quantum system appears only through g_b, and a uniform gap requires only that g_b be positive, which is automatic. However, the size of g_b determines the constant in Theorem 1.1. The characteristic polynomial of $70H_R$ has integer coefficients, and it shows that $g_b \geq [140b(b+1)]^{-5^{b^2+4b}}$, where 5^{b^2+4b} is the dimension of the space on which the bond projections of H_R act nontrivially. The coupling argument below requires very large blocks. This is why the bound on g_b is extremely small, and it is the source of the triple exponential in Theorem 1.1.

2.3 From a window estimate to the gap of the block dynamics

To bound the gap of the block dynamics, we use a standard contraction argument. Suppose that two boundary conditions of a box of side b differ at a single site, and that the two conditional laws in the box can be coupled so that the expected number of sites where they disagree is small compared with b. Then one update of a uniformly chosen block contracts the Lipschitz seminorm with respect to the Hamming distance, and this gives a gap of order one for the block dynamics. After a change of variables that reverses the spins on one sublattice of the box, the conditional laws in a box have the weight $1 + n_x \cdot n_y$ for each bond inside the box and $1 + n_x \cdot \eta_y$ for each bond to a boundary spin η_y. We work in this form from now on.

We build the coupling by following the proof of [35, Proposition 2.1]. Let y be the site where the boundary conditions differ, and let the box have side $b = 2M + 1$. We call the intersections of the box with the boundaries of the squares of radius $\ell, 2\ell, 3\ell, \ldots$ centered at y the *shells*, where $\ell = M^{1/4}$. We try to make the two configurations agree on one shell after another. Once they agree on a shell, the two conditional laws beyond the shell coincide. We then resample the region beyond the shell identically, so that all later disagreements stay within the shell.

The two configurations agree on a shell with probability bounded below by a constant $\delta > 0$, provided two requirements hold. We call the part of a shell at ℓ^∞ distance more than 24 from the complement of the box its *window*. The first requirement is the *window estimate*, which states that the law of the window has a component of mass $p_0 > 0$ that is the same for all boundary conditions. The second requirement is that the configurations on the two ends of the shell, at most 192 sites near the boundary of the box, can be made to agree with probability bounded below. For a finite spin space with a uniform positive lower bound on the conditional weights, the second requirement follows from that lower bound. Here the densities can vanish, and we use an elementary overlap bound (Lemma 4.2), which shows that two conditional laws of K sites overlap in mass at least $(16K)^{-4K}/2$, uniformly in the boundary conditions. With about $M^{1/8}$ trials, the expected number of disagreements is bounded by a constant times $M^{3/4}$ for large M, which is small compared with b.

The window estimate asks only for a common component of positive mass, uniformly in all boundary conditions, and not for a common component of mass close to one. This is what makes it accessible to the arguments of the next two steps, which give a mass p_0 that is positive but extremely small.

2.4 The cut lemma

The window estimate is still a statement about all boundary conditions, and the influence of an arbitrary boundary condition on the square lattice is the difficulty that was left open in [23]. It is natural to change the boundary spins one at a time and to add up their influences on the window. However, this requires the influence of a single boundary spin to be summable over the boundary, which is the kind of decay that we are trying to prove.

We avoid this by changing the region instead of the boundary condition. We call a unit vector attached to a bond that leaves a finite set G a *pin*. A pin τ on the bond from $x \in G$ contributes the factor $1 + n_x \cdot \tau$, and we describe boundary conditions by pins. We say that a site of G is *removed* when its spin and the factors of all its bonds are deleted. Removed sites do not become pins. We call a connected set X of sites that separates the window from the boundary of the box a *cut*. The density of the window spins is the product of their internal bond weights and the partition function on the part of the region outside the window, where the fixed window spins act as additional pins. The normalizing constant is the partition function of the entire region. When the cut is removed, the partition function on the complement factorizes into a part inside the cut, which depends on the window but not on the boundary condition, and a part outside, which depends on the boundary condition but not on the window.

We call the change in the logarithm of the ratio of the partition functions with and without pins, when the cut is removed, the *cut cost*. By the factorization, the density of the window law is, up to normalization, a density independent of the boundary condition times $e^{-\mathrm{cost}_X}$, where X is the cut. A bound on the cut cost by a constant C, uniform in the boundary condition and the window configuration, gives the window estimate with $p_0 = e^{-2C}$ (Lemma 5.4). In particular, the cost need not be small.

Removing a site is a soft modification compared with pinning, because its effect is local and explicit. Let x be a site incident to no pins and with $k \leq 3$ remaining neighbors $u_1, \dots, u_k$. Integrating out the spin at x with the moments of the uniform measure on $\mathbb{S}^2$, we find that

$$Z(\text{with } x) = Z(\text{without } x)\, \mathbb{E}_{\text{without } x}\Big[1 + \frac{1}{3} \sum_{a<a'} n_{u_a} \cdot n_{u_{a'}}\Big], \tag{2.3}$$

both with and without pins, where Z denotes the partition function and $\mathbb{E}$ the corresponding expectation (Lemma 5.3). We call the difference between the correlations $\mathbb{E}[n_{u_a} \cdot n_{u_{a'}}]$

with and without pins the *response* of the pair $(u_a, u_{a'})$. By (2.3), the pins change the cost of removing x only through the responses of the pairs of neighbors of x. Each such pair is at distance 2 and has the removed site x as a common neighbor.

We remove the sites of the cut one at a time, each of them, except the first, next to a site removed earlier. Then every site except the first has at most three remaining neighbors when it is removed, and the cost of its removal is bounded by responses. The first removal costs at most a constant. It remains to show that the responses decay exponentially in the distance to the pins, uniformly in the pins and in the set of removed sites, which is Hypothesis (P) of Section 5. Given this decay, a cut has bounded cost if it keeps a fixed distance from the pins near the ends of the window and moves away from them linearly elsewhere (see Figure 5.1).

2.5 Responses and walk expansions

Recall that a response is the difference between a correlation with pins and the same correlation without pins. We call a walk on the lattice that uses each bond at most once a *trail*. To prove the decay of responses, we expand a correlation $\mathbb{E}[n_u \cdot n_v]$ into trails. The expansion rests on the identity

$$\int (e \cdot n) \prod_{i=1}^{k} (1 + n \cdot w_i) \, \mathrm{d}n = \frac{1}{3} \sum_{i=1}^{k} (e \cdot w_i) \int \prod_{i' \neq i} (1 + n \cdot w_{i'}) \Big(\frac{3}{5} \, \mathrm{d}n + \frac{2}{5} \delta_0 \Big), \tag{2.4}$$

valid for vectors $e, w_1, \dots, w_k \in \mathbb{R}^3$ with $k \leq 4$, where δ_0 denotes the point mass at $0 \in \mathbb{R}^3$; it follows by comparing the moments of both sides. Consider the correlation of $e \cdot n_u$ with a function of the other spins. Integrating out n_u with (2.4) replaces $e \cdot n_u$ by $(e \cdot n_{u'})/3$, summed over the neighbors u' of u. It also uses up the bond $\{u, u'\}$ and gives u the modified measure in parentheses, whose moments of positive degree are $3/5$ times those of $\mathrm{d}n$. Repeating the step at u' and so on, we obtain a sum over trails from u that stop at v or in the boundary layer $\partial_2 G$. A trail of length m carries the weight 3^{-m}, an extra factor $3/5$ for each site it leaves for the second time, and the ratio of the partition functions after and before the trail is traced. On bipartite tori and with Néel boundary conditions, the expansions in [23] have nonnegative terms. The authors bound the two-point function on the torus and the boundary-induced staggered magnetization by weighted numbers of bond-self-avoiding walks, obtaining exponential decay.

With arbitrary pins, the terms have no sign, and the ratio of partition functions along a trail can exceed one. To overcome this obstacle, we compare the expansions with and without pins trail by trail. They have the same trails and the same weights $3^{-m}(3/5)^r$, where r is the number of sites that the trail leaves for the second time. The ratios of partition functions can differ, as can the terminal correlations of trails that end in the boundary layer. Without pins, every term is nonnegative and every ratio is bounded by

one. At a step that leaves a site for the second time, the ratio equals one with and without pins. At a step that leaves a site for the first time, we show that the ratio with pins is bounded by $1/0.961$ when the trail goes straight through the site and by $1/0.935$ when it turns. These bounds are the local inequalities of Section 2.6. Given them, the trails that end in $\partial_2 G$ contribute an amount that is exponentially small in the distance to $\partial_2 G$, provided the weighted number of trails of length m, with the step factors above, decays exponentially in m. We prove this decay with the rate $2499/2500$, which leaves very little room.

Next, we treat the trails that return to v. We write the difference of the two products of ratios along such a trail as a sum over its steps. At one step, the difference of the logarithms of the two ratios is a combination, with coefficients at most $1/6$ and $9/25$, of the responses of the pairs of remaining neighbors of the site that the trail leaves, computed in the region where that site is removed (Lemma 6.8). These pairs are again at distance 2 with a removed common neighbor, and their responses are of the same kind as the one we started from. For an integer d, let $r(d)$ denote the largest absolute value of the response of a pair of this kind at distance at least d from $\partial_2 G$. We obtain a linear inequality of the form

$$r(d) \le \sum_{w} (\text{weight of } w)\, r\big(d - |w|\big) + C\gamma^d, \tag{2.5}$$

where the sum is over the trails w from u that return to v, $|w|$ denotes the length of w, $\gamma < 1$ is the rate of decay of the weighted number of trails, and C is a constant.

There is an apparent circularity here, since we use responses to bound responses. The way out is to work in a finite set G and to consider the supremum of all responses in G, weighted by θ^{-d} for a rate $\theta < 1$ and by a weight for each of the two shapes of pairs. (We call a pair *collinear* if its two sites and their common neighbor lie on a line, and *diagonal* otherwise.) The supremum is over all regions obtained from G by removing sites and tracing trails, and over all pins. Since G is finite, this supremum is finite. Removed sites never become pins, so the distance to the boundary of G is the same for all these regions. The inequality (2.5) then shows that the supremum is at most a constant K_θ times itself plus a constant independent of G. Therefore the supremum is bounded uniformly in G as soon as $K_\theta < 1$.

The constant K_θ is the larger of two weighted sums, one for each shape of pairs, over the trails from u to v in $\mathbb{Z}^2$ that avoid the common neighbor of u and v, in which each trail carries its weight from the expansion, the response coefficients at its first visits, and the step factors of the local inequalities. By continuity in θ, it suffices to prove $K_1 < 1$. With the weights 1 and $5/4$ for collinear and diagonal pairs, we prove that the sums for the two shapes are at most 0.9232 and 0.8994. The margin is small, and it requires the exact geometry of the trails that return. Our bound for the weighted number of all trails

of length m is $10 \cdot (2499/2500)^m$, whose sum over m is 25,000. This bound on all trails is too large to prove $K_1 < 1$.

2.6 The computer-assisted inputs

Three statements are verified by computer. The second and the third use the first only through the two step factors $1/0.961$ and $1/0.935$, and the third also uses the response coefficients $1/6$ and $9/25$.

The first is the family of local inequalities. When a trail leaves a site x for the first time, we bound the ratio of the pinned partition functions after and before the step by an inequality on the sites at distance at most 2 from x. The local pattern records the site statuses and the used bonds at these sites. After the spin at x and the spins at the four sites at distance 2 along the axes are integrated out, the rest of the system appears only through at most twelve vectors of norm at most one. This leaves 7,962 local configurations, after the reflection and identifications of Appendix B.4, and for each of them a polynomial inequality, affine in each of at most eleven unit vectors. We certify each inequality by an explicit Hermitian operator on $(\mathbb{C}^2)^{\otimes k}$ whose expectations in products of spin-$\frac{1}{2}$ coherent states are the values of the polynomial. After the signs of a suitable subset of the vectors are reversed, positivity of this operator implies the inequality. The operator commutes with simultaneous rotations, so in each exchange-symmetry sector it suffices to test one eigenspace of the total spin component. Each test is the positive definiteness of a symmetric integer matrix of size at most 245, which we verify by fraction-free Gaussian elimination [6] in exact arithmetic.

The second is the exponential decay of the weighted number of all trails. We verify it with a positive integer vector v on the 916 possible sequences of the last six steps of a trail. The weighted transfer matrix A_6 satisfies $\mathsf{A}_6\mathsf{v} \le \gamma_{\mathrm{o}}\mathsf{v}$ entrywise, as we check in exact integer arithmetic. The third is the bound $K_1 < 1$. We enumerate the trails of length at most 20 exactly. For a longer trail, we follow the first 20 steps exactly and dominate the continuation by a weighted walk on $\mathbb{Z}^2$ that remembers its position and its last 10 steps. We bound the total weight of its paths to v by a function of its states that satisfies a one-step inequality, and we check this inequality state by state in floating-point arithmetic with directed rounding. The programs and their outputs are provided as ancillary files, which we describe in Appendix E.

2.7 Organization of the paper

Section 3 recalls the polynomial representation and the VBS state and proves Theorem 3.5, the comparison between the quantum gap and the gap of the block dynamics. In Section 4, we prove that the window estimate implies a gap for the block dynamics, through the

coupling argument of [35]. Section 5 introduces the model with pins and removed sites, proves the cut lemma, and deduces the window estimate from the decay of responses. Section 6 sets up the walk expansion and states the local inequalities and the decay of weighted trail counts as hypotheses, and Section 7 proves the decay of responses from these hypotheses and the bound on returning trails. Section 8 verifies the three hypotheses with computer assistance and completes the proof of Theorem 1.1. Appendix A proves that the cut is connected, Appendices B and C contain the certificates for the local inequalities and for the returning trails, Appendix D tracks the constants, Appendix E describes the ancillary files, and Appendix F proves uniqueness of the infinite-volume frustration-free state.

SECTION 3

Coherent states and a classical block dynamics

In this section, we compare the spectral gap of $H_{L,N}$ with the spectral gap of a classical Markov dynamics, the block dynamics of a probability measure $\mu_{L,N}$ on $(\mathbb{S}^2)^\Lambda$. The comparison is Theorem 3.5. In the proof of Theorem 1.1, it is combined with the lower bound on the gap of the block dynamics proved in Section 4 and with the explicit lower bound on the constant g_b of Theorem 3.5 proved in Lemma D.3. We also identify the kernel of $H_{L,N}$ in Corollary 3.3. The heuristics are explained in Section 2.2. The polynomial representation of spin 2, the VBS state, and the classical measure that the VBS state defines through spin coherent states go back to [3, 5, 23]; we recall them in Sections 3.1 and 3.3 in the form that we use. In Section 3.5, we introduce the measures on finite subsets of $\mathbb{Z}^2$ that are used in the rest of the paper, and we relate them to the conditional laws of $\mu_{L,N}$.

Throughout this section, $L, N \geq 3$ are integers, and $\Lambda = \Lambda_{L,N}$ and $\mathcal{E}$ are as in Section 1.2. Since $L, N \geq 3$, every site of Λ belongs to exactly four bonds of $\mathcal{E}$. For $x \in \Lambda$ and $u \in \mathbb{Z}^2$, we write $x + u$ for the sum in Λ, in which u is reduced modulo (L, N). For $R \subset \Lambda$, we let $\mathcal{E}(R)$ denote the set of bonds of $\mathcal{E}$ that meet R. Given a configuration $s = (s_x)_{x\in\Lambda}$ and a set $D \subset \Lambda$, we let $s_D = (s_x)_{x\in D}$ denote the restriction of s to D, and we use the same notation for configurations on subsets of $\mathbb{Z}^2$.

3.1 The polynomial representation

We begin with spin-$\frac{1}{2}$ coherent states. Let $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ denote the vector of Pauli matrices. For $a = (a_1, a_2, a_3) \in \mathbb{S}^2$, we set

$$\zeta(a) = \frac{1}{\sqrt{2(1+a_3)}}\big(1 + a_3, a_1 + \mathrm{i}a_2\big) \in \mathbb{C}^2 \quad \text{if } a_3 \neq -1, \qquad \zeta\big((0,0,-1)\big) = (0,1).$$

Then $\zeta\colon \mathbb{S}^2 \to \mathbb{C}^2$ is measurable, $\zeta(a)$ is a unit vector, and a direct computation shows that

$$\zeta(a)\zeta(a)^* = \frac{1}{2}(\mathrm{Id} + a \cdot \sigma) \qquad \text{for all } a \in \mathbb{S}^2. \tag{3.1}$$

Every orthogonal projection of rank one on $\mathbb{C}^2$ has the form $\frac{1}{2}(\mathrm{Id} + a \cdot \sigma)$ for exactly one $a \in \mathbb{S}^2$. Since such a projection determines the unit vectors in its range up to a factor of modulus one, every unit vector of $\mathbb{C}^2$ has the form $e^{\mathrm{i}\alpha}\zeta(a)$ with $a \in \mathbb{S}^2$ and $\alpha \in \mathbb{R}$. For

$g \in \mathrm{SU}(2)$, we let $\bar{g} \in \mathrm{SO}(3)$ denote the rotation defined by $g(a \cdot \sigma)g^* = (\bar{g}a) \cdot \sigma$ for all $a \in \mathbb{R}^3$. By (3.1), the unit vectors $g\zeta(a)$ and $\zeta(\bar{g}a)$ have the same projection $\frac{1}{2}(\mathrm{Id} + (\bar{g}a) \cdot \sigma)$, and $g\zeta(a) = e^{i\alpha}\zeta(\bar{g}a)$ for some $\alpha \in \mathbb{R}$.

We now realize the spin-2 representation. Let V denote the space of homogeneous polynomials of degree 4 in $z = (z_1, z_2) \in \mathbb{C}^2$, with the inner product

$$\langle \varphi, \psi \rangle_V = 5 \int_{\mathbb{S}^2} \overline{\varphi(\zeta(a))}\, \psi(\zeta(a))\, \mathrm{d}a. \tag{3.2}$$

Since φ and ψ are homogeneous of the same degree, the integrand is a polynomial in the entries of $\zeta(a)\zeta(a)^*$. In particular, it is continuous in a and unchanged when $\zeta(a)$ is multiplied by a complex number of modulus one. The group $\mathrm{SU}(2)$ acts on V by $(U_g\varphi)(z) = \varphi(g^{-1}z)$. Since $g^{-1}\zeta(a)$ is a multiple of $\zeta(\bar{g}^{-1}a)$ by a complex number of modulus one and $\mathrm{d}a$ is invariant under rotations, the inner product (3.2) is invariant under this action. To see that it is positive definite, suppose that $\langle \varphi, \varphi \rangle_V = 0$. Then $\varphi(\zeta(a)) = 0$ for all a by continuity, and φ vanishes on the unit sphere of $\mathbb{C}^2$, since every unit vector has the form $e^{i\alpha}\zeta(a)$. By homogeneity, $\varphi = 0$.

Recall that the action of $\mathrm{SU}(2)$ on the homogeneous polynomials of degree $2j$ in two variables is the irreducible representation of spin j. We use the realization V of spin 2 throughout, and we identify $\mathcal{H}_\Lambda = \bigotimes_{x\in\Lambda} V_x$ with the space of polynomials in $z_\Lambda = (z_x)_{x\in\Lambda}$, where $z_x = (z_{x,1}, z_{x,2}) \in \mathbb{C}^2$, that are homogeneous of degree 4 in each z_x. Here the product vector $\bigotimes_x \varphi_x$ corresponds to the polynomial $\prod_x \varphi_x(z_x)$, and the inner product of $\mathcal{H}_\Lambda$ is the product of the inner products (3.2).

Recall from Section 1.2 that, for a bond $e = \{x, y\}$, P_e is the orthogonal projection of $V_x \otimes V_y$ onto its subspace of total spin 4, extended by the identity. In our realization, $\mathrm{SU}(2)$ acts on $V_x \otimes V_y$ by $U_g \otimes U_g$. By the Clebsch–Gordan rule, $V_x \otimes V_y$ is the orthogonal direct sum of irreducible subspaces of spins $0, 1, 2, 3, 4$, each of which occurs once, and P_e projects onto the subspace of spin 4. For $R \subset \Lambda$, we set

$$H_R = \sum_{e\in\mathcal{E}(R)} P_e, \tag{3.3}$$

and we let Π_R denote the orthogonal projection of $\mathcal{H}_\Lambda$ onto $\ker H_R$. Note that $H_\Lambda = H_{L,N}$.

We fix an orientation (x, y) of every bond $e = \{x, y\}$, and we define the *singlet polynomial* of e by

$$q_e = z_{x,1}z_{y,2} - z_{x,2}z_{y,1}.$$

(The other orientation gives $-q_e$.) Since $\det g = 1$ for $g \in \mathrm{SU}(2)$, we have

$$q_e(gz_x, gz_y) = q_e(z_x, z_y) \qquad \text{for all } g \in \mathrm{SU}(2). \tag{3.4}$$

For a set $\mathcal{F}$ of bonds, we set $q_\mathcal{F} = \prod_{e\in\mathcal{F}} q_e$, and we call

$$\Psi_{\mathrm{VBS}} = q_\mathcal{E} = \prod_{e\in\mathcal{E}} q_e \tag{3.5}$$

the *VBS state*. Since every site belongs to four bonds, Ψ_{VBS} is homogeneous of degree 4 in each z_x, so that $\Psi_{\mathrm{VBS}} \in \mathcal{H}_\Lambda$. Further, $\Psi_{\mathrm{VBS}} \neq 0$, since the polynomial ring $\mathbb{C}[z_\Lambda]$ is an integral domain.

3.2 The kernel of the Hamiltonian

The description of the kernel of $H_{L,N}$ in Corollary 3.3 is standard (see [23, 25]). We include the short proofs of the two lemmas that give it, because we will also use Lemma 3.2 with R a block in the proof of Theorem 3.5.

Lemma 3.1. *Let $e = \{x, y\} \in \mathcal{E}$ and $\varphi \in \mathcal{H}_\Lambda$. Then $P_e\varphi = 0$ if and only if $\varphi = q_e\chi$ for a polynomial χ in z_Λ that is homogeneous of degree 3 in z_x and in z_y and of degree 4 in z_u for every $u \in \Lambda \setminus \{x, y\}$.*

Proof. Let W_4 and W_3 denote the spaces of polynomials in (z_x, z_y) that are homogeneous of degree 4, respectively 3, in each of z_x and z_y, so that $W_4 = V_x \otimes V_y$. The group $\mathrm{SU}(2)$ acts on W_4 and on W_3 by $\chi \mapsto \chi(g^{-1}z_x, g^{-1}z_y)$. We have

$$\mathcal{H}_\Lambda = W_4 \otimes \mathcal{H}',$$

where $\mathcal{H}'$ is the space of polynomials in $(z_u)_{u \neq x,y}$ that are homogeneous of degree 4 in each z_u, and P_e acts on the first factor. The polynomials χ in the statement are the elements of $W_3 \otimes \mathcal{H}'$. It therefore suffices to show that

$$\{\chi \in W_4 : P_e\chi = 0\} = q_e W_3.$$

Multiplication by q_e maps W_3 into W_4 injectively, since $\mathbb{C}[z_x, z_y]$ is an integral domain. The invariance (3.4) of q_e shows that this map commutes with the action of $\mathrm{SU}(2)$. The space of homogeneous polynomials of degree 3 in $z \in \mathbb{C}^2$ is the irreducible representation of spin $\frac{3}{2}$. By the Clebsch–Gordan rule, W_3 is the direct sum of irreducible subspaces of spins $0, 1, 2, 3$, and W_4 is the direct sum of irreducible subspaces of spins $0, \ldots, 4$, each occurring once.

We conclude that $q_e W_3$ is an invariant subspace of W_4 that is a sum of irreducible subspaces of spins $0, 1, 2, 3$. Since each spin occurs only once in W_4, the subspace $q_e W_3$ coincides with the sum of the components of W_4 of spins $0, 1, 2, 3$. This sum is the orthogonal complement of the component of spin 4, which is the kernel of P_e in W_4. □

Lemma 3.2. *Let $R \subset \Lambda$, and let $\varphi \in \mathcal{H}_\Lambda$ satisfy $P_e\varphi = 0$ for all $e \in \mathcal{E}(R)$. Then $\varphi = q_{\mathcal{E}(R)}\chi$ for a polynomial χ in the variables $(z_u)_{u \in \Lambda \setminus R}$.*

Proof. We may assume that $\varphi \neq 0$. We first show that each q_e is irreducible in $\mathbb{C}[z_\Lambda]$. A nontrivial factorization of the homogeneous quadratic polynomial q_e would write it as

a product of two linear forms. Such a product is a quadratic form of rank at most 2, whereas q_e has rank 4 as a quadratic form in $z_{x,1}, z_{x,2}, z_{y,1}, z_{y,2}$. If $e \neq e'$, then q_e and $q_{e'}$ involve different sets of variables and are not scalar multiples of each other.

By Lemma 3.1, q_e divides φ for every $e \in \mathcal{E}(R)$. In the unique factorization domain $\mathbb{C}[z_\Lambda]$, it follows that the product $q_{\mathcal{E}(R)}$ of these pairwise non-associated irreducible polynomials divides φ. Let χ denote the quotient.

It remains to show that χ has degree 0 in z_x for all $x \in R$. Let $x \in R$, and let $\deg_{z_x}$ denote the degree in the pair of variables $z_x = (z_{x,1}, z_{x,2})$, which is additive under multiplication in $\mathbb{C}[z_\Lambda]$. The four bonds that contain x belong to $\mathcal{E}(R)$, and each of them contributes a factor of degree 1 in z_x to $q_{\mathcal{E}(R)}$, while the other factors have degree 0 in z_x. Since φ has degree 4 in z_x, we obtain

$$4 = \deg_{z_x} \varphi = \deg_{z_x} q_{\mathcal{E}(R)} + \deg_{z_x} \chi = 4 + \deg_{z_x} \chi,$$

which gives $\deg_{z_x} \chi = 0$. □

Corollary 3.3. *For all integers $L, N \geq 3$, we have* $\ker H_{L,N} = \mathbb{C}\Psi_{\mathrm{VBS}}$.

Proof. Let $e = \{x, y\} \in \mathcal{E}$. We have $\Psi_{\mathrm{VBS}} = q_e\, q_{\mathcal{E}\setminus\{e\}}$, where $q_{\mathcal{E}\setminus\{e\}}$ is homogeneous of degree 3 in z_x and in z_y and of degree 4 in every other z_u. Then Lemma 3.1 implies that $P_e \Psi_{\mathrm{VBS}} = 0$. As e is arbitrary, $H_{L,N}\Psi_{\mathrm{VBS}} = 0$.

Conversely, let $\varphi \in \ker H_{L,N}$. Then

$$\sum_{e\in\mathcal{E}} \|P_e\varphi\|^2 = \langle \varphi, H_{L,N}\varphi\rangle = 0,$$

so that $P_e\varphi = 0$ for all $e \in \mathcal{E}$. Applying Lemma 3.2 with $R = \Lambda$, for which $\mathcal{E}(R) = \mathcal{E}$, we find that $\varphi = q_{\mathcal{E}}\chi = \chi\Psi_{\mathrm{VBS}}$ for a constant χ. □

3.3 The coherent-state measure

For a polynomial p in z_Λ and a configuration $s \in (\mathbb{S}^2)^\Lambda$, we write $p(\zeta(s)) = p\big((\zeta(s_x))_{x\in\Lambda}\big)$. We define

$$\iota\colon \mathcal{H}_\Lambda \to L^2\Big((\mathbb{S}^2)^\Lambda, \prod_{x\in\Lambda} 5\,\mathrm{d}s_x\Big), \qquad (\iota\varphi)(s) = \varphi(\zeta(s)).$$

By (3.2) and the definition of the inner product of $\mathcal{H}_\Lambda$, the identity $\langle \iota\varphi, \iota\psi\rangle = \langle\varphi,\psi\rangle$ holds for product vectors φ and ψ. Since both sides are sesquilinear and the product vectors span $\mathcal{H}_\Lambda$, it holds for all $\varphi, \psi \in \mathcal{H}_\Lambda$, and ι is an isometry.

We next compute $|\iota\Psi_{\mathrm{VBS}}|^2$. Let $a, a' \in \mathbb{S}^2$. The number $\zeta(a)_1\zeta(a')_2 - \zeta(a)_2\zeta(a')_1$ is the determinant of the 2×2 matrix with columns $\zeta(a)$ and $\zeta(a')$. Its squared modulus is the

Gram determinant $1 - |\langle \zeta(a'), \zeta(a)\rangle|^2$ of these unit vectors. By (3.1) and the identities $\operatorname{Tr}\sigma_j = 0$ and $\operatorname{Tr}(\sigma_j\sigma_k) = 2\delta_{jk}$, we have

$$|\langle \zeta(a'), \zeta(a)\rangle|^2 = \operatorname{Tr}\big(\zeta(a)\zeta(a)^*\zeta(a')\zeta(a')^*\big) = \frac{1}{4}\operatorname{Tr}\big((\mathrm{Id} + a\cdot\sigma)(\mathrm{Id} + a'\cdot\sigma)\big) = \frac{1 + a\cdot a'}{2}.$$

Taking $a = s_x$ and $a' = s_y$, we find that every bond $e = \{x, y\}$ satisfies

$$|q_e(\zeta(s))|^2 = \frac{1 - s_x \cdot s_y}{2}. \tag{3.6}$$

We define the probability measure $\mu_{L,N}$ on $(\mathrm{S}^2)^\Lambda$ by

$$\mu_{L,N}(\,\mathrm{d}s) = \frac{1}{Z_{L,N}} \prod_{\{x,y\}\in\mathcal{E}} (1 - s_x\cdot s_y) \prod_{x\in\Lambda} \mathrm{d}s_x, \qquad Z_{L,N} = \int \prod_{\{x,y\}\in\mathcal{E}} (1 - s_x\cdot s_y) \prod_{x\in\Lambda} \mathrm{d}s_x. \tag{3.7}$$

By (3.5) and (3.6), we have

$$\mu_{L,N}(\,\mathrm{d}s) = \frac{|\iota\Psi_{\mathrm{VBS}}(s)|^2}{\|\Psi_{\mathrm{VBS}}\|^2} \prod_{x\in\Lambda} 5\,\mathrm{d}s_x, \qquad Z_{L,N} = 2^{|\mathcal{E}|}\, 5^{-|\Lambda|}\, \|\Psi_{\mathrm{VBS}}\|^2 > 0.$$

The density in (3.7) is positive outside the set of configurations with $s_x = s_y$ for some bond $\{x, y\}$, which has measure zero. In particular, $\mu_{L,N}$ and the product measure $\prod_x \mathrm{d}s_x$ have the same null sets, and $\iota\Psi_{\mathrm{VBS}} \neq 0$ almost everywhere. We let $L^2(\mu_{L,N})$ denote the space of complex-valued functions that are square integrable with respect to $\mu_{L,N}$.

Lemma 3.4. *Let $\varphi \in \mathcal{H}_\Lambda$, and let $F = \iota\varphi/\iota\Psi_{\mathrm{VBS}}$, which is defined $\mu_{L,N}$-almost everywhere. Then $F \in L^2(\mu_{L,N})$, and*

$$\mu_{L,N}(F) = \frac{\langle \Psi_{\mathrm{VBS}}, \varphi\rangle}{\|\Psi_{\mathrm{VBS}}\|^2}, \qquad \mu_{L,N}(|F|^2) = \frac{\|\varphi\|^2}{\|\Psi_{\mathrm{VBS}}\|^2}.$$

In particular, if φ is orthogonal to Ψ_{VBS}, then

$$\operatorname{Var}_{\mu_{L,N}}(F) = \frac{\|\varphi\|^2}{\|\Psi_{\mathrm{VBS}}\|^2}. \tag{3.8}$$

Proof. By the formula for $\mu_{L,N}$ in terms of $\iota\Psi_{\mathrm{VBS}}$, we have

$$\mu_{L,N}(|F|^2) = \frac{1}{\|\Psi_{\mathrm{VBS}}\|^2} \int |\iota\varphi(s)|^2 \prod_{x\in\Lambda} 5\,\mathrm{d}s_x = \frac{\|\varphi\|^2}{\|\Psi_{\mathrm{VBS}}\|^2},$$

where we used that ι is an isometry. The formula for $\mu_{L,N}(F)$ follows in the same way from the identity $F|\iota\Psi_{\mathrm{VBS}}|^2 = \overline{\iota\Psi_{\mathrm{VBS}}}\,\iota\varphi$, which holds almost everywhere. □

3.4 The block dynamics and the comparison theorem

We now define the heat-bath kernels. Let $R \subset \Lambda$ and $s \in (\mathbb{S}^2)^\Lambda$. We let $\mathcal{Q}_R(s, \cdot)$ denote the probability measure on $(\mathbb{S}^2)^\Lambda$ under which the spins off R are equal to $s_{\Lambda\setminus R}$, and the spins $\tilde{s}_R$ in R have the density proportional to

$$\prod_{\{x,y\}\in\mathcal{E}(R)} (1 - \tilde{s}_x \cdot \tilde{s}_y), \qquad \text{where } \tilde{s}_u = s_u \text{ for } u \in \Lambda \setminus R, \tag{3.9}$$

with respect to $\prod_{x\in R} \mathrm{d}\tilde{s}_x$. As a function of $\tilde{s}_R$, the product (3.9) is continuous and nonnegative, with zeros only on a set of measure zero (where $\tilde{s}_x = \tilde{s}_y$ for some bond $\{x, y\} \in \mathcal{E}(R)$). In particular, its integral over $(\mathbb{S}^2)^R$ is positive for every s. We set

$$(\mathcal{Q}_R F)(s) = \int F(\tilde{s})\, \mathcal{Q}_R(s, \mathrm{d}\tilde{s})$$

for every bounded or nonnegative measurable function F on $(\mathbb{S}^2)^\Lambda$. The density in (3.7) is the product of the factors of the bonds in $\mathcal{E}(R)$, which form (3.9), and of factors that do not involve s_R. By this factorization and Fubini's theorem, $\mathcal{Q}_R F$ is a version of the conditional expectation of F under $\mu_{L,N}$ given $s_{\Lambda\setminus R}$. In particular, $\mathcal{Q}_R$ extends to the orthogonal projection of $L^2(\mu_{L,N})$ onto the closed subspace of functions of $s_{\Lambda\setminus R}$. We call $\mathcal{Q}_R$ the *heat-bath kernel* of R. For $F \in L^2(\mu_{L,N})$, we set $\mathrm{Var}_R F = \mathcal{Q}_R|F - \mathcal{Q}_R F|^2$, so that

$$\mu_{L,N}(\mathrm{Var}_R F) = \|F - \mathcal{Q}_R F\|^2_{L^2(\mu_{L,N})}. \tag{3.10}$$

Let $b \geq 1$ be an integer such that $L, N \geq b$. For $x \in \Lambda$, we call

$$B_x = x + \{1, \dots, b\}^2 = \{x + u : u \in \{1, \dots, b\}^2\}$$

the *block* with corner x; it has b^2 sites. We call the Markov dynamics on $(\mathbb{S}^2)^\Lambda$ in which, for every $x \in \Lambda$, the block B_x is resampled from $\mathcal{Q}_{B_x}$ at rate b^{-2} the *block dynamics*. The generator of the block dynamics is

$$\mathcal{L}_{\mathrm{bl}} = \frac{1}{b^2} \sum_{x\in\Lambda} (\mathcal{Q}_{B_x} - \mathrm{Id}). \tag{3.11}$$

The operator $\mathcal{L}_{\mathrm{bl}}$ is bounded, self-adjoint, and nonpositive on $L^2(\mu_{L,N})$ and annihilates the constant functions. By (3.10), its Dirichlet form is

$$\langle F, -\mathcal{L}_{\mathrm{bl}} F\rangle_{L^2(\mu_{L,N})} = \frac{1}{b^2} \sum_{x\in\Lambda} \mu_{L,N}(\mathrm{Var}_{B_x} F), \qquad F \in L^2(\mu_{L,N}).$$

We define its spectral gap $\mathrm{gap}(\mathcal{L}_{\mathrm{bl}})$ as in Section 1.4.

Next, we introduce the finite quantum system that enters the comparison. We call the set of sites $u \in \mathbb{Z}^2$ with $\mathrm{dist}(u, \{1, \dots, b\}^2) \leq 1$ the *patch*. It has $b^2 + 4b$ sites, and

$2b^2 + 2b$ bonds of $\mathbb{Z}^2$ meet $\{1, \dots, b\}^2$. We let h_b denote the sum of the projections P_e over the bonds e of $\mathbb{Z}^2$ that meet $\{1, \dots, b\}^2$, acting on the tensor product of copies of V over the sites of the patch, where P_e is defined as in Section 3.1. We call h_b the *patch operator*. Since h_b is a nonzero nonnegative operator on a finite-dimensional space, it has a smallest positive eigenvalue, which we denote by g_b.

Now let $L, N \geq b + 4$ and $x \in \Lambda$. The map $u \mapsto x + u$ is injective on the patch, since $L, N \geq b + 2$. It maps $\{1, \dots, b\}^2$ onto B_x and the bonds of $\mathbb{Z}^2$ that meet $\{1, \dots, b\}^2$ bijectively onto $\mathcal{E}(B_x)$, since every bond of $\mathcal{E}(B_x)$ has the form $\{x + u, x + v\}$ with $u \in \{1, \dots, b\}^2$ and $v \sim u$. We identify the patch with its image under this map. Then $\mathcal{H}_\Lambda$ is the tensor product of the space of the patch and the space of the other sites of Λ, and $H_{B_x} = h_b \otimes \mathrm{Id}$. The operator H_{B_x} commutes with Π_{B_x}, and the eigenvalues of H_{B_x} on the range of $\mathrm{Id} - \Pi_{B_x}$ are positive eigenvalues of h_b. This implies that

$$H_{B_x} \geq g_b(\mathrm{Id} - \Pi_{B_x}) \qquad \text{for all } x \in \Lambda. \tag{3.12}$$

An explicit lower bound for g_b is proved in Lemma D.3. We are now ready to state the main result of this section.

Theorem 3.5. *Let $b \geq 1$ be an integer, and let $L, N \geq b + 4$. Then*

$$\mathrm{gap}(H_{L,N}) \geq \frac{b}{b+1}\, g_b\, \mathrm{gap}(\mathcal{L}_{\mathrm{bl}}).$$

Proof. Let $\varphi \in \mathcal{H}_\Lambda$ be nonzero and orthogonal to Ψ_{VBS}. We set $F = \iota\varphi/\iota\Psi_{\mathrm{VBS}}$, and in this proof we write $\mu = \mu_{L,N}$ and $\|\cdot\|$ for the norm of $L^2(\mu)$. Recall that $\mathrm{gap}(H_{L,N})$ is the smallest nonzero eigenvalue of $H_{L,N}$. Using Corollary 3.3, we see that it is the infimum of $\langle \varphi, H_{L,N}\varphi\rangle / \|\varphi\|^2_{\mathcal{H}_\Lambda}$ over all such φ. To prove the theorem, it suffices to show that

$$\mathrm{gap}(\mathcal{L}_{\mathrm{bl}})\, \|\varphi\|^2_{\mathcal{H}_\Lambda} \leq \frac{b+1}{b\, g_b} \langle \varphi, H_{L,N}\varphi\rangle. \tag{3.13}$$

We begin with a single block, which is the main step. Fix $x \in \Lambda$, write $R = B_x$, and set $\psi_R = \Pi_R \varphi$. Then $H_R\psi_R = 0$, and $P_e\psi_R = 0$ for all $e \in \mathcal{E}(R)$, as in the proof of Corollary 3.3. Applying Lemma 3.2, we find that $\psi_R = q_{\mathcal{E}(R)}\chi_R$ for a polynomial χ_R in the variables $(z_u)_{u \in \Lambda\setminus R}$. Observe that $\Psi_{\mathrm{VBS}} = q_{\mathcal{E}(R)}\, q_{\mathcal{E}\setminus\mathcal{E}(R)}$, where $q_{\mathcal{E}\setminus\mathcal{E}(R)}$ is also a polynomial in these variables, since its bonds do not meet R. Dividing, we obtain

$$\frac{\iota\psi_R(s)}{\iota\Psi_{\mathrm{VBS}}(s)} = \frac{\chi_R(\zeta(s))}{q_{\mathcal{E}\setminus\mathcal{E}(R)}(\zeta(s))} \tag{3.14}$$

for all s with $\iota\Psi_{\mathrm{VBS}}(s) \neq 0$, and hence for almost every s. The right side of (3.14) is a function of $s_{\Lambda\setminus R}$. It belongs to $L^2(\mu)$ by Lemma 3.4 applied to ψ_R.

Recall that $\mathcal{Q}_R$ is the orthogonal projection onto the functions of $s_{\Lambda\setminus R}$ in $L^2(\mu)$. Using this fact together with (3.10), Lemma 3.4 applied to $\varphi - \psi_R$, and (3.12), we obtain

$$\begin{aligned}\mu(\mathrm{Var}_R\, F) = \|F - \mathcal{Q}_R F\|^2 \le \Big\|F - \frac{\iota\psi_R}{\iota\Psi_{\mathrm{VBS}}}\Big\|^2 = \frac{\|\varphi - \psi_R\|^2_{\mathcal{H}_\Lambda}}{\|\Psi_{\mathrm{VBS}}\|^2_{\mathcal{H}_\Lambda}} \\ = \frac{\langle\varphi, (\mathrm{Id} - \Pi_R)\varphi\rangle}{\|\Psi_{\mathrm{VBS}}\|^2_{\mathcal{H}_\Lambda}} \le \frac{\langle\varphi, H_R\varphi\rangle}{g_b\,\|\Psi_{\mathrm{VBS}}\|^2_{\mathcal{H}_\Lambda}}.\end{aligned} \tag{3.15}$$

Next, we sum (3.15) over $x \in \Lambda$. We claim that every bond $e \in \mathcal{E}$ meets exactly $b^2 + b$ blocks. We may write $e = \{u, u + (1,0)\}$ with $u \in \Lambda$ by symmetry. Using $L, N \ge b + 1$, we see that $u \in B_x$ for exactly b^2 sites x and $u + (1,0) \in B_x$ for exactly b^2 sites x. The two sites of e lie in B_x for exactly $b(b-1)$ sites x, namely those for which $u = x + v$ with $v \in \{1, \dots, b-1\} \times \{1, \dots, b\}$. By inclusion and exclusion, the number of blocks that meet e is

$$2b^2 - b(b-1) = b^2 + b,$$

which proves the claim. Together with (3.3), the claim implies that

$$\sum_{x\in\Lambda} H_{B_x} = (b^2 + b) H_{L,N}.$$

Summing (3.15) over x, dividing by b^2, and using this identity, we find that

$$\langle F, -\mathcal{L}_{\mathrm{bl}} F\rangle_{L^2(\mu)} = \frac{1}{b^2}\sum_{x\in\Lambda} \mu(\mathrm{Var}_{B_x} F) \le \frac{b+1}{b\, g_b}\,\frac{\langle\varphi, H_{L,N}\varphi\rangle}{\|\Psi_{\mathrm{VBS}}\|^2_{\mathcal{H}_\Lambda}}. \tag{3.16}$$

Finally, by (3.8), the function F is not constant, and

$$\mathrm{gap}(\mathcal{L}_{\mathrm{bl}})\,\frac{\|\varphi\|^2_{\mathcal{H}_\Lambda}}{\|\Psi_{\mathrm{VBS}}\|^2_{\mathcal{H}_\Lambda}} = \mathrm{gap}(\mathcal{L}_{\mathrm{bl}})\,\mathrm{Var}_\mu(F) \le \langle F, -\mathcal{L}_{\mathrm{bl}} F\rangle_{L^2(\mu)},$$

where the inequality follows from the definition of $\mathrm{gap}(\mathcal{L}_{\mathrm{bl}})$. Combining this bound with (3.16), we obtain (3.13), which completes the proof. □

3.5 Local measures

In the rest of the paper, we work with measures on configurations on finite subsets of $\mathbb{Z}^2$ in which the factors $1 - s_x \cdot s_y$ of (3.7) are replaced by $1 + n_x \cdot n_y$. We now define these measures and relate them to the heat-bath kernels. Let $A \subset \mathbb{Z}^2$ be finite. We call a configuration $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2\setminus A}$ a *boundary configuration* of A, and we define the probability measure μ_A^η on $(\mathbb{S}^2)^A$ by

$$\mu_A^\eta(\,\mathrm{d}n) = \frac{1}{Z_A^\eta}\prod_{\substack{\{x,y\}\subset A\\ x\sim y}} (1 + n_x\cdot n_y) \prod_{\substack{x\in A,\, y\notin A\\ x\sim y}} (1 + n_x\cdot\eta_y) \prod_{x\in A} \mathrm{d}n_x, \tag{3.17}$$

where Z_A^η is the integral of the product of the factors over $(\mathbb{S}^2)^A$. The product of the factors vanishes only where $n_x = -n_y$ or $n_x = -\eta_y$ for one of the factors, which is a set of measure zero, so $Z_A^\eta > 0$ for every boundary configuration η. Note that μ_A^η is determined by the spins η_y at the sites $y \notin A$ adjacent to A. For $W \subset A$, we let $\mu_{A,W}^\eta$ denote the marginal of μ_A^η on $(\mathbb{S}^2)^W$.

We will need the following lemma in Section 4. It shows that, after the spins on one of the two sublattices of $\mathbb{Z}^2$ are reversed, the heat-bath kernel of a block becomes a measure of the form (3.17).

Lemma 3.6. *Let $b \geq 1$ be an integer, let $L, N \geq b+4$, let $x \in \Lambda$ and $c \in \mathbb{Z}^2$, and set $T = c + \{1, \dots, b\}^2 \subset \mathbb{Z}^2$. Let $s \in (\mathbb{S}^2)^\Lambda$, and let η be a boundary configuration of T such that*

$$\eta_u = (-1)^{u_1+u_2} s_{x+u-c} \qquad \text{for all } u \in \mathbb{Z}^2 \setminus T \text{ adjacent to } T.$$

Let $\tilde{s}$ be a random configuration with law $\mathcal{Q}_{B_x}(s, \cdot)$. Then the random configuration $\big((-1)^{u_1+u_2}\tilde{s}_{x+u-c}\big)_{u\in T}$ has law μ_T^η.

Proof. The map $u \mapsto x + u - c$ is the map $u \mapsto x + u$ considered before (3.12), composed with the translation by $-c$. It follows that this map sends T onto B_x and the bonds of $\mathbb{Z}^2$ that meet T bijectively onto $\mathcal{E}(B_x)$, and that it is injective on the set of sites at distance at most 1 from T. We set

$$n_u = (-1)^{u_1+u_2}\tilde{s}_{x+u-c}, \qquad u \in T.$$

By (3.9) and the invariance of the uniform measure on $\mathbb{S}^2$ under $a \mapsto -a$, the configuration $(n_u)_{u\in T}$ has a density with respect to $\prod_{u\in T} \mathrm{d}n_u$ that is proportional to

$$\prod_{\substack{\{u,v\} \text{ bond of } \mathbb{Z}^2 \\ \{u,v\}\cap T \neq \emptyset}} (1 - \tilde{s}_{x+u-c} \cdot \tilde{s}_{x+v-c}).$$

For each $v \notin T$ adjacent to T, the site $x + v - c$ lies outside B_x by injectivity on the sites at distance at most 1 from T. Thus $\tilde{s}_{x+v-c} = s_{x+v-c}$ for each such v. Every bond $\{u, v\}$ of $\mathbb{Z}^2$ satisfies $(-1)^{u_1+u_2}(-1)^{v_1+v_2} = -1$, which gives

$$1 - \tilde{s}_{x+u-c} \cdot \tilde{s}_{x+v-c} = \begin{cases} 1 + n_u \cdot n_v & \text{if } u, v \in T, \\ 1 + n_u \cdot \eta_v & \text{if } u \in T \text{ and } v \notin T. \end{cases}$$

These are the factors in (3.17) with $A = T$. □

The signs $(-1)^{u_1+u_2}$ are defined on $\mathbb{Z}^2$ rather than on Λ, and each block is treated with its own translation. For this reason, Lemma 3.6 requires no parity condition on L or N.

SECTION 4

From a window estimate to the gap of the block dynamics

In this section, we prove that the block dynamics of Section 3.4, with blocks of side $b = 2M + 1$, has spectral gap at least $1/2$ for all sufficiently large M, uniformly in $L, N \geq b + 4$. This is Proposition 4.5. We prove it under Hypothesis (W) below, which asks that the laws of certain sets of spins in a large box have a component of positive mass that is common to all boundary configurations. Hypothesis (W) is proved in Theorem 5.9, and Proposition 4.5 is combined with Theorem 3.5 in the proof of Theorem 1.1 in Section 8.4. The heuristics are explained in Section 2.3. We follow the proof of [35, Proposition 2.1], where the single-spin space is finite. For a finite single-spin space and positive weights, two conditional laws on a set of bounded size have an overlap that is bounded below uniformly in the boundary conditions. For the continuous spins here, whose densities can vanish, we prove the corresponding bound in Lemma 4.2.

Throughout this section, $M \geq 1$ is an integer, and we set

$$b = 2M + 1, \qquad \ell = M^{1/4}, \qquad \mathcal{C}_r = \{x \in \mathbb{Z}^2 : |x|_\infty \leq r\} \quad \text{for } r \geq 0.$$

We call $\mathcal{C}_M$ the box; it has b^2 sites. For configurations n and n' on a set D, we let $d_{\mathrm{H}}(n, n') = |\{x \in D : n_x \neq n'_x\}|$ denote their Hamming distance, and for $D' \subset D$, we say that n and n' agree on D' if $n_x = n'_x$ for all $x \in D'$. Given a finite set $A \subset \mathbb{Z}^2$, a configuration n on A, and a boundary configuration η of A, we let $n\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2}$ denote the configuration with $(n\eta)_x = n_x$ for $x \in A$ and $(n\eta)_x = \eta_x$ for $x \notin A$.

4.1 Two properties of the local measures

We begin with the consistency of the local measures (3.17) under resampling.

Lemma 4.1. *Let $A \subset \mathbb{Z}^2$ be finite, let $F \subset A$, and let η be a boundary configuration of A. Let n be a random configuration with law μ_A^η, and let $\mathcal{G}$ be a σ-algebra such that n is $\mathcal{G}$-measurable. Let $\hat{n}$ be a random configuration on A such that $\hat{n}_{A\setminus F} = n_{A\setminus F}$ and the conditional law of $\hat{n}_F$ given $\mathcal{G}$ is $\mu_F^{(n\eta)_{\mathbb{Z}^2\setminus F}}$. Then $\hat{n}$ has law μ_A^η.*

Proof. We write the product of the factors in (3.17) as $w_1(n)\, w_2(n)$, where w_1 is the product of the factors that involve a site of F and w_2 is the product of the other factors. Then

$w_2(n)$ is a function of $n_{A\setminus F}$. As a function of n_F, the product $w_1(n)$ is the product of the factors of $\mu_F^{(n\eta)_{\mathbb{Z}^2\setminus F}}$ in (3.17), so that

$$\int w_1(n) \prod_{x\in F} \mathrm{d}n_x = Z_F^{(n\eta)_{\mathbb{Z}^2\setminus F}}. \tag{4.1}$$

Let f be a bounded measurable function on $(\mathbb{S}^2)^A$, and for a configuration m on F, let $(m, n_{A\setminus F})$ denote the configuration on A that agrees with m on F and with n on $A\setminus F$. By the definition of $\hat{n}$ and Fubini's theorem, we have

$$\begin{aligned}\mathbb{E}[f(\hat{n})] &= \frac{1}{Z_A^\eta} \int \frac{w_2(n)}{Z_F^{(n\eta)_{\mathbb{Z}^2\setminus F}}} \Big(\int w_1(n) \prod_{x\in F} \mathrm{d}n_x \Big) \\ &\qquad \times \Big(\int f(m, n_{A\setminus F})\, w_1(m, n_{A\setminus F}) \prod_{x\in F} \mathrm{d}m_x \Big) \prod_{x\in A\setminus F} \mathrm{d}n_x \\ &= \frac{1}{Z_A^\eta} \int f(m, n_{A\setminus F})\, w_1(m, n_{A\setminus F})\, w_2(n) \prod_{x\in F} \mathrm{d}m_x \prod_{x\in A\setminus F} \mathrm{d}n_x = \mu_A^\eta(f),\end{aligned}$$

where we used (4.1) in the second equality. □

We will apply Lemma 4.1 to pairs of configurations. Fix a finite set $A \subset \mathbb{Z}^2$, a subset $D \subset A$, and boundary configurations η and η' of A. Let (n, n') be a random pair of configurations on A that is measurable with respect to a σ-algebra $\mathcal{G}$. We say that a random pair $(\hat{n}, \hat{n}')$ is obtained from (n, n') by a *coupled resampling* of D if $\hat{n}_{A\setminus D} = n_{A\setminus D}$, $\hat{n}'_{A\setminus D} = n'_{A\setminus D}$, and the conditional law of $(\hat{n}_D, \hat{n}'_D)$ given $\mathcal{G}$ is a coupling of $\mu_D^{(n\eta)_{\mathbb{Z}^2\setminus D}}$ and $\mu_D^{(n'\eta')_{\mathbb{Z}^2\setminus D}}$. If n has law μ_A^η and n' has law $\mu_A^{\eta'}$, then Lemma 4.1, applied to $\hat{n}$ and to $\hat{n}'$, shows that $\hat{n}$ has law μ_A^η and $\hat{n}'$ has law $\mu_A^{\eta'}$.

Next, we bound the overlap of two local measures on a set of bounded size. Let λ be a measure, let p and p' be probability densities with respect to λ, and set $\gamma = \int \min(p, p')\, \mathrm{d}\lambda$. We call the law of the following random pair (m, m') the *maximal coupling* of $p\lambda$ and $p'\lambda$. With probability γ, we draw m from the density $\min(p, p')/\gamma$ and set $m' = m$. With probability $1-\gamma$, we draw m and m' independently from the densities $(p - \min(p, p'))/(1-\gamma)$ and $(p' - \min(p, p'))/(1-\gamma)$. We omit any branch of probability zero. This is a coupling of $p\lambda$ and $p'\lambda$ under which $m = m'$ with probability at least γ. When $p = p'$ almost everywhere, we have $\gamma = 1$ and $m = m'$ almost surely. For an integer $K \geq 1$, we set

$$\varepsilon_K = \frac{1}{2}(16K)^{-4K}. \tag{4.2}$$

Lemma 4.2. *Let $K \geq 1$ be an integer, and let $A \subset \mathbb{Z}^2$ be a finite set with $|A| \leq K$. Let η and η' be boundary configurations of A, and let p^η and $p^{\eta'}$ denote the densities of μ_A^η and $\mu_A^{\eta'}$ with respect to $\prod_{x\in A} \mathrm{d}n_x$. Then*

$$\int \min(p^\eta, p^{\eta'}) \prod_{x\in A} \mathrm{d}n_x \geq \varepsilon_K.$$

In particular, under the maximal coupling of μ_A^η and $\mu_A^{\eta'}$, the two configurations are equal with probability at least ε_K.

Proof. We may assume that $A \neq \emptyset$. The product of the factors in (3.17) has k factors of the form $1 + n_x \cdot n_y$ or $1 + n_x \cdot \eta_y$, with values in $[0, 2]$, where $k \leq 4|A| \leq 4K$, since each factor involves a site of A and each site of A has four neighbors. This gives $Z_A^\eta \leq 2^k$, and the same holds for η'. Under the product measure $\prod_{x\in A} \mathrm{d}n_x$, each factor is uniformly distributed on $[0, 2]$. Indeed, by Archimedes' theorem, the inner product of a uniform random point of $\mathbb{S}^2$ with a unit vector is uniformly distributed on $[-1, 1]$, and for a factor $1 + n_x \cdot n_y$ with $x, y \in A$, we apply this conditionally on n_y.

We set $t = (8K)^{-1}$. Let $\mathcal{Y}$ denote the set of configurations $n \in (\mathbb{S}^2)^A$ at which all factors of p^η and of $p^{\eta'}$ are at least t. The two densities have at most $8K$ factors in total, and each of them is smaller than t on a set of measure $t/2$. By a union bound, the complement of $\mathcal{Y}$ has measure at most $8K \cdot t/2 = 1/2$. (The union bound does not require the factors to be independent.) On $\mathcal{Y}$, we have

$$p^\eta(n) \geq \frac{t^k}{2^k} \geq \left(\frac{t}{2}\right)^{4K} = (16K)^{-4K},$$

and the same bound holds for $p^{\eta'}$. Integrating over $\mathcal{Y}$, we obtain

$$\int \min(p^\eta, p^{\eta'}) \prod_{x\in A} \mathrm{d}n_x \geq (16K)^{-4K} \int \mathbf{1}_{\mathcal{Y}}(n) \prod_{x\in A} \mathrm{d}n_x \geq \frac{1}{2}(16K)^{-4K} = \varepsilon_K. \qquad \square$$

4.2 Shells

We now introduce the sets on which the coupling of Section 4.4 operates. They are a variant of the shells in the proof of [35, Proposition 2.1]. For $v \in \mathbb{Z}^2$ and an integer $r \geq 1$, we call $\{x \in \mathbb{Z}^2 : |x - v|_\infty = r\}$ the *sphere* of radius r around v. It is the boundary of the square $v + [-r, r]^2$, and we call the four segments $\{v_1 \pm r\} \times [v_2 - r, v_2 + r]$ and $[v_1 - r, v_1 + r] \times \{v_2 \pm r\}$ of lattice points its *sides*.

Fix $y \in \mathbb{Z}^2$ with $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$, and set $d_0 = 24$. For integers i with $2 \leq i \leq \ell$, we set $k_i = \lfloor i\ell \rfloor$ and

$$\begin{aligned} F_i &= \{x \in \mathcal{C}_M : |x - y|_\infty > (i-1)\ell\}, & \Sigma_i &= \{x \in \mathcal{C}_M : i\ell - 1 < |x - y|_\infty \leq i\ell\}, \\ \Sigma_i^\partial &= \{x \in \Sigma_i : \mathrm{dist}_\infty(x, \mathbb{Z}^2 \setminus \mathcal{C}_M) \leq d_0\}, & W_i &= \Sigma_i \setminus \Sigma_i^\partial. \end{aligned} \tag{4.3}$$

We call Σ_i the i-th *shell*, Σ_i^∂ its *ends*, and W_i the *window* of the shell Σ_i. Since $|x-y|_\infty$ is an integer, we have

$$\Sigma_i = \{x \in \mathcal{C}_M : |x-y|_\infty = k_i\},$$

which is the part in $\mathcal{C}_M$ of the sphere of radius k_i around y. Observe that $\Sigma_i \subset F_i$, since $k_i > i\ell - 1 \geq (i-1)\ell$. We let I_{corner} denote the set of integers i with $2 \leq i \leq \ell$ such that $\mathrm{dist}_\infty(v, \Sigma_i) < \ell/3$ for some corner site $v \in \{-M, M\}^2$ of $\mathcal{C}_M$. These sets are illustrated in Figure 4.1.

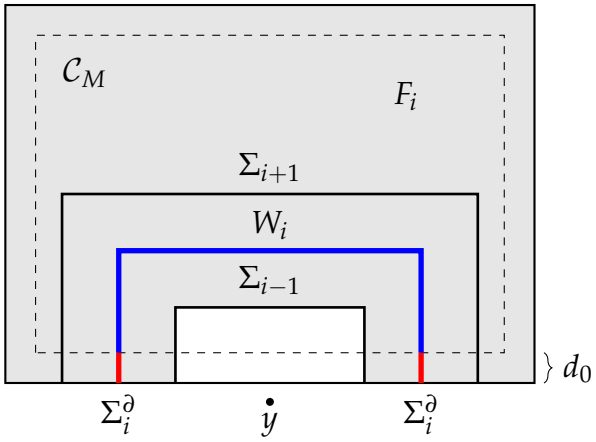


Figure 4.1. The sets of (4.3) for a site y adjacent to the lower side of the box $\mathcal{C}_M$ (schematic, not to scale). The thick lines are the shells Σ_{i-1}, Σ_i, and Σ_{i+1}, the parts in $\mathcal{C}_M$ of the spheres of radii k_{i-1}, k_i, and k_{i+1} around y. The shaded region is F_i. The sites between the dashed line and the boundary of $\mathcal{C}_M$ are those at ℓ^∞ distance at most d_0 from $\mathbb{Z}^2 \setminus \mathcal{C}_M$. The shell Σ_i consists of its two ends Σ_i^∂ (red), which lie in this region, and its window W_i (blue).

We set $M_{\mathrm{g}} = 73^4$. The following lemma collects the properties of the shells that we use below and in Section 5.

Lemma 4.3. *Let $M \geq M_{\mathrm{g}}$, and let $y \in \mathbb{Z}^2$ satisfy $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$. Then the following hold.*

(a) $|I_{\mathrm{corner}}| \leq 4$.

(b) Let $i \in \{2, \dots, \lfloor \ell \rfloor\} \setminus I_{\mathrm{corner}}$, let $j \in \{1,2\}$ and $t \in \{y_j - k_i, y_j + k_i\}$, and suppose that the side

$$S = \{x \in \mathbb{Z}^2 : x_j = t,\ |x_{3-j} - y_{3-j}| \leq k_i\}$$

of the sphere of radius k_i around y meets $\mathcal{C}_M$. Then $|t| \leq M + 1 - \ell/3$. Equivalently, S is at ℓ^∞ distance at least $\ell/3$ from each of the half-planes $\{x \in \mathbb{Z}^2 : x_j \geq M+1\}$ and $\{x \in \mathbb{Z}^2 : x_j \leq -M-1\}$.

(c) $|\Sigma_i^\partial| \leq 8d_0 = 192$ *for all $i \in \{2, \dots, \lfloor \ell \rfloor\} \setminus I_{\mathrm{corner}}$.*

Proof. Since $M \geq M_g$, we have $\ell \geq 73$, which implies that $\lfloor \ell \rfloor > 2\ell/3$ and $\ell/3 > d_0$. For $2 \leq i \leq \ell$, we also have

$$\ell \leq \lfloor 2\ell \rfloor \leq k_i \leq i\ell \leq \ell^2 = M^{1/2} \leq M. \tag{4.4}$$

We begin with (a). Let $2 \leq i < i' \leq \ell$. Using $\lfloor a + a' \rfloor \geq \lfloor a \rfloor + \lfloor a' \rfloor$ for real numbers a and a', we find that

$$k_{i'} - k_i \geq \lfloor (i' - i)\ell \rfloor \geq \lfloor \ell \rfloor > \frac{2\ell}{3}. \tag{4.5}$$

By the triangle inequality, every site v satisfies $\mathrm{dist}_\infty(v, \Sigma_i) \geq \big| |v - y|_\infty - k_i \big|$, and the same holds for i'. A site v with $\mathrm{dist}_\infty(v, \Sigma_i) < \ell/3$ and $\mathrm{dist}_\infty(v, \Sigma_{i'}) < \ell/3$ would satisfy

$$k_{i'} - k_i \leq \big| |v - y|_\infty - k_i \big| + \big| |v - y|_\infty - k_{i'} \big| < \frac{2\ell}{3},$$

which contradicts (4.5). We conclude that each of the four corner sites of $\mathcal{C}_M$ is at distance less than $\ell/3$ from at most one shell, and $|I_{\mathrm{corner}}| \leq 4$.

Next, we prove (b). The symmetries of $\mathbb{Z}^2$ that preserve $\mathcal{C}_M$ map the shells, their sides, the corner sites, and I_{corner} to the corresponding objects for the image of y. Using these symmetries, we may assume that $y_2 = -M - 1$ and $|y_1| \leq M + 1$.

We write $k = k_i$ and consider the four sides in turn. The side with $j = 2$ and $t = y_2 - k$ does not meet $\mathcal{C}_M$. The side with $j = 2$ and $t = y_2 + k$ satisfies

$$-M - 1 + \frac{\ell}{3} \leq t = k - M - 1 \leq -1$$

by (4.4), and $|t| \leq M + 1 - \ell/3$ follows.

Now consider the side with $j = 1$ and $t = y_1 - k$, and suppose that it meets $\mathcal{C}_M$. Since $|y_1| \leq M + 1$, the bounds (4.4) imply that

$$-M \leq t \leq M + 1 - k \leq M + 1 - \frac{\ell}{3}.$$

The site $v' = (t, -M)$ belongs to this side and to $\mathcal{C}_M$, and $|v' - y|_\infty = \max(k, 1) = k$, so that $v' \in \Sigma_i$. Since $i \notin I_{\mathrm{corner}}$, the distance $|v' - (-M, -M)|_\infty = t + M$ from v' to the corner site $(-M, -M)$ is at least $\ell/3$. We obtain $t \geq -M + \ell/3$, which together with the upper bound gives $|t| \leq M + 1 - \ell/3$.

The side with $j = 1$ and $t = y_1 + k$ is treated in the same way, with the corner site $(M, -M)$. The equivalent formulation follows from the fact that S has ℓ^∞ distance $M + 1 - t$ from $\{x : x_j \geq M + 1\}$ and $M + 1 + t$ from $\{x : x_j \leq -M - 1\}$.

Finally, we prove (c). Recall that Σ_i is the part in $\mathcal{C}_M$ of the sphere of radius k_i around y. Let $x \in \Sigma_i^\partial$. Then x lies on a side S of this sphere, with j and t as in (b), and S meets $\mathcal{C}_M$. Since $x \in \mathcal{C}_M$, we have

$$\min\big(M + 1 - |x_1|, M + 1 - |x_2|\big) = M + 1 - |x|_\infty = \mathrm{dist}_\infty(x, \mathbb{Z}^2 \setminus \mathcal{C}_M) \leq d_0.$$

Part (b) implies that $M+1-|x_j| = M+1-|t| \geq \ell/3 > d_0$. The minimum is therefore attained at $3-j$, which gives

$$M+1-d_0 \leq |x_{3-j}| \leq M.$$

Thus x_{3-j} takes at most $2d_0$ values. Since a site of S is determined by its coordinate x_{3-j}, each of the four sides contains at most $2d_0$ sites of Σ_i^∂, and $|\Sigma_i^\partial| \leq 8d_0$. □

4.3 The window estimate

We now state the hypothesis under which we prove the gap of the block dynamics. Given a family of probability measures on a measurable space, a probability measure ν on the same space, and $p \in (0,1]$, we say that $p\nu$ is a *common component* of the family if every measure of the family is bounded below by $p\nu$.

Hypothesis (W). *There exist $p_0 \in (0,1]$ and $M_\star \geq 1$ such that the following holds for all integers $M \geq M_\star$, all $y \in \mathbb{Z}^2$ with $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$, and all integers i with $2 \leq i \leq \ell$ and $i \notin I_{\mathrm{corner}}$. There exists a probability measure $\nu_{M,y,i}$ on $(\mathbb{S}^2)^{W_i}$ such that*

$$\mu^\eta_{F_i,W_i} \geq p_0\, \nu_{M,y,i} \qquad \textit{for every boundary configuration } \eta \textit{ of } F_i.$$

We call Hypothesis (W) the *window estimate*. It states that $p_0\nu_{M,y,i}$ is a common component of the laws $\mu^\eta_{F_i,W_i}$ of the window, over all boundary configurations η of F_i. Note that $\nu_{M,y,i}$ may depend on M, y, and i, but not on η. Assuming Hypothesis (W), we set

$$\delta = p_0\, \varepsilon_{192}, \qquad J = \lfloor M^{1/8} \rfloor, \qquad \Delta_M = 9M^{1/2}(J+5)^2 + b^2(1-\delta)^J. \tag{4.6}$$

4.4 The coupling

We now describe the coupling of two local measures on the box whose boundary configurations differ at one site. Until the end of the proof of Proposition 4.4, we assume Hypothesis (W), we fix $M \geq \max(M_\star, M_{\mathrm{g}})$ and a site $y \in \mathbb{Z}^2$ adjacent to $\mathcal{C}_M$, and we fix boundary configurations η and η' of $\mathcal{C}_M$ with $\eta_u = \eta'_u$ for all $u \neq y$. Since $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$, the sets (4.3) and I_{corner} are defined, and Lemma 4.3 applies.

The shells used. Since $M \geq M_{\mathrm{g}} \geq 3^8$, we have $M^{1/8} \geq 3$ and

$$\ell - 6 = (M^{1/8})^2 - 6 \geq 3M^{1/8} - 6 \geq M^{1/8},$$

which implies that $J \leq \lfloor \ell \rfloor - 6$. By Lemma 4.3(a), the set $\{2, \dots, \lfloor \ell \rfloor - 1\} \setminus I_{\mathrm{corner}}$ has at least $\lfloor \ell \rfloor - 6 \geq J$ elements. Let $i_1 < i_2 < \cdots < i_J$ denote its J smallest elements. For $k \leq J$, the set $\{2, \dots, k+5\}$ is contained in $\{2, \dots, \lfloor \ell \rfloor - 1\}$ and contains at least k elements outside I_{corner}, so that

$$i_k \leq k+5, \qquad 1 \leq k \leq J. \tag{4.7}$$

The trials. We construct the coupling by a finite sequence of coupled resamplings of subsets of $\mathcal{C}_M$, with $A = \mathcal{C}_M$ and the boundary configurations η and η'. We start from a pair (n, n') of independent configurations with laws $\mu^{\eta}_{\mathcal{C}_M}$ and $\mu^{\eta'}_{\mathcal{C}_M}$, and we write (n, n') for the current pair at every stage. In each resampling, $\mathcal{G}$ is the σ-algebra generated by all random variables constructed before it, and the conditional law of the new pair given $\mathcal{G}$ is a function of the current pair. All the kernels used below depend measurably on the current pair, since the densities of the local measures are jointly measurable in the configuration and the boundary configuration.

In every coupled resampling of a set D below, when the laws $\mu_D^{(n\eta)_{\mathbb{Z}^2\setminus D}}$ and $\mu_D^{(n'\eta')_{\mathbb{Z}^2\setminus D}}$ coincide, we sample identically, drawing $\hat{n}_D$ from their common law and setting $\hat{n}'_D = \hat{n}_D$. Recall from Section 3.5 that these two laws coincide when $(n\eta)_u = (n'\eta')_u$ for all sites $u \notin D$ adjacent to D. For $i = i_1, i_2, \dots, i_J$, in this order, we perform the following three coupled resamplings, which we call the *trial* at i.

(a) A coupled resampling of F_i. When the two laws coincide, we sample identically. Otherwise, we write $\xi = (n\eta)_{\mathbb{Z}^2\setminus F_i}$ and $\xi' = (n'\eta')_{\mathbb{Z}^2\setminus F_i}$, which are boundary configurations of F_i, and $\nu = \nu_{M,y,i}$. By Hypothesis (W), which applies since $M \geq M_\star$ and $i \notin I_{\mathrm{corner}}$, we have $\mu^{\xi}_{F_i,W_i} \geq p_0\nu$ and $\mu^{\xi'}_{F_i,W_i} \geq p_0\nu$. With probability p_0, we draw w from ν and set $w' = w$. With probability $1 - p_0$ (which is positive only if $p_0 < 1$), we draw w and w' independently from $(\mu^{\xi}_{F_i,W_i} - p_0\nu)/(1 - p_0)$ and $(\mu^{\xi'}_{F_i,W_i} - p_0\nu)/(1 - p_0)$. We then set $\hat{n}_{W_i} = w$ and $\hat{n}'_{W_i} = w'$, and we draw $\hat{n}_{F_i\setminus W_i}$ and $\hat{n}'_{F_i\setminus W_i}$ independently from $\mu^{w\xi}_{F_i\setminus W_i}$ and $\mu^{w'\xi'}_{F_i\setminus W_i}$. Here $w\xi$ is the boundary configuration of $F_i \setminus W_i$ that agrees with w on W_i and with ξ off F_i.

(b) A coupled resampling of Σ_i^{∂}, by the maximal coupling of the two laws.

(c) A coupled resampling of F_{i+1}. When the two laws coincide, we sample identically. Otherwise, we sample independently.

In (a), the window configuration w has law $\mu^{\xi}_{F_i,W_i}$. By Lemma 4.1 with $A = F_i$, $F = F_i \setminus W_i$, and the boundary configuration ξ, drawing w from $\mu^{\xi}_{F_i,W_i}$ and then the configuration on $F_i \setminus W_i$ from $\mu^{w\xi}_{F_i\setminus W_i}$ produces a configuration on F_i with law $\mu^{\xi}_{F_i}$. Hence $\hat{n}_{F_i}$ has law $\mu^{\xi}_{F_i}$, and likewise $\hat{n}'_{F_i}$ has law $\mu^{\xi'}_{F_i}$, so that (a) is a coupled resampling of F_i. We call the trial at i *successful* if the two configurations agree on Σ_i after (b).

By the discussion after Lemma 4.1 and induction over the resamplings, the final pair (n, n') is a coupling of $\mu^{\eta}_{\mathcal{C}_M}$ and $\mu^{\eta'}_{\mathcal{C}_M}$. The following proposition, which is the main step of this section, bounds its expected Hamming distance.

Proposition 4.4. *Assume Hypothesis (W), and let $M \geq \max(M_\star, M_{\mathrm{g}})$. Let $y \in \mathbb{Z}^2 \setminus \mathcal{C}_M$, and let η and η' be boundary configurations of $\mathcal{C}_M$ such that $\eta_u = \eta'_u$ for all $u \neq y$. Then there exists*

a coupling (n, n') *of* $\mu^{\eta}_{\mathcal{C}_M}$ *and* $\mu^{\eta'}_{\mathcal{C}_M}$ *such that*

$$\mathbb{E}\, d_{\mathrm{H}}(n, n') \le \Delta_M.$$

Further, $\Delta_M / b \to 0$ *as* $M \to \infty$.

Proof. If y is not adjacent to $\mathcal{C}_M$, then $\mu^{\eta}_{\mathcal{C}_M} = \mu^{\eta'}_{\mathcal{C}_M}$ by Section 3.5, and we take $n = n'$. From now on, we assume that y is adjacent to $\mathcal{C}_M$, and we let (n, n') be the final pair of the construction above.

The success probability. We claim that, conditionally on all random variables constructed before the trial at i, the trial at i is successful with probability at least δ. Recall that μ^{η}_{A} is determined by the spins of η at the sites adjacent to A.

Suppose first that the two laws in (a) coincide, so that the configurations agree on F_i after (a). Let $u \notin \Sigma_i^{\partial}$ be adjacent to a site $x \in \Sigma_i^{\partial}$. Since $\ell \ge 2$, we have

$$|u - y|_\infty \ge k_i - 1 > i\ell - 2 \ge (i-1)\ell,$$

and $u \ne y$, since $|x - y|_\infty = k_i \ge 2$. Thus u belongs to F_i, where the configurations agree, or to $\mathbb{Z}^2 \setminus (\mathcal{C}_M \cup \{y\})$, where $\eta_u = \eta'_u$. It follows that the laws in (b) coincide, (b) samples identically, and the trial is successful.

Suppose now that the two laws in (a) differ. Then (a) makes the configurations agree on W_i with conditional probability at least p_0. Further, Lemma 4.3(c) implies that $|\Sigma_i^{\partial}| \le 192$. Applying Lemma 4.2 with $K = 192$ and $A = \Sigma_i^{\partial}$, we find that (b) makes the configurations agree on Σ_i^{∂} with conditional probability at least ε_{192}, given all random variables constructed before (b). The resampling (b) leaves the configurations on W_i unchanged, since $W_i \cap \Sigma_i^{\partial} = \emptyset$.

Combining these two bounds, we conclude that the trial is successful with conditional probability at least $p_0 \varepsilon_{192} = \delta$, as claimed. Multiplying the conditional probabilities of failure over the J trials, we obtain

$$\mathbb{P}(\text{no trial is successful}) \le (1 - \delta)^J. \tag{4.8}$$

Agreement after a success. Suppose that the trial at i is successful. We show that the two configurations agree on $\Sigma_i \cup F_{i+1}$ from the end of this trial on. Let $u \notin D$ be adjacent to a set $D \subset F_{i+1}$. We claim that

$$u \in F_{i+1} \cup \Sigma_i \qquad \text{or} \qquad u \in \mathbb{Z}^2 \setminus (\mathcal{C}_M \cup \{y\}). \tag{4.9}$$

Indeed, a site $u \in \mathcal{C}_M \setminus F_{i+1}$ adjacent to a site of F_{i+1} satisfies $i\ell - 1 < |u - y|_\infty \le i\ell$, which is the defining condition of Σ_i. If $u \notin \mathcal{C}_M$, then $u \ne y$, since the sites of F_{i+1} are at ℓ^∞ distance more than $i\ell \ge 2$ from y.

By (4.9), if the configurations agree on $\Sigma_i \cup F_{i+1}$ before a coupled resampling of D, then the two laws in it coincide, it samples identically, and the configurations agree on $\Sigma_i \cup F_{i+1}$ after it.

Before (c) at i, the configurations agree on Σ_i. By (4.9) with $D = F_{i+1}$, every site $u \notin F_{i+1}$ adjacent to F_{i+1} belongs to Σ_i or to $\mathbb{Z}^2 \setminus (\mathcal{C}_M \cup \{y\})$. The two laws in (c) coincide, and after (c) the configurations agree on $\Sigma_i \cup F_{i+1}$. Every later resampling is a coupled resampling of a set $D \in \{F_{i'}, \Sigma_{i'}^{\partial}, F_{i'+1}\}$ with $i' > i$, and $D \subset F_{i'} \subset F_{i+1}$. By the previous paragraph and induction over the later resamplings, the configurations agree on $\Sigma_i \cup F_{i+1}$ until the end of the construction.

The expected discrepancy. Suppose that some trial is successful, and let i be the smallest index at which a trial is successful. Then (4.7) implies that $i \le i_J \le J + 5$. Since the final configurations agree on F_{i+1}, they can differ only at sites $x \in \mathcal{C}_M$ with $|x - y|_\infty \le i\ell$. Using $i \le J + 5$ and $(J + 5)\ell \ge 1$, we see that the number of these sites is at most

$$\big(2(J+5)\ell + 1\big)^2 \le 9(J+5)^2\ell^2 = 9M^{1/2}(J+5)^2.$$

On the event that no trial is successful, we use the bound $d_{\mathrm{H}}(n, n') \le |\mathcal{C}_M| = b^2$. Combining these bounds with (4.8), we obtain $\mathbb{E}\, d_{\mathrm{H}}(n, n') \le \Delta_M$.

The limit. We have $J + 5 \le 6M^{1/8}$ and $J \ge M^{1/8} - 1$. Using $b \le 3M$ and $1 - \delta \le e^{-\delta}$, we find that

$$\frac{\Delta_M}{b} \le \frac{9M^{1/2} \cdot 36M^{1/4}}{2M} + b\, e^{-\delta(M^{1/8}-1)} \le 162M^{-1/4} + 3Me^{-\delta(M^{1/8}-1)}.$$

Since $\delta = p_0\varepsilon_{192} > 0$, the right side tends to zero as $M \to \infty$. □

4.5 The gap of the block dynamics

We now return to the torus and prove the main result of this section.

Proposition 4.5. *Assume Hypothesis (W). Let $M \ge \max(M_\star, M_{\mathrm{g}})$ be an integer, set $b = 2M + 1$, and let $L, N \ge b + 4$. Then the generator $\mathcal{L}_{\mathrm{bl}}$ of* (3.11) *satisfies*

$$\mathrm{gap}(\mathcal{L}_{\mathrm{bl}}) \ge 1 - \frac{4\Delta_M}{b}. \tag{4.10}$$

In particular, there exists $M_{\mathrm{bl}} \ge \max(M_\star, M_{\mathrm{g}})$, depending only on p_0 and $M_\star$, such that $\mathrm{gap}(\mathcal{L}_{\mathrm{bl}}) \ge 1/2$ for all integers $M \ge M_{\mathrm{bl}}$ and all integers $L, N \ge 2M + 5$.

Proof. The second claim follows from the first, since Δ_M is determined by p_0 and M, and $\Delta_M/b \to 0$ as $M \to \infty$ by Proposition 4.4. To prove (4.10), we write $\mu = \mu_{L,N}$, and we set

$$\mathcal{K} = \frac{1}{|\Lambda|} \sum_{x \in \Lambda} \mathcal{Q}_{B_x}, \qquad \gamma_{\mathrm{bl}} = 1 - \frac{b^2 - 4b\Delta_M}{|\Lambda|}.$$

The operator $\mathcal{K}$ resamples a uniformly chosen block, and (3.11) can be written as

$$\mathcal{L}_{\mathrm{bl}} = \frac{|\Lambda|}{b^2}(\mathcal{K} - \mathrm{Id}). \tag{4.11}$$

Recall from Section 3.4 that $\mathcal{Q}_{B_x}(s,\cdot)$ is defined for every configuration $s \in (\mathbb{S}^2)^\Lambda$. Therefore $\mathcal{K}f$ is a bounded measurable function, defined at every point, for every bounded measurable function f on $(\mathbb{S}^2)^\Lambda$. For such f, we set

$$\mathrm{Lip}(f) = \sup_{s \neq s'} \frac{|f(s) - f(s')|}{d_{\mathrm{H}}(s,s')} \leq 2\|f\|_\infty,$$

where the supremum in $\mathrm{Lip}(f)$ is over all pairs of distinct configurations $s, s' \in (\mathbb{S}^2)^\Lambda$ and $\|f\|_\infty = \sup_{s \in (\mathbb{S}^2)^\Lambda} |f(s)|$ is the pointwise supremum.

Contraction of the Lipschitz constant. We claim that

$$\mathrm{Lip}(\mathcal{K}f) \leq \gamma_{\mathrm{bl}} \,\mathrm{Lip}(f) \tag{4.12}$$

for every bounded measurable function f on $(\mathbb{S}^2)^\Lambda$. Two configurations at Hamming distance k are joined by a sequence of $k+1$ configurations in which consecutive configurations differ at exactly one site. It therefore suffices to show that $|\mathcal{K}f(s) - \mathcal{K}f(s')| \leq \gamma_{\mathrm{bl}} \,\mathrm{Lip}(f)$ when s and s' differ only at one site z.

Fix such s and s', and fix $x \in \Lambda$. For every coupling (ξ, ξ') of $\mathcal{Q}_{B_x}(s,\cdot)$ and $\mathcal{Q}_{B_x}(s',\cdot)$, we have

$$\left|\mathcal{Q}_{B_x} f(s) - \mathcal{Q}_{B_x} f(s')\right| = \left|\mathbb{E}[f(\xi) - f(\xi')]\right| \leq \mathrm{Lip}(f)\,\mathbb{E}\, d_{\mathrm{H}}(\xi,\xi'). \tag{4.13}$$

By (3.9), the law of the spins in B_x under $\mathcal{Q}_{B_x}(s,\cdot)$ is determined by the spins of s at the sites adjacent to B_x. We choose the coupling according to the position of z, in three cases.

- When $z \in B_x$, we have $\mathcal{Q}_{B_x}(s,\cdot) = \mathcal{Q}_{B_x}(s',\cdot)$, and the identical coupling gives $d_{\mathrm{H}}(\xi,\xi') = 0$. Since $L, N \geq b$, this case occurs for exactly b^2 sites x.
- When $z \notin B_x$ and z is not adjacent to B_x, the laws of the block spins under $\mathcal{Q}_{B_x}(s,\cdot)$ and $\mathcal{Q}_{B_x}(s',\cdot)$ coincide. The identical coupling of the block spins gives $d_{\mathrm{H}}(\xi,\xi') = 1$, since $\xi_z = s_z$ and $\xi'_z = s'_z$.
- When $z \notin B_x$ and z is adjacent to B_x, we use Proposition 4.4, as described in the next paragraph, to obtain a coupling with $\mathbb{E}\, d_{\mathrm{H}}(\xi,\xi') \leq \Delta_M + 1$. In this case, $z = x + u$ for a site u of the patch outside $\{1,\dots,b\}^2$, and $x = z - u$ is determined by u. Since there are $4b$ such sites u, this case occurs for at most $4b$ sites x.

We now construct the coupling in the third case. We apply Lemma 3.6 with $c = (-M-1, -M-1)$, for which $T = c + \{1,\dots,b\}^2 = \mathcal{C}_M$. Let η and η' be the boundary configurations of $\mathcal{C}_M$ such that

$$\eta_u = (-1)^{u_1+u_2} s_{x+u-c}, \qquad \eta'_u = (-1)^{u_1+u_2} s'_{x+u-c}$$

for all $u \notin \mathcal{C}_M$ adjacent to $\mathcal{C}_M$, and such that $\eta_u = \eta'_u$ is the same arbitrary element of $\mathbb{S}^2$ at all other sites $u \notin \mathcal{C}_M$. Since the map $u \mapsto x + u - c$ is injective on the sites at distance at most 1 from $\mathcal{C}_M$, the configurations η and η' differ at most at one site $y \notin \mathcal{C}_M$, namely the site adjacent to $\mathcal{C}_M$ with $x + y - c = z$. By Proposition 4.4, there exists a coupling (n, n') of $\mu^{\eta}_{\mathcal{C}_M}$ and $\mu^{\eta'}_{\mathcal{C}_M}$ with $\mathbb{E}\, d_{\mathrm{H}}(n, n') \le \Delta_M$. We define ξ by

$$\xi_{x+u-c} = (-1)^{u_1+u_2} n_u \quad \text{for } u \in \mathcal{C}_M, \qquad \xi_v = s_v \quad \text{for } v \notin B_x,$$

and we define ξ' in the same way from n' and s'.

The map that sends a configuration $\tilde{s}$ equal to s off B_x to $\big((-1)^{u_1+u_2}\tilde{s}_{x+u-c}\big)_{u\in\mathcal{C}_M}$ is a bijection onto $(\mathbb{S}^2)^{\mathcal{C}_M}$, and ξ is the preimage of n under this map. Therefore Lemma 3.6 shows that ξ has law $\mathcal{Q}_{B_x}(s, \cdot)$, and likewise ξ' has law $\mathcal{Q}_{B_x}(s', \cdot)$. Further, $\xi_v = \xi'_v$ for $v \in B_x$ if and only if $n_u = n'_u$ for the corresponding $u \in \mathcal{C}_M$, and off B_x the configurations ξ and ξ' differ only at z. We conclude that

$$\mathbb{E}\, d_{\mathrm{H}}(\xi, \xi') = \mathbb{E}\, d_{\mathrm{H}}(n, n') + 1 \le \Delta_M + 1.$$

We now average (4.13) over x. Let $\mathcal{X}_z$ denote the set of sites x in the third case, so that $|\mathcal{X}_z| \le 4b$, and the second case occurs for $|\Lambda| - b^2 - |\mathcal{X}_z|$ sites x. Using the three cases, we find that

$$\begin{aligned} |\mathcal{K}f(s) - \mathcal{K}f(s')| &\le \frac{\mathrm{Lip}(f)}{|\Lambda|}\Big(|\mathcal{X}_z|(\Delta_M + 1) + |\Lambda| - b^2 - |\mathcal{X}_z|\Big) \\ &\le \mathrm{Lip}(f)\Big(1 - \frac{b^2 - 4b\Delta_M}{|\Lambda|}\Big) = \gamma_{\mathrm{bl}}\,\mathrm{Lip}(f), \end{aligned}$$

which proves (4.12). (Each coupling above is used for one fixed pair (s, s'), so no measurable dependence of the couplings on the pair is needed.)

From contraction to the spectrum. For a finite state space, the passage from (4.12) to a lower bound on the spectral gap is a standard argument (see [30, Theorem 13.1]), which applies (4.12) to an eigenfunction f of $\mathcal{K}$ with eigenvalue λ and obtains $|\lambda|\,\mathrm{Lip}(f) \le \gamma_{\mathrm{bl}}\,\mathrm{Lip}(f)$. (A version for reversible chains on general metric spaces is given in [38].) Here the state space is continuous, $\mathcal{K}$ need not have eigenfunctions, and an element of $L^2(\mu)$ has no Lipschitz constant. We therefore apply (4.12) to the iterates of bounded measurable functions, which are defined at every point because the version of $\mathcal{Q}_{B_x}$ from Section 3.4 is defined for every exterior configuration, and we pass to $L^2(\mu)$ through the spectral theorem.

We may assume that $\gamma_{\mathrm{bl}} < 1$, since otherwise the right side of (4.10) is nonpositive and (4.10) holds trivially.

Let f be a bounded measurable function with $\mu(f) = 0$, and let $t \ge 0$ be an integer. Recall that each $\mathcal{Q}_{B_x} g$ is a version of a conditional expectation under μ. In particular, $\mu(\mathcal{Q}_{B_x} g) = \mu(g)$ for every bounded measurable g, and $\mu(\mathcal{K}^t f) = \mu(f) = 0$.

Iterating (4.12) gives $\operatorname{Lip}(\mathcal{K}^t f) \le 2\gamma_{\mathrm{bl}}^t \|f\|_\infty$. Since $d_{\mathrm{H}}(s,s') \le |\Lambda|$, we obtain, for every $s \in (\mathrm{S}^2)^\Lambda$,

$$|\mathcal{K}^t f(s)| = \left| \int \big(\mathcal{K}^t f(s) - \mathcal{K}^t f(s')\big)\, \mu(\,\mathrm{d}s') \right| \le |\Lambda| \operatorname{Lip}(\mathcal{K}^t f) \le 2|\Lambda| \gamma_{\mathrm{bl}}^t \|f\|_\infty. \tag{4.14}$$

We now pass to $L^2(\mu)$. The function $\mathcal{K}^t f$, defined pointwise, represents the t-th power of the operator $\mathcal{K}$ on $L^2(\mu)$ applied to f, since each $\mathcal{Q}_{B_x}$ maps functions that agree almost everywhere to functions that agree almost everywhere. Each $\mathcal{Q}_{B_x}$ is an orthogonal projection on $L^2(\mu)$, so $\mathcal{K}$ is self-adjoint and $0 \le \mathcal{K} \le \mathrm{Id}$.

For $g \in L^2(\mu)$, let m_g denote the spectral measure of g with respect to $\mathcal{K}$, which is the finite measure on $[0,1]$ such that

$$\langle g, h(\mathcal{K}) g \rangle_{L^2(\mu)} = \int h \,\mathrm{d}m_g \qquad \text{for all bounded Borel functions } h.$$

Let $\gamma' \in (\gamma_{\mathrm{bl}}, 1]$. Using (4.14), we find that

$$(\gamma')^{2t}\, m_f\big((\gamma', 1]\big) \le \int \lambda^{2t}\, m_f(\,\mathrm{d}\lambda) = \|\mathcal{K}^t f\|_{L^2(\mu)}^2 \le 4|\Lambda|^2 \gamma_{\mathrm{bl}}^{2t} \|f\|_\infty^2.$$

Letting $t \to \infty$, we obtain $m_f((\gamma', 1]) = 0$, and letting $\gamma' \downarrow \gamma_{\mathrm{bl}}$, we obtain $m_f((\gamma_{\mathrm{bl}}, 1]) = 0$.

Let $E = \mathbf{1}_{(\gamma_{\mathrm{bl}}, 1]}(\mathcal{K})$ denote the spectral projection of $\mathcal{K}$ on $(\gamma_{\mathrm{bl}}, 1]$, and let $L_0^2(\mu)$ denote the space of functions $g \in L^2(\mu)$ with $\mu(g) = 0$. The previous paragraph shows that

$$\|Ef\|_{L^2(\mu)}^2 = m_f\big((\gamma_{\mathrm{bl}}, 1]\big) = 0$$

for every bounded measurable f with $\mu(f) = 0$.

The bounded measurable functions with mean zero are dense in $L_0^2(\mu)$, as one sees by truncating and recentering, and E is bounded. Therefore $Eg = 0$ for all $g \in L_0^2(\mu)$. Equivalently, the spectral measure m_g of every $g \in L_0^2(\mu)$ is supported on $[0, \gamma_{\mathrm{bl}}]$, which gives

$$\langle g, \mathcal{K} g \rangle_{L^2(\mu)} = \int \lambda\, m_g(\,\mathrm{d}\lambda) \le \gamma_{\mathrm{bl}} \|g\|_{L^2(\mu)}^2, \qquad g \in L_0^2(\mu). \tag{4.15}$$

Conclusion. Let $F \in L^2(\mu)$ be nonconstant, and set $g = F - \mu(F)$. Since $\mathcal{L}_{\mathrm{bl}}$ is self-adjoint and annihilates the constant functions, (4.11) and (4.15) imply that

$$\begin{aligned} \langle F, -\mathcal{L}_{\mathrm{bl}} F \rangle_{L^2(\mu)} &= \frac{|\Lambda|}{b^2} \Big(\|g\|_{L^2(\mu)}^2 - \langle g, \mathcal{K} g \rangle_{L^2(\mu)} \Big) \\ &\ge \frac{|\Lambda|}{b^2} (1 - \gamma_{\mathrm{bl}}) \operatorname{Var}_\mu(F) = \Big(1 - \frac{4\Delta_M}{b}\Big) \operatorname{Var}_\mu(F). \end{aligned}$$

Taking the infimum over F proves (4.10). □

SECTION 5

The cut lemma and the window estimate

In this section, we prove Theorem 5.9, which states that Hypothesis (W) of Section 4 follows from Hypothesis (P) below. Hypothesis (P) is an exponential bound on the effect of boundary conditions on certain pair correlations in a continuous model with pins. Together with Proposition 4.5 and Theorem 3.5, Theorem 5.9 reduces Theorem 1.1 to Hypothesis (P), which we deduce in Sections 6 and 7 from the inputs verified in Section 8. The heuristics behind the argument are explained in Section 2.

The proof has two parts. The first, and the main idea of this section, is an exact identity, Lemma 5.4. Suppose that a set of sites X separates the window W_i from the complement of F_i. Then the density of the law of the spins in W_i is, up to a normalizing constant, the product of a density that is the same for all boundary configurations and the exponential of the change in the logarithm of a ratio of partition functions caused by removing X. In the second part, we bound this change. We remove the sites of X one at a time, in an order in which every site after the first has a neighbor that was removed before it. Each of these removals concerns a site with at most three remaining neighbors, and we bound its cost by the differences of pair correlations that Hypothesis (P) controls. In Section 5.5, we construct for each shell a set X along which these bounds have a sum that is bounded uniformly in M.

5.1 The model with pins

Let $G \subset \mathbb{Z}^2$ be finite. We call a pair (x, z) of sites with $x \in G$, $z \notin G$, and $x \sim z$ a *leg* of G at x. A *pin configuration* on G is a family $\tau = (\tau_{xz})$ of vectors $\tau_{xz} \in \mathrm{S}^2$ indexed by the legs (x, z) of G, and we call τ_{xz} the *pin* on the leg (x, z).

Let $Y \subset G$ and set $D = G \setminus Y$. We call D a *domain* in G, and we call the sites of Y *removed*. A removed site carries no spin, and the integrands of the partition functions below are products of one factor for each bond $\{x, x'\} \subset D$ and one factor for each leg of G at a site of D. In particular, a removed site does not become a pin. A bond $\{x, z\}$ with $x \in D$ and $z \in Y$ is not a leg of G and contributes no factor. Definition 6.1 in Section 6 extends the notion of a domain to sets in which some sites have been visited once by a walk.

For a domain D in G and a pin configuration τ on G, we define the *pinned partition function* Z_D^τ and the *free partition function* Z_D^0 by

$$\begin{aligned} Z_D^\tau &= \int \prod_{\substack{\{x,x'\}\subset D\\ x\sim x'}} (1 + n_x \cdot n_{x'}) \prod_{\substack{(x,z) \text{ leg of } G\\ x\in D}} (1 + \tau_{xz} \cdot n_x) \prod_{x\in D} \mathrm{d}n_x, \\ Z_D^0 &= \int \prod_{\substack{\{x,x'\}\subset D\\ x\sim x'}} (1 + n_x \cdot n_{x'}) \prod_{x\in D} \mathrm{d}n_x, \end{aligned} \tag{5.1}$$

with $Z_\varnothing^\tau = Z_\varnothing^0 = 1$. The integrands are nonnegative and bounded, and they are positive outside a set of measure zero, since $1 + a \cdot b > 0$ for all $a, b \in \mathbb{S}^2$ with $b \neq -a$. In particular, all these partition functions are positive and finite. We let $\mathbb{E}_D^\tau$ and $\mathbb{E}_D^0$ denote the probability measures on $(\mathbb{S}^2)^D$ whose densities with respect to $\prod_{x\in D} \mathrm{d}n_x$ are the integrands in (5.1) divided by Z_D^τ and Z_D^0, respectively, and we use the same symbols for the corresponding expectations. We write $Z_D^\cdot$ and $\mathbb{E}_D^\cdot$ in statements that hold with $\cdot = \tau$ and with $\cdot = 0$. We also define

$$\Xi_D(\tau) = \frac{Z_D^\tau}{Z_D^0}.$$

These quantities depend on G through its legs (we suppress G in the notation, since it will always be clear from the context).

The local measures of Section 3.5 are special cases of these measures. Indeed, for a finite set $A \subset \mathbb{Z}^2$ and $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2\setminus A}$, let τ be the pin configuration on A with $\tau_{xz} = \eta_z$ for every leg (x, z) of A. Comparing (5.1) for the domain A in A with (3.17), we see that $\mu_A^\eta = \mathbb{E}_A^\tau$ and $Z_A^\eta = Z_A^\tau$.

We will use the moments of the uniform measure on $\mathbb{S}^2$. Let n^α denote the α-th coordinate of $n \in \mathbb{S}^2$. For $m \geq 1$ and $\alpha_1, \dots, \alpha_{2m} \in \{1, 2, 3\}$, we have

$$\int_{\mathbb{S}^2} n^{\alpha_1} \cdots n^{\alpha_{2m}} \,\mathrm{d}n = \frac{1}{3 \cdot 5 \cdots (2m+1)} \sum_\pi \prod_{\{r,r'\}\in\pi} \delta_{\alpha_r \alpha_{r'}}, \tag{5.2}$$

where the sum runs over the partitions π of $\{1, \dots, 2m\}$ into pairs. The moments of odd order vanish, since $\mathrm{d}n$ is invariant under $n \mapsto -n$. The formula (5.2) is standard. It follows from Wick's formula for a standard Gaussian vector g in $\mathbb{R}^3$, since $g/|g|$ has law $\mathrm{d}n$ and is independent of $|g|$, and $\mathbb{E}|g|^{2m} = 3 \cdot 5 \cdots (2m+1)$. In particular, we have $\int (a \cdot n)(b \cdot n) \,\mathrm{d}n = (a \cdot b)/3$ for all $a, b \in \mathbb{R}^3$.

We will need two further elementary properties of the measures $\mathbb{E}_D^\cdot$.

Lemma 5.1. *Let D be a domain in G, and let τ be a pin configuration on G.*

1. *Let f be a finite product of scalar products $n_x \cdot n_{x'}$ with $x, x' \in D$. Then $\mathbb{E}_D^0[f] \geq 0$.*
2. *Let $u \in D$. Under $\mathbb{E}_D^\cdot$, the conditional law of n_u given $(n_x)_{x\in D\setminus\{u\}}$ has a density with respect to $\mathrm{d}n_u$ that is proportional to the product of the factors of the integrand of $Z_D^\cdot$ in*

(5.1) *that depend on n_u. If $u, v \in D$ are distinct and not neighbors, then n_u and n_v are conditionally independent under $\mathbb{E}_D^{\cdot}$ given $(n_x)_{x \in D \setminus \{u,v\}}$.*

Proof. We begin with the first part. We expand the product of the bond factors in (5.1), and we write each scalar product as a sum over the three coordinates. This expresses $Z_D^0 \mathbb{E}_D^0[f]$ as a finite sum of integrals of the form

$$\int \prod_{q=1}^{m} n_{x_q}^{\alpha_q} n_{x'_q}^{\alpha_q} \prod_{x \in D} \mathrm{d}n_x = \prod_{x \in D} \int \prod_{q \colon x_q = x} n^{\alpha_q} \prod_{q \colon x'_q = x} n^{\alpha_q} \, \mathrm{d}n,$$

where $m \geq 0$, $x_q, x'_q \in D$, and $\alpha_q \in \{1, 2, 3\}$. By (5.2), every factor on the right side is nonnegative. It follows that $\mathbb{E}_D^0[f] \geq 0$, since $Z_D^0 > 0$.

For the second part, recall that the density of $\mathbb{E}_D^{\cdot}$ is a product of factors, each of which depends on the spins of two neighbors or on the spin of a single site. Under the conditioning, the factors that do not depend on n_u are constant, which gives the first claim; the normalizing integral is positive because the integrand is positive almost everywhere. Now let u and v be distinct sites that are not neighbors. Then every factor depends on at most one of n_u and n_v, and the conditional density of (n_u, n_v) given the other spins is the product of a function of n_u and a function of n_v. □

The first part of Lemma 5.1 is in the spirit of the positivity of staggered correlations used for bipartite tori in [23]. Under $\mathbb{E}_D^\tau$, two isolated sites with one pin each pointing in opposite directions have a negative pair correlation.

We will also need the following bound on the conditional mean of a spin with at most three factors, which will be used in Section 6.

Lemma 5.2. *1. Let $v_1, v_2, v_3 \in \mathbb{B}$. Then the probability measure on $\mathbb{S}^2$ with density proportional to $\prod_{r=1}^3 (1 + n \cdot v_r)$ with respect to $\mathrm{d}n$ has mean of norm at most $3/5$.*

2. Let D be a domain in G, let τ be a pin configuration on G, and let $u \in D$. Suppose that at most three factors of the integrand of Z_D^τ in (5.1) depend on n_u. Then the conditional expectation of n_u under $\mathbb{E}_D^\tau$ given $(n_x)_{x \in D \setminus \{u\}}$ has norm at most $3/5$. If u has at most three neighbors in D, then the same holds under $\mathbb{E}_D^0$.

Proof. We first prove the first part when $v_1 = a$, $v_2 = b$, and $v_3 = c$ are unit vectors. We set

$$\gamma_1 = b \cdot c, \qquad \gamma_2 = a \cdot c, \qquad \gamma_3 = a \cdot b, \qquad \bar{\gamma} = \gamma_1 + \gamma_2 + \gamma_3,$$

and we expand the product $(1 + n \cdot a)(1 + n \cdot b)(1 + n \cdot c)$. By (5.2), the terms of odd degree in n have integral zero, and we have

$$\int n \, (n \cdot a) \, \mathrm{d}n = \frac{a}{3}, \qquad \int n \, (n \cdot a)(n \cdot b)(n \cdot c) \, \mathrm{d}n = \frac{1}{15} \big(a \, (b \cdot c) + b \, (a \cdot c) + c \, (a \cdot b) \big).$$

Integrating term by term, we obtain

$$\int (1+n\cdot a)(1+n\cdot b)(1+n\cdot c)\,\mathrm{d}n = 1+\frac{\bar\gamma}{3}$$

and

$$15\int n\,(1+n\cdot a)(1+n\cdot b)(1+n\cdot c)\,\mathrm{d}n = \mathbf{q}, \qquad \mathbf{q} = (5+\gamma_1)a+(5+\gamma_2)b+(5+\gamma_3)c.$$

Dividing the second identity by 15 times the first, we find that the measure in the first part has mean $\mathbf{q}/(5(3+\bar\gamma))$. Observe that

$$3+2\bar\gamma = |a+b+c|^2 \ge 0, \qquad 3+\bar\gamma \ge \frac{3}{2}.$$

It therefore suffices to show that $|\mathbf{q}| \le 3(3+\bar\gamma)$.

Since a, b, and c are unit vectors, we have

$$|\mathbf{q}|^2 = \sum_{r=1}^{3}(5+\gamma_r)^2 + 2(5+\gamma_1)(5+\gamma_2)\gamma_3 + 2(5+\gamma_1)(5+\gamma_3)\gamma_2 + 2(5+\gamma_2)(5+\gamma_3)\gamma_1.$$

Expanding the two sides, one can check the identity

$$9(3+\bar\gamma)^2 - |\mathbf{q}|^2 = 4\big[(\gamma_1-\gamma_2)^2+(\gamma_1-\gamma_3)^2+(\gamma_2-\gamma_3)^2\big] + 6(1-\gamma_1)(1-\gamma_2)(1-\gamma_3). \tag{5.3}$$

The right side of (5.3) is nonnegative, since $\gamma_r \le 1$ for all r, which completes the proof of the first part for unit vectors.

Next, let $v_1, v_2, v_3 \in \mathbb{B}$. For $r = 1,2,3$, we choose $e_r \in \mathbb{S}^2$ such that $v_r = |v_r|e_r$ (with e_r arbitrary if $v_r = 0$). Then

$$1+n\cdot v_r = \frac{1+|v_r|}{2}(1+n\cdot e_r) + \frac{1-|v_r|}{2}(1-n\cdot e_r).$$

Expanding the product of these three sums, we write the measure $\prod_{r=1}^{3}(1+n\cdot v_r)\,\mathrm{d}n$ as a combination with nonnegative coefficients of the eight measures $\prod_{r=1}^{3}(1+\epsilon_r n\cdot e_r)\,\mathrm{d}n$ with $\epsilon_r \in \{1,-1\}$. After normalization, the probability measure in the first part is a convex combination of the normalizations of these eight measures, whose means have norm at most $3/5$ by the case of unit vectors. Its mean is the corresponding convex combination of these means, and the first part follows.

For the second part, Lemma 5.1 shows that the conditional law of n_u given the other spins has a density proportional to the product of at most three factors $1+n_u\cdot w$, where each w is the spin of a neighbor of u in D or a pin at u. We add factors $1 = 1+n_u\cdot 0$ until there are three factors, and we apply the first part. Under $\mathbb{E}_D^0$, the factors that depend on n_u are those of the bonds from u to its neighbors in D, and the same argument applies. □

5.2 Deleting a site

For a domain D in G and a site $x \in D$, the set $D \setminus \{x\}$ is again a domain in G, and we say that it is obtained from D by *deleting* the site x. The following lemma computes the effect of a deletion on the partition functions when the deleted site has at most three neighbors in D.

Lemma 5.3. *Let D be a domain in G, and let $x \in D$ be a site whose four neighbors all belong to G. Suppose that x has $k \le 3$ neighbors $u_1, \dots, u_k$ in D, and set*

$$I_x = 1 + \frac{1}{3} \sum_{1 \le a < a' \le k} n_{u_a} \cdot n_{u_{a'}}. \tag{5.4}$$

Then $I_x \ge 1/2$. Further, for every pin configuration τ on G, we have

$$Z^{\cdot}_D = Z^{\cdot}_{D\setminus\{x\}}\, \mathbb{E}^{\cdot}_{D\setminus\{x\}}[I_x] \tag{5.5}$$

and

$$\begin{aligned} \bigl|\log \Xi_D(\tau) - \log \Xi_{D\setminus\{x\}}(\tau)\bigr| &= \bigl|\log \mathbb{E}^{\tau}_{D\setminus\{x\}}[I_x] - \log \mathbb{E}^{0}_{D\setminus\{x\}}[I_x]\bigr| \\ &\le \frac{2}{3} \sum_{1 \le a < a' \le k} \bigl|\mathbb{E}^{\tau}_{D\setminus\{x\}}[n_{u_a} \cdot n_{u_{a'}}] - \mathbb{E}^{0}_{D\setminus\{x\}}[n_{u_a} \cdot n_{u_{a'}}]\bigr|. \end{aligned} \tag{5.6}$$

Proof. Since all neighbors of x belong to G, the set G has no leg at x. The spin n_x therefore appears in the integrand of $Z^{\cdot}_D$ only through the k factors $1 + n_x \cdot n_{u_a}$. By (5.2), we have

$$\int \prod_{a=1}^{k} (1 + n \cdot n_{u_a})\, \mathrm{d}n = 1 + \sum_{1 \le a < a' \le k} \int (n \cdot n_{u_a})(n \cdot n_{u_{a'}})\, \mathrm{d}n = I_x,$$

where we used that the terms of degrees one and three have integral zero, and that there is no term of degree four, since $k \le 3$. The remaining factors of the integrand of $Z^{\cdot}_D$ are exactly those of $Z^{\cdot}_{D\setminus\{x\}}$, since the deleted site x does not become a pin. Integrating out n_x first, we arrive at (5.5).

Next, using $|n_{u_a}| = 1$ for all a, we have

$$\sum_{1 \le a < a' \le k} n_{u_a} \cdot n_{u_{a'}} = \frac{1}{2}\Bigl(\Bigl|\sum_{a=1}^{k} n_{u_a}\Bigr|^2 - k\Bigr) \ge -\frac{k}{2}, \qquad I_x \ge 1 - \frac{k}{6} \ge \frac{1}{2}.$$

The equality in (5.6) follows from dividing (5.5) with $\cdot = \tau$ by (5.5) with $\cdot = 0$ and taking logarithms.

For the inequality, note that by (5.4), we have

$$\mathbb{E}^{\tau}_{D\setminus\{x\}}[I_x] - \mathbb{E}^{0}_{D\setminus\{x\}}[I_x] = \frac{1}{3} \sum_{1 \le a < a' \le k} \bigl(\mathbb{E}^{\tau}_{D\setminus\{x\}}[n_{u_a} \cdot n_{u_{a'}}] - \mathbb{E}^{0}_{D\setminus\{x\}}[n_{u_a} \cdot n_{u_{a'}}]\bigr).$$

Since $I_x \ge 1/2$, these expectations belong to $[1/2, \infty)$, where the logarithm is Lipschitz with constant 2. The inequality in (5.6) follows. □

The differences on the right side of (5.6) are the central objects of the rest of the paper. For a domain D in G, a pin configuration τ on G, and sites $u, v \in D$, we call

$$\mathbb{E}_D^\tau[n_u \cdot n_v] - \mathbb{E}_D^0[n_u \cdot n_v]$$

the *response* of the pair (u, v) in D. In this language, Lemma 5.3 states that the pins enter the cost of deleting a site x with $k \le 3$ neighbors in D through the responses of the pairs of neighbors of x, taken in the domain $D \setminus \{x\}$. These pairs are at distance 2, and the deleted site is a removed common neighbor of each of them. Hypothesis (P) in Section 5.4 bounds responses of this kind. In this section, they bound the cost of removing a set that separates the window from the boundary. In Sections 6 and 7, we prove Hypothesis (P) by expanding each response into further responses of the same kind.

5.3 The cut lemma

Throughout this subsection, let $F \subset \mathbb{Z}^2$ be finite, let $W \subset F$ be a set such that every neighbor of a site of W belongs to F, and set $G = F \setminus W$. In the application in Section 5.6, we have $F = F_i$ and $W = W_i$, and we call W the window, as in Section 4. Recall that $\mu_{F,W}^\eta$ denotes the marginal of μ_F^η on W.

Every leg (x, z) of G satisfies $z \in W$ or $z \in \mathbb{Z}^2 \setminus F$. For $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2 \setminus F}$ and $\xi \in (\mathbb{S}^2)^W$, we let $\eta\xi$ denote the pin configuration on G given by

$$(\eta\xi)_{xz} = \xi_z \quad \text{if } z \in W, \qquad (\eta\xi)_{xz} = \eta_z \quad \text{if } z \notin F.$$

We define the internal weight of the window by

$$\Theta_W(\xi) = \prod_{\substack{\{w,w'\} \subset W \\ w \sim w'}} (1 + \xi_w \cdot \xi_{w'}),$$

and, for $X \subset G$, we define the *cut cost*

$$\mathrm{cost}_X(\eta, \xi) = \log \Xi_{G \setminus X}(\eta\xi) - \log \Xi_G(\eta\xi), \tag{5.7}$$

where $G \setminus X$ and G are domains in G. We call a set $X \subset G$ a *cut* if $G \setminus X$ is the disjoint union of two sets G_1 and G_2 such that no bond joins G_1 to G_2, every neighbor of a site of G_1 belongs to F, and every neighbor of a site of G_2 belongs to $\mathbb{Z}^2 \setminus W$.

Lemma 5.4. *Let $X \subset G$ be a cut, and suppose that there exists $C \ge 0$ such that $|\mathrm{cost}_X(\eta, \xi)| \le C$ for all $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2 \setminus F}$ and $\xi \in (\mathbb{S}^2)^W$. Then there exists a probability measure ν on $(\mathbb{S}^2)^W$ such that*

$$\mu_{F,W}^\eta \ge e^{-2C} \nu \qquad \textit{for all } \eta \in (\mathbb{S}^2)^{\mathbb{Z}^2 \setminus F}. \tag{5.8}$$

Proof. Let G_1 and G_2 be as in the definition of a cut. We first compute the density $f_\eta(\xi)$ of $\mu^\eta_{F,W}$ with respect to $\prod_{w\in W} \mathrm{d}\xi_w$. In the density of μ^η_F, we set $n_w = \xi_w$ for $w \in W$. The factors of the bonds inside W give $\Theta_W(\xi)$. Since every neighbor of a site of W belongs to F, every remaining factor belongs to a bond inside G or to a leg of G, and a leg (x,z) of G contributes the factor $1 + (\eta\xi)_{xz} \cdot n_x$. Integrating over $(n_x)_{x\in G}$ yields

$$f_\eta(\xi) = \frac{\Theta_W(\xi)\, Z_G^{\eta\xi}}{Z_F^\eta}. \tag{5.9}$$

Next, we factorize the partition functions of the domain $G \setminus X$. Since no bond joins G_1 to G_2, the integrand of $Z^{\eta\xi}_{G\setminus X}$ is the product of the factors of the bonds inside G_1, of the bonds inside G_2, and of the legs of G at sites of $G_1 \cup G_2$. Every leg of G at a site of G_1 ends in W, since every neighbor of a site of G_1 belongs to F. Similarly, every leg of G at a site of G_2 ends outside F. Let $Z^{\eta\xi}_{G_1}$ and $Z^{\eta\xi}_{G_2}$ denote the partition functions of the domains $G_1 = G \setminus (X \cup G_2)$ and $G_2 = G \setminus (X \cup G_1)$ in G. Then

$$Z^{\eta\xi}_{G\setminus X} = Z^{\eta\xi}_{G_1} Z^{\eta\xi}_{G_2}, \qquad Z^0_{G\setminus X} = Z^0_{G_1} Z^0_{G_2}, \tag{5.10}$$

where $Z^{\eta\xi}_{G_1}$ is a function of ξ alone and $Z^{\eta\xi}_{G_2}$ is a function of η alone.

By the definition (5.7) of the cut cost and by (5.10), we have

$$Z_G^{\eta\xi} = e^{-\mathrm{cost}_X(\eta,\xi)}\, \frac{Z^{\eta\xi}_{G\setminus X}\, Z^0_G}{Z^0_{G\setminus X}} = e^{-\mathrm{cost}_X(\eta,\xi)}\, Z^{\eta\xi}_{G_1}\, \frac{Z^{\eta\xi}_{G_2}\, Z^0_G}{Z^0_{G_1} Z^0_{G_2}}.$$

We set

$$N_1 = \int \Theta_W(\xi)\, Z^{\eta\xi}_{G_1} \prod_{w\in W} \mathrm{d}\xi_w, \qquad \bar f(\xi) = \frac{\Theta_W(\xi)\, Z^{\eta\xi}_{G_1}}{N_1}, \qquad c_\eta = \frac{N_1\, Z^{\eta\xi}_{G_2}\, Z^0_G}{Z^0_{G_1} Z^0_{G_2} Z^\eta_F}.$$

Since $Z^{\eta\xi}_{G_1}$ is a function of ξ alone, $N_1 > 0$ and the probability density $\bar f$ on $(\mathbb{S}^2)^W$ are the same for all η. Since $Z^{\eta\xi}_{G_2}$ is a function of η alone, so is $c_\eta > 0$. Inserting the last two displays into (5.9), we find

$$f_\eta(\xi) = c_\eta\, \bar f(\xi)\, e^{-\mathrm{cost}_X(\eta,\xi)}. \tag{5.11}$$

Integrating (5.11) over ξ and using the assumption that $|\mathrm{cost}_X| \le C$, we find

$$1 = c_\eta \int \bar f(\xi)\, e^{-\mathrm{cost}_X(\eta,\xi)} \prod_{w\in W} \mathrm{d}\xi_w \le c_\eta\, e^C.$$

We conclude that $c_\eta \ge e^{-C}$, and by (5.11), we have $f_\eta(\xi) \ge e^{-2C} \bar f(\xi)$ for all ξ. This is (5.8), with ν the probability measure with density $\bar f$. □

A pin fixes the spin at the other end of a leg, whereas deleting a site with at most three neighbors removes the factors of its bonds and nothing else; by Lemma 5.3, the effect of such a deletion on $\log \Xi$ is a combination of responses, which Hypothesis (P) below makes small far from the pins. This is why Lemma 5.4 compares the boundary configurations through the cost of removing a cut, and not by changing the pins on $\mathbb{Z}^2 \setminus F$ one at a time.

5.4 Hypothesis (P) and the removal lemma

For a finite set $G \subset \mathbb{Z}^2$ and a site $x \in \mathbb{Z}^2$, we call

$$\operatorname{dep}_G(x) = \operatorname{dist}(x, \mathbb{Z}^2 \setminus G)$$

the *depth* of x in G, and we define the boundary layer

$$\partial_2 G = \{z \in G : \operatorname{dep}_G(z) \le 2\}.$$

By the triangle inequality, we have $|\operatorname{dep}_G(x) - \operatorname{dep}_G(z)| \le d(x,z)$ for all $x, z \in \mathbb{Z}^2$. Further, $x \in G$ if and only if $\operatorname{dep}_G(x) \ge 1$, and all four neighbors of x belong to G if and only if $\operatorname{dep}_G(x) \ge 2$. In particular, every leg of G starts at a site of $\partial_2 G$.

Hypothesis (P). *There exist $A \ge 2$ and $\theta \in (0,1)$ such that the following holds. Let $G \subset \mathbb{Z}^2$ be finite, let τ be a pin configuration on G, let $Y \subset G \setminus \partial_2 G$, and let $u, v \in G \setminus Y$ be sites with $d(u,v) = 2$ that have a common neighbor in Y. Then*

$$\bigl|\mathbb{E}^{\tau}_{G\setminus Y}[n_u \cdot n_v] - \mathbb{E}^{0}_{G\setminus Y}[n_u \cdot n_v]\bigr| \le A\,\theta^{\min(\operatorname{dist}(u,\partial_2 G),\, \operatorname{dist}(v,\partial_2 G))}. \tag{5.12}$$

The left side of (5.12) is the absolute value of the response of the pair (u,v) in the domain $G \setminus Y$. Since $|n_u \cdot n_v| \le 1$, every response has absolute value at most 2, and the condition $A \ge 2$ makes (5.12) trivial when $u \in \partial_2 G$ or $v \in \partial_2 G$. Note that the distance in (5.12) is measured to $\partial_2 G$ and not to the removed set Y, although u and v are neighbors of a site of Y. Theorem 7.4 proves Hypothesis (P) from Hypotheses (H1), (H2), and (H3).

We now bound the cost of removing a connected set of sites far from $\mathbb{Z}^2 \setminus G$.

Lemma 5.5. *Assume Hypothesis (P), with constants A and θ. Let $G \subset \mathbb{Z}^2$ be finite, and let τ be a pin configuration on G. Let $X \subset G$ be a nonempty set, and let $x_1, \dots, x_{|X|}$ be an enumeration of X such that $\operatorname{dep}_G(x_j) \ge 3$ for all j and such that x_j has a neighbor in $\{x_1, \dots, x_{j-1}\}$ for all $j \ge 2$. Set $C_\star = \log 8192$. Then*

$$\bigl|\log \Xi_{G\setminus X}(\tau) - \log \Xi_G(\tau)\bigr| \le C_\star + 2A \sum_{j=2}^{|X|} \theta^{\operatorname{dep}_G(x_j)-3}. \tag{5.13}$$

Proof. For $0 \le j \le |X|$, we set $Y_j = \{x_1, \dots, x_j\}$ and $D_j = G \setminus Y_j$, so that $D_0 = G$, $D_{|X|} = G \setminus X$, and D_j is obtained from D_{j-1} by deleting x_j. Then

$$\log \Xi_G(\tau) - \log \Xi_{G\setminus X}(\tau) = \sum_{j=1}^{|X|} \bigl(\log \Xi_{D_{j-1}}(\tau) - \log \Xi_{D_j}(\tau)\bigr). \tag{5.14}$$

Note that all four neighbors of x_j belong to G, since $\mathrm{dep}_G(x_j) \ge 3$. They all belong to $D_0 = G$ when $j = 1$, while at most three of them belong to D_{j-1} when $j \ge 2$, since x_j has a neighbor in Y_{j-1}.

We first bound the summand with $j = 1$. Let $u_1, \dots, u_4$ be the neighbors of x_1, and set

$$\tilde{I}_{x_1} = \int \prod_{a=1}^{4} (1 + n \cdot n_{u_a}) \, \mathrm{d}n.$$

As in the proof of Lemma 5.3, integrating out n_{x_1} first gives

$$Z^{\cdot}_{D_0} = Z^{\cdot}_{D_1} \, \mathbb{E}^{\cdot}_{D_1}[\tilde{I}_{x_1}].$$

Each of the four factors in $\tilde{I}_{x_1}$ is at most 2, which implies $\tilde{I}_{x_1} \le 16$. For a unit vector b, the variable $1 + n \cdot b$ is uniformly distributed on $[0, 2]$ under $\mathrm{d}n$ (this is Archimedes' theorem). In particular, it is smaller than $1/4$ with probability $1/8$. By a union bound over the four factors, the set of n on which every factor is at least $1/4$ has measure at least $1/2$, which gives $\tilde{I}_{x_1} \ge 4^{-4}/2$. Therefore $\mathbb{E}^{\tau}_{D_1}[\tilde{I}_{x_1}]$ and $\mathbb{E}^{0}_{D_1}[\tilde{I}_{x_1}]$ belong to $[1/512, 16]$, and

$$\bigl|\log \Xi_{D_0}(\tau) - \log \Xi_{D_1}(\tau)\bigr| = \bigl|\log \mathbb{E}^{\tau}_{D_1}[\tilde{I}_{x_1}] - \log \mathbb{E}^{0}_{D_1}[\tilde{I}_{x_1}]\bigr| \le \log(16 \cdot 512) = C_\star.$$

Next, let $j \ge 2$. We apply Lemma 5.3 with $D = D_{j-1}$ and $x = x_j$, whose neighbors $u_1, \dots, u_k$ in D_{j-1} number $k \le 3$. This gives

$$\bigl|\log \Xi_{D_{j-1}}(\tau) - \log \Xi_{D_j}(\tau)\bigr| \le \frac{2}{3} \sum_{1 \le a < a' \le k} \bigl|\mathbb{E}^{\tau}_{D_j}[n_{u_a} \cdot n_{u_{a'}}] - \mathbb{E}^{0}_{D_j}[n_{u_a} \cdot n_{u_{a'}}]\bigr|. \tag{5.15}$$

We bound each of the at most three responses on the right side by Hypothesis (P), applied with $Y = Y_j$, $u = u_a$, and $v = u_{a'}$. Its assumptions are satisfied. Indeed, we have $Y_j \subset G \setminus \partial_2 G$ since $\mathrm{dep}_G(x_r) \ge 3$ for all r; the sites u_a and $u_{a'}$ belong to D_j and satisfy $d(u_a, u_{a'}) = 2$; and $x_j \in Y_j$ is a common neighbor of u_a and $u_{a'}$.

To bound the exponent in (5.12), let $z \in \partial_2 G$. Then

$$d(u_a, z) \ge \mathrm{dep}_G(u_a) - \mathrm{dep}_G(z) \ge \mathrm{dep}_G(x_j) - 1 - 2,$$

where we used that dep_G is Lipschitz with constant 1, that $\mathrm{dep}_G(u_a) \ge \mathrm{dep}_G(x_j) - 1$, and that $\mathrm{dep}_G(z) \le 2$. It follows that $\mathrm{dist}(u_a, \partial_2 G) \ge \mathrm{dep}_G(x_j) - 3$, and the same bound

holds for $u_{a'}$. Since $\theta < 1$, Hypothesis (P) bounds each response in (5.15) by $A\theta^{\mathrm{dep}_G(x_j)-3}$ in absolute value. Inserting this into (5.15), we find

$$\bigl|\log \Xi_{D_{j-1}}(\tau) - \log \Xi_{D_j}(\tau)\bigr| \le \frac{2}{3} \cdot 3 \cdot A\theta^{\mathrm{dep}_G(x_j)-3} = 2A\theta^{\mathrm{dep}_G(x_j)-3}.$$

Summing over j in (5.14), and using the bound for $j = 1$, we arrive at (5.13). □

The first removal costs up to $C_\star = \log 8192 < 10$, which is not small. This is harmless, since Lemma 5.4 requires a bounded cut cost, not a small one.

5.5 The cut for a shell

In this subsection, we construct a cut for the window of a shell, and we bound the depths of the sites of the cut. Throughout this subsection, we fix $M \ge M_{\mathrm{g}}$, a site y with $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$, and an index $i \in \{2, \dots, \lfloor \ell \rfloor\} \setminus I_{\mathrm{corner}}$, and we write $k = k_i = \lfloor i\ell \rfloor$. Recall the sets F_i, Σ_i, Σ_i^∂, and W_i from Section 4. We set

$$F = F_i, \qquad W = W_i, \qquad G = F \setminus W.$$

Recall that $M_{\mathrm{g}} = 73^4$, so that $\ell = M^{1/4} \ge 73$. Note that $|y|_\infty = M + 1$, that $2\ell - 1 \le k \le \ell^2$, and that $\Sigma_i = \{x \in \mathcal{C}_M : |x - y|_\infty = k\}$, since $|x - y|_\infty$ is an integer.

For $x \in \mathbb{Z}^2$, we define the *offset* of x from the shell and the *height* of x above the boundary of the box by

$$\mathrm{off}(x) = \bigl||x - y|_\infty - k\bigr|, \qquad \mathrm{hgt}(x) = \mathrm{dist}_\infty(x, \mathbb{Z}^2 \setminus \mathcal{C}_M).$$

We have $\mathrm{hgt}(x) = M + 1 - |x|_\infty$ for $x \in \mathcal{C}_M$, and $\mathrm{hgt}(x) = 0$ for $x \notin \mathcal{C}_M$. By the triangle inequality for $|\cdot|_\infty$, we have

$$|\mathrm{off}(x) - \mathrm{off}(z)| \le |x - z|_\infty, \qquad |\mathrm{hgt}(x) - \mathrm{hgt}(z)| \le |x - z|_\infty \qquad \text{for all } x, z \in \mathbb{Z}^2. \tag{5.16}$$

Recall that $d_0 = 24$. In this notation, $\Sigma_i = \{x \in \mathcal{C}_M : \mathrm{off}(x) = 0\}$ and $W = \{x \in \Sigma_i : \mathrm{hgt}(x) \ge d_0 + 1\}$. We define

$$U = \bigl\{x \in \mathcal{C}_M : \mathrm{off}(x) \le \min\bigl(\mathrm{hgt}(x) - d_0/2,\, \ell/12\bigr)\bigr\}. \tag{5.17}$$

The set U is a band of half-width $\ell/12$ around the shell, which near the boundary of $\mathcal{C}_M$ tapers linearly and ends at height $d_0/2 = 12$ (see Figure 5.1).

We call two sites x and z *$*$-neighbors* if $|x - z|_\infty = 1$. A sequence of sites in which consecutive sites are $*$-neighbors is a *$*$-path*, and a set $A \subset \mathbb{Z}^2$ is *$*$-connected* if any two of its sites are joined by a $*$-path in A. A set is *connected* if any two of its sites are joined by a nearest-neighbor path in it. For $A \subset \mathbb{Z}^2$, we let

$$\partial_\infty A = \{z \in \mathbb{Z}^2 \setminus A : \mathrm{dist}_\infty(z, A) = 1\}$$

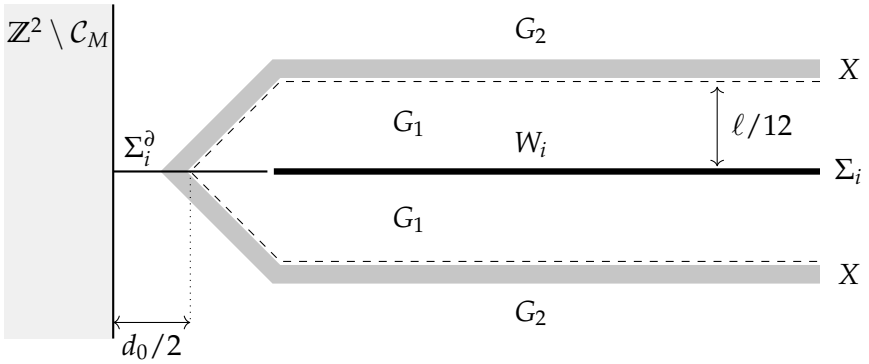


Figure 5.1. A schematic picture, not to scale, of one end of the shell Σ_i near the boundary of the box $\mathcal{C}_M$. The shell is the horizontal segment; its part within ℓ^∞ distance d_0 of $\mathbb{Z}^2 \setminus \mathcal{C}_M$ is Σ_i^∂, and the rest is the window W_i (thick). The region U (dashed boundary) is a band of half-width $\ell/12$ around the shell that tapers linearly and ends at height $d_0/2$. The cut X (shaded) surrounds U and separates G_1, which lies in U, from G_2, which contains the sites of G near the boundary of the box.

denote the outer ℓ^∞ boundary of A. We define

$$\begin{aligned} X_0 &= \partial_\infty U, & X &= \{z \in \mathbb{Z}^2 : \operatorname{dist}(z, X_0) \le 1\}, \\ G_1 &= U \setminus (W \cup X), & G_2 &= G \setminus (U \cup X). \end{aligned} \tag{5.18}$$

The set X_0 surrounds U, but it need not be connected. The set X adds to X_0 the nearest neighbors of its sites, and Lemma 5.8 shows that X is connected. We need this in order to remove X in an order to which Lemma 5.5 applies. The two topological facts that we use, Lemmas A.1 and A.2, are proved in Appendix A, independently of this section.

The shape of U is dictated by the ends of the window. Away from the ends, a band of constant half-width $\ell/12$ keeps the cut at depth of order ℓ. However, the window comes within ℓ^∞ distance $d_0 + 1$ of the pins on $\mathbb{Z}^2 \setminus \mathcal{C}_M$, and near its ends the cut must pass at a bounded depth. Because U tapers, the depth of a site of X_0 near the boundary of $\mathcal{C}_M$ grows linearly with its offset, and Lemma 5.6 shows that a bounded number of sites of X_0 have a given offset. The sum of the bounds of Lemma 5.5 over X is therefore bounded uniformly in M, which is the content of Lemma 5.7. This count near the ends of the window is the only delicate point of the construction.

For the counting, recall from Section 4.2 the sphere $\{x \in \mathbb{Z}^2 : |x - a|_\infty = r\}$ of radius r around a site a and its four sides, the segments $\{a_1 \pm r\} \times [a_2 - r, a_2 + r]$ and $[a_1 - r, a_1 + r] \times \{a_2 \pm r\}$ of lattice points. Thus Σ_i is the part in $\mathcal{C}_M$ of the sphere of radius k around y, and for $1 \le h \le M$, the set $\{x \in \mathcal{C}_M : \operatorname{hgt}(x) = h\}$ is the sphere of radius $M + 1 - h$ around 0.

Lemma 5.6. *Let j and h be integers with $0 \le j \le \ell/12$ and $j + 10 \le h \le j + 11$. Then at most 16 sites $x \in \mathcal{C}_M$ satisfy $\operatorname{off}(x) = j$ and $\operatorname{hgt}(x) = h$.*

Proof. Let Q' denote the sphere of radius $M+1-h$ around 0. Since $\ell \geq 73$, we have

$$1 \leq h \leq \frac{\ell}{12} + 11 \leq M, \qquad Q' = \{x \in \mathcal{C}_M : \mathrm{hgt}(x) = h\}.$$

Every site x in the lemma belongs to Q' and to one of the spheres of radius $k-j$ and $k+j$ around y. It therefore suffices to show that, for $r \in \{k-j, k+j\}$, the sphere Q of radius r around y meets Q' in at most 8 sites. Note that

$$r \geq k-j \geq 2\ell - 1 - \ell/12 \geq \ell/4, \qquad r \leq k+j \leq \ell^2 + \ell/12 \leq M. \tag{5.19}$$

We claim that it suffices to show that no side of Q that meets $\mathcal{C}_M$ lies on a line containing a side of Q'. Indeed, suppose that this holds. A side of Q that does not meet $\mathcal{C}_M$ does not meet $Q' \subset \mathcal{C}_M$. A side of Q that meets $\mathcal{C}_M$ is parallel to two sides of Q' and lies on a different line from each of them. It meets these two sides in no site. Further, it meets each of the two perpendicular sides of Q' in at most one site. Hence every side of Q meets Q' in at most two sites, and Q meets Q' in at most 8 sites.

To verify the condition, we use the symmetries of the problem. The rotations and reflections of $\mathbb{Z}^2$ that fix 0 preserve $\mathcal{C}_M$, hgt, and ℓ^∞ distances, and they map spheres to spheres and sides to sides. The set I_{corner} and the conclusion of part (b) of Lemma 4.3 are invariant under them (see the proof of Lemma 4.3). Since $|y|_\infty = M+1$, we may therefore assume that $y_2 = -M-1$ and $|y_1| \leq M+1$.

For $c \in [-M-1, M+1]$, the vertical line $\{x_1 = c\}$ has ℓ^∞ distance $M+1-c$ from the half-plane $\{x_1 \geq M+1\}$ and $M+1+c$ from the half-plane $\{x_1 \leq -M-1\}$. It contains a side of Q' if and only if one of these two distances is h. The same holds for horizontal lines, with x_2 in place of x_1. We also have

$$j < h \leq \ell/12 + 11 < \ell/4, \tag{5.20}$$

where we used $\ell \geq 73$ in the last inequality. It suffices to show that every side of Q that meets $\mathcal{C}_M$ has distance at most j or at least $\ell/4$ from each of the two half-planes whose boundary lines are parallel to it. We consider the four sides in turn.

The bottom side of Q lies on the line $\{x_2 = -M-1-r\}$, so it does not meet $\mathcal{C}_M$. The top side lies on the line $\{x_2 = -M-1+r\}$, whose distances from the half-planes $\{x_2 \leq -M-1\}$ and $\{x_2 \geq M+1\}$ are r and $2M+2-r$. By (5.19), these distances are at least $\ell/4$.

The right side of Q lies on the line $\{x_1 = y_1 + r\}$. Its distance from the half-plane $\{x_1 \leq -M-1\}$ is $y_1 + r + M + 1 \geq r \geq \ell/4$. To bound its distance $M+1-y_1-r$ from the half-plane $\{x_1 \geq M+1\}$, we consider the right side of the sphere of radius k around y, which lies on the line $\{x_1 = y_1 + k\}$, and we distinguish two cases.

Suppose first that $y_1 + k \leq M$. Since $y_1 \geq -M-1$ and $k \geq 1$, we have

$$-M \leq y_1 + k \leq M, \qquad (y_1 + k, -M) \in \mathcal{C}_M,$$

and the right side $\{x \in \mathbb{Z}^2 : x_1 = y_1 + k,\ |x_2 - y_2| \le k\}$ of the sphere of radius k contains the site $(y_1 + k, -M)$. In particular, this side meets $\mathcal{C}_M$, and part (b) of Lemma 4.3, which applies since $i \notin I_{\text{corner}}$, gives $|y_1 + k| \le M + 1 - \ell/3$, that is, $M + 1 - y_1 - k \ge \ell/3$. Using $r \le k + j$ and $j \le \ell/12$, we obtain

$$M + 1 - y_1 - r \ge M + 1 - y_1 - k - j \ge \frac{\ell}{3} - \frac{\ell}{12} = \frac{\ell}{4}.$$

Suppose next that $y_1 + k \ge M + 1$. Then

$$M + 1 - y_1 - r \le M + 1 - y_1 - k + j \le j.$$

In both cases, the distance is at most j or at least $\ell/4$.

The reflection $x_1 \mapsto -x_1$ preserves our normalization $y_2 = -M - 1$. It maps the left side of Q to the right side of the sphere of radius r around $(-y_1, y_2)$. The argument for the right side therefore applies to the left side as well. This completes the proof. □

Lemma 5.7. *Let M, y, and i satisfy the standing assumptions of this subsection, and let U, X_0, X, G_1, and G_2 be given by* (5.17) *and* (5.18).

1. *We have $W \subset U \subset F$. Every neighbor of a site of W belongs to U, and every neighbor of a site of U belongs to F.*
2. *Every $x \in X_0$ satisfies $\deg_G(x) \ge 4$, every $x \in X$ satisfies $\deg_G(x) \ge 3$, and for all $t \in (0,1)$, we have*

$$\sum_{x \in X} t^{\deg_G(x)} \le 240\, t^3 \Big(7 + \frac{1}{1-t}\Big) + 5(2M+1)^2\, t^{\ell/12 - 4}. \tag{5.21}$$

3. *We have $X \subset G$, and X is a cut, with $G \setminus X$ the disjoint union of the sets G_1 and G_2 in* (5.18).

Proof. We prove the three parts in turn. Recall from (5.17) that every $x' \in U$ satisfies $x' \in \mathcal{C}_M$, $\operatorname{off}(x') \le \ell/12$, and $\operatorname{hgt}(x') \ge \operatorname{off}(x') + 12$.

The first part. Let $x \in W$. Then $\operatorname{off}(x) = 0$ and $\operatorname{hgt}(x) \ge 25$, which implies $x \in U$. Let z be a neighbor of x. By (5.16), we have $\operatorname{off}(z) \le 1$ and $\operatorname{hgt}(z) \ge 24$, so that $z \in \mathcal{C}_M$ and

$$\operatorname{off}(z) \le 1 \le \min\big(\operatorname{hgt}(z) - 12,\ \ell/12\big),$$

where we used $\ell \ge 12$. This shows that $z \in U$.

Now let $x \in U$. Then $\operatorname{hgt}(x) \ge 12$, and every neighbor of x belongs to $\mathcal{C}_M$. Every site $z \in \mathcal{C}_M \setminus F$ satisfies

$$|z - y|_\infty \le (i-1)\ell = i\ell - \ell < k + 1 - \ell, \tag{5.22}$$

where we used $k > i\ell - 1$. Suppose that x had a neighbor $z \in \mathcal{C}_M \setminus F$. Then (5.16) and (5.22) would give

$$\operatorname{off}(x) \ge k - |x - y|_\infty \ge k - |z - y|_\infty - 1 > \ell - 2,$$

which contradicts $\mathrm{off}(x) \le \ell/12$. Hence every neighbor of x belongs to F. Finally, we have

$$|x-y|_\infty \ge k - \mathrm{off}(x) \ge k - \ell/12 > i\ell - 1 - \ell/12 > (i-1)\ell,$$

which shows that $x \in F$. This proves the first part.

A lower bound for the depth. We now turn to the second part. Let $x \in X_0$, and set $j = \mathrm{off}(x)$. We have

$$\mathbb{Z}^2 \setminus G = (\mathbb{Z}^2 \setminus \mathcal{C}_M) \cup (\mathcal{C}_M \setminus F) \cup W,$$

and the graph distance dominates the ℓ^∞ distance. Therefore $\mathrm{dep}_G(x)$ is at least the smallest of the ℓ^∞ distances from x to these three sets, which satisfy

$$\mathrm{dist}_\infty(x, \mathbb{Z}^2 \setminus \mathcal{C}_M) = \mathrm{hgt}(x), \qquad \mathrm{dist}_\infty(x, \mathcal{C}_M \setminus F) > \ell - 1 - j,$$
$$\mathrm{dist}_\infty(x, W) \ge \max(j, 25 - \mathrm{hgt}(x)).$$

The first identity is the definition of hgt. The second bound follows from (5.22), since $|x-z|_\infty \ge |x-y|_\infty - |z-y|_\infty$ and $|x-y|_\infty \ge k - j$. The third follows from (5.16), since every $w \in W$ satisfies $\mathrm{off}(w) = 0$ and $\mathrm{hgt}(w) \ge 25$. We conclude that

$$\mathrm{dep}_G(x) \ge \min\bigl\{\max\bigl(j,\, 25 - \mathrm{hgt}(x)\bigr),\, \mathrm{hgt}(x),\, \ell - 1 - j\bigr\}. \tag{5.23}$$

The neighbor in U. Since $x \in \partial_\infty U$, there exists $x' \in U$ such that $|x - x'|_\infty = 1$. Using $\mathrm{off}(x') \le \ell/12$ and $\mathrm{hgt}(x') \ge \mathrm{off}(x') + 12$, together with (5.16), we obtain

$$j \le \ell/12 + 1, \qquad \mathrm{hgt}(x) \ge \mathrm{hgt}(x') - 1 \ge \mathrm{off}(x') + 11 \ge j + 10. \tag{5.24}$$

In particular, $\mathrm{hgt}(x) > 0$, and $X_0 \subset \mathcal{C}_M$. We now distinguish two cases.

Sites far from the shell. Suppose that $j > \ell/12$. By (5.24), we have $\mathrm{hgt}(x) > \ell/12 + 10$ and

$$\ell - 1 - j \ge \ell - 2 - \ell/12 > \ell/12.$$

Inserting these bounds into (5.23), we find $\mathrm{dep}_G(x) > \ell/12$. Since $X_0 \subset \mathcal{C}_M$, there are at most $(2M+1)^2$ such sites, and

$$\sum_{\substack{x \in X_0 \\ \mathrm{off}(x) > \ell/12}} t^{\mathrm{dep}_G(x)} \le (2M+1)^2\, t^{\ell/12}. \tag{5.25}$$

Sites near the shell. Suppose that $j \le \ell/12$. Since $x \in \mathcal{C}_M \setminus U$ and $\mathrm{off}(x) \le \ell/12$, the definition (5.17) implies that $\mathrm{off}(x) > \mathrm{hgt}(x) - 12$, that is, $\mathrm{hgt}(x) \le j + 11$. Together with (5.24), this gives

$$\mathrm{hgt}(x) \in \{j + 10,\, j + 11\}. \tag{5.26}$$

Since $j \le \ell/12$ and $\ell \ge 12$, we have

$$\ell - 1 - j \ge \frac{11\ell}{12} - 1 \ge \frac{\ell}{12} + 6.$$

Inserting this bound and (5.26) into (5.23), we find

$$\operatorname{dep}_G(x) \geq \min\bigl\{\max(j, 14-j),\, j+10,\, \ell/12+6\bigr\} \geq \max(j, 6). \tag{5.27}$$

We now count these sites. By (5.26) and Lemma 5.6, applied with $h = j+10$ and $h = j+11$, at most 32 sites of X_0 have offset j, for each integer $j \in [0, \ell/12]$. Combining this with (5.27), we find

$$\sum_{\substack{x \in X_0 \\ \operatorname{off}(x) \leq \ell/12}} t^{\operatorname{dep}_G(x)} \leq 32 \sum_{j \geq 0} t^{\max(j,6)} = 32\Bigl(7t^6 + \frac{t^7}{1-t}\Bigr) \leq 32\, t^6\Bigl(7 + \frac{1}{1-t}\Bigr). \tag{5.28}$$

Conclusion of the second part. In the two cases, we have shown that $\operatorname{dep}_G(x) \geq 6$ for all $x \in X_0$. Every site of X is within graph distance 1 of a site of X_0, and dep_G is Lipschitz with constant 1, which implies $\operatorname{dep}_G(z) \geq 5$ for all $z \in X$. Adding (5.25) and (5.28), we find

$$\sum_{x \in X_0} t^{\operatorname{dep}_G(x)} \leq 32\, t^6\Bigl(7 + \frac{1}{1-t}\Bigr) + (2M+1)^2\, t^{\ell/12}.$$

The set $X_0 \subset \mathcal{C}_M$ is finite, and the second part of Lemma A.2 gives

$$\sum_{z \in X} t^{\operatorname{dep}_G(z)} \leq 5t^{-1} \sum_{x \in X_0} t^{\operatorname{dep}_G(x)} \leq 160\, t^5\Bigl(7 + \frac{1}{1-t}\Bigr) + 5(2M+1)^2\, t^{\ell/12-1}.$$

Since $0 < t < 1$, this implies (5.21). The bounds on the depths in the second part hold as well.

The third part. By the second part, every site of X has positive depth, which means that $X \subset G$. By the first part, $U \subset F$, and therefore $U \cap G = U \setminus W$. It follows that $G_1 = (U \cap G) \setminus X$ and $G_2 = (G \setminus U) \setminus X$ partition $G \setminus X$.

Let $x \in G_1$, and let z be a neighbor of x with $z \notin U$. Then $\operatorname{dist}_\infty(z, U) = 1$, so that $z \in X_0 \subset X$ and $z \notin G_2$. This shows that no bond joins G_1 to G_2.

By the first part, every neighbor of a site of $G_1 \subset U$ belongs to F, and every neighbor of a site of W belongs to U. Since G_2 is disjoint from U, every neighbor of a site of G_2 belongs to $\mathbb{Z}^2 \setminus W$, and X is a cut. □

The proof gives better constants than those in (5.21). We state (5.21) in the form used in Appendix D.

Lemma 5.8. *Let M, y, and i satisfy the standing assumptions of this subsection. Then the set X of* (5.18) *is connected.*

Proof. We will show that U is nonempty and $*$-connected, and that $\mathbb{Z}^2 \setminus U$ is connected, hence $*$-connected. Since $U \subset \mathcal{C}_M$ is finite, Lemma A.1 then shows that $X_0 = \partial_\infty U$ is

nonempty and $*$-connected, and the first part of Lemma A.2 shows that X is connected. Observe that

$$U = \big\{x \in \mathbb{Z}^2 : \mathrm{off}(x) \le \ell/12,\ \mathrm{hgt}(x) - \mathrm{off}(x) \ge 12\big\}, \tag{5.29}$$

since $\mathrm{hgt}(x) \ge 12$ implies $x \in \mathcal{C}_M$. We set $m = M - 11$, so that $\mathrm{hgt}(x) \ge 12$ if and only if $|x|_\infty \le m$. Then

$$12 \le 2\ell - 1 \le k \le \ell^2 < m. \tag{5.30}$$

Moving towards the shell. Let $x \in U$. Note that $x \ne y$, since $y \notin \mathcal{C}_M$. Suppose that $|x - y|_\infty > k$. We move every coordinate x_a with $|x_a - y_a| = |x - y|_\infty$ one step towards y_a. This is a $*$-step that decreases $|x - y|_\infty$ by one. Suppose instead that $|x - y|_\infty < k$. We move one coordinate x_a with $|x_a - y_a| = |x - y|_\infty$ one step away from y_a. This is a nearest-neighbor step that increases $|x - y|_\infty$ by one.

After each of these steps, off has decreased by one and hgt has changed by at most one. The difference hgt $-$ off has not decreased, and the new site belongs to U by (5.29). Repeating these steps, we join x inside U to a site of

$$K = \{x \in \mathbb{Z}^2 : |x - y|_\infty = k,\ |x|_\infty \le m\} = \{x \in \mathbb{Z}^2 : |x - y|_\infty = k,\ \mathrm{hgt}(x) \ge 12\},$$

and $K \subset U$ by (5.29). It remains to show that K is nonempty and connected.

The set K. For $a \in \{1, 2\}$, the interval $[y_a - k, y_a + k]$ meets $[-m, m]$, since $|y_a| \le M + 1$ and $k \ge 12 = M + 1 - m$. Further, at least one of the endpoints $y_a - k$ and $y_a + k$ belongs to $[-m, m]$. Otherwise we would have $y_a - k < -m$ and $y_a + k > m$, which implies $k > m$ and contradicts (5.30). We choose such an endpoint c_a for $a = 1, 2$. Then $c = (c_1, c_2)$ is a corner of the square $y + [-k, k]^2$, and $c \in K$.

The intersection of the square $y + [-k, k]^2$ with $[-m, m]^2$ is the product of the two intervals $[y_a - k, y_a + k] \cap [-m, m]$, each of which contains c_a. The two sides of the square through c meet $[-m, m]^2$ in segments that contain c. Each of the two other sides meets $[-m, m]^2$, if at all, in a segment that contains its common corner with one of the first two sides, and this corner belongs to K. Thus K is a union of at most four segments, each of which meets one of the two segments through c, and K is connected.

The complement of U. The set $\mathbb{Z}^2 \setminus \mathcal{C}_M$ is connected and disjoint from U. Let $x \in \mathcal{C}_M \setminus U$. We join x to $\mathbb{Z}^2 \setminus \mathcal{C}_M$ by a nearest-neighbor path in $\mathbb{Z}^2 \setminus U$, and we consider three cases.

Suppose first that $|x - y|_\infty > k + \ell/12$. We repeatedly move a coordinate x_a with $|x_a - y_a| = |x - y|_\infty$ one step away from y_a. Each step increases $|x - y|_\infty$, and off stays larger than $\ell/12$ until the path leaves $\mathcal{C}_M$. In particular, the path stays outside U.

Suppose next that $|x - y|_\infty < k - \ell/12$. We repeatedly move a coordinate x_a with $|x_a - y_a| = |x - y|_\infty$ one step towards y_a. The value of $|x - y|_\infty$ never increases, and off stays larger than $\ell/12$ until the path reaches $y \notin \mathcal{C}_M$. In particular, the path stays outside U.

Otherwise, we have $\mathrm{off}(x) \le \ell/12$, and (5.29) gives $\mathrm{hgt}(x) - \mathrm{off}(x) < 12$. Then $x \neq 0$. Indeed, if $\mathrm{off}(0) \le \ell/12$, then

$$\mathrm{hgt}(0) - \mathrm{off}(0) \ge M + 1 - \frac{\ell}{12} > 12.$$

We repeatedly move a coordinate x_a with $|x_a| = |x|_\infty$ one step away from 0. Until the path leaves $\mathcal{C}_M$, each step decreases hgt by one and changes off by at most one. Thus hgt $-$ off does not increase, and the path stays outside U by (5.29).

In all three cases, the path ends in $\mathbb{Z}^2 \setminus \mathcal{C}_M$, and we conclude that $\mathbb{Z}^2 \setminus U$ is connected. This completes the proof. □

5.6 Proof of the window estimate

We now prove the main result of this section.

Theorem 5.9. *Assume Hypothesis (P), with constants $A \ge 2$ and $\theta \in (0,1)$, and set*

$$C_{\mathrm{cut}} = C_\star + 480A\Big(7 + \frac{1}{1-\theta}\Big) + 10A, \tag{5.31}$$

with $C_\star = \log 8192$ as in Lemma 5.5. Then Hypothesis (W) holds with $p_0 = e^{-2C_{\mathrm{cut}}}$ and with a constant $M_\star$ that depends only on θ.

Proof. Since $(2M+1)^2\theta^{M^{1/4}/12-7} \to 0$ as $M \to \infty$, there exists M_1, depending only on θ, such that

$$(2M+1)^2\,\theta^{M^{1/4}/12-7} \le 1 \qquad \text{for all } M \ge M_1. \tag{5.32}$$

We set

$$M_\star = \max(M_{\mathrm{g}}, M_1).$$

Let $M \ge M_\star$, let y satisfy $\mathrm{dist}_\infty(y, \mathcal{C}_M) = 1$, and let $i \in \{2, \dots, \lfloor \ell \rfloor\} \setminus I_{\mathrm{corner}}$, where $\ell = M^{1/4}$. Hypothesis (W) asks for a probability measure $\nu_{M,y,i}$ on $(\mathbb{S}^2)^{W_i}$ such that $\mu^{\eta}_{F_i,W_i} \ge p_0\nu_{M,y,i}$ for all $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2\setminus F_i}$.

Since $M \ge M_{\mathrm{g}}$, the standing assumptions of Section 5.5 hold. We set $F = F_i$, $W = W_i$, and $G = F \setminus W$, and we define U, X_0, X, G_1, and G_2 by (5.17) and (5.18). By the first part of Lemma 5.7, every neighbor of a site of W belongs to F, and the setting of Section 5.3 applies. By the third part of Lemma 5.7, the set X is a cut.

We now bound the cut cost. Fix $\eta \in (\mathbb{S}^2)^{\mathbb{Z}^2\setminus F}$ and $\xi \in (\mathbb{S}^2)^W$. By Lemma 5.8, the finite set X is connected. We fix $x_1 \in X$, and we enumerate the sites of X as $x_1, \dots, x_{|X|}$ in order of nondecreasing length of the shortest nearest-neighbor path in X from x_1 (a breadth-first order). Then every x_j with $j \ge 2$ has a neighbor among $x_1, \dots, x_{j-1}$, and

$\mathrm{dep}_G(x_j) \geq 3$ for all j by the second part of Lemma 5.7. Applying Lemma 5.5 with $\tau = \eta\xi$, we obtain

$$|\mathrm{cost}_X(\eta,\xi)| \leq C_\star + 2A \sum_{j=2}^{|X|} \theta^{\mathrm{dep}_G(x_j)-3} \leq C_\star + 2A\theta^{-3} \sum_{x\in X} \theta^{\mathrm{dep}_G(x)}.$$

Using (5.21) with $t = \theta$, together with (5.32), we find

$$|\mathrm{cost}_X(\eta,\xi)| \leq C_\star + 480A\Big(7 + \frac{1}{1-\theta}\Big) + 10A(2M+1)^2\,\theta^{\ell/12-7} \leq C_{\mathrm{cut}}. \tag{5.33}$$

By (5.33), the cut X satisfies the assumption of Lemma 5.4 with $C = C_{\mathrm{cut}}$. This lemma implies that there exists a probability measure ν on $(\mathbb{S}^2)^W$ such that $\mu^\eta_{F,W} \geq e^{-2C_{\mathrm{cut}}}\nu$ for all η. Setting $\nu_{M,y,i} = \nu$ proves Hypothesis (W) with $p_0 = e^{-2C_{\mathrm{cut}}}$. The constant $M_\star$ depends only on θ, since M_{g} is an absolute constant. □

Remark 5.10. The proof of Theorem 5.9 uses the exponential form of (5.12) only through the sum over X. Suppose that the right side of (5.12) is replaced by $f(\min(\mathrm{dist}(u,\partial_2 G), \mathrm{dist}(v,\partial_2 G)))$ for a nonincreasing function $f\colon \{0,1,2,\dots\} \to [0,\infty)$. Then the proofs of Lemma 5.5 and Theorem 5.9 bound the cut cost by $C_\star + 2\sum_{x\in X} f(\mathrm{dep}_G(x) - 3)$. By the proof of Lemma 5.7, the sites of X within distance 1 of a site $x \in X_0$ with $\mathrm{off}(x) \leq \ell/12$ contribute a bounded amount if f is summable. The remaining sites, at most $5(2M+1)^2 \leq 45\ell^8$ in number, have depth larger than $\ell/12 - 1$, and they contribute a bounded amount if $f(d) = O(d^{-8})$. Theorem 7.4 gives exponential decay, so we do not need this generality.

SECTION 6

Walk expansions

In this section and in Section 7, we prove that the hypotheses (H1), (H2), and (H3) imply Hypothesis (P); this implication is Theorem 7.4. The three hypotheses are verified with computer assistance in Section 8, and Hypothesis (P) is the input of Theorem 5.9. In this section, we expand a pair correlation of the model with pins of Section 5 into a sum over walks that start at one site of the pair and stop at the other site or near the boundary of the base set. Each step of a walk integrates out one spin and changes the domain of integration. We show that the pins change the weight of a step only through the responses of pairs of neighbors of the site that the walk leaves, and these pairs are of the same kind as the pairs in Hypothesis (P). We also state the local hypothesis (H1), which bounds the weight of a single step, and the growth hypothesis (H2) for the walks. The heuristics behind the argument are explained in Section 2.

6.1 Domains

Throughout this section and Section 7, $G \subset \mathbb{Z}^2$ denotes a finite set and τ denotes a pin configuration on G. Recall from Section 5 the legs of G, the depth $\mathrm{dep}_G(x) = \mathrm{dist}(x, \mathbb{Z}^2 \setminus G)$, and the boundary layer $\partial_2 G = \{z \in G : \mathrm{dep}_G(z) \le 2\}$. Recall also that the site of every leg has depth one, and in particular belongs to $\partial_2 G$.

Each step of our expansion modifies the domain of integration. We record these modifications with Definition 6.1 below, which extends the notion of a domain in G of Section 5.1. Before stating it, we define the measures that the sites will carry. We regard the uniform measure $\mathrm{d}n$ on $\mathbb{S}^2$ as a probability measure on $\mathbb{S}^2 \cup \{0\}$ with $\mathrm{d}n(\{0\}) = 0$, we let δ_0 denote the unit mass at $0 \in \mathbb{R}^3$, and we define the probability measures

$$\lambda_0 = \mathrm{d}n, \qquad \lambda_1 = \frac{3}{5}\,\mathrm{d}n + \frac{2}{5}\delta_0 \tag{6.1}$$

on $\mathbb{S}^2 \cup \{0\}$.

Definition 6.1. A *domain* D on G consists of a *status* $\mathrm{st}_D(x) \in \{0, 1, 2\}$ for each site $x \in G$ and a set of bonds $\{x, y\}$ with $x, y \in G$, which we call the *used* bonds of D. As in Section 5.1, we call a site of status 2 removed. We call a bond $\{x, y\}$ with $x, y \in G$ *present* in D if it is not used and neither x nor y is removed. We require that every site of $\partial_2 G$ has status 0 and that every site of status 1 has at most two present bonds. We let $\mathcal{D}_{\mathrm{cons}}(G)$

denote the set of domains on G in which every site of status 0 has no used bond and every site of status 1 has exactly two used bonds, and we call these domains *consistent*.

The statuses count the passages of the walks of our expansion through a site. A site of status 0 has not been visited, a site of status 1 has been visited once, and a site of status 2 has been visited twice or deleted. A walk uses the bonds that it traverses, and a used bond counts at its two ends. When a walk passes once through a site, it uses two bonds at this site, and the measure λ_1 will replace the factors of these two bonds. (The weight $3/5$ of λ_1 on $\mathbb{S}^2$ will produce the factor $3/5$ of a revisit in Proposition 6.6.)

In a domain D, each non-removed site x carries a spin $n_x \in \mathbb{S}^2 \cup \{0\}$. For a bounded measurable function f of these spins, we define

$$Z_D^\tau[f] = \int f \prod_{\{x,y\} \text{ present}} (1 + n_x \cdot n_y) \prod_{(x,z) \text{ leg of } G} (1 + \tau_{xz} \cdot n_x) \prod_{x \in G:\, \mathrm{st}_D(x) \neq 2} \lambda_{\mathrm{st}_D(x)}(\mathrm{d}n_x), \quad (6.2)$$

and we let $Z_D^0[f]$ denote the same integral without the factors of the legs. The factors of the legs make sense, since the site of every leg belongs to $\partial_2 G$ and therefore has status 0. We set $Z_D^\cdot = Z_D^\cdot[1]$ and $\mathbb{E}_D^\cdot[f] = Z_D^\cdot[f]/Z_D^\cdot$, where here and below the dot stands for τ or 0. Since all spins and pins have norm at most one, the factors in (6.2) are nonnegative. They are positive almost everywhere on the event that all spins belong to $\mathbb{S}^2$, which has positive probability under the product of the site measures. We conclude that $Z_D^\cdot > 0$ and that $\mathbb{E}_D^\cdot$ is a probability measure.

Let $Y \subset G \setminus \partial_2 G$. We identify the domain $G \setminus Y$ in G of Section 5.1 with the domain on G in which the sites of Y are removed, the other sites have status 0, and the set of used bonds is empty. This domain is consistent. The present bonds of this domain are the bonds $\{x, y\}$ with $x, y \in G \setminus Y$, and the sites of all legs of G belong to $G \setminus Y$. Therefore, the integrals (6.2) for this domain are those of (5.1). For a domain D on G and non-removed sites u and v, we call $\mathbb{E}_D^\tau[n_u \cdot n_v] - \mathbb{E}_D^0[n_u \cdot n_v]$ the response of the pair (u, v) in D, as in Section 5.2.

We will need the following elementary extension of Lemmas 5.1 and 5.2 to domains.

Lemma 6.2. *Let D be a domain on G, and let the dot stand for τ or 0.*

1. *For all $k \geq 1$ and $a_1, \dots, a_k \in \{1, 2, 3\}$, we have*
$$\int n^{a_1} \cdots n^{a_k} \, \lambda_1(\mathrm{d}n) = \frac{3}{5} \int n^{a_1} \cdots n^{a_k} \, \mathrm{d}n.$$
2. *Let f be a finite product of factors $n_x \cdot n_y$, where x and y are non-removed sites. Then $\mathbb{E}_D^0[f] \geq 0$.*
3. *Let $x \neq y$ be non-removed sites such that $\{x, y\}$ is not a present bond. Then, under $\mathbb{E}_D^\cdot$, the spins n_x and n_y are conditionally independent given the spins of all other non-removed sites.*

4. *Let x be a non-removed site with $\mathrm{dep}_G(x) \geq 2$ and at most three present bonds. Then, under $\mathbb{E}_D^{\cdot}$, the conditional mean of n_x given the spins of all other non-removed sites has norm at most $3/5$.*

Proof. The measure δ_0 integrates every monomial of positive degree to zero, which proves (1).

For (2), we expand the factors of the present bonds in $Z_D^0[f]$ as $\prod(1 + n_x \cdot n_y) = \sum_B \prod_{\{x,y\} \in B} n_x \cdot n_y$, where B runs over the sets of present bonds. Every term of the resulting sum is the integral of a product of scalar products, and we expand each scalar product in coordinates. For each choice of the indices, the integral is a product, over the non-removed sites, of moments of λ_0 or λ_1, and these moments are nonnegative by (5.2) and part (1). Summing over the indices and over B, we find that $Z_D^0[f] \geq 0$, and (2) follows from $Z_D^0 > 0$.

For (3), observe that every factor of (6.2) involves at most one of the spins n_x and n_y. The conditional density of (n_x, n_y) given the other spins is therefore the product of a function of n_x and a function of n_y.

For (4), let $b_1, \dots, b_k$, with $k \leq 3$, denote the spins of the sites joined to x by present bonds, and set $b_i = 0$ for $k < i \leq 3$. The site x is not the site of a leg, since $\mathrm{dep}_G(x) \geq 2$. The conditional law of n_x given the other spins is thus proportional to $\prod_{i=1}^3 (1 + n \cdot b_i)\, \lambda_{\mathrm{st}_D(x)}(\mathrm{d}n)$, and the vectors b_i belong to $\mathbb{B}$. By part (1) of Lemma 5.2, the normalization ν of $\prod_{i=1}^3 (1 + n \cdot b_i)\, \mathrm{d}n$ has mean of norm at most $3/5$. This is the claim when $\mathrm{st}_D(x) = 0$. When $\mathrm{st}_D(x) = 1$, the conditional law is a convex combination of ν and δ_0, since the product is one at $n = 0$. Its mean also has norm at most $3/5$. □

6.2 Trails and traced domains

We now define the walks of the expansion and the domains that they leave behind.

Definition 6.3. Let D be a domain on G, and let $z \in G$ be a non-removed site. A *trail* in D from z is a sequence $w = (v_0, \dots, v_m)$ of sites, with $m \geq 0$ and $v_0 = z$, such that for all $j \in \{1, \dots, m\}$ the bond $\{v_{j-1}, v_j\}$ is present in $D[w_{j-1}]$. Here $w_j = (v_0, \dots, v_j)$, and the pairs $D[w_j]$ of statuses and used bonds are defined recursively by $D[w_0] = D$, with $D[w_j]$ obtained from $D[w_{j-1}]$ by adding $\{v_{j-1}, v_j\}$ to the used bonds and raising the status of v_{j-1} by one. We call $|w| = m$ the *length* of w and w_j its *prefixes*, and we call $D[w] = D[w_m]$ the *traced domain* of w.

Since a present bond has two non-removed ends, the site v_{j-1} has status 0 or 1 in $D[w_{j-1}]$ and status 1 or 2 in $D[w_j]$. Each step adds a bond to the used bonds. Therefore, the bonds $\{v_{j-1}, v_j\}$ of a trail are pairwise distinct, and the length of a trail is bounded by the number of bonds $\{x, y\}$ with $x, y \in G$. We will show in Lemma 6.4 that the pairs $D[w_j]$ are domains for the trails that we use.

Let $w = (v_0, \dots, v_m)$ be a trail in D. For $1 \le j \le m$, we call the move from v_{j-1} to v_j the *step* j of w, and we call v_{j-1} its *passage site*. Step j is a *first visit* if $\mathrm{st}_{D[w_{j-1}]}(v_{j-1}) = 0$ and a *revisit* if $\mathrm{st}_{D[w_{j-1}]}(v_{j-1}) = 1$. We let $\mathrm{FV}(w) \subset \{1, \dots, m\}$ denote the set of first-visit steps of w, and we let $\mathrm{rev}_D(w) = m - |\mathrm{FV}(w)|$ denote the number of revisits. Note that a step from a site of status 1 in D is a revisit even if w has not visited this site before. For $T \subset G$ such that $z \notin T$, we let $\mathcal{T}_D(z, T)$ denote the set of trails $w = (v_0, \dots, v_m)$ in D from z such that $m \ge 1$, $v_m \in T$, and $v_i \notin T$ for $0 \le i < m$. This set is finite.

The following lemma records the properties of traced domains that we use. Its proof is a routine induction along the trail.

Lemma 6.4. *Let D' be a domain on G, let $T \subset G$ be a set that contains $\partial_2 G$, let $z \in G \setminus T$ be a non-removed site with at most three present bonds in D', and let $w = (v_0, \dots, v_m)$ be a trail in D' from z such that $v_i \notin T$ for $0 \le i < m$. Then $D'[w_j]$ is a domain on G for $0 \le j \le m$. Further, let $2 \le j \le m$. Then the bond $\{v_{j-2}, v_{j-1}\}$ is used in $D'[w_{j-1}]$, and if step j is a revisit, then $\{v_{j-1}, v_j\}$ is the only present bond at v_{j-1} in $D'[w_{j-1}]$.*

Proof. Along w, the sites that change their status are the passage sites $v_0, \dots, v_{m-1}$, which do not belong to $T \supset \partial_2 G$. In particular, every site of $\partial_2 G$ has status 0 in $D'[w_j]$ for all j. By the definition of the traced domains, the bond $\{v_{j-2}, v_{j-1}\}$ is added to the used bonds at step $j-1$, which proves the second statement.

We show by induction on j that every site of status 1 in $D'[w_j]$ has at most two present bonds. This holds for $j = 0$, since D' is a domain. Suppose that it holds for $j-1$. The bonds that are present in $D'[w_j]$ are present in $D'[w_{j-1}]$, and v_{j-1} is the only site whose status changes at step j.

Suppose first that v_{j-1} has status 0 in $D'[w_{j-1}]$. Then it has at most three present bonds there, by assumption for $j = 1$ and by the second statement for $j \ge 2$. The bond $\{v_{j-1}, v_j\}$ becomes used at step j, and v_{j-1} obtains the status 1 and at most two present bonds.

Suppose next that v_{j-1} has status 1 in $D'[w_{j-1}]$. Then it is removed in $D'[w_j]$. This completes the induction.

Now let $2 \le j \le m$ be such that step j is a revisit, and let $y = v_{j-1}$. The status of y changes only at the steps that leave y, and step $j-1$ leaves $v_{j-2} \ne y$. We conclude that y has status 1 in $D'[w_{j-2}]$ as well. In this domain, y has at most two present bonds, and $\{v_{j-2}, y\}$ is one of them, since step $j-1$ uses it. This bond is used in $D'[w_{j-1}]$. Therefore, y has at most one present bond in $D'[w_{j-1}]$, and $\{y, v_j\}$ is present there, since step j uses it. □

For the rest of this section, we let D be a domain on G, $u \in G \setminus \partial_2 G$ a non-removed site, v_{-1} a neighbor of u such that the bond $\{v_{-1}, u\}$ is not present in D, and $v \ne u$ a non-removed site, and we set $T = \partial_2 G \cup \{v\}$. We assume that u has at most one present bond if $\mathrm{st}_D(u) = 1$. In Section 7, v_{-1} will be the removed common neighbor of a pair

(u, v). For a trail $w = (v_0, \dots, v_m)$ in D from u and $1 \le j \le m$, we call v_{j-2} the *arrival site* of step j and $\{v_{j-2}, v_{j-1}\}$ its *arrival bond*; for $j = 1$, they are v_{-1} and $\{v_{-1}, u\}$.

Since the bond $\{v_{-1}, u\}$ is not present in D, the site u has at most three present bonds. We may therefore apply Lemma 6.4 with $D' = D$ and $z = u$, and we find that $D[w_j]$ is a domain for all $w \in \mathcal{T}_D(u, T)$ and $0 \le j \le |w|$. The arrival bond of each step is not present before that step, by assumption for $j = 1$ and by Lemma 6.4 for $j \ge 2$. If step 1 is a revisit, then u has status 1 and $\{u, v_1\}$ is its only present bond, by the assumption on u. Figure 6.1 shows an example of a trail and its traced domain.

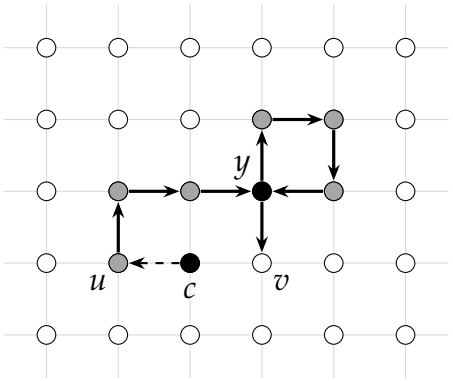


Figure 6.1. A trail $w = (v_0, \dots, v_8)$ in D from $u = (-1, 0)$ to $v = (1, 0)$, shown on a portion of the base set G. In D, the site $c = (0, 0)$ is removed, all other sites shown have status 0, and no bond is used. The dashed arrow shows the arrival site $v_{-1} = c$ of u. The circles show the statuses in the traced domain $D[w]$, with white for status 0, gray for status 1, and black for status 2. The trail leaves the site $y = (1, 1)$ twice, at steps 4 and 8. Step 8 is a revisit and all other steps are first visits.

6.3 The step identity and the pair walk identity

The walk expansion rests on the following identity, which integrates out one spin at the cost of one step of a trail. Its proof is a direct computation with the moments (5.2).

Lemma 6.5. *Let $a, b_1, \dots, b_k \in \mathbb{R}^3$. If $k \le 4$, then*

$$\int (a \cdot n) \prod_{i=1}^{k} (1 + n \cdot b_i) \, \mathrm{d}n = \sum_{i=1}^{k} \frac{a \cdot b_i}{3} \int \prod_{i' \ne i} (1 + n \cdot b_{i'}) \, \lambda_1(\mathrm{d}n). \tag{6.3}$$

If $k \le 2$, then

$$\int (a \cdot n) \prod_{i=1}^{k} (1 + n \cdot b_i) \, \lambda_1(\mathrm{d}n) = \frac{3}{5} \sum_{i=1}^{k} \frac{a \cdot b_i}{3}. \tag{6.4}$$

Proof. We expand the products and integrate term by term with (5.2); the terms of odd degree in n vanish. Let $k \le 4$. Only the terms of degree two and four contribute to the left side of (6.3), since its terms have degree at most five. We find that the left side of (6.3) is

$$\frac{1}{3}\sum_{i=1}^{k} a \cdot b_i + \frac{1}{15}\sum_{i<i'<i''} \big[(a \cdot b_i)(b_{i'} \cdot b_{i''}) + (a \cdot b_{i'})(b_i \cdot b_{i''}) + (a \cdot b_{i''})(b_i \cdot b_{i'})\big]. \tag{6.5}$$

Each product on the right side of (6.3) has at most three factors. By part (1) of Lemma 6.2 and (5.2), we have

$$\int \prod_{i' \neq i} (1 + n \cdot b_{i'})\, \lambda_1(\mathrm{d}n) = 1 + \frac{1}{5} \sum_{\substack{i'<i'' \\ i',i'' \neq i}} b_{i'} \cdot b_{i''}.$$

Multiplying by $(a \cdot b_i)/3$ and summing over i, we obtain (6.5) again, since each triple $i < i' < i''$ arises from three choices of the index outside the pair. This gives (6.3).

Next, let $k \le 2$. The integrand on the left side of (6.4) has degree at most three in n, and part (1) of Lemma 6.2 and (5.2) give

$$\int (a \cdot n) \prod_{i=1}^{k} (1 + n \cdot b_i)\, \lambda_1(\mathrm{d}n) = \frac{3}{5}\sum_{i=1}^{k} \int (a \cdot n)(n \cdot b_i)\, \mathrm{d}n = \frac{3}{5}\sum_{i=1}^{k} \frac{a \cdot b_i}{3}. \qquad \square$$

In (6.3), the factor $1 + n \cdot b_i$ of one bond is replaced by $(a \cdot b_i)/3$, and the remaining factors are integrated against λ_1. In (6.4), the remaining factors disappear, and the factor $3/5$ appears. Expansions of the correlations of the classical AKLT model into bond-self-avoiding walks with weight $3^{-|w|}$, obtained by integrating out one site at a time with the moment identities of the sphere, go back to Kennedy, Lieb, and Tasaki [23]. They used them to prove exponential decay of the two-point function on bipartite square-lattice tori. For Néel boundary conditions they bounded the staggered magnetization by the boundary size times a function that decays exponentially in the distance to the boundary. Lemma 6.5 and the following proposition are versions of their expansion in domains with pins and removed sites.

Proposition 6.6. *Let D, u, v, and T be as above. Then for $\cdot \in \{\tau, 0\}$, we have*

$$\mathbb{E}^{\cdot}_{D}[n_u \cdot n_v] = \sum_{w \in \mathcal{T}_D(u,T)} 3^{-|w|} \Big(\frac{3}{5}\Big)^{\mathrm{rev}_D(w)} \varrho^{\cdot}_{D}(w)\, \mathbb{E}^{\cdot}_{D[w]}[n_{v_{|w|}} \cdot n_v], \qquad \varrho^{\cdot}_{D}(w) = \frac{Z^{\cdot}_{D[w]}}{Z^{\cdot}_{D}}. \tag{6.6}$$

If $w \in \mathcal{T}_D(u,T)$ ends at $v_{|w|} = v$, then $\mathbb{E}^{\cdot}_{D[w]}[n_{v_{|w|}} \cdot n_v] = \mathbb{E}^{\cdot}_{D[w]}[|n_v|^2]$, which equals 1 if $\mathrm{st}_D(v) = 0$, and $3/5$ if $\mathrm{st}_D(v) = 1$ and v has exactly one present bond in D.

Proof. For $\beta \in \{1,2,3\}$, we write n^β for the β-th coordinate of n. We first prove the following claim. Let D' be a domain on G in which v is non-removed, and let $z \in G \setminus T$ be a non-removed site with at most three present bonds in D'. Then for all β,

$$Z^{\cdot}_{D'}[n_z^\beta n_v^\beta] = \sum_{w \in \mathcal{T}_{D'}(z,T)} 3^{-|w|} \Big(\frac{3}{5}\Big)^{\mathrm{rev}_{D'}(w)} Z^{\cdot}_{D'[w]}[n^\beta_{v_{|w|}} n_v^\beta]. \tag{6.7}$$

By Lemma 6.4, the traced domains on the right side are domains, and so is $D'[(z,y)]$ for every site y such that $\{z,y\}$ is present in D'. We prove (6.7) by induction on the number of present bonds of D'. When z has no present bond, the two sides of (6.7) vanish; this covers the base case, in which D' has no present bond.

We begin by integrating out n_z. Let $y_1, \dots, y_k$ be the sites joined to z by present bonds, so that $k \le 3$. Recall that $z \notin \partial_2 G$ is not the site of a leg, and note that $v \neq z$, since $z \notin T$. The factors of $Z^{\cdot}_{D'}[n_z^\beta n_v^\beta]$ that involve n_z are therefore n_z^β and $1 + n_z \cdot n_{y_i}$ for $1 \le i \le k$.

When $\mathrm{st}_{D'}(z) = 0$, we apply (6.3) with a the β-th unit vector and $b_i = n_{y_i}$. When $\mathrm{st}_{D'}(z) = 1$, we have $k \le 2$, and we apply (6.4). In each case, the i-th term of the result, together with the factors that do not involve n_z, is the integrand of $Z^{\cdot}_{D'[(z,y_i)]}[n^\beta_{y_i} n_v^\beta]$, multiplied by $1/3$ in the first case and by $(1/3)(3/5)$ in the second. We obtain

$$Z^{\cdot}_{D'}[n_z^\beta n_v^\beta] = \sum_{i=1}^{k} \frac{1}{3} \Big(\frac{3}{5}\Big)^{\mathbf{1}\{\mathrm{st}_{D'}(z)=1\}} Z^{\cdot}_{D'[(z,y_i)]}[n^\beta_{y_i} n_v^\beta]. \tag{6.8}$$

We now treat the terms of (6.8) one by one. For $y_i \in T$, the i-th term is the term of the trail (z, y_i) in (6.7). For $y_i \notin T$, we let $D'' = D'[(z,y_i)]$, which is a domain with fewer present bonds than D'. The site $v \in T$ remains non-removed because only the status of $z \notin T$ changes. In D'', the site y_i is non-removed and has at most three present bonds, since $\{z,y_i\}$ is used. By the induction hypothesis, (6.7) holds for D'' and y_i. The trails in $\mathcal{T}_{D'}(z,T)$ whose first step goes to y_i are the sequences $(z, v'_0, \dots, v'_{m'})$ with $w'' = (v'_0, \dots, v'_{m'}) \in \mathcal{T}_{D''}(y_i, T)$. For such a trail w, the recursive definition of the traced domains gives

$$D'[w] = D''[w''], \qquad |w| = |w''| + 1, \qquad \mathrm{rev}_{D'}(w) = \mathrm{rev}_{D''}(w'') + \mathbf{1}\{\mathrm{st}_{D'}(z) = 1\}.$$

Inserting (6.7) for D'' into (6.8), we find that (6.7) holds for D'.

We now apply (6.7) with $D' = D$ and $z = u$, which is allowed, since $u \notin T$ and u has at most three present bonds. Summing over β and dividing by $Z^{\cdot}_D$, we obtain (6.6).

Finally, let $w \in \mathcal{T}_D(u,T)$ end at v. Since v is not a passage site of w, its status in $D[w]$ is $\mathrm{st}_D(v)$. When this status is 0, we have $|n_v| = 1$ almost surely. Suppose that $\mathrm{st}_D(v) = 1$ and that v has exactly one present bond in D. The last step of w uses a bond at v that is present in D, and v has no present bond in $D[w]$. Further, v is not the site of a leg, since a site of status 1 does not belong to $\partial_2 G$. Under $\mathbb{E}^{\cdot}_{D[w]}$, the spin n_v is therefore independent of the other spins and has law λ_1, and $\mathbb{E}^{\cdot}_{D[w]}[|n_v|^2] = 3/5$. □

6.4 Step ratios and responses

We now compute how the pins affect the factor $\varrho^{\cdot}_D(w)$ in (6.6), one step at a time. Let $w = (v_0, \dots, v_m) \in \mathcal{T}_D(u, T)$. For $1 \le j \le m$, we define the *step ratios*

$$\rho^{\cdot}_j = \frac{Z^{\cdot}_{D[w_j]}}{Z^{\cdot}_{D[w_{j-1}]}}, \qquad \text{so that} \qquad \varrho^{\cdot}_D(w) = \prod_{j=1}^{m} \rho^{\cdot}_j. \tag{6.9}$$

We suppress the dependence on w in this notation. Let $1 \le j \le m$, and write $x = v_{j-1}$ for the passage site and $o = v_j$ for the *out-neighbor* of step j. We call the sites y such that $\{x, y\}$ is present in $D[w_{j-1}]$ the *remaining neighbors* of x at step j. The arrival bond of step j is not present in $D[w_{j-1}]$ (by Lemma 6.4 and the remark after it), so the remaining neighbors of x belong to the set of the three other neighbors of x, and o is one of them. We let f and g denote the two neighbors of x other than v_{j-2} and o, in either order. We let D^-_j denote the domain obtained from $D[w_{j-1}]$ by changing the status of x to 2, and we define

$$I_{\mathrm{cur}} = 1 + \frac{1}{3}(n_f \cdot n_g + n_f \cdot n_o + n_g \cdot n_o), \qquad I_{\mathrm{next}} = 1 + \frac{1}{5} n_f \cdot n_g, \tag{6.10}$$

with the convention that n_f is replaced by 0 if f is not a remaining neighbor of x at step j, and likewise for g. These are functions of the spins of the remaining neighbors, which are not removed in D^-_j.

Lemma 6.7. *Let $w \in \mathcal{T}_D(u, T)$ and $1 \le j \le |w|$, and let x, o, f, and g be as above.*

1. *If step j is a revisit, then $\rho^{\tau}_j = \rho^0_j = 1$.*
2. *If step j is a first visit, then D^-_j is a domain on G, and for $\cdot \in \{\tau, 0\}$,*

$$Z^{\cdot}_{D[w_{j-1}]} = Z^{\cdot}_{D^-_j} \mathbb{E}^{\cdot}_{D^-_j}[I_{\mathrm{cur}}], \qquad Z^{\cdot}_{D[w_j]} = Z^{\cdot}_{D^-_j} \mathbb{E}^{\cdot}_{D^-_j}[I_{\mathrm{next}}], \qquad \rho^{\cdot}_j = \frac{\mathbb{E}^{\cdot}_{D^-_j}[I_{\mathrm{next}}]}{\mathbb{E}^{\cdot}_{D^-_j}[I_{\mathrm{cur}}]}. \tag{6.11}$$

Proof. Recall that the passage site x does not belong to $T \supset \partial_2 G$, so that x is not the site of a leg. Suppose first that step j is a revisit. By Lemma 6.4 and the remark after it, x has status 1 in $D[w_{j-1}]$, and $\{x, o\}$ is its only present bond. In $D[w_j]$, the site x is removed, and all other statuses and present bonds are those of $D[w_{j-1}]$. Integrating out n_x in $Z^{\cdot}_{D[w_{j-1}]}$ and using

$$\int (1 + n \cdot n_o)\, \lambda_1(\mathrm{d}n) = 1,$$

we obtain $Z^{\cdot}_{D[w_j]}$, which proves (1).

Suppose now that step j is a first visit. The site $x \notin \partial_2 G$ is removed in D^-_j, and removing a site can only decrease the set of present bonds, so that D^-_j is a domain. The

factors of (6.2) for $D[w_{j-1}]$ that involve n_x are those of the bonds from x to its remaining neighbors y, of which there are at most three. In $D[w_j]$, the site x has status 1, and its present bonds join it to its remaining neighbors other than o. By (5.2) and part (1) of Lemma 6.2, we have

$$\int \prod_y (1 + n \cdot n_y)\, \mathrm{d}n = I_{\mathrm{cur}}, \qquad \int \prod_{y \neq o} (1 + n \cdot n_y)\, \lambda_1(\mathrm{d}n) = I_{\mathrm{next}},$$

where the products run over the remaining neighbors y of x. Here the terms of degree one and three vanish, each term of degree two contributes $(n_y \cdot n_{y'})/3$ in the first integral and $(n_y \cdot n_{y'})/5$ in the second, and there is no term of degree four. The other factors and site measures of $D[w_{j-1}]$ and of $D[w_j]$ are those of D_j^-, since the present bonds of D_j^- are the present bonds of $D[w_{j-1}]$ that do not contain x. Integrating out n_x first, we obtain the first two identities in (6.11), and the third follows from them. □

A first-visit step j is *straight* if $o - x = x - v_{j-2}$ and a *turn* otherwise. For $j = 1$, this refers to the arrival site v_{-1} of u. For a first-visit step j, we define

$$\Gamma_j = \log \rho_j^0 - \log \rho_j^\tau, \tag{6.12}$$

and for remaining neighbors $a \neq b$ of x at step j, we let

$$r_{ab} = \mathbb{E}^\tau_{D_j^-}[n_a \cdot n_b] - \mathbb{E}^0_{D_j^-}[n_a \cdot n_b] \tag{6.13}$$

denote the response of the pair (a, b) in D_j^-. The following lemma plays for a site that a trail passes the role that Lemma 5.3 plays for a deleted site. In the two lemmas, the effect of the pins is expressed through the responses of the pairs formed by neighbors of the site, computed in a domain in which the site is removed. This is the main point of the expansion, since these pairs are again at distance 2 and have a removed common neighbor.

Lemma 6.8. *Let $w \in \mathcal{T}_D(u, T)$, let step j of w be a first visit, and let x, o, f, and g be as above. Then $\rho_j^0 \leq 1$. Suppose in addition that $\rho_j^\tau \leq 200/187$.*

1. *If f, g, and o are remaining neighbors of x at step j, then there exist $\alpha_j \in [0, 1/6]$ and $\beta_j \in (0, 9/25]$ such that*
$$\Gamma_j = \alpha_j r_{fg} + \beta_j (r_{fo} + r_{go}). \tag{6.14}$$
2. *If $y \in \{f, g\}$ and o are the only remaining neighbors of x at step j, then there exists $\beta_j' \in (0, 9/25]$ such that $\Gamma_j = \beta_j' r_{yo}$.*
3. *If o is the only remaining neighbor of x at step j, then $\Gamma_j = 0$.*

Proof. For $\cdot \in \{\tau, 0\}$, we abbreviate $E^{\cdot}_{\mathrm{cur}} = \mathbb{E}^{\cdot}_{D_j^-}[I_{\mathrm{cur}}]$ and $E^{\cdot}_{\mathrm{next}} = \mathbb{E}^{\cdot}_{D_j^-}[I_{\mathrm{next}}]$, so that $\rho^{\cdot}_j = E^{\cdot}_{\mathrm{next}}/E^{\cdot}_{\mathrm{cur}}$ by Lemma 6.7.

Step 1: the free ratio. By (6.10), we have

$$\begin{aligned} E^0_{\mathrm{cur}} - E^0_{\mathrm{next}} &= \frac{2}{15}\mathbb{E}^0_{D_j^-}[n_f \cdot n_g] + \frac{1}{3}\mathbb{E}^0_{D_j^-}[n_f \cdot n_o + n_g \cdot n_o] \geq 0, \\ E^0_{\mathrm{next}} &= 1 + \frac{1}{5}\mathbb{E}^0_{D_j^-}[n_f \cdot n_g] \geq 1, \end{aligned} \tag{6.15}$$

where the inequalities follow from part (2) of Lemma 6.2. We conclude that

$$E^0_{\mathrm{cur}} \geq E^0_{\mathrm{next}} \geq 1, \qquad \rho^0_j = \frac{E^0_{\mathrm{next}}}{E^0_{\mathrm{cur}}} \leq 1.$$

Further, the assumption $\rho^\tau_j \leq 200/187$ means that

$$E^\tau_{\mathrm{cur}} \geq \frac{187}{200} E^\tau_{\mathrm{next}}. \tag{6.16}$$

Step 2: a lower bound for E^τ_{next} in case (1). In D_j^-, the sites f and g are not removed, and each of them has at most three present bonds, since its bond to the removed site x is not present. The depth is 1-Lipschitz and $\mathrm{dep}_G(x) \geq 3$, so that $\mathrm{dep}_G(f) \geq 2$ and $\mathrm{dep}_G(g) \geq 2$. Further, f and g are not neighbors. We condition on the spins of all non-removed sites other than f and g. By part (3) of Lemma 6.2, the conditional expectation of $n_f \cdot n_g$ is the scalar product of the conditional means of n_f and n_g, and each of these conditional means is also the conditional mean given all spins other than its own. By part (4) of Lemma 6.2, these conditional means have norm at most $3/5$. Taking expectations, we find that $\mathbb{E}^\tau_{D_j^-}[n_f \cdot n_g] \geq -(3/5)^2$, and

$$E^\tau_{\mathrm{next}} \geq 1 - \frac{1}{5}\Big(\frac{3}{5}\Big)^2 = \frac{116}{125}. \tag{6.17}$$

Step 3: interpolation in case (1). For $t \in [0,1]$, we define

$$E_{\mathrm{cur}}(t) = (1-t)E^0_{\mathrm{cur}} + tE^\tau_{\mathrm{cur}}, \qquad E_{\mathrm{next}}(t) = (1-t)E^0_{\mathrm{next}} + tE^\tau_{\mathrm{next}}.$$

These numbers are positive, by (6.15), (6.16), and (6.17). By (6.10) and (6.13), we have

$$E^\tau_{\mathrm{cur}} - E^0_{\mathrm{cur}} = \frac{1}{3}(r_{fg} + r_{fo} + r_{go}), \qquad E^\tau_{\mathrm{next}} - E^0_{\mathrm{next}} = \frac{1}{5} r_{fg}.$$

Writing the differences of logarithms as integrals of their derivatives in t, we obtain

$$\Gamma_j = \big(\log E^\tau_{\mathrm{cur}} - \log E^0_{\mathrm{cur}}\big) - \big(\log E^\tau_{\mathrm{next}} - \log E^0_{\mathrm{next}}\big) = \int_0^1 \Big(\frac{r_{fg} + r_{fo} + r_{go}}{3E_{\mathrm{cur}}(t)} - \frac{r_{fg}}{5E_{\mathrm{next}}(t)}\Big)\mathrm{d}t,$$

which is (6.14) with

$$\alpha_j = \int_0^1 \Big(\frac{1}{3E_{\mathrm{cur}}(t)} - \frac{1}{5E_{\mathrm{next}}(t)}\Big)\mathrm{d}t, \qquad \beta_j = \int_0^1 \frac{\mathrm{d}t}{3E_{\mathrm{cur}}(t)} > 0. \tag{6.18}$$

Step 4: the bounds in case (1). Since all spins have norm at most one, we have

$$5I_{\mathrm{next}} - 3I_{\mathrm{cur}} = 2 - n_f \cdot n_o - n_g \cdot n_o \geq 0.$$

Taking expectations, we find that $5E_{\mathrm{next}}(t) \geq 3E_{\mathrm{cur}}(t)$ for all t, and $\alpha_j \geq 0$. Next, the inequality $E_{\mathrm{cur}}(t) \geq (187/200)E_{\mathrm{next}}(t)$ holds at $t = 0$ by (6.15) and at $t = 1$ by (6.16). Both sides are affine in t, and the inequality holds for all $t \in [0,1]$. Since $y \mapsto 1/y$ is convex on $(0,\infty)$, we also have

$$\int_0^1 \frac{\mathrm{d}t}{(1-t)a + ta'} \leq \frac{1}{2}\Big(\frac{1}{a} + \frac{1}{a'}\Big), \qquad a, a' > 0. \tag{6.19}$$

Using $E_{\mathrm{cur}}(t) \geq (187/200)E_{\mathrm{next}}(t)$ together with (6.19), we obtain

$$\alpha_j \leq \Big(\frac{200}{561} - \frac{1}{5}\Big)\int_0^1 \frac{\mathrm{d}t}{E_{\mathrm{next}}(t)} \leq \frac{439}{2805}\cdot\frac{1}{2}\Big(1 + \frac{125}{116}\Big) = \frac{105799}{650760} < \frac{1}{6},$$

where in the second inequality we used $E^0_{\mathrm{next}} \geq 1$ and (6.17). Similarly, (6.19) gives

$$\beta_j \leq \frac{1}{6}\Big(\frac{1}{E^0_{\mathrm{cur}}} + \frac{1}{E^\tau_{\mathrm{cur}}}\Big) \leq \frac{1}{6}\Big(1 + \frac{125\cdot 200}{116\cdot 187}\Big) = \frac{3891}{10846} < \frac{9}{25},$$

where in the second inequality we used $E^0_{\mathrm{cur}} \geq 1$, (6.16), and (6.17).

Step 5: cases (2) and (3). In case (2), our convention gives $I_{\mathrm{cur}} = 1 + (n_y \cdot n_o)/3$ and $I_{\mathrm{next}} = 1$. Then $E_{\mathrm{next}} = 1$, and (6.15) and (6.16) give $E^0_{\mathrm{cur}} \geq 1$ and $E^\tau_{\mathrm{cur}} \geq 187/200$. With $E_{\mathrm{cur}}(t)$ as above, we find

$$\Gamma_j = \log E^\tau_{\mathrm{cur}} - \log E^0_{\mathrm{cur}} = \beta'_j\, r_{yo}, \qquad \beta'_j = \int_0^1 \frac{\mathrm{d}t}{3E_{\mathrm{cur}}(t)},$$
$$0 < \beta'_j \leq \frac{1}{6}\Big(1 + \frac{200}{187}\Big) = \frac{129}{374} < \frac{9}{25},$$

where we used (6.19) in the second line. In case (3), we have $I_{\mathrm{cur}} = I_{\mathrm{next}} = 1$, so that $\rho^\tau_j = \rho^0_j = 1$ and $\Gamma_j = 0$. □

We now bound the difference of the products (6.9) with and without pins by the quantities Γ_j. We call a bijection $\varsigma\colon \{1,\dots,|\,\mathrm{FV}(w)|\} \to \mathrm{FV}(w)$ an *enumeration* of $\mathrm{FV}(w)$.

Lemma 6.9. *Let $w \in \mathcal{T}_D(u,T)$, and let $\bar\rho_j \geq 1$, for $j \in \mathrm{FV}(w)$, be numbers such that $\rho^\tau_j \leq \bar\rho_j$ for all $j \in \mathrm{FV}(w)$. Then for every enumeration ς of $\mathrm{FV}(w)$, we have*

$$\big|\varrho^\tau_D(w) - \varrho^0_D(w)\big| \leq \sum_{k=1}^{|\,\mathrm{FV}(w)|} |\Gamma_{\varsigma(k)}| \prod_{l=1}^{k} \bar\rho_{\varsigma(l)}. \tag{6.20}$$

Proof. Recall from part (1) of Lemma 6.7 that the step ratios of the revisits equal one, with and without pins. Let $q = |\mathrm{FV}(w)|$. We have

$$\varrho_D^\tau(w) - \varrho_D^0(w) = \prod_{k=1}^{q} \rho_{\varsigma(k)}^\tau - \prod_{k=1}^{q} \rho_{\varsigma(k)}^0 = \sum_{k=1}^{q} \Big(\prod_{l<k} \rho_{\varsigma(l)}^\tau \Big) (\rho_{\varsigma(k)}^\tau - \rho_{\varsigma(k)}^0) \Big(\prod_{l>k} \rho_{\varsigma(l)}^0 \Big).$$

For $j \in \mathrm{FV}(w)$, we have $\rho_j^\tau \le \bar\rho_j$ by assumption and $\rho_j^0 \le 1 \le \bar\rho_j$ by Lemma 6.8. By the mean value theorem for the exponential function, this implies that

$$|\rho_j^\tau - \rho_j^0| \le |\log \rho_j^\tau - \log \rho_j^0| \max(\rho_j^\tau, \rho_j^0) \le |\Gamma_j| \, \bar\rho_j.$$

Inserting these bounds into the previous identity, we obtain (6.20). □

6.5 Patterns and the local hypothesis

The hypothesis $\rho_j^\tau \le 200/187$ of Lemma 6.8 is a real restriction, since pins can make a step ratio larger than one. We now reduce it to a finite family of inequalities, which form Hypothesis (H1). The ratio ρ_j^τ depends on the whole domain. However, we will show that, after the passage site and four further sites are integrated out, a bound on it follows from an inequality that involves only the sites at distance at most 2 from the passage site and twelve vectors of norm at most one. This reduction to finitely many local configurations is what makes the step bounds accessible to a computer-assisted proof.

For $x \in \mathbb{Z}^2$, we define the *template* $\Diamond(x) = \{z \in \mathbb{Z}^2 : d(z,x) \le 2\}$. In the template $\Diamond(0)$, we call 0 the *middle site* of the template (it corresponds to the passage site), we call the four neighbors $\mathsf{S} = (0,-1)$, $\mathsf{E} = (1,0)$, $\mathsf{N} = (0,1)$, and $\mathsf{W} = (-1,0)$ of 0 the *axial sites*, the four sites $(\pm1,\pm1)$ the *corners*, and the four sites $(\pm2,0)$ and $(0,\pm2)$ the *tips*. The tip of the axial site z is $2z$, and we call the bond $\{z,2z\}$ the *tip bond* of z. The set R of the eight axial sites and corners is the *ring*, and we call its elements ring sites. Each corner is a neighbor of two axial sites, and we call the eight bonds between a corner and an axial site the *ring bonds*; they form a cycle through R. Each corner y has two neighbors outside $\Diamond(0)$, and we call the two pairs (y,y') formed by y and such a neighbor y' the *slots* of y. We enumerate the eight slots as $k = 1,\dots,8$, in a fixed order, and we let $z(k)$ denote the corner of the slot k. Figure 6.2 shows the template.

Definition 6.10. A *pattern* π consists of an out-neighbor $o_\pi \in \{\mathsf{E},\mathsf{N},\mathsf{W}\}$, a status $\mathrm{st}_\pi(z) \in \{0,1,2\}$ for each ring site z, and a set of bonds, which we call the used bonds of π, each of which has an end at 0 or at a ring site of status 0 or 1. We require that $\mathrm{st}_\pi(o_\pi) \ne 2$, and we regard the middle site 0 as a site of status 0. As in Definition 6.1, a ring site of status 2 is removed, and a ring bond is present in π if it is not used and neither of its ends is removed. We call the tip bond of a non-removed axial site *retained* if it is not used. We let $\mathcal{P}_{\mathrm{ext}}$ denote the set of patterns π such that

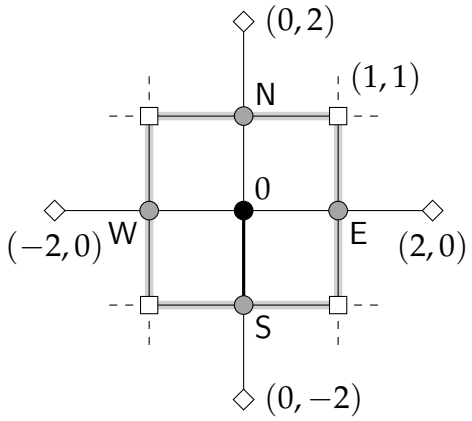


Figure 6.2. The template $\Diamond(x)$ after the translation that takes the passage site x to the middle site 0 (black) and the rotation that takes the arrival site to $\mathsf{S} = (0, -1)$. The bond $\{\mathsf{S}, 0\}$, the image of the arrival bond, is drawn thick. Gray circles: the axial sites S, E, N, and W. Squares: the corners $(\pm 1, \pm 1)$. Diamonds: the tips $(\pm 2, 0)$ and $(0, \pm 2)$, each joined to its axial site by its tip bond. The shaded cycle of ring bonds marks the ring R. The dashed stubs are the slots, which lead from each corner to its two neighbors $(\pm 2, \pm 1)$ and $(\pm 1, \pm 2)$ outside the template.

1. the only used bond of π at 0 is $\{0, \mathsf{S}\}$;
2. the arrival site S has status 1 or 2;
3. every ring site of status 0 has no used bond; and
4. every ring site of status 1 has exactly two used bonds,

and we call the elements of $\mathcal{P}_{\text{ext}}$ *extended patterns*. We call a pattern straight if $o_\pi = \mathsf{N}$ and a turn otherwise.

The tips have no status in a pattern. Let $w = (v_0, \dots, v_m) \in \mathcal{T}_D(u, T)$, and let step j of w be a first visit with passage site x. Since $x \notin \partial_2 G$, the template $\Diamond(x)$ is contained in G. We transport the out-neighbor $o = v_j$, the statuses of the ring sites of $\Diamond(x)$, and the used bonds of $D[w_{j-1}]$ that have an end at x or at a non-removed ring site of $\Diamond(x)$ to $\Diamond(0)$, by the translation by $-x$ followed by the rotation of $\mathbb{Z}^2$ about 0 that maps $v_{j-2} - x$ to S. The image of o belongs to $\{\mathsf{E}, \mathsf{N}, \mathsf{W}\}$ and is not removed, and the middle site has status 0, so that these data form a pattern. We call it the *pattern of step* j. It is straight if and only if step j is straight.

We now define the ring functional. The *exterior parameters* are $\mathbf{a} = (\mathbf{u}, \mathbf{m}) \in \mathbb{B}^{12}$, where $\mathbf{u} = (u_z)$ consists of one vector for each of the four axial sites z and $\mathbf{m} = (m_k)$ of one vector for each of the eight slots k. Let π be a pattern. Each non-removed ring site z carries a spin $n_z \in \mathbb{S}^2 \cup \{0\}$. For a bounded measurable function h of these spins, we define

$$\mathcal{Z}_\pi^{\mathbf{a}}[h] = \int h \prod_{\{y,z\} \text{ present in } \pi} (1 + n_y \cdot n_z) \prod_{z \text{ axial, tip bond retained}} \Big(1 + \frac{3}{5} u_z \cdot n_z\Big) \times \prod_{k:\, \mathrm{st}_\pi(z(k)) \neq 2} (1 + m_k \cdot n_{z(k)}) \prod_{z \in \mathsf{R}:\, \mathrm{st}_\pi(z) \neq 2} \lambda_{\mathrm{st}_\pi(z)}(\mathrm{d}n_z), \quad (6.21)$$

where the first product runs over the ring bonds that are present in π. Every slot of a non-removed corner has a factor, whether or not its bond is used. In the proof of Lemma 6.11, a missing factor is represented by $m_k = 0$. For a pattern π, we let f and g denote the two sites of $\{\mathsf{E}, \mathsf{N}, \mathsf{W}\} \setminus \{o_\pi\}$, and we define I_{cur} and I_{next} by (6.10) with $o = o_\pi$, where n_f is replaced by 0 if f is removed, and likewise for g. For a first-visit step whose pattern belongs to $\mathcal{P}_{\mathrm{ext}}$, this agrees with the convention of Section 6.4, as we will see in the proof of Lemma 6.11.

Hypothesis (H1). *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0, 1]$. For every $\pi \in \mathcal{P}_{\mathrm{ext}}$ and every $\mathbf{a} \in \mathbb{B}^{12}$, we have*

$$\mathcal{Z}_\pi^{\mathbf{a}}[I_{\mathrm{cur}} - c_\pi I_{\mathrm{next}}] \geq 0, \quad (6.22)$$

where $c_\pi = c_{\mathrm{s}}$ if π is straight and $c_\pi = c_{\mathrm{t}}$ if π is a turn.

For this hypothesis, and for (H2) and (H3) below, we say that the hypothesis holds at $(c_{\mathrm{s}}, c_{\mathrm{t}})$ when we need to specify the thresholds. Each exterior parameter occurs in at most one factor of (6.21), so that the left side of (6.22) is affine in each of the twelve vectors. Since an affine function on $\mathbb{B}$ attains its minimum on $\mathbb{S}^2$, it suffices to check (6.22) for $\mathbf{a} \in (\mathbb{S}^2)^{12}$. The class $\mathcal{P}_{\mathrm{ext}}$ is finite. Modulo the reflection $(z_1, z_2) \mapsto (-z_1, z_2)$ and the duplicate relation defined in Appendix B.4, it consists of 7,962 classes, of which 2,696 are straight and 5,266 are turns. Theorem 8.5 proves (H1) at $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$.

Lemma 6.11. *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0, 1]$, and assume that (H1) holds at $(c_{\mathrm{s}}, c_{\mathrm{t}})$. Let $w \in \mathcal{T}_D(u, T)$, and let step j of w be a first visit whose pattern π belongs to $\mathcal{P}_{\mathrm{ext}}$. Then $\rho_j^\tau \leq 1/c_\pi$.*

Proof. Let $x = v_{j-1}$ be the passage site, and write $D' = D[w_{j-1}]$ and $D'' = D[w_j]$. We may apply a translation and a rotation of $\mathbb{Z}^2$ to G, τ, D, and w, since this does not change the step ratios, and we may therefore assume that $x = 0$ and that the arrival site of step j is S. By the definition of the pattern of step j, the statuses of the ring sites and the used bonds of D' at the middle site and at the non-removed ring sites are then those of π.

Step 1: splitting the integral. We divide the factors of (6.2) for D' into those that involve no site of $\Diamond(0)$ and those that involve at least one site of $\Diamond(0)$, and we let n_{out} denote the spins of the non-removed sites outside $\Diamond(0)$. Integrating first over the spins of the non-removed sites of $\Diamond(0)$, we write $Z_{D'}^\tau$ as the integral, with respect to the site measures of n_{out}, of the product of the factors of the first kind and of a function $Q'(n_{\mathrm{out}})$. We define $Q''(n_{\mathrm{out}})$ in the same way for D''. The step changes only the status of 0 and the bond $\{0, o\}$, and the factors of the first kind and the site measures outside $\Diamond(0)$ are the

same for D' and D''. These factors are nonnegative. It therefore suffices to show that $Q'(n_{\mathrm{out}}) \geq c_\pi Q''(n_{\mathrm{out}})$ for all values of n_{out}.

Step 2: the middle site and the tips. By condition (1) of Definition 6.10, the remaining neighbors of the middle site at step j are the non-removed sites among E, N, and W. In particular, the conventions for I_{cur} and I_{next} in Section 6.4 and in the definition of the ring functional agree. As in the proof of Lemma 6.7, integrating out n_0 gives the factor I_{cur} in Q' and I_{next} in Q''.

Next, let $t = 2z$ be a non-removed tip, with axial site z. Its only neighbor in $\Diamond(0)$ is z, and its other factors, at most three, have the form $1 + n_t \cdot b$, where b is a component of n_{out} or a pin. We let $A_t > 0$ denote the integral of the product of these other factors against $\lambda_{\mathrm{st}(t)}$, and b_t the mean of the corresponding probability measure. Integrating out n_t gives $A_t(1 + b_t \cdot n_z)$ if $\{z, t\}$ is present in D', and A_t otherwise. As in the proof of part (4) of Lemma 6.2, we have $|b_t| \leq 3/5$. We set $u_z = 5b_t/3$ if the tip bond of z is present in D', and $u_z = 0$ otherwise. The tips are not neighbors of each other, of the corners, or of the middle site, so that they are integrated out independently, and they give the same factors in Q' and Q''.

Step 3: the ring. After these integrations, the remaining integrand involves the ring spins through the factors of the present ring bonds, the factors $1 + (3/5)u_z \cdot n_z$ of the axial sites whose tip bond is retained in π, and the factors of the slots of the non-removed corners. The axial sites carry no legs, since their depth is at least 2. For the slot $k = (y, y')$ of a non-removed corner y, we set $m_k = \tau_{yy'}$ if $y' \notin G$, $m_k = n_{y'}$ if $\{y, y'\}$ is present in D', and $m_k = 0$ otherwise. All these vectors belong to $\mathbb{B}$.

When the tip bond of z is retained in π but the tip is removed, we have $u_z = 0$, and the factor of z is one. Comparing with (6.21), we find

$$Q'(n_{\mathrm{out}}) = K\, \mathcal{Z}^{\mathbf{a}}_\pi[I_{\mathrm{cur}}], \qquad Q''(n_{\mathrm{out}}) = K\, \mathcal{Z}^{\mathbf{a}}_\pi[I_{\mathrm{next}}],$$

where $\mathbf{a} = (\mathbf{u}, \mathbf{m})$ and $K > 0$ is the product of the constants A_t over the non-removed tips. By (H1), we conclude that $Q'(n_{\mathrm{out}}) \geq c_\pi Q''(n_{\mathrm{out}})$. Integrating over n_{out}, we conclude that $Z^\tau_{D'} \geq c_\pi Z^\tau_{D''}$, which is the claim. □

6.6 The trail weight and the growth hypothesis

The second hypothesis bounds the number of all trails of $\mathbb{Z}^2$, weighted by the factors that the expansion attaches to their steps. We use the terms of Sections 6.2 and 6.4 also for sequences of sites of $\mathbb{Z}^2$, as follows. A trail of $\mathbb{Z}^2$ is a sequence $w = (v_0, \ldots, v_m)$ of sites of $\mathbb{Z}^2$ such that $v_{j-1} \sim v_j$ for $1 \leq j \leq m$ and the bonds $\{v_{j-1}, v_j\}$ are pairwise distinct, and its length is $|w| = m$. We consider such trails together with a neighbor v_{-1} of v_0, the arrival site of v_0. For $1 \leq j \leq m$, step j of w is a revisit if $v_{j-1} \in \{v_0, \ldots, v_{j-2}\}$ and a first visit otherwise. A first visit is straight if $v_j - v_{j-1} = v_{j-1} - v_{j-2}$ and a turn

otherwise. These notions agree with those of Sections 6.2 and 6.4 when w is a trail in a domain in which all sites of w have status 0, since the status of a site in $D[w_{j-1}]$ is then the number of earlier steps that left it. We let $\mathrm{FV}(w)$ denote the set of first-visit steps of w and $\mathrm{rev}(w) = m - |\mathrm{FV}(w)|$ the number of revisits.

Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0,1]$. We define the *trail weight*

$$\mathrm{wt}(w) = \prod_{j=1}^{m} \mathrm{wt}_j, \qquad \mathrm{wt}_j = \begin{cases} 1/5 & \text{if step } j \text{ is a revisit,} \\ 1/(3c_{\mathrm{s}}) & \text{if step } j \text{ is a straight first visit,} \\ 1/(3c_{\mathrm{t}}) & \text{if step } j \text{ is a turning first visit.} \end{cases} \tag{6.23}$$

The factor $1/5 = (1/3)(3/5)$ of a revisit is its weight in (6.6), with step ratio one. The factor of a first visit is its weight $1/3$ in (6.6), multiplied by the bound $1/c_\pi$ of Lemma 6.11 on its step ratio.

Hypothesis (H2). *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0,1]$. There exist $C_{\mathrm{o}} < \infty$ and $\gamma_{\mathrm{o}} \in (0,1)$ such that for all $n \geq 0$, all $v_0 \in \mathbb{Z}^2$, and all neighbors v_{-1} of v_0, we have*

$$\sum_{w} \mathrm{wt}(w) \leq C_{\mathrm{o}} \gamma_{\mathrm{o}}^{n}, \tag{6.24}$$

where the sum runs over the trails $w = (v_0, \dots, v_n)$ of $\mathbb{Z}^2$ of length n with $v_1 \neq v_{-1}$ when $n \geq 1$, and $\mathrm{wt}(w)$ is computed with the arrival site v_{-1}. For $n = 0$, the sum consists of the length-zero trail of weight 1.

The condition $v_1 \neq v_{-1}$ means that the first step does not reverse the direction of arrival. Proposition 8.4 proves (H2) at $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$, with $\gamma_{\mathrm{o}} = 2499/2500$ and $C_{\mathrm{o}} = 10$.

SECTION 7

Contraction of the boundary response

The main result of this section is Theorem 7.4, which states that Hypothesis (P) holds if (H1), (H2), and (H3) hold at the same pair of thresholds $(c_{\mathrm{s}}, c_{\mathrm{t}})$ with $\min(c_{\mathrm{s}}, c_{\mathrm{t}}) \geq 187/200$. Together with Theorem 8.1, it gives Hypothesis (P), the input of Theorem 5.9. We work on one finite set G and apply the pair walk identity of Proposition 6.6 to a response. The trails that return to the second site of the pair produce responses of new pairs in new domains. By Lemma 7.2, these pairs belong to the same class as the original pair. By (H3), their total weighted contribution is bounded by a constant smaller than one times the supremum of all such responses. The trails that end near the boundary contribute an amount that decays exponentially in the distance of the pair from $\partial_2 G$, by (H2). The boundary layer $\partial_2 G$ is the same for all domains on G, and this gives a bound that is uniform in G. The heuristics are explained in Section 2. As in Section 6, $G \subset \mathbb{Z}^2$ denotes a finite set and τ a pin configuration on G.

7.1 Pair classes

We begin by defining the class of pairs whose responses we bound.

Definition 7.1. Let $D \in \mathcal{D}_{\mathrm{cons}}(G)$ and $\mathfrak{t} \in \{\mathrm{coll}, \mathrm{diag}\}$. We let $\mathcal{J}_{\mathfrak{t}}(D)$ denote the set of ordered pairs $p = (u, v)$ of non-removed sites such that

1. $v - u \in \{(\pm 2, 0), (0, \pm 2)\}$ if $\mathfrak{t} = \mathrm{coll}$, and $v - u \in \{(\pm 1, \pm 1)\}$ if $\mathfrak{t} = \mathrm{diag}$;
2. u and v have a common neighbor of status 2; and
3. each of u and v has status 0, or has status 1 and at most one present bond.

We call the pairs in $\mathcal{J}_{\mathrm{coll}}(D)$ *collinear* and the pairs in $\mathcal{J}_{\mathrm{diag}}(D)$ *diagonal*, and we call $\mathfrak{t}$ the *type* of p. For each pair $p = (u, v)$ in $\mathcal{J}_{\mathrm{coll}}(D) \cup \mathcal{J}_{\mathrm{diag}}(D)$, we fix a common neighbor c of u and v of status 2, and we call it the *center* of p.

A collinear pair has one common neighbor and a diagonal pair has two. Either common neighbor can serve as the center of a diagonal pair when both are removed. For $p = (u, v) \in \mathcal{J}_{\mathrm{coll}}(D) \cup \mathcal{J}_{\mathrm{diag}}(D)$, we define the response of p and its distance from the boundary layer by

$$r_D^{\tau}(p) = \mathbb{E}_D^{\tau}[n_u \cdot n_v] - \mathbb{E}_D^{0}[n_u \cdot n_v], \qquad d(p) = \min\bigl(\mathrm{dist}(u, \partial_2 G), \mathrm{dist}(v, \partial_2 G)\bigr). \tag{7.1}$$

Since all spins have norm at most one, we have $|r^\tau_D(p)| \le 2$.

The pairs of Hypothesis (P) belong to these classes. Indeed, let $Y \subset G \setminus \partial_2 G$. Let $u, v \in G \setminus Y$ satisfy $d(u,v) = 2$ and have a common neighbor in Y. Recall from Section 6.1 that the domain $G \setminus Y$ is consistent and that its integrals are those of (5.1). The sites u and v have status 0 in it, so $(u,v) \in \mathcal{J}_{\mathrm{coll}}(G \setminus Y) \cup \mathcal{J}_{\mathrm{diag}}(G \setminus Y)$. In this notation, Hypothesis (P) asks for the bound $|r^\tau_{G\setminus Y}((u,v))| \le A\theta^{d((u,v))}$. Therefore, a bound $|r^\tau_D(p)| \le A\theta^{d(p)}$ for all $D \in \mathcal{D}_{\mathrm{cons}}(G)$ and all $p \in \mathcal{J}_{\mathrm{coll}}(D) \cup \mathcal{J}_{\mathrm{diag}}(D)$, for all finite G and all τ, implies Hypothesis (P) with the same constants.

7.2 Closure of the pair classes

Let $D \in \mathcal{D}_{\mathrm{cons}}(G)$ and $\mathfrak{t} \in \{\mathrm{coll}, \mathrm{diag}\}$. Let $p = (u,v) \in \mathcal{J}_{\mathfrak{t}}(D)$ with center c, and suppose that $u \notin \partial_2 G$. We let $\hat{D}$ denote the domain obtained from D by adding the bond $\{c,u\}$ to the used bonds. Since c is removed, this bond is not present in D, so $\hat{D}$ and D have the same statuses and the same present bonds. In particular, $\hat{D}$ is a domain, $Z^\cdot_{\hat{D}}[f] = Z^\cdot_D[f]$ for all f, and the trails in $\hat{D}$ from u are the trails in D from u, with the same statuses along them. The additional used bond only serves to count the used bonds at u. We set $T = \partial_2 G \cup \{v\}$.

The assumptions of Section 6.2 hold for the domain $\hat{D}$, the site u, the arrival site $v_{-1} = c$ of u, and the site v. We have $u \notin T$ by construction, and Definition 7.1 makes u and v non-removed. The removed center c makes $\{c,u\}$ absent from the present bonds. If u has status 1, then it has at most one present bond by condition (3) of Definition 7.1. In this section, the traced domains, steps, arrival sites, remaining neighbors, domains D^-_j, and patterns of steps of Section 6 refer to these choices. In Figure 6.1, the pair (u,v) is collinear, with center c.

Lemma 7.2. *Let D, $p = (u,v)$, c, $\hat{D}$, and T be as above, and let $w = (v_0, \dots, v_m) \in \mathcal{T}_{\hat{D}}(u,T)$. Then for $0 \le j \le m$, the following hold, where $v_{-1} = c$.*

1. *In $\hat{D}[w_j]$, every site of status 0 other than v_j has no used bond, and every site of status 1 other than v_j has exactly two used bonds.*
2. *If v_j has status 0 in $\hat{D}[w_j]$, then its only used bond in $\hat{D}[w_j]$ is $\{v_{j-1}, v_j\}$, and v_{j-1} has status 1 or 2 in $\hat{D}[w_j]$.*
3. *If $j < m$, then the domain D^-_{j+1} is consistent.*
4. *If $j < m$ and step $j+1$ is a first visit, then the pattern of step $j+1$ belongs to $\mathcal{P}_{\mathrm{ext}}$. Further, every pair (a,b) of distinct remaining neighbors of v_j at step $j+1$ belongs to $\mathcal{J}_{\mathrm{coll}}(D^-_{j+1}) \cup \mathcal{J}_{\mathrm{diag}}(D^-_{j+1})$, and v_j is a common neighbor of a and b of status 2 in D^-_{j+1}.*

Proof. Along w, the status of a site changes only when the site is left, from 0 to 1 at a first visit and from 1 to 2 at a revisit. Step j adds the bond $\{v_{j-1}, v_j\}$ to the used bonds, which

changes the number of used bonds only at v_{j-1} and v_j.

We prove (1) and (2) by induction on j. Recall that $\hat{D}$ differs from $D \in \mathcal{D}_{\mathrm{cons}}(G)$ only by the used bond $\{c, u\}$, and that c is removed. Every site other than u therefore satisfies the conditions in (1) for $j = 0$. Suppose that u has status 0. Since D is consistent, u has no used bond in D, and its only used bond in $\hat{D}$ is $\{c, u\}$, whose end c has status 2. We conclude that (1) and (2) hold for $j = 0$.

Suppose that (1) and (2) hold for some $j < m$. We first consider the site v_j. When v_j has status 0 in $\hat{D}[w_j]$, part (2) shows that it has status 1 and exactly two used bonds in $\hat{D}[w_{j+1}]$, namely $\{v_{j-1}, v_j\}$ and $\{v_j, v_{j+1}\}$. When v_j has status 1 in $\hat{D}[w_j]$, it is removed in $\hat{D}[w_{j+1}]$.

We next consider the site $v_{j+1} \neq v_j$, which is not removed, since $\{v_j, v_{j+1}\}$ is present in $\hat{D}[w_j]$. When v_{j+1} has status 0, part (1) shows that it has no used bond in $\hat{D}[w_j]$. Its only used bond in $\hat{D}[w_{j+1}]$ is then $\{v_j, v_{j+1}\}$, and v_j has status 1 or 2 in $\hat{D}[w_{j+1}]$. The statuses and used bonds of the other sites stay the same at step $j + 1$, and (1) and (2) hold for $j + 1$.

We now prove (3). Let $j < m$. The site v_j does not belong to $T \supset \partial_2 G$, and every bond that is present in D^-_{j+1} is present in $\hat{D}[w_j]$. Therefore, D^-_{j+1} is a domain. In D^-_{j+1}, the site v_j is removed, and the statuses and used bonds of all other sites are those of $\hat{D}[w_j]$. By (1), D^-_{j+1} is consistent.

We now prove (4). Let $j < m$ be such that step $j + 1$ is a first visit. Its passage site $x = v_j$ has status 0 in $\hat{D}[w_j]$. By (2), the only used bond at x is the arrival bond $\{v_{j-1}, x\}$, and the arrival site v_{j-1} has status 1 or 2. The ring sites of $\Diamond(x)$ are different from x. By (1), every ring site of status 0 has no used bond and every ring site of status 1 has exactly two used bonds. Transported to $\Diamond(0)$, these are the conditions (1)–(4) of Definition 6.10, and the pattern of step $j + 1$ belongs to $\mathcal{P}_{\mathrm{ext}}$.

Since the only used bond at x leads to v_{j-1}, the remaining neighbors of x at step $j + 1$ are the non-removed neighbors of x other than v_{j-1}. Let $a \neq b$ be two of them. Then $d(a, b) = 2$, the sites a and b are not removed in D^-_{j+1}, and their common neighbor x is removed in D^-_{j+1}.

Suppose that a has status 1. By (1), it has exactly two used bonds in $\hat{D}[w_j]$, and $\{a, x\}$ is not one of them, since the only used bond at x leads to $v_{j-1} \neq a$. The site a does not belong to $\partial_2 G$, so that its four bonds join it to sites of G, and the bond $\{a, x\}$ is not present in D^-_{j+1}, since x is removed. We conclude that a has at most one present bond in D^-_{j+1}. The same holds for b. Since D^-_{j+1} is consistent by (3), it follows that $(a, b) \in \mathcal{J}_{\mathrm{coll}}(D^-_{j+1}) \cup \mathcal{J}_{\mathrm{diag}}(D^-_{j+1})$. □

We record the types of the pairs in part (4). Let step j be a first visit with passage site x and out-neighbor o. Let f and g be the two neighbors of x other than the arrival site and o. At a straight first visit, f and g lie on opposite sides of x, so (f, g) is collinear, and

(f,o) and (g,o) are diagonal. At a turn, one of f and g, say f, is opposite to the arrival site, and g is opposite to o; then (f,g) and (f,o) are diagonal, and (g,o) is collinear. By Lemma 6.8, the pair that does not contain o carries the coefficient $\alpha_j \le 1/6$, and the pairs that contain o carry coefficients at most $9/25$.

7.3 Weighted trail sums

We now define the weighted sums over trails of $\mathbb{Z}^2$ that will dominate the contribution of the trails that end at the second site of a pair. Let $\mathrm{t} \in \{\mathrm{coll}, \mathrm{diag}\}$. In this subsection, we write $c = (0,0)$, $u = (-1,0)$, and $v = v^{\mathrm{t}}$, where $v^{\mathrm{coll}} = (1,0)$ and $v^{\mathrm{diag}} = (0,1)$, so that (u,v) is a pair of type t with center c. We let $\mathcal{T}_n^{\mathrm{t}}$ denote the set of trails $w = (v_0, \dots, v_n)$ of $\mathbb{Z}^2$ of length n that do not visit c, with $v_0 = u$, $v_n = v$, and $v_i \ne v$ for $0 \le i < n$. We use the notions of Section 6.6 with the arrival site $v_{-1} = c$ of u.

Let $w \in \mathcal{T}_n^{\mathrm{t}}$. Recall from Section 6.6 that step j of w, for $1 \le j \le n$, is the move from v_{j-1} to v_j, that its arrival site is v_{j-2}, and that it is a revisit if $v_{j-1} \in \{v_0, \dots, v_{j-2}\}$. There are $|\mathrm{FV}(w)|$ first-visit steps and $\mathrm{rev}(w)$ revisits.

We set $\bar\alpha = 1/6$ and $\bar\beta = 9/25$, the bounds on the coefficients in Lemma 6.8. Let $j \in \mathrm{FV}(w)$. The three neighbors of v_{j-1} other than the arrival site v_{j-2} form three unordered pairs, which we call the *pairs at step j*. We assign the coefficient $\bar\alpha$ to the pair at step j that does not contain v_j, and the coefficient $\bar\beta$ to the two pairs that contain v_j. We call a pair at step j *dead* if it contains c or a site that w has left twice before step j (a site that occurs twice among $v_0, \dots, v_{j-2}$), and *live* otherwise. We use the *type weights*

$$h = (h_{\mathrm{coll}}, h_{\mathrm{diag}}) = (1, 5/4), \tag{7.2}$$

we let $h(p) = h_{\mathrm{t}'}$ denote the type weight of a pair p of type t', and we set $h_{\min} = 1$ and $h_{\max} = 5/4$. For $j \in \mathrm{FV}(w)$, we define

$$\kappa(j) = \sum_{p \text{ live at step } j} (\text{coefficient of } p)\, h(p), \tag{7.3}$$

and we set $\bar\rho_j = 1/c_{\mathrm{s}}$ if step j is straight and $\bar\rho_j = 1/c_{\mathrm{t}}$ if it is a turn. For each $w \in \mathcal{T}_n^{\mathrm{t}}$, let ς_w be an enumeration of $\mathrm{FV}(w)$, as in Section 6.4. For $\mathcal{T} \subset \mathcal{T}_n^{\mathrm{t}}$, we define

$$\mathsf{m}(\mathcal{T}) = \sum_{w \in \mathcal{T}} 3^{-n} \Big(\frac{3}{5}\Big)^{\mathrm{rev}(w)} \sum_{k=1}^{|\mathrm{FV}(w)|} \kappa(\varsigma_w(k)) \prod_{l=1}^{k} \bar\rho_{\varsigma_w(l)}, \qquad \mathsf{m}_{\mathrm{t}}(n) = \mathsf{m}(\mathcal{T}_n^{\mathrm{t}}), \tag{7.4}$$

and for $0 < t \le 1$, we define the normalized weighted sums

$$K_{\mathrm{t}}(t) = \frac{1}{h_{\mathrm{t}}} \sum_{n \ge 1} \mathsf{m}_{\mathrm{t}}(n)\, t^{-n-1} \in [0, \infty]. \tag{7.5}$$

These quantities depend on c_{s} and c_{t} through the numbers $\bar\rho_j$, and on the family of enumerations (ς_w).

The term of a trail w in (7.4) is the bound that Lemma 6.9 and Lemma 6.8 give for a trail that returns to the second site of a pair, with the response coefficients replaced by $\bar\alpha$ and $\bar\beta$ and each response replaced by its type weight. The dead pairs are left out, since in the expansion they contain a removed site and have no response. The factor t^{-n-1} in (7.5) accounts for the distance to $\partial_2 G$ that a pair can lose along a trail of length n.

Hypothesis (H3). *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0,1]$, and let the sums* (7.4) *be formed with $\bar\alpha = 1/6$, $\bar\beta = 9/25$, the type weights $h = (1, 5/4)$, and a fixed family of enumerations ς_w of the sets $\mathrm{FV}(w)$ of first-visit steps, chosen as in Appendix C. There exists $K_\star < 1$ such that $K_{\mathrm{coll}}(1) \le K_\star$ and $K_{\mathrm{diag}}(1) \le K_\star$.*

Since the terms of (7.5) are nonnegative, the bound $K_{\mathrm{t}}(1) \le K_\star$ is equivalent to the bounds $\sum_{n=1}^{N} \mathsf{m}_{\mathrm{t}}(n) \le K_\star h_{\mathrm{t}}$ for all $N \ge 1$, and it can be verified without knowing in advance that the series converges. Our arguments hold for every family of enumerations; the family of Appendix C is chosen to make the sums small.

Lemma 7.3. *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0,1]$, and assume that (H2) and (H3) hold at $(c_{\mathrm{s}}, c_{\mathrm{t}})$. Then for $\mathrm{t} \in \{\mathrm{coll}, \mathrm{diag}\}$, the function K_{t} is finite and continuous on $(\gamma_{\mathrm{o}}, 1]$. In particular, there exists $\theta \in (\gamma_{\mathrm{o}}, 1)$ such that*

$$K_\theta = \max\bigl(K_{\mathrm{coll}}(\theta), K_{\mathrm{diag}}(\theta)\bigr) < 1. \tag{7.6}$$

Proof. The finiteness of K_{t} is a routine consequence of (H2). Let $w \in \mathcal{T}_n^{\mathrm{t}}$. Every prefix product in (7.4) is bounded by the product of $\bar\rho_j$ over all $j \in \mathrm{FV}(w)$, since $\bar\rho_j \ge 1$. Further, $\kappa(j) \le h_{\max}(\bar\alpha + 2\bar\beta)$ for all j, and $|\mathrm{FV}(w)| \le n$. The term of w in (7.4) is therefore bounded by

$$n\, h_{\max}(\bar\alpha + 2\bar\beta)\, 3^{-n} \Big(\frac{3}{5}\Big)^{\mathrm{rev}(w)} \prod_{j \in \mathrm{FV}(w)} \bar\rho_j = n\, h_{\max}(\bar\alpha + 2\bar\beta)\, \mathrm{wt}(w),$$

where $\mathrm{wt}(w)$ is the trail weight (6.23) with the arrival site c; here we used $\mathrm{wt}_j = \bar\rho_j/3$ for $j \in \mathrm{FV}(w)$ and $\mathrm{wt}_j = (1/3)(3/5)$ for a revisit. Every $w \in \mathcal{T}_n^{\mathrm{t}}$ is a trail of $\mathbb{Z}^2$ of length n from u, and its first step does not go to the arrival site c. Applying (H2) with $v_0 = u$ and $v_{-1} = c$, and using $h_{\mathrm{t}} \ge h_{\min}$, we obtain

$$\frac{\mathsf{m}_{\mathrm{t}}(n)}{h_{\mathrm{t}}} \le \frac{h_{\max}}{h_{\min}} (\bar\alpha + 2\bar\beta)\, n\, C_{\mathrm{o}} \gamma_{\mathrm{o}}^n \le 2n C_{\mathrm{o}} \gamma_{\mathrm{o}}^n, \tag{7.7}$$

where in the last inequality we used $(5/4)(1/6 + 18/25) = 133/120 < 2$.

Let $t_0 \in (\gamma_{\mathrm{o}}, 1]$. For $t \in [t_0, 1]$, the terms of (7.5) are continuous in t and bounded by $2nC_{\mathrm{o}}\gamma_{\mathrm{o}}^n t_0^{-n-1}$, which is summable in n. Hence the series (7.5) converges uniformly on $[t_0, 1]$, and K_{t} is finite and continuous on $(\gamma_{\mathrm{o}}, 1]$. By (H3), $K_{\mathrm{t}}(1) \le K_\star < 1$ for $\mathrm{t} \in \{\mathrm{coll}, \mathrm{diag}\}$, and the existence of θ follows from the continuity at $t = 1$. □

7.4 The contraction

We now state the main result of this section.

Theorem 7.4. *Let $c_{\mathrm{s}}, c_{\mathrm{t}} \in [187/200, 1]$, and assume that (H1), (H2), and (H3) hold at $(c_{\mathrm{s}}, c_{\mathrm{t}})$. Then there exist $A \geq 2$ and $\theta \in (0,1)$ such that for every finite $G \subset \mathbb{Z}^2$, every pin configuration τ on G, every $D \in \mathcal{D}_{\mathrm{cons}}(G)$, and every $p \in \mathcal{J}_{\mathrm{coll}}(D) \cup \mathcal{J}_{\mathrm{diag}}(D)$, we have*

$$|r_D^\tau(p)| \leq A\theta^{d(p)}. \tag{7.8}$$

In particular, Hypothesis (P) holds with these constants.

The proof occupies the rest of this section. We first split the response of a pair (u, v) into the contributions of the trails that end at v and of the trails that end in $\partial_2 G$. For a domain D' and a trail w in D', we set $a_{D'}(w) = 3^{-|w|}(3/5)^{\mathrm{rev}_{D'}(w)}$, the coefficient of w in (6.6).

Lemma 7.5. *Let $D \in \mathcal{D}_{\mathrm{cons}}(G)$, let $\mathfrak{t} \in \{\mathrm{coll}, \mathrm{diag}\}$, and let $p = (u, v) \in \mathcal{J}_{\mathfrak{t}}(D)$ with center c and $d(p) \geq 1$. Let $\hat{D}$ and T be as in Section 7.2, let $\mathcal{T}_{\mathrm{ret}}$ denote the set of trails in $\mathcal{T}_{\hat{D}}(u, T)$ that end at v, and let $\mathcal{T}_{\mathrm{pin}} = \mathcal{T}_{\hat{D}}(u, T) \setminus \mathcal{T}_{\mathrm{ret}}$. Then $r_D^\tau(p) = \mathrm{Y}_{\mathrm{ret}} + \mathrm{Y}_{\mathrm{pin}}$, where*

$$\begin{aligned} \mathrm{Y}_{\mathrm{ret}} &= \sum_{w \in \mathcal{T}_{\mathrm{ret}}} a_{\hat{D}}(w)\big(\varrho_{\hat{D}}^\tau(w) - \varrho_{\hat{D}}^0(w)\big)\mathbb{E}^0_{\hat{D}[w]}[|n_v|^2], \\ \mathrm{Y}_{\mathrm{pin}} &= \sum_{w \in \mathcal{T}_{\mathrm{pin}}} a_{\hat{D}}(w)\Big(\varrho_{\hat{D}}^\tau(w)\,\mathbb{E}^\tau_{\hat{D}[w]}[n_{v_{|w|}} \cdot n_v] - \varrho_{\hat{D}}^0(w)\,\mathbb{E}^0_{\hat{D}[w]}[n_{v_{|w|}} \cdot n_v]\Big). \end{aligned} \tag{7.9}$$

Moreover, $\mathbb{E}^0_{\hat{D}[w]}[|n_v|^2] \leq 1$ for all $w \in \mathcal{T}_{\mathrm{ret}}$, and every trail in $\mathcal{T}_{\mathrm{pin}}$ ends in $\partial_2 G$.

Proof. Since $d(p) \geq 1$, the sites u and v do not belong to $\partial_2 G$, and the construction of Section 7.2 applies. The domains $\hat{D}$ and D define the same integrals, and we have

$$r_D^\tau(p) = \mathbb{E}^\tau_{\hat{D}}[n_u \cdot n_v] - \mathbb{E}^0_{\hat{D}}[n_u \cdot n_v].$$

We expand the two terms with Proposition 6.6, applied to $\hat{D}$, u, v, and T. The two expansions run over the same set $\mathcal{T}_{\hat{D}}(u, T)$, with the same coefficients $a_{\hat{D}}(w)$.

Let $w \in \mathcal{T}_{\mathrm{ret}}$. By condition (3) of Definition 7.1, the site v has status 0, or status 1 and at most one present bond. In the first case, and in the second case when v has exactly one present bond, Proposition 6.6 shows that the last factor of the term of w is $\mathbb{E}^\cdot_{\hat{D}[w]}[|n_v|^2] \in \{1, 3/5\}$, with the same value for $\cdot = \tau$ and $\cdot = 0$. When v has status 1 and no present bond, the set $\mathcal{T}_{\mathrm{ret}}$ is empty, since the last step of a trail that ends at v uses a bond at v that is present in $\hat{D}$. Subtracting the two expansions gives $r_D^\tau(p) = \mathrm{Y}_{\mathrm{ret}} + \mathrm{Y}_{\mathrm{pin}}$. Finally, a trail in $\mathcal{T}_{\mathrm{pin}}$ ends in $T \setminus \{v\} \subset \partial_2 G$. □

The next lemma compares the returning trails in $\hat{D}$ with the trails of $\mathcal{T}_n^{\mathfrak{t}}$ that enter the sums (7.4). This comparison is the subtle point of the argument. A step of a returning trail can be a revisit in $\hat{D}$ and a first visit in $\mathbb{Z}^2$, a pair can be absent from the expansion and live in (7.4), and the coefficients of Lemma 6.8 must be compared with $\bar{\alpha}$ and $\bar{\beta}$; in each case, the trail sums (7.4) give the larger contribution.

Lemma 7.6. *Let D, $\mathfrak{t}$, $p = (u, v)$, c, $\hat{D}$, and $\mathcal{T}_{\mathrm{ret}}$ be as in Lemma 7.5. Then there exists an isometry of $\mathbb{Z}^2$ that maps c to $(0,0)$, u to $(-1,0)$, and v to $v^{\mathfrak{t}}$. Fix such an isometry, and for $w \in \mathcal{T}_{\mathrm{ret}}$, let $\tilde{w} = (\tilde{v}_0, \dots, \tilde{v}_n)$ denote the image of $w = (v_0, \dots, v_n)$ under it. Then the following hold.*

1. *We have $\tilde{w} \in \mathcal{T}_n^{\mathfrak{t}}$, and the map $w \mapsto \tilde{w}$ is injective on $\mathcal{T}_{\mathrm{ret}}$.*
2. *The set $\mathrm{FV}(w)$ of first-visit steps of w in $\hat{D}$ is contained in the set $\mathrm{FV}(\tilde{w})$ of first-visit steps of $\tilde{w}$ in the sense of Section 6.6, and $\mathrm{rev}_{\hat{D}}(w) \geq \mathrm{rev}(\tilde{w})$. A step $j \in \mathrm{FV}(w)$ is straight for w if and only if it is straight for $\tilde{w}$.*
3. *Let $j \in \mathrm{FV}(w)$. The image of every pair of distinct remaining neighbors of v_{j-1} at step j is a live pair at step j of $\tilde{w}$, of the same type. It contains $\tilde{v}_j$ if and only if the original pair contains v_j.*

Proof. The vectors $c - u$ and $v - c$ are unit vectors, which are equal if $\mathfrak{t} = \mathrm{coll}$ and orthogonal if $\mathfrak{t} = \mathrm{diag}$. Hence an isometry with the stated properties exists. It preserves neighbors, the distinction between straight steps and turns, and the types of pairs. Further, it maps the arrival site c of u to the arrival site $(0,0)$ of $(-1,0)$.

For (1), observe that every site of w is non-removed in $\hat{D}$, since $v_0 = u$ and v_j is an end of a bond that is present in $\hat{D}[w_{j-1}]$ for $j \geq 1$. In particular, w does not visit c. The bonds of w are pairwise distinct, $v_n = v$, and $v_i \neq v$ for $i < n$. Hence $\tilde{w} \in \mathcal{T}_n^{\mathfrak{t}}$. An isometry is injective, and so is the map $w \mapsto \tilde{w}$.

For (2), let $1 \leq j \leq n$ be such that step j of $\tilde{w}$ is a revisit, so that $v_{j-1} = v_i$ for some $i \leq j - 2$. Then step $i + 1$ of w left v_{j-1} and raised its status, so v_{j-1} has status 1 in $\hat{D}[w_{j-1}]$, and step j is a revisit of w in $\hat{D}$. This proves the inclusion $\mathrm{FV}(w) \subset \mathrm{FV}(\tilde{w})$, and the inequality follows from $\mathrm{rev}_{\hat{D}}(w) = n - |\mathrm{FV}(w)|$ and $\mathrm{rev}(\tilde{w}) = n - |\mathrm{FV}(\tilde{w})|$. The steps in $\mathrm{FV}(\tilde{w}) \setminus \mathrm{FV}(w)$ leave sites that have status 1 in $\hat{D}$ and that w has not left before. The statement on straight steps holds since the isometry preserves them.

For (3), let $j \in \mathrm{FV}(w)$. Let $a \neq b$ be remaining neighbors of $x = v_{j-1}$ at step j. They are neighbors of x other than the arrival site v_{j-2}, so their images form a pair at step j of $\tilde{w}$. This pair does not contain $(0,0)$, since c is removed and remaining neighbors are not. Suppose that it contains a site that $\tilde{w}$ has left twice before step j. The corresponding site of w has then been left twice by w before step j, so it has status 2 in $\hat{D}[w_{j-1}]$ and is not a remaining neighbor. This contradiction shows that the pair is live. The remaining statements hold since the isometry preserves types and maps v_j to $\tilde{v}_j$. □

We now bound the contribution of the returning trails.

Lemma 7.7. *Let $c_s, c_t \in [187/200, 1]$, and assume that (H1) holds at (c_s, c_t). Let $\mathcal{M} \geq 0$ and $\theta \in (0,1]$ be such that*

$$|r^{\tau}_{D'}(p')| \leq \mathcal{M} h(p') \theta^{d(p')} \quad \text{for all } D' \in \mathcal{D}_{\mathrm{cons}}(G) \text{ and all } p' \in \mathcal{J}_{\mathrm{coll}}(D') \cup \mathcal{J}_{\mathrm{diag}}(D'). \tag{7.10}$$

Let D, $\mathfrak{t}$, $p = (u, v)$, $\hat{D}$, $\mathcal{T}_{\mathrm{ret}}$, and Υ_{ret} be as in Lemma 7.5, and let $d = d(p)$. Then for all $n \geq 1$,

$$\sum_{\substack{w \in \mathcal{T}_{\mathrm{ret}} \\ |w| = n}} a_{\hat{D}}(w) \left| \varrho^{\tau}_{\hat{D}}(w) - \varrho^{0}_{\hat{D}}(w) \right| \leq \mathcal{M} \theta^{d-n} \mathsf{m}_{\mathfrak{t}}(n). \tag{7.11}$$

If $K_{\mathfrak{t}}(\theta) < \infty$, then $|\Upsilon_{\mathrm{ret}}| \leq \mathcal{M} h_{\mathfrak{t}} \theta^{d+1} K_{\mathfrak{t}}(\theta)$.

Proof. Let $w = (v_0, \dots, v_n) \in \mathcal{T}_{\mathrm{ret}}$, and let $\tilde{w}$ be as in Lemma 7.6. We let $\mathrm{FV}(w)$ denote the set of first-visit steps of w in $\hat{D}$. For $j \in \mathrm{FV}(\tilde{w})$, we write $\kappa(j)$ and $\bar{\rho}_j$ for the quantities of Section 7.3, computed for $\tilde{w}$.

Step 1: the step bounds. Let $j \in \mathrm{FV}(w)$. By part (2) of Lemma 7.6, step j of $\tilde{w}$ is a first visit of the same type as step j of w. By part (4) of Lemma 7.2, the pattern π of step j belongs to $\mathcal{P}_{\mathrm{ext}}$. Using Lemma 6.11 and the assumption that $c_s, c_t \in [187/200, 1]$, we obtain

$$\rho^{\tau}_j \leq \frac{1}{c_\pi} = \bar{\rho}_j, \qquad 1 \leq \bar{\rho}_j \leq \frac{200}{187}.$$

Therefore, Lemma 6.8 applies at step j, and Lemma 6.9 applies with these numbers $\bar{\rho}_j$, $j \in \mathrm{FV}(w)$.

Step 2: the responses at a step. Let $j \in \mathrm{FV}(w)$, let $x = v_{j-1}$, and let $o = v_j$. We define $\kappa'(j)$ as the sum, over the unordered pairs $\{a, b\}$ of distinct remaining neighbors of x at step j, of $\bar{\alpha}\, h((a,b))$ if $o \notin \{a, b\}$ and of $\bar{\beta}\, h((a,b))$ if $o \in \{a, b\}$. By part (4) of Lemma 7.2, $D^-_j \in \mathcal{D}_{\mathrm{cons}}(G)$, and each such pair belongs to $\mathcal{J}_{\mathrm{coll}}(D^-_j) \cup \mathcal{J}_{\mathrm{diag}}(D^-_j)$.

By (7.10), we have $|r_{ab}| \leq \mathcal{M} h((a,b)) \theta^{d((a,b))}$. The sites a and b are neighbors of x, and the prefix w_{j-1} joins u to x in $j-1$ steps. Thus both $d(u,a)$ and $d(u,b)$ are at most $j \leq n$. Since $z \mapsto \mathrm{dist}(z, \partial_2 G)$ is 1-Lipschitz, we obtain

$$d((a,b)) \geq \mathrm{dist}(u, \partial_2 G) - n \geq d - n, \qquad \theta^{d((a,b))} \leq \theta^{d-n}.$$

Lemma 6.8 bounds $|\Gamma_j|$ by the sum of $|r_{ab}|$ times $\alpha_j \leq \bar{\alpha}$ for the pair without o, and times $\beta_j \leq \bar{\beta}$ or $\beta'_j \leq \bar{\beta}$ for the pairs with o. Therefore,

$$|\Gamma_j| \leq \mathcal{M} \theta^{d-n} \kappa'(j) \leq \mathcal{M} \theta^{d-n} \kappa(j), \tag{7.12}$$

where the second inequality follows from part (3) of Lemma 7.6, since $\kappa'(j)$ is a sum of some of the nonnegative terms of $\kappa(j)$.

Step 3: the enumeration. Let $q = |\mathrm{FV}(\tilde{w})|$, and let $\varsigma_{\tilde{w}}$ be the enumeration of $\mathrm{FV}(\tilde{w})$ in (7.4). We list the elements of $\mathrm{FV}(w) \subset \mathrm{FV}(\tilde{w})$ in the order in which they occur in $\varsigma_{\tilde{w}}(1), \dots, \varsigma_{\tilde{w}}(q)$, and we let ς denote the resulting enumeration of $\mathrm{FV}(w)$. Let $1 \le k \le |\mathrm{FV}(w)|$, and let k' be such that $\varsigma_{\tilde{w}}(k') = \varsigma(k)$. Then $\{\varsigma(1), \dots, \varsigma(k)\} \subset \{\varsigma_{\tilde{w}}(1), \dots, \varsigma_{\tilde{w}}(k')\}$, and since $\bar{\rho}_j \ge 1$ for all $j \in \mathrm{FV}(\tilde{w})$, we have

$$\prod_{l=1}^{k} \bar{\rho}_{\varsigma(l)} \le \prod_{l=1}^{k'} \bar{\rho}_{\varsigma_{\tilde{w}}(l)}. \tag{7.13}$$

Step 4: conclusion. By Lemma 6.9 with the enumeration ς, (7.12), and (7.13), we have

$$\bigl|\varrho^{\tau}_{\hat{D}}(w) - \varrho^{0}_{\hat{D}}(w)\bigr| \le \mathcal{M}\theta^{d-n} \sum_{k=1}^{|\mathrm{FV}(w)|} \kappa(\varsigma(k)) \prod_{l=1}^{k} \bar{\rho}_{\varsigma(l)} \le \mathcal{M}\theta^{d-n} \sum_{k'=1}^{q} \kappa(\varsigma_{\tilde{w}}(k')) \prod_{l=1}^{k'} \bar{\rho}_{\varsigma_{\tilde{w}}(l)}.$$

In the second inequality, the term k on the left is bounded by the term k' with $\varsigma_{\tilde{w}}(k') = \varsigma(k)$ on the right, and the other terms on the right are nonnegative. By part (2) of Lemma 7.6, we also have $a_{\hat{D}}(w) \le 3^{-n}(3/5)^{\mathrm{rev}(\tilde{w})}$. Multiplying the two bounds, we find that $a_{\hat{D}}(w)|\varrho^{\tau}_{\hat{D}}(w) - \varrho^{0}_{\hat{D}}(w)|$ is bounded by $\mathcal{M}\theta^{d-n}$ times the term of $\tilde{w}$ in (7.4). Summing over $w \in \mathcal{T}_{\mathrm{ret}}$ with $|w| = n$ and using part (1) of Lemma 7.6, we obtain (7.11).

Finally, if $K_{\mathfrak{t}}(\theta) < \infty$, Lemma 7.5 and (7.11) imply that

$$|\mathrm{Y}_{\mathrm{ret}}| \le \sum_{n \ge 1} \mathcal{M}\theta^{d-n} \mathsf{m}_{\mathfrak{t}}(n) = \mathcal{M}\theta^{d+1} \sum_{n \ge 1} \mathsf{m}_{\mathfrak{t}}(n)\theta^{-n-1} = \mathcal{M} h_{\mathfrak{t}} \theta^{d+1} K_{\mathfrak{t}}(\theta),$$

where we used $\mathbb{E}^{0}_{\hat{D}[w]}[|n_v|^2] \le 1$ and the definition (7.5). □

We now bound the contribution of the trails that end in $\partial_2 G$.

Lemma 7.8. *Let $c_{\mathfrak{s}}, c_{\mathfrak{t}} \in (0,1]$, and assume that (H1) and (H2) hold at $(c_{\mathfrak{s}}, c_{\mathfrak{t}})$. Let D, $\mathfrak{t}$, $p = (u, v)$, c, $\hat{D}$, $\mathcal{T}_{\mathrm{pin}}$, and $\mathrm{Y}_{\mathrm{pin}}$ be as in Lemma 7.5, and let $d = d(p)$. Then*

$$|\mathrm{Y}_{\mathrm{pin}}| \le C_{\mathrm{pin}} \gamma_{\mathrm{o}}^{d}, \qquad C_{\mathrm{pin}} = \frac{2C_{\mathrm{o}}}{1 - \gamma_{\mathrm{o}}}. \tag{7.14}$$

Proof. Since all spins have norm at most one, we have $|\mathbb{E}^{\cdot}_{\hat{D}[w]}[n_{v_{|w|}} \cdot n_v]| \le 1$, and (7.9) gives

$$|\mathrm{Y}_{\mathrm{pin}}| \le \sum_{w \in \mathcal{T}_{\mathrm{pin}}} a_{\hat{D}}(w) \bigl(\varrho^{\tau}_{\hat{D}}(w) + \varrho^{0}_{\hat{D}}(w)\bigr). \tag{7.15}$$

Let $w = (v_0, \dots, v_n) \in \mathcal{T}_{\mathrm{pin}}$. We regard w as a trail of $\mathbb{Z}^2$ with the arrival site c of u, and we let $\mathrm{wt}(w)$ be its trail weight (6.23). We claim that

$$a_{\hat{D}}(w)\, \varrho^{\cdot}_{\hat{D}}(w) \le \mathrm{wt}(w), \qquad \cdot \in \{\tau, 0\}. \tag{7.16}$$

To prove (7.16), recall from part (1) of Lemma 6.7 that the step ratios of the revisits in $\hat{D}$ equal one. Let j be a first-visit step of w in $\hat{D}$. By part (4) of Lemma 7.2 and Lemma 6.11, we have $\rho_j^\tau \le 1/c_\pi$, where π is the pattern of step j, and Lemma 6.8 gives $\rho_j^0 \le 1 \le 1/c_\pi$. We conclude that $a_{\hat{D}}(w)\varrho_{\hat{D}}^{\cdot}(w)$ is bounded by the product over the steps j of w of the factor $1/5$ if step j is a revisit in $\hat{D}$, and $1/(3c_\pi)$ if it is a first visit in $\hat{D}$.

We now compare these factors with wt_j. When step j leaves a site that w has left before, it is a revisit in $\hat{D}$, and the two factors are $1/5$. Otherwise, $\mathrm{wt}_j = 1/(3c_\mathrm{s})$ or $1/(3c_\mathrm{t})$, according to the type of step j, and the factor above is either $1/(3c_\pi) = \mathrm{wt}_j$ or $1/5 < 1/3 \le \mathrm{wt}_j$. This gives (7.16).

The trails in $\mathcal{T}_{\mathrm{pin}}$ are distinct trails of $\mathbb{Z}^2$ from u. Their first step avoids the removed site c, and they end in $\partial_2 G$. Their length is therefore at least $\mathrm{dist}(u, \partial_2 G) \ge d$. Inserting (7.16) into (7.15) and applying (H2) with $v_0 = u$ and $v_{-1} = c$, we obtain

$$|\mathrm{Y}_{\mathrm{pin}}| \le 2 \sum_{n \ge d} \mathrm{C_o} \gamma_\mathrm{o}^n = \mathrm{C}_{\mathrm{pin}} \gamma_\mathrm{o}^d,$$

which proves (7.14). □

Proof of Theorem 7.4. Choose $\theta \in (\gamma_\mathrm{o}, 1)$ as in Lemma 7.3, so that $K_\theta < 1$. This choice uses only the functions K_{coll} and K_{diag} and the constant γ_o. Let $G \subset \mathbb{Z}^2$ be finite. We define the weighted supremum

$$\mathcal{M}_G = \sup_{\tau, D, p} \frac{|r_D^\tau(p)|}{h(p)\theta^{d(p)}}, \tag{7.17}$$

where the supremum runs over all pin configurations τ on G, all $D \in \mathcal{D}_{\mathrm{cons}}(G)$, and all $p \in \mathcal{J}_{\mathrm{coll}}(D) \cup \mathcal{J}_{\mathrm{diag}}(D)$, and where the supremum of the empty set is 0. Since $|r_D^\tau(p)| \le 2$, $h(p) \ge 1$, and $d(p)$ is bounded by the diameter of G, the supremum $\mathcal{M}_G$ is finite. We work with a single set G because all responses that occur below have the same boundary layer $\partial_2 G$, which makes the following bound uniform.

Let τ, D, and $p = (u, v) \in \mathcal{J}_\mathrm{t}(D)$ be as in (7.17), and let $d = d(p)$. When $d = 0$, we have

$$\frac{|r_D^\tau(p)|}{h(p)\theta^{d(p)}} \le \frac{2}{h_{\min}}.$$

Suppose now that $d \ge 1$. By the definition of $\mathcal{M}_G$, the bound (7.10) holds with $\mathcal{M} = \mathcal{M}_G$, for the pin configuration τ. Since $c_\mathrm{s}, c_\mathrm{t} \in [187/200, 1]$, we may apply Lemma 7.5, Lemma 7.7, and Lemma 7.8, and we obtain

$$|r_D^\tau(p)| \le |\mathrm{Y}_{\mathrm{ret}}| + |\mathrm{Y}_{\mathrm{pin}}| \le \mathcal{M}_G h_\mathrm{t} \theta^{d+1} K_\mathrm{t}(\theta) + \mathrm{C}_{\mathrm{pin}} \gamma_\mathrm{o}^d \le K_\theta \mathcal{M}_G h_\mathrm{t} \theta^d + \mathrm{C}_{\mathrm{pin}} \theta^d,$$

where in the last inequality we used $\theta \le 1$, (7.6), and $\gamma_\mathrm{o} < \theta$. Dividing by $h(p)\theta^d = h_\mathrm{t}\theta^d \ge h_{\min}\theta^d$, we obtain

$$\frac{|r_D^\tau(p)|}{h(p)\theta^{d(p)}} \le K_\theta \mathcal{M}_G + \frac{\mathrm{C}_{\mathrm{pin}}}{h_{\min}}.$$

Taking the supremum over τ, D, and p, we find that

$$\mathcal{M}_G \leq \max\Big(\frac{2}{h_{\min}},\, K_\theta \mathcal{M}_G + \frac{C_{\text{pin}}}{h_{\min}}\Big). \tag{7.18}$$

Since $\mathcal{M}_G < \infty$ and $K_\theta < 1$, the bound (7.18) implies that

$$\mathcal{M}_G \leq \frac{1}{h_{\min}} \max\Big(2, \frac{C_{\text{pin}}}{1 - K_\theta}\Big).$$

Using $h(p) \leq h_{\max}$ and the definition (7.17), we obtain (7.8) with

$$A = \frac{h_{\max}}{h_{\min}} \max\Big(2, \frac{C_{\text{pin}}}{1 - K_\theta}\Big).$$

The constant A is independent of G and τ, and $A \geq 2$. Finally, Section 7.1 shows that (7.8) for all finite G implies Hypothesis (P) with the same constants A and θ. This completes the proof. □

SECTION 8

The computer-assisted inputs

In this section, we prove Theorem 8.1, according to which Hypotheses (H1), (H2), and (H3) hold at the thresholds $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$, and in Section 8.4 we complete the proof of Theorem 1.1. Together with Theorem 7.4, Theorem 8.1 gives Hypothesis (P), and the remaining implications of our argument were proved in Sections 3 to 5. For the heuristics behind the three hypotheses, we refer to Section 2.

The three hypotheses are verified by computations of three different kinds. Hypothesis (H1) consists of the polynomial inequalities (6.22), one for each of the 7,962 patterns, and we verify them in exact integer arithmetic. Hypothesis (H2) follows from an inequality for a vector of positive integers indexed by the 916 histories of six steps defined below, which we also verify in exact arithmetic. Hypothesis (H3) is a bound on two series with nonnegative terms, which we verify by a floating-point computation with directed rounding [19]. The proofs of the bounds for (H1) and (H3) are long, and we give them in Appendix B and Appendix C. These appendices rely on the definitions and results of Sections 6 and 7, which precede them. The programs, the certificates, and the recorded outputs are the ancillary files described in Appendix E, and we refer to them by their names `anc/...`.

Theorem 8.1. *Let* $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$. *Then Hypothesis (H1) holds, Hypothesis (H2) holds with* $\gamma_{\mathrm{o}} = 2499/2500$ *and* $C_{\mathrm{o}} = 10$, *and Hypothesis (H3) holds with* $K_\star = 0.92320$ *and the enumerations* ς_w *of Appendix C.2, that is,*

$$K_{\mathfrak{t}}(1) \le 0.92320 \qquad \textit{for } \mathfrak{t} \in \{\mathrm{coll}, \mathrm{diag}\}. \tag{8.1}$$

We prove Theorem 8.1 at the end of Section 8.3. Of the three hypotheses, only (H1) becomes harder when the thresholds increase.

Remark 8.2. The trail weight $\mathrm{wt}(w)$ in (H2) and, for a fixed enumeration ς_w, the mass $\mathsf{m}(\{w\})$ in (H3) are sums of products of the numbers $1/5$, $1/c_{\mathrm{s}}$, $1/c_{\mathrm{t}}$, $\bar{\alpha}$, $\bar{\beta}$, h_{coll}, and h_{diag}, with nonnegative coefficients. They are nonincreasing functions of c_{s} and of c_{t}. For the short trails, the rule of Definition C.1 takes the minimum of the masses over all enumerations, which preserves this monotonicity. The enumeration of a longer trail is independent of the thresholds. Therefore, if (H1) holds at a pair $(c'_{\mathrm{s}}, c'_{\mathrm{t}})$ such that $c'_{\mathrm{s}} \ge 961/1000$ and $c'_{\mathrm{t}} \ge 187/200$, then (H1), (H2), and (H3) hold at $(c'_{\mathrm{s}}, c'_{\mathrm{t}})$, with the

constants of Theorem 8.1 and the enumerations of Definition C.1 evaluated at the new thresholds. Thus Theorem 7.4 applies at this pair.

8.1 The trail growth

We now prove (H2). Recall from Section 6.6 that, for a trail $w = (v_0, \dots, v_n)$ of $\mathbb{Z}^2$ with an arrival site v_{-1} of v_0, the trail weight (6.23) is the product of the step weights wt_j. Here $\mathrm{wt}_j = 1/5$ if step j is a revisit ($v_{j-1} \in \{v_0, \dots, v_{j-2}\}$), and $\mathrm{wt}_j = 1/(3c_{\mathrm{s}})$ or $\mathrm{wt}_j = 1/(3c_{\mathrm{t}})$ if step j is a straight or a turning first visit. Hypothesis (H2) asks for the bound (6.24) on the sum of $\mathrm{wt}(w)$ over the trails of length n with given v_0 and v_{-1}, subject to $v_1 \neq v_{-1}$ when $n \geq 1$.

The weight of a step depends on the whole past of the trail, through the revisits. To bound these sums, we keep only the last m steps of the past. We can then detect a revisit of a site that was visited within the last m steps, and we treat every other step as a first visit. This can only increase the weight of a step, since $1/5 < 1/3 \leq \min(1/(3c_{\mathrm{s}}), 1/(3c_{\mathrm{t}}))$. Finite-memory transfer matrices and automata were used in [4, 42] to bound the connective constant of self-avoiding walks. In [23], the bond-self-avoiding walks of the square lattice were counted in blocks of four steps, which bounds their number by a constant times $79^{n/4}$. However, $79^{1/4} > 2.98 > 3c_{\mathrm{t}}$, so this count together with the largest step weight $1/(3c_{\mathrm{t}})$ bounds the sums in (H2) only by a quantity that grows with n. We will see that a memory of six steps suffices at our thresholds.

Definition 8.3. Let $m \geq 1$. A *history* of memory m is a sequence $\omega = (\mathsf{e}_1, \dots, \mathsf{e}_m)$ of unit vectors of $\mathbb{Z}^2$ such that the path $z_0 = 0$, $z_k = z_{k-1} + \mathsf{e}_k$ for $1 \leq k \leq m$, uses m distinct bonds $\{z_{k-1}, z_k\}$. We let Ω_m denote the set of histories of memory m. A unit vector e is an *admissible step* from ω if $\{z_m, z_m + \mathsf{e}\}$ is not one of the bonds $\{z_{k-1}, z_k\}$, $1 \leq k \leq m$. For an admissible step e from ω, we call

$$\omega\mathsf{e} = (\mathsf{e}_2, \dots, \mathsf{e}_m, \mathsf{e})$$

the *successor* of ω by e, and we define the *kind* of e to be r if $z_m \in \{z_0, \dots, z_{m-1}\}$, to be s if $z_m \notin \{z_0, \dots, z_{m-1}\}$ and $\mathsf{e} = \mathsf{e}_m$, and to be t otherwise.

Note that the successor $\omega\mathsf{e}$ belongs to Ω_m (its path is a translate of $(z_1, \dots, z_m, z_m + \mathsf{e})$). Suppose that ω consists of the last m steps of a trail. Then the kind r means that the step from the endpoint of the trail leaves a site that the trail visited within its last m steps. The kinds s and t mean that it leaves another site, by a straight step or by a turn, respectively. We let A_m denote the matrix indexed by $\Omega_m \times \Omega_m$ whose entries are, for all $\omega \in \Omega_m$ and

all admissible steps e from ω,

$$\mathsf{A}_m(\omega, \omega\mathsf{e}) = \begin{cases} 1/5 & \text{if } \mathsf{e} \text{ has kind r,} \\ 1/(3c_{\mathrm{s}}) & \text{if } \mathsf{e} \text{ has kind s,} \\ 1/(3c_{\mathrm{t}}) & \text{if } \mathsf{e} \text{ has kind t,} \end{cases} \tag{8.2}$$

and whose other entries vanish. Since e is the last entry of $\omega\mathsf{e}$, (8.2) defines each entry at most once.

Proposition 8.4. *Let* $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$ *and* $\gamma_{\mathrm{o}} = 2499/2500$. *There exists a vector* $\mathsf{v} = (\mathsf{v}(\omega))_{\omega \in \Omega_6}$ *of positive integers such that*

$$\mathsf{A}_6\mathsf{v} \le \gamma_{\mathrm{o}}\mathsf{v} \quad \textit{entrywise}, \qquad \frac{\max_\omega \mathsf{v}(\omega)}{\min_\omega \mathsf{v}(\omega)} \, (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-6} \le 10. \tag{8.3}$$

Further, Hypothesis (H2) holds with $\gamma_{\mathrm{o}} = 2499/2500$ *and* $C_{\mathrm{o}} = 10$.

Proof. We first verify (8.3), and we then deduce (H2) from it.

The vector. The vector v is stored in the file `anc/h2/h2_vector.json`. The script `anc/h2/h2_check.py` generates the 916 histories of Ω_6 and their admissible steps from Definition 8.3. It then verifies the two inequalities (8.3) in integer arithmetic, after multiplying the three weights $1/5$, $1/(3c_{\mathrm{s}}) = 1000/2883$, and $1/(3c_{\mathrm{t}}) = 200/561$ by their common denominator $5 \cdot 2883 \cdot 561$.

The integer and rational comparisons establish (8.3). For reference, the rounded diagnostics in `anc/h2/h2_output.log` give

$$\max_{\omega \in \Omega_6} \frac{(\mathsf{A}_6\mathsf{v})(\omega)}{\mathsf{v}(\omega)} \approx 0.999563271649, \qquad \frac{\max_\omega \mathsf{v}(\omega)}{\min_\omega \mathsf{v}(\omega)} \, (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-6} \approx 8.049712.$$

Reduction to histories. We now deduce (H2). Fix $v_0 \in \mathbb{Z}^2$ and a neighbor v_{-1} of v_0. For $n \ge 0$, let Q_n denote the sum on the left side of (6.24), so that $Q_0 = 1$. For $n \ge 6$ and $\omega \in \Omega_6$, let $Q_n(\omega)$ denote the part of this sum over the trails whose last six steps $(v_{n-5} - v_{n-6}, \dots, v_n - v_{n-1})$ form ω, so that

$$Q_n = \sum_{\omega \in \Omega_6} Q_n(\omega).$$

We claim that for all $n \ge 6$ and all $\omega' \in \Omega_6$,

$$Q_{n+1}(\omega') \le \sum_{\omega \in \Omega_6} Q_n(\omega) \, \mathsf{A}_6(\omega, \omega'). \tag{8.4}$$

To prove (8.4), let $w = (v_0, \dots, v_{n+1})$ be a trail counted in $Q_{n+1}(\omega')$. Its prefix $w_n = (v_0, \dots, v_n)$ is counted in $Q_n(\omega)$, where ω is formed by the last six steps of

w_n, and $\omega' = \omega\mathsf{e}$ with $\mathsf{e} = v_{n+1} - v_n$. The step e is admissible from ω, since the path of ω is a translate of $(v_{n-6}, \dots, v_n)$ and w does not use a bond twice. We have $\mathrm{wt}(w) = \mathrm{wt}(w_n)\,\mathrm{wt}_{n+1}$, and we now show that $\mathrm{wt}_{n+1} \le \mathsf{A}_6(\omega, \omega\mathsf{e})$.

We distinguish three cases, and we claim that

$$\mathrm{wt}_{n+1} \begin{cases} = 1/5 = \mathsf{A}_6(\omega,\omega\mathsf{e}) & \text{if } \mathsf{e} \text{ has kind r,} \\ = 1/5 < \mathsf{A}_6(\omega,\omega\mathsf{e}) & \text{if } \mathsf{e} \text{ has kind s or t and } v_n \in \{v_0,\dots,v_{n-7}\}, \\ = \mathsf{A}_6(\omega,\omega\mathsf{e}) & \text{otherwise.} \end{cases} \tag{8.5}$$

In the first case, $v_n \in \{v_{n-6}, \dots, v_{n-1}\}$, so that step $n+1$ is a revisit. In the second case, the same holds, and $1/5 < 1/3 \le \min(1/(3c_{\mathrm{s}}), 1/(3c_{\mathrm{t}}))$. In the remaining case, step $n+1$ is a first visit. Its arrival direction $v_n - v_{n-1}$ coincides with the last entry e_6 of ω, so that the step is straight if and only if e has kind s. This proves (8.5). Distinct trails w give distinct pairs (w_n, e), and summing over w proves (8.4).

Conclusion. Iterating (8.4) and using the first inequality of (8.3), we find for all $j \ge 0$ that

$$\begin{aligned} Q_{6+j} &\le \frac{1}{\min \mathsf{v}} \sum_{\omega} Q_{6+j}(\omega)\mathsf{v}(\omega) \le \frac{1}{\min \mathsf{v}} \sum_{\omega} Q_6(\omega)(\mathsf{A}_6^j \mathsf{v})(\omega) \\ &\le \frac{\gamma_{\mathrm{o}}^j}{\min \mathsf{v}} \sum_{\omega} Q_6(\omega)\mathsf{v}(\omega) \le \frac{\max \mathsf{v}}{\min \mathsf{v}}\, \gamma_{\mathrm{o}}^j Q_6, \end{aligned} \tag{8.6}$$

where we used that all entries of A_6 and all $Q_n(\omega)$ are nonnegative.

Next, a trail has at most three continuations at each step, including the first, where the step to v_{-1} is excluded. Every step weight is at most $1/(3c_{\mathrm{t}})$, since $c_{\mathrm{t}} \le c_{\mathrm{s}} \le 1$. We conclude that

$$Q_r \le 3^r (3c_{\mathrm{t}})^{-r} = c_{\mathrm{t}}^{-r} \qquad \text{for all } r \ge 0. \tag{8.7}$$

Since $c_{\mathrm{t}}\gamma_{\mathrm{o}} < 1$, the function $r \mapsto (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-r}$ is increasing. For $n < 6$, we deduce from (8.7) that

$$Q_n \le (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-n}\gamma_{\mathrm{o}}^n \le (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-6}\gamma_{\mathrm{o}}^n.$$

For $n \ge 6$, (8.6) with $j = n-6$ and (8.7) with $r = 6$ imply that

$$Q_n \le \frac{\max \mathsf{v}}{\min \mathsf{v}}\, \gamma_{\mathrm{o}}^{n-6} c_{\mathrm{t}}^{-6} = \frac{\max \mathsf{v}}{\min \mathsf{v}}\, (c_{\mathrm{t}}\gamma_{\mathrm{o}})^{-6}\gamma_{\mathrm{o}}^n.$$

Using the second inequality of (8.3), and $\max \mathsf{v} \ge \min \mathsf{v}$ for $n < 6$, we obtain $Q_n \le 10\gamma_{\mathrm{o}}^n$ for all $n \ge 0$. This is (H2) with $\gamma_{\mathrm{o}} = 2499/2500$ and $C_{\mathrm{o}} = 10$. □

8.2 The weighted trail sums

Recall that Hypothesis (H3) concerns the two scalar sums $K_{\mathrm{coll}}(1)$ and $K_{\mathrm{diag}}(1)$ of (7.5), which are defined with the type weights $h = (1, 5/4)$ and with enumerations ς_w of the

sets $\mathrm{FV}(w)$ of first-visit steps. We choose the enumerations in Appendix C.2. For trails of length at most 20, ς_w minimizes the contribution of w to (7.4), and for longer trails it lists the straight first visits before the turns. Each of these enumerations is a function of w alone, as Lemma 7.3 requires. We bound the contribution of each type in two parts.

The trails of length at most 20 are enumerated exactly, by depth-first search, and their contributions are summed with upward rounding. A trail of length $n > 20$ is followed exactly for 20 steps. Its continuation is dominated by a weighted walk on $\mathbb{Z}^2$ whose state consists of a phase, which keeps track of the enumeration of the first-visit steps, the history of the last ten steps, and the position. The main difference from the proof of (H2) is that the walk keeps its position. The walk stops when it reaches v. Its steps into c have weight zero, so that it retains the constraint that the trail ends at v. A bound by all trails of a given length, closed or not, would be far too weak here, since the growth rate γ_{o} in (H2) is close to one.

By Lemma C.5, the total weight of the paths of the walk that end at v is bounded by every nonnegative function F of its states such that $\mathsf{g} + \mathsf{TF} \le \mathsf{F}$, where T is the one-step operator of the walk and g its reward at v. Lemma C.6 gives such a function in terms of finitely many numbers, namely its values at the states with position in the box $\mathcal{C}_{70}$ and a function of the phase and the history, which serves at the positions outside the box (there, the coefficients of the walk no longer depend on the position). A program proposes these numbers. It then verifies the inequalities of Lemma C.6 using directed rounding, as described in Appendix C.6. The result is

$$K_{\mathrm{coll}}(1) \le 0.923198, \qquad K_{\mathrm{diag}}(1) \le 0.899348, \tag{8.8}$$

which is Proposition C.7. The certified numbers are collected in Table C.1, and the programs are described in Appendix C.6.

8.3 The local inequalities

Recall the class $\mathcal{P}_{\mathrm{ext}}$ of extended patterns from Definition 6.10, the ring functional $\mathcal{Z}^{\mathbf{a}}_{\pi}$ of (6.21) with its twelve exterior parameters $\mathbf{a} = (\mathbf{u}, \mathbf{m}) \in \mathbb{B}^{12}$, and the polynomials I_{cur} and I_{next} of (6.10). Hypothesis (H1) is the family of inequalities (6.22). By Lemma 6.11, it implies the bound $\rho_j^{\tau} \le 1/c_{\pi}$ for the step ratio at every first visit whose pattern π belongs to $\mathcal{P}_{\mathrm{ext}}$.

Theorem 8.5. *Let $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$. Then for all $\pi \in \mathcal{P}_{\mathrm{ext}}$ and all $\mathbf{a} \in \mathbb{B}^{12}$,*

$$\mathcal{Z}^{\mathbf{a}}_{\pi}[I_{\mathrm{cur}} - c_{\pi} I_{\mathrm{next}}] \ge 0, \tag{8.9}$$

where $c_{\pi} = c_{\mathrm{s}}$ if the pattern π is straight and $c_{\pi} = c_{\mathrm{t}}$ if it is a turn. In particular, Hypothesis (H1) holds with $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$.

The proof of Theorem 8.5 is computer-assisted and is carried out in integer and rational arithmetic; we give it in Appendix B. We now describe its structure. Fix a pattern π, and let $c = c_\pi$. The middle site and the tips of the template were integrated out in the definition (6.21), so that $\Phi_c(\mathbf{a}) = \mathcal{Z}_\pi^{\mathbf{a}}[I_{\mathrm{cur}} - cI_{\mathrm{next}}]$ is a polynomial in the twelve vectors of $\mathbf{a}$ that is affine in each of them. A corner without present ring bonds contributes a positive factor, which we delete. We may also integrate out a corner with exactly one present ring bond, which is then represented by a single vector of norm at most one; see Lemma B.3. This leads to a reduced polynomial $\hat{\Phi}_c$ of at most eleven vectors, and Corollary B.5 shows that it suffices to prove that $\hat{\Phi}_c \geq 0$ when all its vectors are unit vectors.

The main step is a representation of $\hat{\Phi}_c$ by an operator. By Lemma B.6, $\hat{\Phi}_c(\mathbf{b})$ is a diagonal matrix element of an explicit Hermitian operator, taken at a tensor product of spin-1/2 coherent states, after the vectors with indices in an inversion set $\mathcal{I}$ have been replaced by their negatives. It suffices to show that this operator is positive semidefinite. The inversion is needed, since a polynomial that is nonnegative on spheres can have an operator with a negative eigenvalue. The operator commutes with the action of $\mathrm{SU}(2)$ by Lemma B.8. It therefore suffices to test it on one eigenspace of the total spin component S^3_{tot}, for each choice of exchange symmetry at the corners of status 1. See Lemma B.10. After multiplication by an integer, each test asks whether a symmetric integer matrix M_ϵ of size at most 245 is positive definite, which we decide by fraction-free Gaussian elimination [6]. See Proposition B.11. A numerical minimization guides the choice of inversion sets, and Appendix B.4 gives the bounds needed for exact coefficient arithmetic.

The certificates are stored in `anc/h1/certificates.jsonl`, one for each of the 7,962 patterns of `anc/h1/patterns.json`. The script `anc/h1/check_certificate_file.py` checks that every pattern has exactly one certificate and that the certificate of π carries the threshold c_π. The checker `anc/h1/verify_all.py` verifies every certificate from the stored data alone. Its output is recorded in `anc/h1/verify_all_output.log`. The programs, the further checks, and the code that must be trusted are described in Appendix B.

Proof of Theorem 8.1. Theorem 8.5 gives (H1). Proposition 8.4 gives (H2) with $\gamma_{\mathrm{o}} = 2499/2500$ and $C_{\mathrm{o}} = 10$. We now apply Proposition C.7, with the enumerations ς_w of Appendix C.2, and we obtain the bounds (8.8). These bounds are at most $0.92320 < 1$, which proves (8.1). The terms of the series (7.5) are nonnegative, so that its partial sums at $t = 1$ are bounded by its sum, and (H3) holds with $K_\star = 0.92320$. □

8.4 Proof of the main theorem

We now prove Theorem 1.1 by combining Theorem 8.1 with the implications proved in Sections 3 to 7. The explicit values of the constants are verified in Appendix D.

Proof of Theorem 1.1. We pass from the computer-assisted inputs to Hypothesis (P), then to Hypothesis (W), to the block dynamics, and finally to the quantum gap.

Hypothesis (P). Theorem 8.1 shows that Hypotheses (H1), (H2), and (H3) hold at the pair $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$, for which $\min(c_{\mathrm{s}}, c_{\mathrm{t}}) = 187/200$. Then Theorem 7.4 implies that Hypothesis (P) holds with some constants $A \geq 2$ and $\theta \in (0,1)$.

Hypothesis (W). Next, Theorem 5.9 shows that Hypothesis (P) implies Hypothesis (W), with constants p_0 and $M_\star$ that depend only on A and θ.

The gap of $\mathcal{L}_{\mathrm{bl}}$. Applying Proposition 4.5, we find that Hypothesis (W) implies the following. For every integer $M \geq \max(M_\star, M_{\mathrm{g}})$, with $b = 2M+1$, and all $L, N \geq b+4$, the bound (4.10) holds, with Δ_M as in (4.6), and its right side is at least $1/2$ for all $M \geq M_{\mathrm{bl}}$. We fix $M \geq \max(M_\star, M_{\mathrm{g}})$ such that $4\Delta_M \leq b/2$ (for instance, $M \geq M_{\mathrm{bl}}$), so that for all $L, N \geq b+4$,

$$\operatorname{gap}(\mathcal{L}_{\mathrm{bl}}) \geq \frac{1}{2}. \tag{8.10}$$

We set $L_0 = b+4$.

The quantum gap. Let $L, N \geq L_0$. Using Theorem 3.5 and (8.10), we obtain

$$\operatorname{gap}(H_{L,N}) \geq \frac{b}{b+1}\, g_b \operatorname{gap}(\mathcal{L}_{\mathrm{bl}}) \geq \frac{g_b}{4}, \tag{8.11}$$

where we used that $b/(b+1) \geq 1/2$. Recall that g_b is the smallest positive eigenvalue of the patch operator h_b, which is the same operator for all $L, N \geq b+4$. Lemma D.3 gives an explicit positive lower bound for g_b.

The kernel. Since $L, N \geq 3$, the kernel of $H_{L,N}$ is spanned by Ψ_{VBS}, by Corollary 3.3.

The constants. In Appendix D, we make the constants of the previous steps explicit. Lemma D.1 shows that (P) holds with $A = 10^{13}$ and $\theta = 1 - 10^{-20}$, and Lemma D.2 shows that (W) then holds with $M_\star = 10^{180}$ and $p_0 = \exp(-2 \cdot 10^{36})$. We also show there that $M = \lceil \exp(10^{40}) \rceil$ satisfies $M \geq \max(M_\star, M_{\mathrm{g}})$ and $4\Delta_M \leq b/2$. We may therefore take this value of M in the third step, which gives $L_0 = 2\lceil \exp(10^{40}) \rceil + 5$. For this choice, Lemma D.3 and (8.11) imply that $\operatorname{gap}(H_{L,N}) \geq c$ with $c = \exp[-\exp(\exp(10^{41}))]$; see (D.18). This completes the proof. □

APPENDIX A

Connectivity of the cut

In this appendix, we prove the two topological facts used in Section 5.5. The first, Lemma A.1, states that the outer ℓ^∞ boundary of a finite $*$-connected set with $*$-connected complement is itself $*$-connected. The second, Lemma A.2, states that adding to a $*$-connected set the nearest neighbors of its sites gives a connected set; it also compares sums over the two sets. Recall from Section 5.5 the definitions of $*$-neighbors, $*$-paths, $*$-connected and connected sets, and of the outer ℓ^∞ boundary

$$\partial_\infty A = \{z \in \mathbb{Z}^2 \setminus A : \mathrm{dist}_\infty(z, A) = 1\}.$$

We also recall the depth dep_G from Section 5.4. Note that a connected set is also $*$-connected.

For connected subsets of $\mathbb{Z}^d$, results on the connectivity of their exterior boundaries go back to [24] and [9]. The graph-theoretic criterion of [44, Lemma 2 in the arXiv version] applies to the graph on $\mathbb{Z}^2$ whose edges join $*$-neighbors. Since the complement of A is $*$-connected, every site of $\partial_\infty A$ is reachable from a fixed site outside A by a $*$-path that avoids A. The criterion gives Lemma A.1 once we show that the cycle space of the graph is generated by its triangles. We give a self-contained proof.

A.1 The outer boundary of a set

In this subsection, we call an unordered pair of $*$-neighbors an *edge*. The finite sets of edges form a vector space over $\mathbb{F}_2$, with the symmetric difference as addition. We call a finite set of edges *even* if every site of $\mathbb{Z}^2$ belongs to an even number of edges of the set. We call a set of three sites that are pairwise $*$-neighbors a *triangle*, and we call the set of its three edges its *boundary*. Every triangle lies in a square $(a, b) + \{0, 1\}^2$. The boundary of a triangle is even, and so is every sum of boundaries of triangles.

Lemma A.1. *Let $A \subset \mathbb{Z}^2$ be finite, nonempty, and $*$-connected, and suppose that $\mathbb{Z}^2 \setminus A$ is $*$-connected. Then $\partial_\infty A$ is nonempty and $*$-connected.*

Proof. The proof has three steps. In the first, we show that the boundaries of triangles generate all even sets of edges; this is the only place where we use the geometry of $\mathbb{Z}^2$. The second and third steps deduce the lemma from this fact by a parity argument.

Step 1. We show that every even set of edges is a sum of boundaries of triangles. Let $\mathcal{F}$ be an even set of edges. Every diagonal edge can be written uniquely as $\{p, p+(1,\epsilon)\}$ with $p \in \mathbb{Z}^2$ and $\epsilon \in \{1,-1\}$. For each such edge of $\mathcal{F}$, we add to $\mathcal{F}$ the boundary of the triangle $\{p, p+(1,0), p+(1,\epsilon)\}$, whose only diagonal edge is $\{p, p+(1,\epsilon)\}$. The resulting set $\mathcal{F}'$ is even and consists of bonds.

For $(i,j) \in \mathbb{Z}^2$, let $\alpha_{i,j} \in \mathbb{F}_2$ be equal to 1 if and only if the bond $\{(i,j),(i+1,j)\}$ belongs to $\mathcal{F}'$, and let $\beta_{i,j} \in \mathbb{F}_2$ be equal to 1 if and only if the bond $\{(i,j),(i,j+1)\}$ belongs to $\mathcal{F}'$. We set

$$\gamma_{i,j} = \sum_{r>j} \alpha_{i,r} \in \mathbb{F}_2.$$

We claim that

$$\sum_{r\in\mathbb{Z}} \alpha_{i,r} = 0 \qquad \text{for all } i \in \mathbb{Z}. \tag{A.1}$$

Indeed, the left side of (A.1) is the parity of the number of bonds of $\mathcal{F}'$ that join the half-plane $\{x_1 \le i\}$ to its complement. This number has the same parity as the sum, over the sites of the half-plane, of the number of bonds of $\mathcal{F}'$ that contain the site, and this sum is even because $\mathcal{F}'$ is even. Since $\mathcal{F}'$ is finite, (A.1) implies that $\gamma_{i,j} = 0$ for all but finitely many (i,j).

For $(i,j) \in \mathbb{Z}^2$, let $\partial\square_{i,j}$ denote the set of the four bonds of the square $(i,j)+\{0,1\}^2$. We now show that

$$\mathcal{F}' = \sum_{(i,j)\in\mathbb{Z}^2} \gamma_{i,j}\,\partial\square_{i,j}. \tag{A.2}$$

The bond $\{(i,j),(i+1,j)\}$ belongs to $\partial\square_{i,j}$ and to $\partial\square_{i,j-1}$, and to no other set $\partial\square_{i',j'}$. Its coefficient on the right side of (A.2) is

$$\gamma_{i,j-1} + \gamma_{i,j} = \alpha_{i,j}.$$

The bond $\{(i,j),(i,j+1)\}$ belongs to $\partial\square_{i,j}$ and to $\partial\square_{i-1,j}$. The site (i,r) belongs to an even number of bonds of $\mathcal{F}'$, which means that $\alpha_{i-1,r} + \alpha_{i,r} = \beta_{i,r-1} + \beta_{i,r}$ for all r. Only finitely many $\beta_{i,r}$ are nonzero, and the coefficient of the bond is therefore

$$\gamma_{i-1,j} + \gamma_{i,j} = \sum_{r>j} (\alpha_{i-1,r} + \alpha_{i,r}) = \sum_{r>j} (\beta_{i,r-1} + \beta_{i,r}) = \beta_{i,j},$$

where the last sum telescopes. This proves (A.2).

Finally, each set $\partial\square_{i,j}$ is the sum of the boundaries of the two triangles $\{(i,j),(i+1,j),(i+1,j+1)\}$ and $\{(i,j),(i,j+1),(i+1,j+1)\}$, since their common edge cancels. Together with (A.2), this shows that $\mathcal{F}'$ is a sum of boundaries of triangles. The same holds for $\mathcal{F}$, which differs from $\mathcal{F}'$ by a sum of boundaries of triangles.

Step 2. Let $\partial_{\mathrm{e}}A$ denote the set of edges with one site in A and the other in $\mathbb{Z}^2 \setminus A$. This set is finite, since A is finite. It is nonempty, since a site of A with the largest first

coordinate has a $*$-neighbor outside A. We call two edges of $\partial_e A$ *linked* if they belong to the boundary of one triangle. We will show that any two edges of $\partial_e A$ are joined by a chain of edges of $\partial_e A$ in which consecutive edges are linked.

Suppose that this fails. Then $\partial_e A$ is the disjoint union of two nonempty sets $\mathcal{E}_1$ and $\mathcal{E}_2$ such that no edge of $\mathcal{E}_1$ is linked to an edge of $\mathcal{E}_2$. Given a triangle, let m denote the number of its sites in A. Exactly $m(3-m) \in \{0,2\}$ edges of its boundary belong to $\partial_e A$, and when there are two, they are linked and lie in the same set $\mathcal{E}_q$. Hence the boundary of every triangle contains an even number of edges of $\mathcal{E}_1$. The map $\mathcal{F} \mapsto |\mathcal{F} \cap \mathcal{E}_1| \bmod 2$ is linear on finite sets of edges and vanishes on the boundaries of triangles. By Step 1, it vanishes on every even set of edges.

We now construct an even set of edges on which this map does not vanish. For $q = 1, 2$, we choose an edge $\{a_q, b_q\} \in \mathcal{E}_q$ with $a_q \in A$ and $b_q \notin A$. By assumption, there exist a $*$-path from a_1 to a_2 in A and a $*$-path from b_2 to b_1 in $\mathbb{Z}^2 \setminus A$. Together with the edges $\{a_2, b_2\}$ and $\{b_1, a_1\}$, these paths form a closed walk. Let $\mathcal{F}$ be the set of edges that this walk traverses an odd number of times. The set $\mathcal{F}$ is even, because the walk is closed.

Every edge of the first path has its two sites in A, and every edge of the second path has its two sites outside A. We conclude that

$$\mathcal{F} \cap \partial_e A = \big\{\{a_1, b_1\}, \{a_2, b_2\}\big\}, \qquad |\mathcal{F} \cap \mathcal{E}_1| = 1,$$

which contradicts the fact that this map vanishes on every even set of edges.

Step 3. A site $z \notin A$ belongs to $\partial_\infty A$ if and only if it has a $*$-neighbor in A, which means that z is the site outside A of some edge of $\partial_e A$. In particular, $\partial_\infty A$ is nonempty. The sites outside A of two linked edges lie in one triangle, and they are equal or $*$-neighbors. Consequently, a chain of linked edges as in Step 2 gives a $*$-path in $\partial_\infty A$ between the sites outside A of its first and last edges. By Step 2, $\partial_\infty A$ is $*$-connected. □

A.2 Thickening

Lemma A.2. *Let $X_0 \subset \mathbb{Z}^2$ be finite, and let $X = \{z \in \mathbb{Z}^2 : \operatorname{dist}(z, X_0) \le 1\}$.*

1. *If X_0 is nonempty and $*$-connected, then X is connected.*
2. *Let $G \subset \mathbb{Z}^2$ be finite, and let $t \in (0,1)$. Then*

$$\sum_{z \in X} t^{\deg_G(z)} \le 5t^{-1} \sum_{x \in X_0} t^{\deg_G(x)}.$$

Proof. For $x \in \mathbb{Z}^2$, we let

$$C(x) = \{z \in \mathbb{Z}^2 : \operatorname{dist}(z, x) \le 1\}$$

denote the set consisting of x and its four neighbors. Each set $C(x)$ is connected and has five sites, and X is the union of the sets $C(x)$ with $x \in X_0$.

We begin with the first part. Let $x, x' \in X_0$ be $*$-neighbors. When $x \sim x'$, we have $x' \in C(x) \cap C(x')$. Otherwise, there exist $\epsilon_1, \epsilon_2 \in \{1, -1\}$ such that

$$x' - x = (\epsilon_1, \epsilon_2), \qquad x + (\epsilon_1, 0) \in C(x) \cap C(x').$$

In the two cases, the union $C(x) \cup C(x')$ is connected. Any two sites of X_0 are joined by a $*$-path in X_0, and the corresponding sets $C(x)$ form a chain in which consecutive sets intersect. It follows that X is connected.

We now prove the second part. Let $z \in X$, and choose $x \in X_0$ with $z \in C(x)$. Using the fact that dep_G is Lipschitz with constant 1 for the graph distance, together with $t < 1$, we obtain

$$t^{\mathrm{dep}_G(z)} \leq t^{\mathrm{dep}_G(x)-1}.$$

Summing this bound over $z \in C(x)$ and over $x \in X_0$, and using $|C(x)| = 5$, we arrive at the claim. □

APPENDIX B

Certificates for the local inequalities

In this appendix, we prove Theorem 8.5. Recall that it asserts that

$$\mathcal{Z}_\pi^{\mathbf{a}}[I_{\mathrm{cur}} - c_\pi I_{\mathrm{next}}] \geq 0 \qquad \text{for all } \pi \in \mathcal{P}_{\mathrm{ext}} \text{ and } \mathbf{a} \in \mathbb{B}^{12}, \tag{B.1}$$

where $c_\pi = 961/1000$ if π is straight and $c_\pi = 187/200$ if π is a turn. This is Hypothesis (H1) at the thresholds $(c_{\mathrm{s}}, c_{\mathrm{t}}) = (961/1000, 187/200)$, one of the three inputs of Theorem 8.1, and we refer to Section 2 for the role of these inequalities in the argument.

The proof is computer-assisted and uses exact tests of integer matrices. We begin in Appendix B.1 by integrating out the corners with at most one present ring bond. This reduces (B.1) to the nonnegativity of a polynomial in at most eleven unit vectors that is affine in each of them. The main idea, in Appendix B.2, is to write this polynomial as the expectation of an explicit Hermitian operator in a product of spin-1/2 coherent states, after the signs of some of the vectors have been reversed. The symmetries of the operator then reduce its positivity to a few eigenspaces of the total spin component, and in Appendix B.3 we turn each of these tests into a test of positive definiteness of a symmetric integer matrix. In Appendix B.4, we describe the programs and the certificate file and complete the proof of Theorem 8.5. The checks in Appendix B.5 corroborate the computation but are not part of the proof, and Appendix B.6 lists what a reader must trust.

Throughout this appendix, we fix a pattern $\pi \in \mathcal{P}_{\mathrm{ext}}$ and a rational number $c = c_1/c_2$, where c_1 and c_2 are positive integers; for Theorem 8.5, we take $c = c_\pi$. We use the notation of Section 6 for the template $\Diamond(0)$, its middle site 0, its ring R, which consists of the axial sites E, N, W, and S and the corners $(\pm 1, \pm 1)$, its tips, and its slots. Let $o = o_\pi$ be the out-neighbor of π, and let f and g be the two sites of $\{\mathsf{E}, \mathsf{N}, \mathsf{W}\} \setminus \{o\}$. For a non-removed ring site y, we abbreviate $\lambda_y = \lambda_{\mathrm{st}_\pi(y)}$. Superscripts on a vector $b \in \mathbb{R}^3$ denote its coordinates, $b = (b^1, b^2, b^3)$, and we set $b^0 = 1$. For $Q \in \mathrm{O}(3)$ and $\mathbf{b} = (b_1, \dots, b_k) \in (\mathbb{R}^3)^k$, we set $Q\mathbf{b} = (Qb_1, \dots, Qb_k)$.

B.1 The ring polynomial and leaf corners

Recall from Section 6 that every ring bond joins an axial site to a corner; in particular, the corners are pairwise non-adjacent. Recall also that the integrand of the ring functional (6.21) is the product of h, of the factors $1 + n_y \cdot n_z$ of the present ring bonds $\{y, z\}$, of the

factors $1 + (3/5)u_z \cdot n_z$ of the axial sites z whose tip bond is retained, and of the two slot factors $1 + m_k \cdot n_{z(k)}$ of every non-removed corner $z(k)$, and that it is integrated against the measures $\lambda_y(\mathrm{d}n_y)$ of the non-removed ring sites y. Note that in I_{cur} and I_{next}, the spin n_f is replaced by 0 if f is removed, and likewise for g. We define the *ring polynomial*

$$\Phi_c(\mathbf{a}) = \mathcal{Z}_\pi^{\mathbf{a}}[I_{\mathrm{cur}} - cI_{\mathrm{next}}], \qquad \mathbf{a} \in (\mathbb{R}^3)^{12}, \tag{B.2}$$

where for $\mathbf{a} \notin \mathbb{B}^{12}$ the right side is defined by the formula (6.21). In this notation, Theorem 8.5 asserts that $\Phi_{c_\pi} \geq 0$ on $\mathbb{B}^{12}$ for all $\pi \in \mathcal{P}_{\mathrm{ext}}$.

Lemma B.1. *Let $c \in \mathbb{R}$. Then Φ_c is a polynomial in the coordinates of $\mathbf{a}$ that is affine in each of the twelve vectors of $\mathbf{a}$, and $\Phi_c(Q\mathbf{a}) = \Phi_c(\mathbf{a})$ for all $Q \in \mathrm{O}(3)$ and $\mathbf{a} \in (\mathbb{R}^3)^{12}$.*

Proof. Each vector of $\mathbf{a}$ occurs in at most one factor of the integrand in (6.21), which is affine in it, and I_{cur} and I_{next} do not involve $\mathbf{a}$. Expanding the integrand, we see that $\Phi_c(\mathbf{a})$ is a finite sum of moments of the product measure $\prod_y \lambda_y$, each multiplied by a product of coordinates of the vectors of $\mathbf{a}$ that contains at most one coordinate of each vector. The first assertion follows.

To prove the invariance, let $Q \in \mathrm{O}(3)$, let $\mathbf{n}$ denote the family of spins of the non-removed ring sites, and let $\mathcal{G}(\mathbf{n}, \mathbf{a})$ denote the integrand of (6.21) with $h = I_{\mathrm{cur}} - cI_{\mathrm{next}}$. Every factor of $\mathcal{G}$ is a function of scalar products between spins and vectors of $\mathbf{a}$, so that $\mathcal{G}(Q\mathbf{n}, Q\mathbf{a}) = \mathcal{G}(\mathbf{n}, \mathbf{a})$. Using the invariance of λ_0 and λ_1 under Q and changing variables $n_y \mapsto Qn_y$ for every non-removed ring site y, we obtain

$$\Phi_c(Q\mathbf{a}) = \int \mathcal{G}(\mathbf{n}, Q\mathbf{a}) \prod_y \lambda_y(\mathrm{d}n_y) = \int \mathcal{G}(Q\mathbf{n}, Q\mathbf{a}) \prod_y \lambda_y(\mathrm{d}n_y) = \Phi_c(\mathbf{a}). \qquad \square$$

We will use the following elementary fact several times.

Lemma B.2. *Let $k \geq 1$, and let f be a real polynomial on $(\mathbb{R}^3)^k$ that is affine in each of its k arguments. If $f \geq 0$ on $(\mathbb{S}^2)^k$, then $f \geq 0$ on $\mathbb{B}^k$.*

Proof. We show by induction on $l \in \{0, \dots, k\}$ that $f \geq 0$ on $\mathbb{B}^l \times (\mathbb{S}^2)^{k-l}$. The case $l = 0$ is the assumption.

Suppose that the claim holds for $l - 1$, and let $\mathbf{b} \in \mathbb{B}^l \times (\mathbb{S}^2)^{k-l}$. Choose a unit vector e such that $b_l = |b_l| e$, and let $\mathbf{b}_+$ and $\mathbf{b}_-$ be obtained from $\mathbf{b}$ by replacing b_l with e and $-e$, respectively. Since f is affine in its l-th argument and

$$b_l = \frac{1 + |b_l|}{2}\, e + \frac{1 - |b_l|}{2}\, (-e),$$

we have

$$f(\mathbf{b}) = \frac{1 + |b_l|}{2} f(\mathbf{b}_+) + \frac{1 - |b_l|}{2} f(\mathbf{b}_-) \geq 0,$$

where the inequality follows from the induction hypothesis, applied to $\mathbf{b}_+$ and $\mathbf{b}_-$. $\square$

We now integrate out the spin of a corner that has at most one present ring bond.

Lemma B.3. *Let $s \in \{0,1\}$ and $m, m' \in \mathbb{S}^2$, and set $r = 1/2$ if $s = 0$ and $r = 1/3$ if $s = 1$. Then*

$$\int (1 + m \cdot n)(1 + m' \cdot n)\, \lambda_s(\mathrm{d}n) > 0, \tag{B.3}$$

and there exist $A > 0$ and $b \in \mathbb{B}$, depending only on s, m, and m', such that

$$\int (1 + n' \cdot n)(1 + m \cdot n)(1 + m' \cdot n)\, \lambda_s(\mathrm{d}n) = A(1 + r\, b \cdot n') \qquad \text{for all } n' \in \mathbb{R}^3. \tag{B.4}$$

Proof. Set $t = m \cdot m'$. Recall that the moments of $\mathrm{d}n$ of odd degree vanish and that

$$\int (a \cdot n)(a' \cdot n)\, \mathrm{d}n = \frac{a \cdot a'}{3}, \qquad a, a' \in \mathbb{R}^3.$$

Expanding the products and using these facts, we find for all $n' \in \mathbb{R}^3$ that

$$\int (1 + m \cdot n)(1 + m' \cdot n)\, \mathrm{d}n = A_0, \qquad \int (1 + n' \cdot n)(1 + m \cdot n)(1 + m' \cdot n)\, \mathrm{d}n = A_0 + V \cdot n', \tag{B.5}$$

where

$$A_0 = 1 + \frac{t}{3}, \qquad V = \frac{m + m'}{3}.$$

Since $\lambda_1 = (3/5)\, \mathrm{d}n + (2/5)\delta_0$, the same two integrals with λ_1 in place of $\mathrm{d}n$ are

$$A_1 \quad \text{and} \quad A_1 + \frac{3}{5} V \cdot n', \qquad \text{where } A_1 = \frac{3A_0 + 2}{5}. \tag{B.6}$$

The bound $t \geq -1$ implies that $A_0 \geq 2/3$ and $A_1 \geq 4/5$, and (B.3) follows.

We now prove (B.4). Using $|m + m'|^2 = 2 + 2t$, which holds since m and m' are unit vectors, we find

$$A_0^2 - 4|V|^2 = \Big(1 + \frac{t}{3}\Big)^2 - \frac{8 + 8t}{9} = \frac{(1 - t)^2}{9} \geq 0,$$

so that $|V| \leq A_0/2$. For $s = 0$, we set

$$A = A_0, \qquad b = \frac{2V}{A_0}.$$

Then $|b| \leq 1$, and (B.5) yields (B.4) with $r = 1/2$.

For $s = 1$, we set

$$A = A_1, \qquad b = \frac{9V}{3A_0 + 2}, \qquad \text{so that} \qquad Arb = \frac{3A_0 + 2}{15} \cdot \frac{9V}{3A_0 + 2} = \frac{3}{5} V \quad \text{for } r = \frac{1}{3}.$$

Then (B.6) gives (B.4). Further, using $|V| \leq A_0/2$, we have

$$|b| \leq \frac{9A_0}{2(3A_0 + 2)} \leq 1,$$

where the last inequality is equivalent to $A_0 \leq 4/3$, which holds since $t \leq 1$. □

We call a non-removed corner an *isolated corner* if it has no present ring bond, and a *leaf corner* if it has exactly one present ring bond. Recall that every non-removed corner carries two slot factors in (6.21). Using Lemma B.3, we can integrate out the spins of the isolated corners and of the leaf corners, and this leads to the following two representations of Φ_c.

Definition B.4. The *unreduced representation* of π consists of the following *groups* of vectors in $\mathbb{R}^3$, each vector with a radius and a ring site that carries it.

1. For each non-removed axial site z whose tip bond is retained, one vector, carried by z, with radius $3/5$.
2. For each non-removed corner z that is not isolated, two vectors, carried by z, with radius 1.

The *reduced representation* of π is obtained from the unreduced one by replacing, for each leaf corner z, the two vectors of z by one vector, carried by the axial site that is joined to z by its present ring bond, with radius $1/2$ if z has status 0 and radius $1/3$ if z has status 1. In both representations, we call a group of one vector a *single vector* and a group of two vectors a *pair*, and we say that the corner z of item (2) carries its pair. We number the vectors $1, \dots, k$, and we let r_i denote the radius of the vector i and z_i the site that carries it. The *retained sites* are the non-removed ring sites other than the isolated corners and, in the reduced representation, the leaf corners.

For $\mathbf{b} = (b_1, \dots, b_k) \in (\mathbb{R}^3)^k$ and a function h of the spins of the retained sites, we define

$$\hat{Z}[h](\mathbf{b}) = \int h \prod_{\{y,z\}} (1 + n_y \cdot n_z) \prod_{i=1}^{k} (1 + r_i\, b_i \cdot n_{z_i}) \prod_{y} \lambda_y(\mathrm{d}n_y), \tag{B.7}$$

where the first product runs over the present ring bonds whose ends are retained and the last over the retained sites. Every non-removed axial site is retained, so that I_{cur} and I_{next} are functions of the spins of the retained sites. We set $\hat{Z}_{\mathrm{cur}} = \hat{Z}[I_{\mathrm{cur}}]$ and $\hat{Z}_{\mathrm{next}} = \hat{Z}[I_{\mathrm{next}}]$, and we define

$$\hat{\Phi}_c = \hat{Z}_{\mathrm{cur}} - c\hat{Z}_{\mathrm{next}}. \tag{B.8}$$

These polynomials on $(\mathbb{R}^3)^k$ depend on the representation, which will always be clear from the context. The proof of Lemma B.1 applies verbatim to (B.7). It shows that $\hat{Z}_{\mathrm{cur}}$ and $\hat{Z}_{\mathrm{next}}$ are affine in each of the k vectors and that $\hat{Z}[h](Q\mathbf{b}) = \hat{Z}[h](\mathbf{b})$ for $h \in \{I_{\mathrm{cur}}, I_{\mathrm{next}}\}$, $Q \in \mathrm{O}(3)$, and $\mathbf{b} \in (\mathbb{R}^3)^k$.

Corollary B.5. *Fix one of the two representations of π. If $\hat{\Phi}_c(\mathbf{b}) \geq 0$ for all $\mathbf{b} \in (\mathbb{S}^2)^k$, then $\Phi_c(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in \mathbb{B}^{12}$.*

Proof. By Lemma B.1, the ring polynomial Φ_c is affine in each vector of $\mathbf{a}$. By Lemma B.2, applied with $k = 12$ and $f = \Phi_c$, it suffices to show that $\Phi_c(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in (\mathbb{S}^2)^{12}$. Fix such an $\mathbf{a}$.

We begin by integrating out the spins of the corners that are not retained. The spin of a corner z occurs in (6.21) only in the two slot factors of z and in the factors of the present ring bonds at z, since I_{cur} and I_{next} involve only spins of axial sites. The corners are pairwise non-adjacent, so that we may integrate out their spins one corner at a time.

Let z be a corner that is not retained, and let $m, m' \in \mathbb{S}^2$ be its slot vectors. By (B.3), an isolated corner z contributes the factor

$$A_z = \int (1 + m \cdot n)(1 + m' \cdot n)\, \lambda_z(\mathrm{d}n) > 0.$$

In the reduced representation, let z be a leaf corner, with present ring bond $\{z, y\}$. Applying (B.4) with $n' = n_y$, we have

$$\int (1 + n_y \cdot n)(1 + m \cdot n)(1 + m' \cdot n)\, \lambda_z(\mathrm{d}n) = A_z(1 + r\, b_z \cdot n_y)$$

for some $A_z > 0$ and $b_z \in \mathbb{B}$, where r is the radius of the vector that replaces z in the reduced representation.

Now let $\mathbf{b} \in \mathbb{B}^k$ be the family with the entry u_z for the vector of item (1) of Definition B.4 at an axial site z, the entry b_z for the vector that replaces a leaf corner z in the reduced representation, and the corresponding slot vectors for the vectors of a pair. Comparing with (B.7), we find

$$\mathcal{Z}^{\mathbf{a}}_{\pi}[h] = A\, \hat{Z}[h](\mathbf{b}), \qquad h \in \{I_{\mathrm{cur}}, I_{\mathrm{next}}\}, \qquad \text{where } A = \prod_z A_z > 0,$$

and the product runs over the corners that are not retained ($A = 1$ if there are none). The factor A is the same for $h = I_{\mathrm{cur}}$ and $h = I_{\mathrm{next}}$, so that

$$\Phi_c(\mathbf{a}) = A\, \hat{\Phi}_c(\mathbf{b}).$$

By assumption, $\hat{\Phi}_c \geq 0$ on $(\mathbb{S}^2)^k$, and Lemma B.2, applied with $f = \hat{\Phi}_c$, extends this to $\mathbb{B}^k$. (For $k = 0$, the polynomial $\hat{\Phi}_c$ is a constant.) The claim follows. □

B.2 The inversion certificate

Recall from Section 3 the Pauli matrices $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ and the spin-1/2 coherent states $\zeta(a) \in \mathbb{C}^2$, $a \in \mathbb{S}^2$, which satisfy (3.1). We let $\sigma_0 = \mathrm{Id}$ be the 2×2 identity matrix, and for $a \in \mathbb{R}^3$ we set $a \cdot \sigma = \sum_{j=1}^{3} a^j \sigma_j$. In the standard basis $e_{\uparrow} = (1, 0)$, $e_{\downarrow} = (0, 1)$ of $\mathbb{C}^2$ (the eigenvectors of σ_3 for the eigenvalues 1 and -1), we have

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The inner product $\langle \varphi, \psi \rangle = \varphi^* \psi$ is linear in the second argument. On $(\mathbb{C}^2)^{\otimes k}$, we use the *standard basis*, which consists of the tensor products of $e_\uparrow$ and $e_\downarrow$. For an operator A on $\mathbb{C}^2$ and $1 \le i \le k$, we let $A^{(i)}$ denote the operator on $(\mathbb{C}^2)^{\otimes k}$ that acts as A on the i-th factor and as the identity on the others. For $\mathbf{b} \in (\mathbb{S}^2)^k$, we set $\zeta(\mathbf{b}) = \zeta(b_1) \otimes \cdots \otimes \zeta(b_k)$.

Let $k \ge 0$, and let f be a real polynomial on $(\mathbb{R}^3)^k$ that is affine in each argument. Expanding f in one argument at a time, we see that there exist real numbers $f_{\mathbf{j}}$, indexed by $\mathbf{j} = (j_1, \dots, j_k) \in \{0,1,2,3\}^k$, such that

$$f(\mathbf{b}) = \sum_{\mathbf{j}} f_{\mathbf{j}} \prod_{i=1}^{k} b_i^{j_i}, \qquad \mathbf{b} \in (\mathbb{R}^3)^k. \tag{B.9}$$

The numbers $f_{\mathbf{j}}$ are unique, since the products on the right side are distinct monomials in the coordinates of $\mathbf{b}$. We call them the *coefficients* of f, and we define the Hermitian operator

$$\mathcal{O}(f) = \sum_{\mathbf{j}} f_{\mathbf{j}}\, \sigma_{\mathbf{j}} \quad \text{on } (\mathbb{C}^2)^{\otimes k}, \qquad \text{where } \sigma_{\mathbf{j}} = \sigma_{j_1} \otimes \cdots \otimes \sigma_{j_k}. \tag{B.10}$$

For $k = 0$, the polynomial f is a constant, $(\mathbb{C}^2)^{\otimes 0} = \mathbb{C}$, and $\mathcal{O}(f) = f$. For a set $\mathcal{I} \subset \{1, \dots, k\}$, which we call an *inversion set*, let $\mathbf{b}^{\mathcal{I}}$ be obtained from $\mathbf{b}$ by replacing b_i with $-b_i$ for all $i \in \mathcal{I}$, and set $f^{\mathcal{I}}(\mathbf{b}) = f(\mathbf{b}^{\mathcal{I}})$. Observe that, by (B.9), the coefficients of $f^{\mathcal{I}}$ are

$$f^{\mathcal{I}}_{\mathbf{j}} = (-1)^{|\{i \in \mathcal{I} : j_i \neq 0\}|} f_{\mathbf{j}}. \tag{B.11}$$

Lemma B.6. *Let $k \ge 0$, let f be a real polynomial on $(\mathbb{R}^3)^k$ that is affine in each argument, and let $\mathcal{I} \subset \{1, \dots, k\}$. Then*

$$f(\mathbf{b}) = \big\langle \zeta(\mathbf{b}^{\mathcal{I}}), \mathcal{O}(f^{\mathcal{I}})\, \zeta(\mathbf{b}^{\mathcal{I}}) \big\rangle \qquad \textit{for all } \mathbf{b} \in (\mathbb{S}^2)^k. \tag{B.12}$$

In particular, if $\mathcal{O}(f^{\mathcal{I}})$ is positive semidefinite, then $f \ge 0$ on $\mathbb{B}^k$.

Proof. Let $a \in \mathbb{S}^2$ and $j \in \{0,1,2,3\}$. Using (3.1) and the relations $\operatorname{Tr}(\sigma_j \sigma_l) = 2\delta_{jl}$ for $j, l \in \{0,1,2,3\}$, we have

$$\langle \zeta(a), \sigma_j \zeta(a) \rangle = \operatorname{Tr}\left(\sigma_j \zeta(a) \zeta(a)^*\right) = \frac{1}{2} \operatorname{Tr}\Big(\sigma_j \sigma_0 + \sum_{l=1}^{3} a^l \sigma_j \sigma_l\Big) = a^j. \tag{B.13}$$

Now let $\tilde{f}$ be a real polynomial on $(\mathbb{R}^3)^k$ that is affine in each argument, and let $\mathbf{b}' \in (\mathbb{S}^2)^k$. Applying (B.13) in each factor of the tensor product, we find

$$\langle \zeta(\mathbf{b}'), \mathcal{O}(\tilde{f})\, \zeta(\mathbf{b}') \rangle = \sum_{\mathbf{j}} \tilde{f}_{\mathbf{j}} \prod_{i=1}^{k} \langle \zeta(b_i'), \sigma_{j_i} \zeta(b_i') \rangle = \sum_{\mathbf{j}} \tilde{f}_{\mathbf{j}} \prod_{i=1}^{k} (b_i')^{j_i} = \tilde{f}(\mathbf{b}').$$

We apply this with $\tilde{f} = f^{\mathcal{I}}$ and $\mathbf{b}' = \mathbf{b}^{\mathcal{I}}$, which lies in $(\mathbb{S}^2)^k$. Then (B.12) follows from

$$f^{\mathcal{I}}(\mathbf{b}^{\mathcal{I}}) = f\big((\mathbf{b}^{\mathcal{I}})^{\mathcal{I}}\big) = f(\mathbf{b}).$$

For the last assertion, note that (B.12) writes $f(\mathbf{b})$ as the expectation of $\mathcal{O}(f^{\mathcal{I}})$ in a unit vector. When $\mathcal{O}(f^{\mathcal{I}})$ is positive semidefinite, we therefore have $f \geq 0$ on $(\mathbb{S}^2)^k$, and for $k \geq 1$, Lemma B.2 extends this to $\mathbb{B}^k$. □

The inversion is what makes Lemma B.6 useful, already for two vectors. To see this, let X be the operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$ that exchanges the two factors, $\mathsf{X}(\varphi \otimes \psi) = \psi \otimes \varphi$. It is straightforward to check on the standard basis that

$$\sum_{j=0}^{3} \sigma_j \otimes \sigma_j = 2\mathsf{X}. \tag{B.14}$$

Recall that X has the eigenvalue 1 on the symmetric subspace $\mathrm{Sym}^2\mathbb{C}^2$, which is spanned by $e_\uparrow \otimes e_\uparrow$, $e_\uparrow \otimes e_\downarrow + e_\downarrow \otimes e_\uparrow$, and $e_\downarrow \otimes e_\downarrow$, and the eigenvalue -1 on the antisymmetric subspace $\bigwedge^2 \mathbb{C}^2$, which is spanned by $e_\uparrow \otimes e_\downarrow - e_\downarrow \otimes e_\uparrow$.

Example B.7. Let $k = 2$ and $f(b_1, b_2) = 1 + b_1 \cdot b_2$, which is nonnegative on $(\mathbb{S}^2)^2$ and vanishes when $b_2 = -b_1$. By (B.11) and (B.14), we have

$$\mathcal{O}(f) = \sigma_0 \otimes \sigma_0 + \sum_{j=1}^{3} \sigma_j \otimes \sigma_j = 2\mathsf{X}, \qquad \mathcal{O}(f^{\{1\}}) = \sigma_0 \otimes \sigma_0 - \sum_{j=1}^{3} \sigma_j \otimes \sigma_j = 2\sigma_0 \otimes \sigma_0 - 2\mathsf{X}.$$

The operator $\mathcal{O}(f)$ has the eigenvalue -2 on $\bigwedge^2 \mathbb{C}^2$, so that Lemma B.6 with $\mathcal{I} = \emptyset$ does not apply, although $f \geq 0$. However, $\mathcal{O}(f^{\{1\}})$ has the eigenvalues 0 on $\mathrm{Sym}^2\mathbb{C}^2$ and 4 on $\bigwedge^2 \mathbb{C}^2$, and Lemma B.6 with $\mathcal{I} = \{1\}$ gives $f \geq 0$ on $\mathbb{B}^2$. (The operator $\sum_{j=1}^{3} \sigma_j \otimes \sigma_j = 2\mathsf{X} - \sigma_0 \otimes \sigma_0$ is the Heisenberg coupling of two spins 1/2, with the eigenvalue 1 on the triplet $\mathrm{Sym}^2\mathbb{C}^2$ and -3 on the singlet $\bigwedge^2 \mathbb{C}^2$.)

We now fix one of the two representations of π, and we apply Lemma B.6 to $\hat{\Phi}_c$. We will always take the inversion set to be a union of groups, which means that the two vectors of a pair are either both in $\mathcal{I}$ or both outside $\mathcal{I}$. For a corner z that carries a pair, with vectors i and i', let X_z be the operator on $(\mathbb{C}^2)^{\otimes k}$ that exchanges the factors i and i', and let

$$\Pi_z^{\mathrm{sym}} = \frac{1}{2}(\mathrm{Id} + \mathsf{X}_z)$$

be the orthogonal projection onto the vectors that are invariant under X_z. On the factors i and i', it acts as the orthogonal projection of $\mathbb{C}^2 \otimes \mathbb{C}^2$ onto $\mathrm{Sym}^2\mathbb{C}^2$.

We will need the following lemma, which collects the properties of the operator that the test uses. Parts (3) and (4) distinguish the two statuses of the corners, since the measure λ_1 has an atom at 0.

Lemma B.8. *Fix one of the two representations of π, let f be a real linear combination of $\hat{Z}_{\mathrm{cur}}$ and $\hat{Z}_{\mathrm{next}}$, and let $\mathcal{I}$ be an inversion set that is a union of groups. Then the following hold.*

1. *The matrix of* $\mathcal{O}(f^{\mathcal{I}})$ *in the standard basis is real and symmetric.*
2. *For all* $g \in \mathrm{SU}(2)$, *we have* $g^{\otimes k}\mathcal{O}(f^{\mathcal{I}}) = \mathcal{O}(f^{\mathcal{I}})g^{\otimes k}$.
3. *If* z *is a corner of status* 0 *that carries a pair, then* $\Pi_z^{\mathrm{sym}}\mathcal{O}(f^{\mathcal{I}}) = \mathcal{O}(f^{\mathcal{I}}) = \mathcal{O}(f^{\mathcal{I}})\Pi_z^{\mathrm{sym}}$.
4. *If* z *is a corner of status* 1 *that carries a pair, then* $\mathsf{X}_z\mathcal{O}(f^{\mathcal{I}}) = \mathcal{O}(f^{\mathcal{I}})\mathsf{X}_z$.

In the proof of Lemma B.8, we will use four properties of the map $f \mapsto \mathcal{O}(f)$, each of which follows from (B.10) and the uniqueness of the coefficients. First, the map is linear. Second, suppose that $\{1, \dots, k\}$ is the disjoint union of sets J and J', and that $f(\mathbf{b}) = f_1(\mathbf{b}_J)f_2(\mathbf{b}_{J'})$, where $\mathbf{b}_J = (b_i)_{i\in J}$ and $\mathbf{b}_{J'} = (b_i)_{i\in J'}$. Then the coefficients of f are the products of those of f_1 and f_2, and $\mathcal{O}(f)$ is the tensor product of $\mathcal{O}(f_1)$, acting on the factors in J, and $\mathcal{O}(f_2)$, acting on the factors in J'. In particular, for one vector, $a \in \mathbb{R}^3$, and $r \in \mathbb{R}$, we have

$$\mathcal{O}(b \mapsto 1 + r\, a \cdot b) = \sigma_0 + r\, a \cdot \sigma. \tag{B.15}$$

Third, if $f = \int f_n \, \mu(\mathrm{d}n)$ for a finite measure μ and polynomials f_n whose coefficients are bounded measurable functions of n, then $\mathcal{O}(f) = \int \mathcal{O}(f_n)\, \mu(\mathrm{d}n)$. Fourth, let f and $\mathcal{I}$ be as in Lemma B.8. As noted after (B.8), we have $f(Q\mathbf{b}) = f(\mathbf{b})$ for all $Q \in \mathrm{O}(3)$. Since $(Q\mathbf{b})^{\mathcal{I}} = Q(\mathbf{b}^{\mathcal{I}})$, this implies that

$$f^{\mathcal{I}}(Q\mathbf{b}) = f^{\mathcal{I}}(\mathbf{b}) \qquad \text{for all } Q \in \mathrm{O}(3) \text{ and } \mathbf{b} \in (\mathbb{R}^3)^k. \tag{B.16}$$

Proof of Lemma B.8(2). Let $g \in \mathrm{SU}(2)$, and recall from Section 3 that its rotation $\bar{g} \in \mathrm{SO}(3)$ is defined by

$$g(a \cdot \sigma)g^* = (\bar{g}a) \cdot \sigma, \qquad a \in \mathbb{R}^3.$$

Let R be the 4×4 matrix, indexed by $\{0, 1, 2, 3\}$, such that $R_{00} = 1$, $R_{0j} = R_{j0} = 0$, and $R_{lj} = \bar{g}_{lj}$ for $l, j \in \{1, 2, 3\}$. Taking for a the unit vectors of $\mathbb{R}^3$, we obtain

$$g\sigma_j g^* = \sum_{l=0}^{3} R_{lj}\sigma_l, \qquad j \in \{0, 1, 2, 3\}. \tag{B.17}$$

Inserting (B.17) into (B.10), we find

$$g^{\otimes k}\mathcal{O}(f^{\mathcal{I}})(g^*)^{\otimes k} = \sum_{\mathbf{l}} \Big(\sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}} \prod_{i=1}^{k} R_{l_i j_i} \Big) \sigma_{\mathbf{l}}, \tag{B.18}$$

where $\mathbf{l}$ and $\mathbf{j}$ run over $\{0, 1, 2, 3\}^k$.

On the other hand, for $b \in \mathbb{R}^3$ and $j \in \{0, 1, 2, 3\}$, we have

$$(\bar{g}^{\mathsf{T}}b)^j = \sum_{l=0}^{3} R_{lj} b^l,$$

so that

$$f^{\mathcal{I}}(\bar{g}^{\mathsf{T}}\mathbf{b}) = \sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}} \prod_{i=1}^{k} \sum_{l_i=0}^{3} R_{l_i j_i} b_i^{l_i} = \sum_{\mathbf{l}} \Big(\sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}} \prod_{i=1}^{k} R_{l_i j_i} \Big) \prod_{i=1}^{k} b_i^{l_i}.$$

By (B.16), the left side coincides with $f^{\mathcal{I}}(\mathbf{b})$, and the uniqueness of the coefficients implies that

$$\sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}} \prod_{i=1}^{k} R_{l_i j_i} = f^{\mathcal{I}}_{\mathbf{l}} \qquad \text{for all } \mathbf{l} \in \{0,1,2,3\}^k.$$

Inserting this into (B.18), we have $g^{\otimes k} \mathcal{O}(f^{\mathcal{I}}) (g^*)^{\otimes k} = \mathcal{O}(f^{\mathcal{I}})$, which is (2), since $g^* = g^{-1}$. □

Proof of Lemma B.8(1). The operator $\mathcal{O}(f^{\mathcal{I}})$ is Hermitian, since its coefficients are real and the matrices σ_j are Hermitian.

To show that its matrix is real, observe that

$$\bar{\sigma}_j = \sigma_j \quad \text{for } j \neq 2, \qquad \bar{\sigma}_2 = -\sigma_2,$$

where the bar denotes the entrywise complex conjugate. Let $D = \operatorname{diag}(1,-1,1)$, the element of $\mathrm{O}(3)$ that reverses the second coordinate. Then the entrywise complex conjugate of the matrix of $\mathcal{O}(f^{\mathcal{I}})$ is

$$\sum_{\mathbf{j}} (-1)^{|\{i : j_i = 2\}|} f^{\mathcal{I}}_{\mathbf{j}} \, \sigma_{\mathbf{j}} = \mathcal{O}\big(\mathbf{b} \mapsto f^{\mathcal{I}}(D\mathbf{b})\big) = \mathcal{O}(f^{\mathcal{I}}),$$

where we used (B.16) in the last step. We conclude that the matrix of $\mathcal{O}(f^{\mathcal{I}})$ is real, and since it is Hermitian, it is symmetric. □

Proof of Lemma B.8(3). Let z be a corner of status 0 that carries a pair, with vectors i and i'. Since $\mathcal{I}$ is a union of groups, the vectors i and i' are either both in $\mathcal{I}$ or both outside it, and we let the sign $\pm$ be $-$ in the first case and $+$ in the second. The spin n_z occurs in (B.7) only in the factors of the vectors i and i' and in the factors of the ring bonds at z, and the vectors b_i and $b_{i'}$ occur only in the factors $1 + b_i \cdot n_z$ and $1 + b_{i'} \cdot n_z$. Integrating first over all spins other than n_z, and recalling that $\lambda_z = \mathrm{d}n$, we can write

$$f^{\mathcal{I}}(\mathbf{b}) = \int (1 \pm b_i \cdot n)(1 \pm b_{i'} \cdot n)\, H_n(\mathbf{b}')\, \mathrm{d}n, \tag{B.19}$$

where $\mathbf{b}' = (b_l)_{l \neq i,i'}$, and H_n is a real polynomial in $\mathbf{b}'$ that is affine in each argument and whose coefficients are polynomials in n. Applying the second and third properties above and (B.15) to (B.19), we find

$$\mathcal{O}(f^{\mathcal{I}}) = \int (\sigma_0 \pm n \cdot \sigma)^{(i)} (\sigma_0 \pm n \cdot \sigma)^{(i')} \, \widetilde{\mathcal{O}}(H_n)\, \mathrm{d}n, \tag{B.20}$$

where $\widetilde{\mathcal{O}}(H_n)$ acts as $\mathcal{O}(H_n)$ on the factors other than i and i' and as the identity on these two.

We now show that Π_z^{sym} leaves every integrand in (B.20) unchanged, from the left and from the right. Let $n \in \mathbb{S}^2$. By (3.1) with $a = \pm n$, we have $\sigma_0 \pm n \cdot \sigma = 2\zeta(\pm n)\zeta(\pm n)^*$, so that

$$(\sigma_0 \pm n \cdot \sigma)^{(i)}(\sigma_0 \pm n \cdot \sigma)^{(i')} = 4P_n,$$

where P_n is the orthogonal projection onto $\zeta(\pm n) \otimes \zeta(\pm n)$ on the factors i and i', tensored with the identity on the others. The vector $\zeta(\pm n) \otimes \zeta(\pm n)$ is invariant under the exchange of the two factors, which gives

$$\Pi_z^{\mathrm{sym}} P_n = P_n = P_n \Pi_z^{\mathrm{sym}}.$$

Further, Π_z^{sym} commutes with $\widetilde{\mathcal{O}}(H_n)$, since they act on different factors. The measure $\mathrm{d}n$ is carried by $\mathbb{S}^2$, and (3) follows from (B.20). (For a corner of status 1, the atom of λ_1 at 0 contributes a term to the analogue of (B.20) in which the factors i and i' carry $\sigma_0 \otimes \sigma_0$, and (3) fails.) □

Proof of Lemma B.8(4). Let z be a corner of status 1 that carries a pair, with vectors i and i', and let the sign $\pm$ be as in the proof of (3). The vectors b_i and $b_{i'}$ occur in $f^{\mathcal{I}}$ only through the product $(1 \pm b_i \cdot n_z)(1 \pm b_{i'} \cdot n_z)$, which is symmetric in b_i and $b_{i'}$. In particular, exchanging b_i and $b_{i'}$ leaves $f^{\mathcal{I}}$ unchanged, and the uniqueness of the coefficients implies that $f^{\mathcal{I}}_{\mathbf{j}} = f^{\mathcal{I}}_{\mathbf{j}'}$, where $\mathbf{j}'$ is obtained from $\mathbf{j}$ by exchanging j_i and $j_{i'}$. Since $\mathsf{X}_z \sigma_{\mathbf{j}} \mathsf{X}_z = \sigma_{\mathbf{j}'}$, we obtain

$$\mathsf{X}_z \mathcal{O}(f^{\mathcal{I}}) \mathsf{X}_z = \sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}} \sigma_{\mathbf{j}'} = \sum_{\mathbf{j}} f^{\mathcal{I}}_{\mathbf{j}'} \sigma_{\mathbf{j}'} = \mathcal{O}(f^{\mathcal{I}}),$$

which is (4), since $\mathsf{X}_z^2 = \mathrm{Id}$. □

Remark B.9. The second part of Lemma B.8 suggests how to choose $\mathcal{I}$. Let f be as in Lemma B.8, let $e \in \mathbb{S}^2$, and let $\mathbf{b} \in (\mathbb{S}^2)^k$ be such that $b_i \in \{e, -e\}$ for all i and the two vectors of every pair are equal. Let $\mathcal{I} = \{i : b_i = -e\}$, a union of groups, so that $\zeta(\mathbf{b}^{\mathcal{I}}) = \zeta(e)^{\otimes k}$. Up to a scalar factor, this is the only eigenvector of $(1/2)\sum_{i=1}^k (e \cdot \sigma)^{(i)}$ for its largest eigenvalue $k/2$. Differentiating the identity of part (2) along $g_t = \exp(\mathrm{i}t\, e \cdot \sigma/2)$ at $t = 0$, we see that this operator commutes with $\mathcal{O}(f^{\mathcal{I}})$. It follows that $\zeta(e)^{\otimes k}$ is an eigenvector of $\mathcal{O}(f^{\mathcal{I}})$, with the eigenvalue $f(\mathbf{b})$ by (B.12). Consequently, if a minimizer $\mathbf{b}$ of $\hat{Z}_{\mathrm{cur}}/\hat{Z}_{\mathrm{next}}$ on $(\mathbb{S}^2)^k$ has this form, then $\hat{\Phi}_c(\mathbf{b})$ is an eigenvalue of $\mathcal{O}(\hat{\Phi}_c^{\mathcal{I}})$, and this eigenvalue vanishes when c is the minimum of the ratio. The polynomial of the example above is also invariant under simultaneous rotations. Its minimizer $\mathbf{b} = (e, -e)$ leads to $\mathcal{I} = \{2\}$, for which $f^{\{2\}} = f^{\{1\}}$. In the computation, the inversion set is chosen in this way from a numerical minimizer, which is only approximately of this form; see Appendix B.4. The choice of $\mathcal{I}$ affects only whether the test succeeds, not its validity.

We now use the symmetries of Lemma B.8 to reduce the test to small subspaces. Let

$$S^3_{\mathrm{tot}} = \frac{1}{2}\sum_{i=1}^{k} \sigma_3^{(i)}$$

be the total spin component. Let $\Pi_{\mathcal{V}}$ be the product of the projections Π_z^{sym} over the corners z of status 0 that carry a pair, and let $\mathcal{V}$ be its range. These projections act on different factors, so that $\Pi_{\mathcal{V}}$ is the orthogonal projection onto $\mathcal{V}$, and $\mathcal{V}$ is the tensor product, over the groups, of $\mathbb{C}^2$ for a single vector, $\mathrm{Sym}^2\mathbb{C}^2$ for the pair of a corner of status 0, and $\mathbb{C}^2 \otimes \mathbb{C}^2$ for the pair of a corner of status 1. We call a family $\epsilon = (\epsilon_z)$ of signs $\epsilon_z \in \{1, -1\}$, indexed by the corners z of status 1 that carry a pair, a *sign assignment*, and we set

$$\mathcal{V}_\epsilon = \{\psi \in \mathcal{V} : \mathsf{X}_z\psi = \epsilon_z\psi \text{ for every corner } z \text{ of status 1 that carries a pair}\}.$$

Equivalently, $\mathcal{V}_\epsilon$ is the tensor product in which the factor $\mathbb{C}^2 \otimes \mathbb{C}^2$ of such a corner z is replaced by $\mathrm{Sym}^2\mathbb{C}^2$ if $\epsilon_z = 1$ and by $\bigwedge^2 \mathbb{C}^2$ if $\epsilon_z = -1$. The space $\mathcal{V}$ is the orthogonal direct sum of the spaces $\mathcal{V}_\epsilon$. Let $\kappa_\circ = 0$ if the number of single vectors is even and $\kappa_\circ = 1/2$ if it is odd, and let

$$\mathcal{V}_\epsilon^\circ = \{\psi \in \mathcal{V}_\epsilon : S^3_{\mathrm{tot}}\psi = \kappa_\circ\psi\}.$$

Note that for $g \in \mathrm{SU}(2)$, the operator $g \otimes g$ commutes with the exchange of the two factors of $\mathbb{C}^2 \otimes \mathbb{C}^2$, so that $g^{\otimes k}$ commutes with every X_z and every Π_z^{sym}. Therefore $\mathcal{V}$ and every $\mathcal{V}_\epsilon$ are invariant under $g^{\otimes k}$ and, by the same argument, under S^3_{tot}.

Lemma B.10. *Let ϵ be a sign assignment, and let $\mathcal{O}'$ be a Hermitian operator on $\mathcal{V}_\epsilon$ such that $\mathcal{O}' g^{\otimes k}\psi = g^{\otimes k}\mathcal{O}'\psi$ for all $g \in \mathrm{SU}(2)$ and $\psi \in \mathcal{V}_\epsilon$. Then $\mathcal{O}'$ maps $\mathcal{V}_\epsilon^\circ$ into itself, and the restriction of $\mathcal{O}'$ to $\mathcal{V}_\epsilon^\circ$ has the same set of eigenvalues as $\mathcal{O}'$. In particular, $\mathcal{O}'$ is positive definite if and only if its restriction to $\mathcal{V}_\epsilon^\circ$ is positive definite.*

Proof. We first determine the eigenvalues of S^3_{tot} on $\mathcal{V}_\epsilon$. Recall that $\sigma_3 e_\uparrow = e_\uparrow$ and $\sigma_3 e_\downarrow = -e_\downarrow$, so that the operator $(\sigma_3 \otimes \sigma_0 + \sigma_0 \otimes \sigma_3)/2$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$ has the eigenvalues 1, 0, and -1 on the vectors $e_\uparrow \otimes e_\uparrow$, $e_\uparrow \otimes e_\downarrow + e_\downarrow \otimes e_\uparrow$, and $e_\downarrow \otimes e_\downarrow$ of $\mathrm{Sym}^2\mathbb{C}^2$, and the eigenvalue 0 on $\bigwedge^2 \mathbb{C}^2$. Taking tensor products of these eigenvectors and of $e_\uparrow$, $e_\downarrow$ for the single vectors, we obtain a basis of $\mathcal{V}_\epsilon$ consisting of eigenvectors of S^3_{tot}. Each eigenvalue is a sum of one number $\pm 1/2$ for each single vector and one integer for each pair, so that every eigenvalue of S^3_{tot} on $\mathcal{V}_\epsilon$ lies in $\kappa_\circ + \mathbb{Z}$.

Next, the operators $\sigma_3^{(i)}$ commute, so that

$$g_t = \exp(\mathrm{i}t\sigma_3/2) \in \mathrm{SU}(2) \qquad \text{satisfies} \qquad g_t^{\otimes k} = \exp(\mathrm{i}tS^3_{\mathrm{tot}}), \qquad t \in \mathbb{R}.$$

By assumption, $\mathcal{O}'$ commutes with the restriction of $g_t^{\otimes k}$ to $\mathcal{V}_\epsilon$. Differentiating at $t = 0$, we find that $\mathcal{O}'$ commutes with the restriction of S^3_{tot} to $\mathcal{V}_\epsilon$. In particular, $\mathcal{O}'$ maps

the eigenspace $\mathcal{V}_\epsilon^\circ$ into itself, and every eigenvalue of the restriction of $\mathcal{O}'$ to $\mathcal{V}_\epsilon^\circ$ is an eigenvalue of $\mathcal{O}'$.

Conversely, let θ be an eigenvalue of $\mathcal{O}'$, and let $E \subset \mathcal{V}_\epsilon$ be its eigenspace. The space E is invariant under $g^{\otimes k}$ for all $g \in \mathrm{SU}(2)$, since $\mathcal{O}'$ commutes with $g^{\otimes k}$. Let $W \subset E$ be an irreducible invariant subspace with $W \neq \{0\}$.

By the standard classification of the irreducible representations of $\mathrm{SU}(2)$, there exists $j \in \{0, 1/2, 1, 3/2, \dots\}$ such that the eigenvalues of S^3_{tot} on W are $j, j-1, \dots, -j$; here we used that S^3_{tot} is $-\mathrm{i}$ times the derivative of $g_t^{\otimes k}$ at $t = 0$. These eigenvalues lie in $\kappa_\circ + \mathbb{Z}$ by the first step. Therefore $j - \kappa_\circ$ is a nonnegative integer, and $\kappa_\circ$ is one of the eigenvalues $j, j-1, \dots, -j$. It follows that $W \cap \mathcal{V}_\epsilon^\circ \neq \{0\}$, and every nonzero vector of $W \cap \mathcal{V}_\epsilon^\circ$ is an eigenvector of the restriction of $\mathcal{O}'$ to $\mathcal{V}_\epsilon^\circ$ with eigenvalue θ. This completes the proof. □

We can now state the certificate.

Proposition B.11. *Fix one of the two representations of π, and let $\mathcal{I}$ be an inversion set that is a union of groups. Suppose that for every sign assignment ϵ,*

$$\langle \psi, \mathcal{O}(\hat{\Phi}_c^{\mathcal{I}})\, \psi \rangle > 0 \qquad \text{for all } \psi \in \mathcal{V}_\epsilon^\circ \setminus \{0\}. \tag{B.21}$$

Then $\Phi_c(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in \mathbb{B}^{12}$.

Proof. Let $\mathcal{O} = \mathcal{O}(\hat{\Phi}_c^{\mathcal{I}})$. We apply Lemma B.8 with $f = \hat{\Phi}_c$, which is a real linear combination of $\hat{Z}_{\mathrm{cur}}$ and $\hat{Z}_{\mathrm{next}}$. Applying part (3) to every corner of status 0 that carries a pair, we obtain

$$\mathcal{O} = \Pi_{\mathcal{V}} \mathcal{O} \Pi_{\mathcal{V}}, \tag{B.22}$$

so that $\mathcal{O}$ maps $\mathcal{V}$ into itself. By part (4), we have $\mathcal{O}\mathsf{X}_z = \mathsf{X}_z\mathcal{O}$ for every corner z of status 1 that carries a pair, so that $\mathcal{O}$ maps every $\mathcal{V}_\epsilon$ into itself.

Let ϵ be a sign assignment, and let $\mathcal{O}_\epsilon$ be the restriction of $\mathcal{O}$ to $\mathcal{V}_\epsilon$. The operator $\mathcal{O}_\epsilon$ is Hermitian. Using part (2) and the invariance of $\mathcal{V}_\epsilon$ under $g^{\otimes k}$, we see that $\mathcal{O}_\epsilon$ satisfies the assumption of Lemma B.10. By that lemma, $\mathcal{O}_\epsilon$ maps $\mathcal{V}_\epsilon^\circ$ into itself. Its restriction to $\mathcal{V}_\epsilon^\circ$ is positive definite by (B.21), and Lemma B.10 shows that $\mathcal{O}_\epsilon$ is positive definite. The space $\mathcal{V}$ is the orthogonal direct sum of the spaces $\mathcal{V}_\epsilon$, and $\mathcal{O}$ maps each of them into itself. We conclude that the restriction of $\mathcal{O}$ to $\mathcal{V}$ is positive definite.

Now let $\mathbf{b} \in (\mathbb{S}^2)^k$. We apply Lemma B.6 with $f = \hat{\Phi}_c$ and the inversion set $\mathcal{I}$, and we obtain

$$\hat{\Phi}_c(\mathbf{b}) = \langle \zeta(\mathbf{b}^{\mathcal{I}}), \mathcal{O}\, \zeta(\mathbf{b}^{\mathcal{I}}) \rangle = \langle \Pi_{\mathcal{V}} \zeta(\mathbf{b}^{\mathcal{I}}), \mathcal{O}\, \Pi_{\mathcal{V}} \zeta(\mathbf{b}^{\mathcal{I}}) \rangle \geq 0,$$

where we used (B.22) in the second step and the positive definiteness of $\mathcal{O}$ on $\mathcal{V}$ in the last. By Corollary B.5, this implies that $\Phi_c \geq 0$ on $\mathbb{B}^{12}$. □

(The proof works equally well if the strict inequality in (B.21) is replaced by $\geq$; the program checks the strict inequality.) The reduction to the spaces $\mathcal{V}_\epsilon^\circ$ makes the test much

cheaper. For the patterns of Appendix B.4, the space $\mathcal{V}$ has dimension up to $2^{11} = 2048$, while the spaces $\mathcal{V}_\epsilon^\circ$ have dimension at most 245.

B.3 The test in integer arithmetic

We now turn (B.21) into a statement about integer matrices. Let $\mathfrak{d}$ be a positive integer such that the polynomials $3\mathfrak{d}\hat{Z}_{\text{cur}}$ and $5\mathfrak{d}\hat{Z}_{\text{next}}$ have integer coefficients. Such an integer exists, since the coefficients of $\hat{Z}_{\text{cur}}$ and $\hat{Z}_{\text{next}}$ are rational. (By (B.7), they are finite sums of products of the radii r_i, of the coefficients $1/3$ and $1/5$ of I_{cur} and I_{next}, and of moments of λ_0 and λ_1, which are rational by (B.28) below.) The value of $\mathfrak{d}$ that the programs use is given in Appendix B.4. By (B.8), we have

$$15\mathfrak{d}c_2\,\hat{\Phi}_c = 5c_2\,(3\mathfrak{d}\hat{Z}_{\text{cur}}) - 3c_1\,(5\mathfrak{d}\hat{Z}_{\text{next}}). \tag{B.23}$$

Let $\mathcal{B}$ be the set of the tensor products, over the groups, of one vector for each group, chosen among $e_\uparrow$ and $e_\downarrow$ for a single vector, among

$$e_\uparrow \otimes e_\uparrow, \qquad e_\uparrow \otimes e_\downarrow + e_\downarrow \otimes e_\uparrow, \qquad e_\downarrow \otimes e_\downarrow$$

for the pair of a corner of status 0, and among these three vectors and $e_\uparrow \otimes e_\downarrow - e_\downarrow \otimes e_\uparrow$ for the pair of a corner of status 1. Then $\mathcal{B}$ is a basis of $\mathcal{V}$, its elements have integer coordinates in the standard basis, and each of them is an eigenvector of S^3_{tot} and of every X_z. For a sign assignment ϵ, the space $\mathcal{V}_\epsilon^\circ$ is, by its definition, a joint eigenspace of these commuting operators. It follows that the set $\mathcal{B}_\epsilon$ of the elements of $\mathcal{B}$ that lie in $\mathcal{V}_\epsilon^\circ$ is a basis of $\mathcal{V}_\epsilon^\circ$. We define the matrices

$$\mathsf{A}_\epsilon^{\text{cur}} = \Big(\big\langle \beta, \mathcal{O}\big((3\mathfrak{d}\hat{Z}_{\text{cur}})^{\mathcal{I}}\big)\beta'\big\rangle\Big)_{\beta,\beta'\in\mathcal{B}_\epsilon}, \qquad \mathsf{A}_\epsilon^{\text{next}} = \Big(\big\langle \beta, \mathcal{O}\big((5\mathfrak{d}\hat{Z}_{\text{next}})^{\mathcal{I}}\big)\beta'\big\rangle\Big)_{\beta,\beta'\in\mathcal{B}_\epsilon}.$$

Their entries lie in $\mathbb{Z}[\mathrm{i}]$, since the coefficients of the polynomials are integers and the entries of $\sigma_0, \dots, \sigma_3$, β, and β' lie in $\mathbb{Z}[\mathrm{i}]$. By Lemma B.8(1), applied with $f = 3\mathfrak{d}\hat{Z}_{\text{cur}}$ and with $f = 5\mathfrak{d}\hat{Z}_{\text{next}}$, the entries are real and the matrices are symmetric, so that $\mathsf{A}_\epsilon^{\text{cur}}$ and $\mathsf{A}_\epsilon^{\text{next}}$ are symmetric integer matrices. We set

$$\mathsf{M}_\epsilon = 5c_2\mathsf{A}_\epsilon^{\text{cur}} - 3c_1\mathsf{A}_\epsilon^{\text{next}}. \tag{B.24}$$

By (B.23) and the linearity of $f \mapsto \mathcal{O}(f^{\mathcal{I}})$, we have

$$\mathsf{M}_\epsilon = 15\mathfrak{d}c_2\Big(\big\langle \beta, \mathcal{O}(\hat{\Phi}_c^{\mathcal{I}})\beta'\big\rangle\Big)_{\beta,\beta'\in\mathcal{B}_\epsilon}. \tag{B.25}$$

Lemma B.12. *Fix one of the two representations of π, let $\mathcal{I}$ be an inversion set that is a union of groups, and let ϵ be a sign assignment. Then* (B.21) *holds for this ϵ if and only if the symmetric integer matrix M_ϵ is positive definite.*

Proof. Every $\psi \in \mathcal{V}_\epsilon^\circ$ can be written uniquely as

$$\psi = \sum_{\beta \in \mathcal{B}_\epsilon} x_\beta \beta, \qquad x = (x_\beta)_{\beta \in \mathcal{B}_\epsilon} \in \mathbb{C}^{\mathcal{B}_\epsilon},$$

and $\psi \neq 0$ if and only if $x \neq 0$. Write $x = x' + \mathrm{i}x''$ with x' and x'' real. Using (B.25) and the fact that M_ϵ is real and symmetric, we have

$$15\mathfrak{d}c_2 \langle \psi, \mathcal{O}(\hat{\Phi}_c^{\mathcal{I}})\, \psi \rangle = x^* \mathsf{M}_\epsilon x = (x')^{\mathsf{T}} \mathsf{M}_\epsilon x' + (x'')^{\mathsf{T}} \mathsf{M}_\epsilon x''.$$

The right side is positive for all nonzero x if and only if M_ϵ is positive definite. □

Combining Proposition B.11 and Lemma B.12, we see that once a representation and an inversion set have been chosen, the inequality (B.1) for π follows from the positive definiteness of the matrices M_ϵ for all sign assignments ϵ. We test positive definiteness with a standard criterion. For a symmetric $n \times n$ matrix M and $1 \leq j \leq n$, let $\mathsf{M}_{[j]}$ denote its upper left $j \times j$ submatrix, and set $\det \mathsf{M}_{[0]} = 1$. Recall that, by Sylvester's criterion [18], a real symmetric matrix M is positive definite if and only if $\det \mathsf{M}_{[j]} > 0$ for $j = 1, \ldots, n$.

To compute these minors, we use fraction-free Gaussian elimination [6]. Let M be a symmetric $n \times n$ integer matrix, and set $\mathsf{M}^{(0)} = \mathsf{M}$. For $l = 1, \ldots, n-1$ and $l < a, b \leq n$, define

$$\mathsf{M}_{ab}^{(l)} = \frac{\mathsf{M}_{ll}^{(l-1)} \mathsf{M}_{ab}^{(l-1)} - \mathsf{M}_{al}^{(l-1)} \mathsf{M}_{lb}^{(l-1)}}{\det \mathsf{M}_{[l-1]}}. \tag{B.26}$$

By Sylvester's determinant identity [6], as long as the minors $\det \mathsf{M}_{[1]}, \ldots, \det \mathsf{M}_{[l-1]}$ are nonzero, $\mathsf{M}_{ab}^{(l)}$ is the determinant of the submatrix of M with rows $1, \ldots, l, a$ and columns $1, \ldots, l, b$. In particular, the pivots are the leading principal minors,

$$\mathsf{M}_{jj}^{(j-1)} = \det \mathsf{M}_{[j]}, \qquad 1 \leq j \leq n, \tag{B.27}$$

every division in (B.26) is exact, and every $\mathsf{M}_{ab}^{(l)}$ is an integer. Further, $\mathsf{M}_{ab}^{(l)} = \mathsf{M}_{ba}^{(l)}$, since the two submatrices are transposes of each other. The program computes the pivots (B.27) in the order $j = 1, \ldots, n$ with unbounded integers, updates only the entries $\mathsf{M}_{ab}^{(l)}$ with $a \leq b$, and rejects M at the first pivot that is not positive. When it accepts M, all leading principal minors are positive, and M is positive definite.

B.4 The computation

In this section, we describe the patterns, the programs, and the certificate file, and we complete the proof of Theorem 8.5. All files named below are ancillary files of this paper in the directory `anc/h1`.

Patterns. The reflection $(z_1, z_2) \mapsto (-z_1, z_2)$ of the template maps $\mathcal{P}_{\text{ext}}$ onto itself. It fixes S and N, exchanges E and W, and preserves the type of a pattern. It maps the ring functional of a pattern to that of its image, after the corresponding relabeling of the twelve vectors of **a**, so that it suffices to treat the straight patterns and the turns with $o = \mathsf{E}$ (the turns with $o = \mathsf{W}$ are their reflections). Further, we call two patterns *duplicates* if they have the same out-neighbor, the same removed sites among E, N, and W, the same present ring bonds, the same retained tip bonds, and the same status at every ring site that is a non-removed site among E, N, and W or an end of a present ring bond or of a retained tip bond. The functions I_{cur} and I_{next} of two duplicates are the same. Their ring functionals differ only in the factors of the remaining non-removed ring sites, those without present ring bonds and retained tip bonds. Such a site is either a corner, which contributes the factor $1 + m \cdot m'/3$ or $1 + m \cdot m'/5$ if it has status 0 or 1 and slot vectors m and m', or the arrival site S, which contributes the factor 1. We conclude that the ring polynomials of two duplicates differ by a factor that is a positive function of $\mathbf{a} \in \mathbb{B}^{12}$.

The file `patterns.json` lists 7,962 patterns, of which 2,696 are straight and 5,266 are turns with $o = \mathsf{E}$. For each of them, it records the out-neighbor, the non-removed sites among E, N, and W, the present ring bonds, the retained tip bonds, and the statuses of the ring sites at which these bonds end. The program `enumerate_patterns.py`, which was written from Definition 6.10 independently of the program that produced `patterns.json`, finds 106,101 assignments of statuses and used bonds that satisfy the four numbered conditions of Definition 6.10, before choosing an out-neighbor. For each assignment, it adds each choice of a non-removed site among E, N, and W as an out-neighbor. The resulting patterns have exactly the 7,962 classes in `patterns.json` modulo the reflection and duplicates. Consequently, for every pattern of $\mathcal{P}_{\text{ext}}$, the pattern or its reflection is a duplicate of a pattern of `patterns.json` of the same type.

The coefficients. The module `ring.py` reads a pattern from `patterns.json` and reads the tip radius from the environment variable `TIPRAD`, with the default value 3/5. It also contains a floating-point evaluator and minimizer of $\hat{Z}_{\text{cur}}/\hat{Z}_{\text{next}}$, which only `run_prod.py` uses. The module `cert.py` constructs the retained sites, bonds, and groups for the chosen representation, computes the integer coefficients of $3\mathfrak{d}\hat{Z}_{\text{cur}}$ and $5\mathfrak{d}\hat{Z}_{\text{next}}$, and forms and tests the matrices M_ϵ for the chosen inversion set. For each retained site y, it computes the moment tensor of λ_y, with one index in $\{0, 1, 2, 3\}$ for each factor of (B.7) at y and, if $y \in \{\mathsf{E}, \mathsf{N}, \mathsf{W}\}$, for the bond from y to the middle site. It uses exact rational arithmetic and the formula

$$\int (n^1)^{p_1} (n^2)^{p_2} (n^3)^{p_3} \, \mathrm{d}n = \frac{(p_1 - 1)!! \, (p_2 - 1)!! \, (p_3 - 1)!!}{(p_1 + p_2 + p_3 + 1)!!} \tag{B.28}$$

for even integers $p_1, p_2, p_3 \geq 0$, where $(-1)!! = 1$; the integral vanishes if one of the p_l is odd. The moments of λ_1 of positive degree are 3/5 times those of $\mathrm{d}n$. (Formula (B.28)

follows by writing a standard Gaussian vector of $\mathbb{R}^3$ as the product of its norm and an independent uniform unit vector.)

The program multiplies each moment tensor by the least common denominator of its entries. For a single vector from a retained tip bond, it multiplies the factor $1 + (3/5)\, b_i \cdot n$ by 5. For a reduced leaf, it represents the single-vector field on the neighboring axial site by a diagonal transfer tensor at the leaf. If its radius is $r_i = 1/q$ with $q \in \{2, 3\}$, the tensor after scaling is $\mathrm{diag}(q, 1, 1, 1)$. It joins the index of the leaf vector to that of the ring bond, giving $q(1 + r_i\, b_i \cdot n)$ upon contraction. The integer $\mathfrak{d}$ is the product of the local scales.

Since $1 + n_y \cdot n_z = \sum_{j=0}^{3} n_y^j n_z^j$, the coefficients of $\mathfrak{d}\hat{Z}[1]$ are obtained by contracting the integer tensors along the present ring bonds, in the order E, $(1,1)$, N, $(-1,1)$, W, $(-1,-1)$, S, $(1,-1)$ of the ring, with the indices of the bonds to the middle site set to 0. For a pair $\{a', a''\}$ of non-removed sites in $\{\mathsf{E}, \mathsf{N}, \mathsf{W}\}$, the coefficients of $\mathfrak{d}\hat{Z}[n_{a'} \cdot n_{a''}]$ are obtained in the same way, with the indices of the bonds from a' and a'' to the middle site set to a common value $j \in \{1, 2, 3\}$ and summed over j. By (6.10), this gives

$$3\mathfrak{d}\hat{Z}_{\mathrm{cur}} = 3\mathfrak{d}\hat{Z}[1] + \sum_{\{a', a''\}} \mathfrak{d}\hat{Z}[n_{a'} \cdot n_{a''}], \qquad 5\mathfrak{d}\hat{Z}_{\mathrm{next}} = 5\mathfrak{d}\hat{Z}[1] + \mathfrak{d}\hat{Z}[n_f \cdot n_g],$$

where the sum runs over the pairs of non-removed sites in $\{\mathsf{E}, \mathsf{N}, \mathsf{W}\}$, and the last term is absent if f or g is removed.

We bound the integers that occur in these contractions. A moment tensor is of degree at most four. By (B.28), its least common denominator divides 15 for λ_0 and 25 for λ_1. Its scaled entries are nonnegative and at most 25, since the moments have absolute value at most one. The transfer tensor of a reduced leaf has entries of size at most 3. Each of the at most four retained tip factors contributes an additional scale of at most 5. Thus any product of local entries is at most $25^8 5^4$.

Fix the exterior indices and the indices of the bonds to the middle site in a coefficient. At a ring site, once the index of the incoming ring bond is fixed, the parity of the three coordinate exponents determines at most one index of the outgoing bond for which the moment is nonzero. Indeed, index 0 leaves all parities unchanged, and each of indices 1, 2, and 3 changes a different parity. A missing bond has index 0. The transfer tensor of a reduced leaf satisfies the same rule, since its leaf-vector index equals its single ring-bond index. For the closed ring this leaves at most four contributing assignments, one for each possible index on the first bond. For a partial path contraction, the index at its first end is fixed, and the same rule determines the rest of the assignment. Since the summands are nonnegative, all intermediate products and sums are bounded by

$$M_{\mathrm{coeff}} = 4 \cdot 25^8 5^4.$$

In the coefficients of $3\mathfrak{d}\hat{Z}_{\mathrm{cur}}$ there are three copies of the partition coefficient and at most nine pair-index contributions. The coefficients of $5\mathfrak{d}\hat{Z}_{\mathrm{next}}$ use five copies and at most

three contributions. Both sums are bounded by $12M_{\mathrm{coeff}} < 2^{53}$, including every partial sum. This proves that the 64-bit integer tensor contractions and coefficient sums are free of overflow. The checker also guards their magnitudes by run-time assertions using binary64 copies of the contractions. The bounds above ensure exact integer representation in these copies.

The matrices. The program then forms the matrices of Appendix B.3. It changes the signs of the coefficients according to (B.11), replaces each index j by the matrix σ_j, with real and imaginary parts kept separately, and passes to the basis $\mathcal{B}$ by an integer change of basis on the factors of each pair. This is again done in 64-bit integers. Before each single-factor Pauli substitution, the magnitude of each input is checked to be less than 2^{61}, and each output is a signed sum of at most two inputs. Before each pair change of basis, the magnitude is checked to be less than 2^{58}, and each resulting entry is a signed sum of at most four inputs. If a magnitude assertion fails, the program repeats the assembly with unbounded integers.

The program checks exactly that the imaginary parts vanish and that the matrices are symmetric. For every corner of status 0 that carries a pair, it checks that the vector $e_\uparrow \otimes e_\downarrow - e_\downarrow \otimes e_\uparrow$ of that pair is annihilated from the left and from the right. The program converts the maximum absolute residual of each of these three checks to binary64 for comparison with zero. These residuals are integers, and the conversion preserves whether they vanish. It also checks that the entries joining two elements of $\mathcal{B}$ with different eigenvalues of S^3_{tot} or of some X_z vanish. These checks follow from Lemma B.8 and serve to catch errors in the assembly. Finally, for every sign assignment ϵ, the program extracts the rows and columns of $\mathcal{B}_\epsilon$, forms M_ϵ by (B.24) with unbounded integers, and tests it by the elimination of Appendix B.3.

Production and checking. We produced the certificates with the program `run_prod.py`. For each pattern, it minimizes the ratio $\hat{Z}_{\mathrm{cur}}/\hat{Z}_{\mathrm{next}}$ of the unreduced representation numerically, by coordinate descent. In each step, the ratio is a quotient of two affine functions of one vector, which the program minimizes over that vector in closed form. The program then lets $\mathcal{I}$ consist of the groups whose vectors have a negative total component along the principal axis of the minimizer, as in Remark B.9. Next, it estimates the largest c for which the forms in (B.21) are positive semidefinite. When this value is not sufficiently above the threshold, it changes $\mathcal{I}$ by a greedy search over the groups or passes to the unreduced representation. The numerical minimization and eigenvalue estimates guide these choices.

The program `verify_all.py` is the checker. It reads each record of the certificate file and recomputes the integer coefficients, the matrices M_ϵ, and the test with `cert.py`, from the stored data alone and without any optimization. It refuses to run if assertions are disabled. Its output is the number of certificates verified and the list of failures. The

program `check_certificate_file.py` checks the relation between the certificate file and `patterns.json`, which `verify_all.py` does not check; see Appendix B.6. The program `closed_exact.py` is imported by `verify_all.py` and `run_prod.py`. It implements a different kind of certificate, which does not occur in the certificate file.

The certificate file. The file `certificates.jsonl` contains one record for each of the 7,962 patterns. A record consists of the index of the pattern in `patterns.json`, the threshold c as a fraction c_1/c_2, the representation, the inversion set $\mathcal{I}$, the tip radius $3/5$, and the value `covered` of a field that names the ring functional (6.21), in which every non-removed corner carries two slot factors. The reduced representation is used for all patterns except two, which use the unreduced one. The number of vectors is $k \leq 11$, there are at most 16 sign assignments for a pattern, and the matrices M_ϵ have size at most 245. A record also contains floating-point information, such as the value of c at which some M_ϵ becomes singular, which the checker does not use. The program `verify_all.py` accepts all 7,962 certificates, in about 15 to 30 CPU-minutes, and its output is recorded in `anc/h1/verify_all_output.log`. Producing the certificates took about 34 CPU-minutes.

Proof of Theorem 8.5. Let $\pi \in \mathcal{P}_{\mathrm{ext}}$. By the paragraph on patterns, there exists a pattern π' of `patterns.json` of the same type as π such that π or its reflection is a duplicate of π'. The ring polynomial of π at $\mathbf{a}$ is then a positive multiple of the ring polynomial of π' at a relabeling of $\mathbf{a}$, and $c_\pi = c_{\pi'}$. It therefore suffices to prove (B.1) for the patterns of `patterns.json`.

The program `check_certificate_file.py` verifies that `certificates.jsonl` contains exactly one record for each pattern π' of `patterns.json`, with the threshold $c_{\pi'}$, the tip radius $3/5$, and the ring functional (6.21). For each record, `verify_all.py` verifies that M_ϵ is positive definite for every sign assignment ϵ, with the representation and the inversion set of the record and with $c = c_{\pi'}$. By Lemma B.12, the hypothesis (B.21) of Proposition B.11 holds, and Proposition B.11 gives $\mathcal{Z}^{\mathbf{a}}_{\pi'}[I_{\mathrm{cur}} - c_{\pi'} I_{\mathrm{next}}] \geq 0$ for all $\mathbf{a} \in \mathbb{B}^{12}$. This completes the proof. □

B.5 Checks that are not part of the proof

We also made the following checks. They are not part of the proof, and their programs are not among the ancillary files, but they corroborate the computation of Appendix B.4 at the points where an error would be hardest to see.

The first check concerns the polynomials themselves. An independent evaluator, written from the definitions, starts from the template before the middle site and the tips are integrated out. It keeps these sites as spins, with three slot vectors at each tip whose bond is present. Every site is integrated with a randomly rotated product quadrature rule that is exact for polynomials of degree at most 4, with an additional node at 0 for

the atom of λ_1. At three random configurations of the slot vectors for every pattern, it agrees with the ratio $\hat{Z}_{\mathrm{cur}}/\hat{Z}_{\mathrm{next}}$ computed from the integer coefficients of `cert.py` in the representation of the certificate, to a relative error of at most $2.3 \cdot 10^{-15}$. This comparison checks the model, namely the sites, bonds, measures, and radii that enter the polynomials, including the integration of the tips in Lemma 6.11. Where this was tested, changing the status of a single site changed the ratio by at least $5 \cdot 10^{-5}$ in relative terms. However, a change of one unit in a single integer coefficient can be invisible at this precision when $\mathfrak{d}$ is large.

The second check concerns the exactness of the coefficients. For all 7,962 certificates, the integer matrices of `cert.py` on $\mathcal{V}$, written in the basis $\mathcal{B}$, commute exactly with the raising and lowering operators of the total spin, written in the same basis. This verifies the hypothesis of Lemma B.10 directly, in addition to its proof in Lemma B.8. In 30 random trials, a single coefficient of $3\mathfrak{d}\hat{Z}_{\mathrm{cur}}$ that multiplies a nonconstant monomial was changed by 1, and in all 30 trials the perturbed operator failed the exact checks of Appendix B.4 or this commutation.

The third check concerns Lemma B.10. For 63 patterns, the matrices were tested exactly on every eigenspace of S^3_{tot} in every $\mathcal{V}_\epsilon$, and they were positive definite on all of them. These are the 13 tightest patterns and 50 patterns with corners of status 1 that carry a pair, among them the 10 patterns with the largest operators.

The fourth check consists of negative controls. The stored certificates of the tightest patterns, which are accepted at their threshold, are rejected when the threshold is raised $2 \cdot 10^{-5}$ above an independent numerical minimum of the ratio. The same holds for 20 random patterns.

Finally, a second implementation, written independently of the programs of Appendix B.4, confirms all 7,962 certificates, with the thresholds determined from the patterns themselves. It shares with them only the definitions, the order of the vectors needed to read the stored inversion sets, and the idea of the inversion. It computes the polynomials by exact symbolic integration, site by site, with the moment formula (B.28) in integer arithmetic. It assembles the operators on the whole space $(\mathbb{C}^2)^{\otimes k}$ from exchange operators, using (B.14), and checks exactly that their entries between different eigenvalues of S^3_{tot} vanish. On every eigenspace of S^3_{tot}, it proves positivity by an exact integer congruence. It rounds a scaled inverse of a floating-point Cholesky factor of the block to an integer triangular matrix, computes the congruent matrix exactly, and checks that it is strictly diagonally dominant with a positive diagonal. For the patterns with a corner of status 0 that carries a pair, it checks exactly that the vectors $e_\uparrow \otimes e_\downarrow - e_\downarrow \otimes e_\uparrow$ of these pairs are annihilated, and it proves positive definiteness on the orthogonal complement. It therefore shows that $\mathcal{O}(\hat{\Phi}^{\mathcal{I}}_c)$ is positive semidefinite on all of $(\mathbb{C}^2)^{\otimes k}$, so that its conclusion rests on Lemma B.6 and Corollary B.5 alone, and not on Lemmas B.8 and B.10.

B.6 What must be trusted

A reader who accepts the lemmas of this appendix must trust the following points.

1. The list `patterns.json` represents $\mathcal{P}_{\mathrm{ext}}$ modulo the reflection and duplicates. Two independent enumerations from the definition of the class, one of which is `enumerate_patterns.py`, confirm this, class by class.
2. The certificate file contains each pattern of `patterns.json` exactly once, with the threshold of its type, the tip radius $3/5$, and the ring functional (6.21). The checker verifies each record at the threshold written in it, without checking these properties; the program `check_certificate_file.py` checks them.
3. The program `cert.py` computes the coefficients of $3\mathfrak{d}\hat{Z}_{\mathrm{cur}}$ and $5\mathfrak{d}\hat{Z}_{\mathrm{next}}$ correctly. This part has about 130 lines, together with the parts of `ring.py` that read a pattern. Its output agrees with the independent evaluator of Appendix B.5 and passes the exact commutation check there.
4. The program `cert.py` assembles the matrices M_ϵ and selects the bases $\mathcal{B}_\epsilon$ correctly. This part has about 110 lines, and its output satisfies the exact checks of Appendix B.4.
5. The fraction-free elimination of Appendix B.3 is implemented correctly. It has about 20 lines. It agrees with an exact LDL^{T} decomposition in rational arithmetic on 600 random symmetric integer test matrices, which include positive definite, singular positive semidefinite, and indefinite matrices with entries up to 2^{200}.
6. The integer and rational arithmetic of Python and the integer tensor contraction of the library `numpy` are correct in the absence of overflow. The coefficient bound above excludes overflow in the contractions and coefficient sums. The signed operator assembly uses separate integer magnitude assertions and falls back to unbounded arithmetic when needed. The checker must be run with assertions enabled (without the option `-O` of Python). The program `verify_all.py` refuses to run otherwise, and `check_certificate_file.py` checks this as well.
7. The tip radius, a parameter of the programs, must be left at its value $3/5$, which means that the environment variable `TIPRAD` must be unset or equal to $3/5$. With a smaller value, the checker would verify a weaker inequality. The program `check_certificate_file.py` checks this.

The proof is valid for any representation and inversion set for which the checker accepts the certificate. In particular, it remains valid if the numerical minimization misses the true minimum.

APPENDIX C

The tail of the weighted trail sums

In this appendix, we prove Proposition C.7, which gives the bound $K_{\mathrm{t}}(1) \le 0.92320$ of Theorem 8.1 for both types, and we fix the enumerations ς_w of the first-visit steps that Hypothesis (H3) leaves open. The trails of length at most 20 are enumerated exactly and their weights are bounded using upward rounding. A longer trail is followed exactly for 20 steps, and we dominate its continuation by a weighted walk on $\mathbb{Z}^2$ that remembers a phase, its last ten steps, and its position. We bound the total weight of the paths of this walk that end at v by a supersolution, that is, a nonnegative function on the states of the walk that dominates its own image under one step of the walk plus the reward at v. Such a function is determined by finitely many numbers, and a program verifies the defining inequality with directed rounding. Since the walk keeps its position and must end at v, the bound retains the constraint that the trails close at v (see Section 8.2).

C.1 Reduction to two scalar sums

Throughout this appendix, we fix a type $\mathrm{t} \in \{\mathrm{coll}, \mathrm{diag}\}$, and we write $c = (0,0)$, $u = (-1,0)$, and $v = v^{\mathrm{t}}$, so that (u,v) is a pair of type t with center c. Recall from Section 7.3 that $\mathcal{T}_n^{\mathrm{t}}$ is the set of trails $w = (v_0, \dots, v_n)$ of $\mathbb{Z}^2$ that do not visit c, with $v_0 = u$, $v_n = v$, and $v_i \ne v$ for $0 \le i < n$, and that we use the arrival site $v_{-1} = c$. The thresholds $c_{\mathrm{s}}, c_{\mathrm{t}} \in (0,1]$ appear through the numbers $1/c_{\mathrm{s}}$ and $1/c_{\mathrm{t}}$, and we use their values $(961/1000, 187/200)$ only in Appendix C.5.

We use the notions of Sections 6.6 and 7.3 for trails of $\mathbb{Z}^2$: the steps $j \in \{1, \dots, n\}$ of a trail, the set $\mathrm{FV}(w)$ of first-visit steps, the number $\mathrm{rev}(w)$ of revisits, straight steps and turns, the pairs at a step, dead and live pairs, the weighted coefficient sum $\kappa(j)$ of (7.3), and the numbers $\bar{\rho}_j \in \{1/c_{\mathrm{s}}, 1/c_{\mathrm{t}}\}$. These definitions apply verbatim to every trail $(v_0, \dots, v_k)$ of $\mathbb{Z}^2$ that does not visit c, with $v_0 = u$ and the arrival site $v_{-1} = c$, whether or not it ends at v. The type of step j (revisit, straight first visit, or turning first visit) and the number $\kappa(j)$ are determined by $(v_0, \dots, v_j)$. For $w \in \mathcal{T}_n^{\mathrm{t}}$, we let $\mathsf{m}(\{w\})$ denote the term of w in (7.4), that is,

$$\mathsf{m}(\{w\}) = 3^{-n} \Big(\frac{3}{5}\Big)^{\mathrm{rev}(w)} \sum_{k=1}^{|\mathrm{FV}(w)|} \kappa(\varsigma_w(k)) \prod_{l=1}^{k} \bar{\rho}_{\varsigma_w(l)}. \tag{C.1}$$

By (7.5), we have

$$h_{\mathsf{t}} K_{\mathsf{t}}(1) = \sum_{n \geq 1} \mathsf{m}_{\mathsf{t}}(n). \tag{C.2}$$

Since $d(u, v) = 2$ and $\mathbb{Z}^2$ is bipartite, every trail from u to v has even length, and $\mathsf{m}_{\mathsf{t}}(n) = 0$ for odd n. The first step of a trail of $\mathcal{T}_n^{\mathsf{t}}$ avoids $c = u + \mathsf{E}$, so that its direction belongs to $\{\mathsf{N}, \mathsf{W}, \mathsf{S}\}$. For $\mathsf{e}_0 \in \{\mathsf{N}, \mathsf{W}, \mathsf{S}\}$, we let $\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}$ denote the set of trails of $\mathcal{T}_n^{\mathsf{t}}$ whose first step is e_0, and we set $a_0 = u + \mathsf{e}_0$. Then

$$\mathsf{m}_{\mathsf{t}}(n) = \sum_{\mathsf{e}_0 \in \{\mathsf{N},\mathsf{W},\mathsf{S}\}} \mathsf{m}(\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}). \tag{C.3}$$

By (C.2), the bound (8.1) of Theorem 8.1 is equivalent to the two scalar inequalities

$$\sum_{n \geq 1} \mathsf{m}_{\mathrm{coll}}(n) \leq 0.92320, \qquad \sum_{n \geq 1} \mathsf{m}_{\mathrm{diag}}(n) \leq \frac{5}{4} \cdot 0.92320. \tag{C.4}$$

C.2 The enumerations

We now fix the enumerations ς_w in (H3). For trails of length at most 20, we use an enumeration that minimizes (C.1). This enumeration depends on the whole trail, so that a walk with a bounded memory cannot follow it. For longer trails, we list the straight first-visit steps before the turns, which a walk can follow with the four phases of Appendix C.3.

Definition C.1. Let $w \in \mathcal{T}_n^{\mathsf{t}}$. If $n \leq 20$, we let ς_w list the steps $j \in \mathrm{FV}(w)$ in increasing order of the quotient $(\bar{\rho}_j - 1)/(\bar{\rho}_j \kappa(j))$, which we set equal to $+\infty$ if $\kappa(j) = 0$, with ties broken by increasing j. If $n > 20$, we let ς_w list the straight steps of $\mathrm{FV}(w)$ in increasing order, followed by the turning steps of $\mathrm{FV}(w)$ in increasing order.

For fixed thresholds, each ς_w is determined by w. We now explain why, for $n \leq 20$, the enumeration ς_w minimizes (C.1) over all enumerations of $\mathrm{FV}(w)$. Let i and j be consecutive in an enumeration, with i before j, and let Π be the product of $\bar{\rho}$ over the steps that precede them. Exchanging i and j changes the sum in (C.1) by

$$\Pi\big[\bar{\rho}_j \kappa(j) + \bar{\rho}_j \bar{\rho}_i \kappa(i) - \bar{\rho}_i \kappa(i) - \bar{\rho}_i \bar{\rho}_j \kappa(j)\big] = \Pi\big[\bar{\rho}_i \kappa(i)(\bar{\rho}_j - 1) - \bar{\rho}_j \kappa(j)(\bar{\rho}_i - 1)\big], \tag{C.5}$$

since the terms of the other steps do not change. Observe that the right side is nonpositive when the quotient of i is larger than that of j, and that it vanishes when the quotients are equal. Every enumeration can be brought into the order of Definition C.1 by a finite sequence of exchanges of consecutive steps, each of which exchanges two steps whose quotients are in decreasing order, or equal quotients in the wrong order of indices. None of these exchanges increases (C.1).

C.3 Kinds and phases

We now write (C.1), for $n > 20$, as a sum of products with one factor for each step. We define the *kind* of a step of a trail to be s if it is a straight first visit, t if it is a turning first visit, and r if it is a revisit. (For a step $j \geq 2$ of a trail $(v_0, \dots, v_k)$, this is the kind of the admissible step $v_j - v_{j-1}$ in the sense of Definition 8.3, from the history formed by the steps $1, \dots, j-1$.) We use four *phases*, denoted t^-, t^+, s^-, and s^+. Each term of (C.1) belongs to one step $j = \varsigma_w(k)$ of $\mathrm{FV}(w)$. Heuristically, the phases t^- and t^+ are used for the steps before and after j when j is a turn, and the phases s^- and s^+ when j is straight. For a phase ϑ and a kind k, we define the *phase weight* $\varpi_\vartheta(\mathrm{k})$ by the table

ϑ	$\varpi_\vartheta(\mathrm{s})$	$\varpi_\vartheta(\mathrm{t})$	$\varpi_\vartheta(\mathrm{r})$
t^-	$1/(3c_\mathrm{s})$	$1/(3c_\mathrm{t})$	$1/5$
t^+	$1/(3c_\mathrm{s})$	$1/3$	$1/5$
s^-	$1/(3c_\mathrm{s})$	$1/3$	$1/5$
s^+	$1/3$	$1/3$	$1/5$

(C.6)

Since $c_\mathrm{s}, c_\mathrm{t} \leq 1$, the smallest phase weight is $1/5$, and $\varpi_\vartheta(\mathrm{s}), \varpi_\vartheta(\mathrm{t}) \geq 1/3$ for every phase ϑ.

Let $w = (v_0, \dots, v_k)$ be a trail of $\mathbb{Z}^2$ that does not visit c, with $v_0 = u$, and let k_j denote the kind of its step j. For $1 \leq j \leq k$ and phases ϑ, ϑ', we define the weight of step j for the transition from ϑ to ϑ' by

$$\varpi_j(\vartheta, \vartheta') = \begin{cases} \varpi_\vartheta(\mathrm{k}_j) & \text{if } \vartheta' = \vartheta, \\ \kappa(j)/(3c_\mathrm{t}) & \text{if } (\vartheta, \vartheta') = (\mathrm{t}^-, \mathrm{t}^+) \text{ and } \mathrm{k}_j = \mathrm{t}, \\ \kappa(j)/(3c_\mathrm{s}) & \text{if } (\vartheta, \vartheta') = (\mathrm{s}^-, \mathrm{s}^+) \text{ and } \mathrm{k}_j = \mathrm{s}, \\ 0 & \text{otherwise.} \end{cases} \tag{C.7}$$

Note that $\varpi_j(\vartheta, \vartheta')$ is determined by $(v_0, \dots, v_j)$. A transition from t^- to t^+, or from s^- to s^+, marks the step j of a term of (C.1).

Lemma C.2. *Let $n > 20$ and $w \in \mathcal{T}_n^\mathrm{t}$, and let ς_w be as in Definition C.1. Then*

$$\mathsf{m}(\{w\}) = \sum_{\vartheta_0 \in \{\mathrm{t}^-, \mathrm{s}^-\}} \sum_{\vartheta_1, \dots, \vartheta_{n-1}} \sum_{\vartheta_n \in \{\mathrm{t}^+, \mathrm{s}^+\}} \prod_{j=1}^{n} \varpi_j(\vartheta_{j-1}, \vartheta_j). \tag{C.8}$$

Proof. By (C.7), a sequence of phases can have positive weight in (C.8) only if it changes exactly once, at a step j, from t^- to t^+ with $\mathrm{k}_j = \mathrm{t}$, or from s^- to s^+ with $\mathrm{k}_j = \mathrm{s}$. Each step $j \in \mathrm{FV}(w)$ determines exactly one candidate sequence of this form, whose weight is zero when $\kappa(j) = 0$. It therefore suffices to show that the product in (C.8) for the sequence determined by j coincides with the term of j in (C.1), which is

$$3^{-n} \Big(\frac{3}{5}\Big)^{\mathrm{rev}(w)} \kappa(j) \prod_{l=1}^{k} \bar{\rho}_{\varsigma_w(l)}, \qquad j = \varsigma_w(k). \tag{C.9}$$

Note that $3^{-n}(3/5)^{\mathrm{rev}(w)}$ is the product, over the steps i of w, of $1/5$ if i is a revisit and $1/3$ if i is a first visit.

Suppose first that j is a turn. By Definition C.1, the steps up to j in the enumeration are all straight steps of $\mathrm{FV}(w)$ and the turning steps $i \in \mathrm{FV}(w)$ with $i \le j$. Therefore (C.9) is the product over the steps i of w of the factors

$$\begin{cases} 1/5 = \varpi_{\mathrm{t}^{\mp}}(\mathrm{r}) & \text{if } \mathrm{k}_i = \mathrm{r}, \\ 1/(3c_{\mathrm{s}}) = \varpi_{\mathrm{t}^{\mp}}(\mathrm{s}) & \text{if } \mathrm{k}_i = \mathrm{s}, \\ 1/(3c_{\mathrm{t}}) = \varpi_{\mathrm{t}^{-}}(\mathrm{t}) & \text{if } \mathrm{k}_i = \mathrm{t} \text{ and } i < j, \\ \kappa(j)/(3c_{\mathrm{t}}) = \varpi_j(\mathrm{t}^{-}, \mathrm{t}^{+}) & \text{if } i = j, \\ 1/3 = \varpi_{\mathrm{t}^{+}}(\mathrm{t}) & \text{if } \mathrm{k}_i = \mathrm{t} \text{ and } i > j, \end{cases}$$

where we used (C.6), and $\mathrm{t}^{\mp}$ stands for t^{-} if $i < j$ and for t^{+} if $i > j$. This is the product in (C.8) for the sequence with $\vartheta_i = \mathrm{t}^{-}$ for $i < j$ and $\vartheta_i = \mathrm{t}^{+}$ for $i \ge j$.

Suppose next that j is straight. Then the steps up to j in the enumeration are the straight steps $i \in \mathrm{FV}(w)$ with $i \le j$. In the same way, (C.9) is the product over the steps i of w of the factors

$$\begin{cases} 1/5 = \varpi_{\mathrm{s}^{\mp}}(\mathrm{r}) & \text{if } \mathrm{k}_i = \mathrm{r}, \\ 1/3 = \varpi_{\mathrm{s}^{\mp}}(\mathrm{t}) & \text{if } \mathrm{k}_i = \mathrm{t}, \\ 1/(3c_{\mathrm{s}}) = \varpi_{\mathrm{s}^{-}}(\mathrm{s}) & \text{if } \mathrm{k}_i = \mathrm{s} \text{ and } i < j, \\ \kappa(j)/(3c_{\mathrm{s}}) = \varpi_j(\mathrm{s}^{-}, \mathrm{s}^{+}) & \text{if } i = j, \\ 1/3 = \varpi_{\mathrm{s}^{+}}(\mathrm{s}) & \text{if } \mathrm{k}_i = \mathrm{s} \text{ and } i > j, \end{cases}$$

with the analogous meaning of $\mathrm{s}^{\mp}$. Summing over j proves (C.8). □

C.4 The stationary walk

In this section and the next, we fix, in addition to $\mathfrak{t}$, a first direction $\mathsf{e}_0 \in \{\mathsf{N}, \mathsf{W}, \mathsf{S}\}$, and we recall that $a_0 = u + \mathsf{e}_0$. We use the histories of memory 10 of Definition 8.3, with their admissible steps, successors, and kinds. We have $|\Omega_{10}| = 60{,}092$.

The walk below cannot determine which pairs are dead. Instead, it uses the following two upper bounds for $\kappa(j)$. Let $x \in \mathbb{Z}^2$, and let e' and e be unit vectors with $\mathsf{e} \ne -\mathsf{e}'$. Consider the three neighbors of x other than $x - \mathsf{e}'$, and the three pairs that they form, with the coefficient $\bar{\alpha}$ for the pair that does not contain $x + \mathsf{e}$ and $\bar{\beta}$ for the two pairs that contain it. We define $\kappa^{\mathrm{c}}(x, \mathsf{e}', \mathsf{e})$ to be the sum of the coefficient of p times $h(p)$ over the pairs p that do not contain c, and we define $\kappa^{\circ}(\mathsf{e}', \mathsf{e})$ to be the same sum over all three pairs. A short computation gives

$$\kappa^{\circ}(\mathsf{e}', \mathsf{e}) = \begin{cases} \bar{\alpha} h_{\mathrm{coll}} + 2\bar{\beta} h_{\mathrm{diag}} & \text{if } \mathsf{e} = \mathsf{e}', \\ \bar{\beta} h_{\mathrm{coll}} + (\bar{\alpha} + \bar{\beta}) h_{\mathrm{diag}} & \text{if } \mathsf{e} \ne \pm\mathsf{e}'. \end{cases} \tag{C.10}$$

Indeed, for a straight step the pair that does not contain $x + \mathsf{e}$ is collinear and the other two pairs are diagonal. For a turn, the pair formed by $x + \mathsf{e}$ and $x - \mathsf{e}$ is collinear, and the other two pairs are diagonal. Now let w be a trail as in Appendix C.3, and let $j \geq 2$ be a step of $\mathrm{FV}(w)$, with arrival direction $\mathsf{e}' = v_{j-1} - v_{j-2}$ and direction $\mathsf{e} = v_j - v_{j-1}$. The pairs at step j are the three pairs above for $x = v_{j-1}$, with the same coefficients, and every live pair avoids c. Since all coefficients and type weights are nonnegative, we have

$$\kappa(j) \leq \kappa^{\mathsf{c}}(v_{j-1}, \mathsf{e}', \mathsf{e}) \leq \kappa^{\circ}(\mathsf{e}', \mathsf{e}). \tag{C.11}$$

Further, $\kappa^{\mathsf{c}}(x, \mathsf{e}', \mathsf{e}) = \kappa^{\circ}(\mathsf{e}', \mathsf{e})$ unless c is a neighbor of x.

Definition C.3. A *state* is a triple (ϑ, ω, x) of a phase ϑ, a history $\omega \in \Omega_{10}$, and a site $x \in \mathbb{Z}^2$. We define the nonnegative matrix $\mathsf{T} = \mathsf{T}_{\mathsf{t},\mathsf{e}_0}$, indexed by pairs of states, as follows. Its entries vanish except for the entries $\mathsf{T}(\vartheta, \omega, x; \vartheta', \omega\mathsf{e}, x + \mathsf{e})$, where e is an admissible step from ω and ϑ' is a phase. Let k denote the kind of e as a step from ω, and let e_{10} denote the last entry of ω. Then the entry $\mathsf{T}(\vartheta, \omega, x; \vartheta', \omega\mathsf{e}, x + \mathsf{e})$ is given by the first of the following rules that applies.

(a) If $x \in \{c, v\}$ or $x + \mathsf{e} = c$, the entry is 0.

(b) If $x \in \{u, a_0\}$, the entry is 0 if $\{x, x + \mathsf{e}\} = \{u, a_0\}$ or $\vartheta' \neq \vartheta$, and it is $1/5$ otherwise.

(c) The entry is $\varpi_\vartheta(\mathsf{k})$ if $\vartheta' = \vartheta$, it is $\kappa^{\mathsf{c}}(x, \mathsf{e}_{10}, \mathsf{e})/(3c_{\mathsf{t}})$ if $(\vartheta, \vartheta') = (\mathsf{t}^-, \mathsf{t}^+)$ and $\mathsf{k} = \mathsf{t}$, it is $\kappa^{\mathsf{c}}(x, \mathsf{e}_{10}, \mathsf{e})/(3c_{\mathsf{s}})$ if $(\vartheta, \vartheta') = (\mathsf{s}^-, \mathsf{s}^+)$ and $\mathsf{k} = \mathsf{s}$, and it is 0 otherwise.

For a nonnegative function F on the states, we define

$$(\mathsf{TF})(\vartheta, \omega, x) = \sum_{\mathsf{e}} \sum_{\vartheta'} \mathsf{T}(\vartheta, \omega, x; \vartheta', \omega\mathsf{e}, x + \mathsf{e})\, \mathsf{F}(\vartheta', \omega\mathsf{e}, x + \mathsf{e}), \tag{C.12}$$

where e runs over the admissible steps from ω and ϑ' over the phases. We define the *reward* g by $\mathsf{g}(\vartheta, \omega, x) = 1$ if $x = v$ and $\vartheta \in \{\mathsf{t}^+, \mathsf{s}^+\}$, and $\mathsf{g}(\vartheta, \omega, x) = 0$ otherwise. Finally, we define the nonnegative matrix T°, indexed by pairs (ϑ, ω), whose entry $\mathsf{T}^{\circ}(\vartheta, \omega; \vartheta', \omega\mathsf{e})$ is given by rule (c) with $\kappa^{\circ}(\mathsf{e}_{10}, \mathsf{e})$ in place of $\kappa^{\mathsf{c}}(x, \mathsf{e}_{10}, \mathsf{e})$, and whose other entries vanish. It acts on nonnegative functions of (ϑ, ω) as in (C.12).

The sum (C.12) has at most 12 nonzero terms, so that TF is finite whenever F is. The walk with the transition weights T remembers its last ten steps, its position, and the first bond $\{u, a_0\}$ of the trail. By rule (a), the walk stops at v, where it collects the reward g. Its steps into c have weight zero. Rule (b) concerns the sites u and a_0. A trail of $\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}$ leaves them at the steps 1 and 2, so that every later step from these sites is a revisit, although the history of ten steps may not show it. The operator T° is the one-step operator of the same walk without its position. At the sites x with $|x|_\infty \geq 3$, rules (a) and (b) do not apply and $\kappa^{\mathsf{c}} = \kappa^{\circ}$, so that $\mathsf{T}(\vartheta, \omega, x; \vartheta', \omega\mathsf{e}, x + \mathsf{e}) = \mathsf{T}^{\circ}(\vartheta, \omega; \vartheta', \omega\mathsf{e})$.

The main step of this appendix is Lemma C.4 below, which dominates the trails of length $n > 20$ by the walk. We next describe how such a trail starts. We let $\mathcal{T}_{\mathrm{pre}}$ denote

the set of trails $w' = (v'_0, \dots, v'_{20})$ of $\mathbb{Z}^2$ that visit neither c nor v, with $v'_0 = u$ and $v'_1 = a_0$. For $w' \in \mathcal{T}_{\mathrm{pre}}$, we let $\omega(w') \in \Omega_{10}$ denote the history formed by the last ten steps of w', and for a phase ϑ we define

$$\mathsf{P}_\vartheta(w') = \sum_{\vartheta_0 \in \{\mathsf{t}^-, \mathsf{s}^-\}} \sum_{\vartheta_1, \dots, \vartheta_{19}} \prod_{j=1}^{20} \varpi_j(\vartheta_{j-1}, \vartheta_j), \qquad \vartheta_{20} = \vartheta, \tag{C.13}$$

where ϖ_j is the weight (C.7) of step j of w'. For $n > 20$ and $w \in \mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}$, the prefix $w_{20} = (v_0, \dots, v_{20})$ belongs to $\mathcal{T}_{\mathrm{pre}}$, by the definition of $\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}$.

Lemma C.4. *For all $n > 20$,*

$$\mathsf{m}(\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}) \le \sum_{w' \in \mathcal{T}_{\mathrm{pre}}} \sum_{\vartheta} \mathsf{P}_\vartheta(w') \, (\mathsf{T}^{n-20}\mathsf{g})(\vartheta, \omega(w'), v'_{20}). \tag{C.14}$$

Proof. We split each trail after its step 20, and we then compare the later steps with the steps of the walk.

Splitting after step 20. Let $w = (v_0, \dots, v_n)$ be a trail of $\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}$, and let $w' = w_{20}$. The weights ϖ_j of the steps $j \le 20$ of w and of w' coincide, since they are determined by $(v_0, \dots, v_{20})$. Splitting each product in (C.8) after step 20, we obtain from Lemma C.2 that

$$\mathsf{m}(\{w\}) = \sum_{\vartheta} \mathsf{P}_\vartheta(w') \sum_{\vartheta_{21}, \dots, \vartheta_{n-1}} \sum_{\vartheta_n \in \{\mathsf{t}^+, \mathsf{s}^+\}} \prod_{j=21}^{n} \varpi_j(\vartheta_{j-1}, \vartheta_j), \qquad \vartheta_{20} = \vartheta. \tag{C.15}$$

Domination of one step. For $10 \le j \le n$, let $\omega_j \in \Omega_{10}$ denote the history formed by the last ten steps of $w_j = (v_0, \dots, v_j)$, so that $\omega_{20} = \omega(w')$ and $\omega_j = \omega_{j-1}\mathsf{e}$ with $\mathsf{e} = v_j - v_{j-1}$. We claim that for $21 \le j \le n$ and all phases ϑ and ϑ',

$$\varpi_j(\vartheta, \vartheta') \le \mathsf{T}(\vartheta, \omega_{j-1}, v_{j-1}; \vartheta', \omega_j, v_j). \tag{C.16}$$

The step e is admissible from ω_{j-1}, since the path of ω_{j-1} is a translate of $(v_{j-11}, \dots, v_{j-1})$ and w does not use a bond twice. Rule (a) does not apply, since w does not visit c and $v_{j-1} \ne v$.

Suppose first that $v_{j-1} \in \{u, a_0\}$. Then step j is a revisit, since w left u at step 1 and a_0 at step 2, and $\varpi_j(\vartheta, \vartheta')$ equals $1/5$ if $\vartheta' = \vartheta$ and 0 otherwise. Moreover, $\{v_{j-1}, v_j\} \ne \{u, a_0\}$, since w used the bond $\{u, a_0\}$ at step 1. By rule (b), the two sides of (C.16) are equal.

Suppose next that $v_{j-1} \notin \{u, a_0\}$, so that rule (c) applies, and let k be the kind of e as a step from ω_{j-1}.

We distinguish three cases. If $\mathsf{k} = \mathsf{r}$, then $v_{j-1} \in \{v_{j-11}, \dots, v_{j-2}\}$, and step j is a revisit. If $\mathsf{k} \in \{\mathsf{s}, \mathsf{t}\}$ and step j is a revisit, then $v_{j-1} \in \{v_0, \dots, v_{j-12}\}$. Otherwise, step j is a first visit. Its arrival direction $v_{j-1} - v_{j-2}$ is the last entry of ω_{j-1}, so that its kind

as a step of w is k. Writing $\mathsf{T}_j = \mathsf{T}(\vartheta, \omega_{j-1}, v_{j-1}; \vartheta', \omega_j, v_j)$, and using (C.7), rule (c), and (C.11), we find in the three cases that

$$\varpi_j(\vartheta, \vartheta') \begin{cases} = \mathsf{T}_j \in \{0, 1/5\} & \text{if } \mathrm{k} = \mathrm{r}, \\ = \mathbf{1}_{\vartheta'=\vartheta}/5 \le \varpi_\vartheta(\mathrm{k})\mathbf{1}_{\vartheta'=\vartheta} \le \mathsf{T}_j & \text{if } \mathrm{k} \in \{\mathrm{s}, \mathrm{t}\} \text{ and step } j \text{ is a revisit}, \\ \le \mathsf{T}_j & \text{otherwise.} \end{cases}$$

In the second case, we used that $\varpi_\vartheta(\mathrm{s}), \varpi_\vartheta(\mathrm{t}) \ge 1/3$. In the third case, the two sides are equal for $\vartheta' = \vartheta$ and for the transitions with weight zero. For the transitions from t^- to t^+ and from s^- to s^+, the inequality follows from (C.11), which completes the proof of (C.16).

Summation. Fix $w' \in \mathcal{T}_{\mathrm{pre}}$ and a phase ϑ. We consider the map that sends a trail $w \in \mathcal{T}_n^{\mathrm{t},\mathrm{e}_0}$ with $w_{20} = w'$, together with phases $\vartheta_{21}, \dots, \vartheta_n$ such that $\vartheta_n \in \{\mathrm{t}^+, \mathrm{s}^+\}$, to the sequence of states $(\vartheta_j, \omega_j, v_j)$, $20 \le j \le n$, where $\vartheta_{20} = \vartheta$. This map is injective, since the states determine $v_{20}, \dots, v_n$ and the phases. Every sequence in its image starts at $(\vartheta, \omega(w'), v'_{20})$ and ends at a state with reward 1. Therefore (C.16) implies that

$$\begin{aligned} &\sum_{\substack{w \in \mathcal{T}_n^{\mathrm{t},\mathrm{e}_0} \\ w_{20} = w'}} \sum_{\vartheta_{21}, \dots, \vartheta_{n-1}} \sum_{\vartheta_n \in \{\mathrm{t}^+, \mathrm{s}^+\}} \prod_{j=21}^{n} \varpi_j(\vartheta_{j-1}, \vartheta_j) \\ &\qquad \le \sum_{(\vartheta'_j, \omega'_j, x_j)_{21 \le j \le n}} \prod_{j=21}^{n} \mathsf{T}(\vartheta'_{j-1}, \omega'_{j-1}, x_{j-1}; \vartheta'_j, \omega'_j, x_j)\, \mathsf{g}(\vartheta'_n, \omega'_n, x_n) \\ &\qquad = (\mathsf{T}^{n-20}\mathsf{g})(\vartheta, \omega(w'), v'_{20}), \end{aligned} \tag{C.17}$$

where the sum in the second line runs over all sequences of states, and $(\vartheta'_{20}, \omega'_{20}, x_{20}) = (\vartheta, \omega(w'), v'_{20})$. The equality in the third line follows from (C.12) by induction on $n - 20$. Summing (C.15) over the trails w with $w_{20} = w'$, inserting (C.17), and summing over w', we obtain (C.14). □

Lemma C.5. *Let* F *be a nonnegative function on the states such that* $\mathsf{g} + \mathsf{TF} \le \mathsf{F}$. *Then for all* $K \ge 0$,

$$\sum_{k=0}^{K} \mathsf{T}^k \mathsf{g} \le \mathsf{F}. \tag{C.18}$$

Proof. We argue by induction on K. For $K = 0$, (C.18) reads $\mathsf{g} \le \mathsf{F}$, which holds since $\mathsf{TF} \ge 0$. Suppose that (C.18) holds for some $K \ge 0$. Since the entries of T are nonnegative, we obtain

$$\sum_{k=0}^{K+1} \mathsf{T}^k \mathsf{g} = \mathsf{g} + \mathsf{T} \sum_{k=0}^{K} \mathsf{T}^k \mathsf{g} \le \mathsf{g} + \mathsf{TF} \le \mathsf{F},$$

which completes the induction. □

Lemma C.5 is the standard comparison principle for supersolutions of a linear equation with a nonnegative matrix. The next lemma constructs a function as in Lemma C.5 from its values at the sites of a box and from a function of the phase and the history. Recall that $\mathcal{C}_R = \{x \in \mathbb{Z}^2 : |x|_\infty \le R\}$.

Lemma C.6. *Let $R \ge 1$. Let Z be a positive function of (ϑ, ω) such that*

$$\mathsf{D} = \mathsf{Z} - \mathsf{T}^\circ \mathsf{Z} \tag{C.19}$$

is positive. Let U be a nonnegative function on the states, and let $\epsilon_{\mathrm{out}} \ge 0$. Suppose that $\mathsf{U}(\vartheta, \omega, x) = 0$ if $x \notin \mathcal{C}_R$ or $x = c$, that $\mathsf{U}(\vartheta, \omega, v) = \mathsf{g}(\vartheta, \omega, v)$, and that for all phases ϑ and all $\omega \in \Omega_{10}$,

$$\begin{aligned} (\mathsf{TU})(\vartheta, \omega, x) &\le \mathsf{U}(\vartheta, \omega, x) \quad && \text{for } x \in \mathcal{C}_R \setminus \{c, v\}, \\ (\mathsf{TU})(\vartheta, \omega, x) &\le \epsilon_{\mathrm{out}} \mathsf{D}(\vartheta, \omega) && \text{for } x \notin \mathcal{C}_R. \end{aligned} \tag{C.20}$$

Then the function $\mathsf{F} = \mathsf{U} + \epsilon_{\mathrm{out}} \mathsf{Z}$, where Z is regarded as a function of the states that does not depend on the site, satisfies $\mathsf{g} + \mathsf{TF} \le \mathsf{F}$.

Proof. We begin by comparing T with T°. We claim that for all states (ϑ, ω, x), all admissible steps e from ω, and all phases ϑ',

$$\mathsf{T}(\vartheta, \omega, x; \vartheta', \omega\mathsf{e}, x + \mathsf{e}) \le \mathsf{T}^\circ(\vartheta, \omega; \vartheta', \omega\mathsf{e}). \tag{C.21}$$

The entries of rule (a) vanish. Rule (b) gives the entry $1/5$ only for $\vartheta' = \vartheta$, and $1/5 \le \varpi_\vartheta(\mathsf{k})$ for all kinds k by (C.6). Rule (c) differs from the definition of T° through the replacement of κ° by $\kappa^{\mathsf{c}} \le \kappa^\circ$. This proves (C.21), which implies that

$$(\mathsf{TZ})(\vartheta, \omega, x) \le (\mathsf{T}^\circ \mathsf{Z})(\vartheta, \omega) = \mathsf{Z}(\vartheta, \omega) - \mathsf{D}(\vartheta, \omega) \tag{C.22}$$

for all states. Note also that $c, v \in \mathcal{C}_R$, since $R \ge 1$, and that g vanishes at the states with $x \ne v$.

We now verify $\mathsf{g} + \mathsf{TF} \le \mathsf{F}$ at a state (ϑ, ω, x). Suppose first that $x \in \mathcal{C}_R \setminus \{c, v\}$. Then $\mathsf{g}(\vartheta, \omega, x) = 0$. Using (C.20) and (C.22), we obtain

$$(\mathsf{TF})(\vartheta, \omega, x) \le \mathsf{U}(\vartheta, \omega, x) + \epsilon_{\mathrm{out}}\big(\mathsf{Z}(\vartheta, \omega) - \mathsf{D}(\vartheta, \omega)\big) \le \mathsf{F}(\vartheta, \omega, x).$$

Next, suppose that $x \notin \mathcal{C}_R$. Then $\mathsf{g}(\vartheta, \omega, x) = 0$ and $\mathsf{U}(\vartheta, \omega, x) = 0$, and we have

$$(\mathsf{TF})(\vartheta, \omega, x) \le \epsilon_{\mathrm{out}} \mathsf{D}(\vartheta, \omega) + \epsilon_{\mathrm{out}}\big(\mathsf{Z}(\vartheta, \omega) - \mathsf{D}(\vartheta, \omega)\big) = \mathsf{F}(\vartheta, \omega, x).$$

Finally, suppose that $x \in \{c, v\}$. Then $(\mathsf{TF})(\vartheta, \omega, x) = 0$ by rule (a). Since $\mathsf{U} = \mathsf{g}$ at v and $\mathsf{g} = 0$ at c, we have

$$\mathsf{g}(\vartheta, \omega, x) \le \mathsf{U}(\vartheta, \omega, x) \le \mathsf{F}(\vartheta, \omega, x),$$

which completes the proof. □

The function TU vanishes at the sites outside $\mathcal{C}_R$ that are not adjacent to $\mathcal{C}_R$, since U vanishes outside $\mathcal{C}_R$. The second condition in (C.20) therefore concerns only the sites adjacent to $\mathcal{C}_R$, and (C.20) is a finite set of inequalities.

C.5 The certified bound

We now state the bound, with the numerical values of the parameters.

Proposition C.7. *Let* $(c_s, c_t) = (961/1000, 187/200)$, $\bar{\alpha} = 1/6$, $\bar{\beta} = 9/25$, *and* $h = (1, 5/4)$, *and let* ς_w *be the enumerations of Definition C.1. Then the entries of Table C.1 are upper bounds for the quantities in its first column. In particular,*

$$K_{\mathrm{coll}}(1) \le 0.923198 \qquad \textit{and} \qquad K_{\mathrm{diag}}(1) \le 0.899348, \tag{C.23}$$

and $K_t(1) \le 0.92320$ *for* $t \in \{\mathrm{coll}, \mathrm{diag}\}$.

	$t = \mathrm{coll}$	$t = \mathrm{diag}$
$\sum_{n \le 20} \mathsf{m}_t(n)$	0.582353	0.823644
$\sum_{n > 20} \mathsf{m}(\mathcal{T}_n^{t,\mathsf{N}})$	0.111332	0.078222
$\sum_{n > 20} \mathsf{m}(\mathcal{T}_n^{t,\mathsf{W}})$	0.118181	0.114607
$\sum_{n > 20} \mathsf{m}(\mathcal{T}_n^{t,\mathsf{S}})$	0.111332	0.107712
$\sum_{n \ge 1} \mathsf{m}_t(n) = h_t K_t(1)$	0.923198	1.124185
$K_t(1)$	0.923198	0.899348

Table C.1. Certified upper bounds at $(c_s, c_t) = (961/1000, 187/200)$, $\bar{\alpha} = 1/6$, $\bar{\beta} = 9/25$, and $h = (1, 5/4)$, with the enumerations of Definition C.1. The entries of the first four rows are rounded up in the last digit, and the last two rows are computed from the rounded entries. For the collinear type, the rows for the first directions N and S are equal by symmetry.

Proof. We treat the trails of length at most 20 and the longer trails separately.

Trails of length at most 20. The program `anc/h3/tailmem5.py` enumerates the trails of $\mathcal{T}_n^t$ for $n \le 20$ by depth-first search. It sorts the first-visit steps by binary64 approximations to the quotients of Definition C.1, and uses upward rounding to evaluate (C.1) in the resulting order. For each trail, the exact value in this order is at least the minimum attained by the order of Definition C.1. Thus the evaluation yields a valid upper bound even if rounding changes the order. Summing the bounds with upward rounding gives the first row of Table C.1.

Longer trails. We treat the five cases $(\mathrm{coll}, \mathsf{N})$, $(\mathrm{coll}, \mathsf{W})$, $(\mathrm{diag}, \mathsf{N})$, $(\mathrm{diag}, \mathsf{W})$, and $(\mathrm{diag}, \mathsf{S})$ for (t, e_0). In each recorded run, the Bellman program constructed a positive function Z of (ϑ, ω) such that $\mathsf{D} > 0$, a function U on the states with the vanishing and boundary properties of Lemma C.6 for $R = 70$, and a constant ϵ_{out}. It then verified the inequalities (C.20) as described in Appendix C.6. By Lemma C.6, the function $\mathsf{F} = \mathsf{U} + \epsilon_{\mathrm{out}}\mathsf{Z}$ satisfies $\mathsf{g} + \mathsf{T}\mathsf{F} \le \mathsf{F}$.

Let $K \geq 1$. Summing (C.14) over $20 < n \leq 20 + K$ and using Lemma C.5, we find that

$$\begin{aligned}\sum_{n=21}^{20+K} \mathsf{m}(\mathcal{T}_n^{\mathsf{t},\mathsf{e}_0}) &\leq \sum_{w' \in \mathcal{T}_{\mathrm{pre}}} \sum_{\vartheta} \mathsf{P}_\vartheta(w') \sum_{k=1}^{K} (\mathsf{T}^k \mathsf{g})(\vartheta, \omega(w'), v'_{20}) \\ &\leq \sum_{w' \in \mathcal{T}_{\mathrm{pre}}} \sum_{\vartheta} \mathsf{P}_\vartheta(w')\, \mathsf{F}(\vartheta, \omega(w'), v'_{20}).\end{aligned} \tag{C.24}$$

Letting $K \to \infty$, we obtain the same bound for the sum over all $n > 20$. The recorded runs evaluated the right side of (C.24) with upward rounding. Rounding each reported bound up to six decimal places gives the corresponding entry of Table C.1. In the five runs, we had $\epsilon_{\mathrm{out}} \leq 4.6 \cdot 10^{-6}$, and the term with $\epsilon_{\mathrm{out}}\mathsf{Z}$ contributed less than $9.2 \cdot 10^{-4}$ to the right side of (C.24).

Symmetry. For $\mathsf{t} = \mathrm{coll}$, the reflection $(x_1, x_2) \mapsto (x_1, -x_2)$ fixes c, u, and v. It maps $\mathcal{T}_n^{\mathrm{coll},\mathsf{N}}$ bijectively onto $\mathcal{T}_n^{\mathrm{coll},\mathsf{S}}$. Further, it preserves the arrival site c of u, the kinds of the steps, the types of the pairs, and the dead pairs. It therefore preserves the coefficient sums $\kappa(j)$ and the enumerations of Definition C.1. Thus $\mathsf{m}(\mathcal{T}_n^{\mathrm{coll},\mathsf{N}}) = \mathsf{m}(\mathcal{T}_n^{\mathrm{coll},\mathsf{S}})$ for all n, and the fourth row of the first column equals the second.

Conclusion. By (C.3), the fifth row is the sum of the first four. By (C.2), the last row is the fifth divided by h_{t}, which gives (C.23). The two bounds in (C.23) are smaller than 0.92320. This completes the proof. □

C.6 What is computed, and what must be trusted

Programs. The program `anc/h3/bellman.py` treats one case $(\mathsf{t}, \mathsf{e}_0)$. It generates the histories of Ω_{10} and their admissible steps from Definition 8.3. It computes an approximation of $\sum_{k \geq 0} (\mathsf{T}^\circ)^k \mathbf{1}$, which serves as Z, and a lower bound for the slack $\mathsf{D} = \mathsf{Z} - \mathsf{T}^\circ \mathsf{Z}$. In all five recorded runs the minimum of this lower bound was $0.999\ldots$. It enumerates the trails of $\mathcal{T}_{\mathrm{pre}}$ by depth-first search (between $2.3 \cdot 10^8$ and $3.6 \cdot 10^8$ of them in the recorded runs, depending on the case) and accumulates the weights $\mathsf{P}_\vartheta(w')$ by phase, history, and endpoint. It then initializes U with Z rounded up to binary32 at the states with position in $\mathcal{C}_{70} \setminus \{c, v\}$, with zero at c, and with g at v. It proposes smaller values by sweeping the states with position in $\mathcal{C}_{70} \setminus \{c, v\}$, taking the minimum of the currently stored value and the rounded value of TU at each state. The current program uses atomic accesses to the shared entries of U during the parallel sweeps.

The proof uses the final verification pass in one thread, which recomputes every row of the first inequality of (C.20) from the stored U. For each state at a site just outside the box and adjacent to it, the program divides an upper bound for TU by a positive lower bound for D. It takes ϵ_{out} to be the maximum of these upward-rounded quotients, which ensures the second inequality of (C.20). Finally, it bounds the right side of (C.24) with upward rounding. The recorded runs took about ten minutes on four cores and used 1.3

gigabytes of memory per case. The command lines are listed in `anc/h3/jobs.yaml`, and the historical outputs used for Table C.1 are stored in `anc/h3/*.log`. In these files, the collinear type is labeled B and the diagonal type C, the first directions N, W, S are labeled 1, 2, 3, and the phases t^-, s^-, t^+, s^+ are labeled A, B, C, D. A verification run must report zero violated rows and end with `PASS`.

The program does not store U for each history. It merges the histories into 3,744 classes, such that the histories of a class have the same last entry and the same admissible steps with the same kinds, and such that the successors by the same step of two histories of a class lie in the same class. By Definition C.3, the entries of T and T° depend on the history only through these data. Hence a function of the class that satisfies the inequalities of Lemma C.6 on classes satisfies them on histories. The program checks the defining property of the classes for every history.

Arithmetic. Upward rounding gives upper bounds for the sums and products of nonnegative numbers used for the prefix weights, the operator rows, and the final bound [19, 45]. To bound D from below, the Bellman program subtracts the upward-rounded value of $T^\circ Z$ from Z with downward rounding. In the two proof programs, `tailmem5.py` and `bellman.py`, the decimal inputs for c_s and c_t are moved down to the adjacent binary64 numbers after conversion, and those for $\bar{\alpha}$ and $\bar{\beta}$ are moved up. The coefficient inputs are 0.16666666666666667 and 0.36, so that the binary64 numbers used are at least $1/6$ and $9/25$. The type weights 1 and $5/4$ are exact binary64 numbers. The optional comparison program `exact75.py` applies a relative margin to the converted thresholds and coefficients. The function U is stored in binary32. In the final pass, each row TU is evaluated in binary64 with upward rounding and compared with the stored value of U, so that an accepted row holds for the exact coefficients. The C library calls that set the rounding direction are checked for success. See Appendix E.5 for the compiler settings and the assumptions about directed rounding.

Checks. We also made the following empirical comparisons.

1. We independently enumerated the short trails in `exact75.py`, starting from (7.4). For every length $n \le 20$ with a nonzero sum, the sums from `exact75.py` and `tailmem5.py` differ by at most $9 \cdot 10^{-13}$ in relative terms.
2. For $n = 22$, the right side of (C.14), evaluated by `anc/h3/bellman.py`, agrees to a relative error of at most $3 \cdot 10^{-6}$ with the value computed by a separately written dynamic program in `anc/h3/tailmem5.py`. This second walk runs on a periodic quotient of $\mathbb{Z}^2$ and discards steps from u or a_0 when the history already records a visit to that site.
3. The same quotient walk gives values for $22 \le n \le 108$ in the recorded runs. Over these lengths the quotient can count paths that reach a periodic translate of v. The sums of the recorded values in the five cases lie between 94.8% and 95.1% of the

corresponding entries of Table C.1. The Bellman bound in (C.24) applies to the walk on $\mathbb{Z}^2$.

What must be trusted. Granting the lemmas of this appendix, the bounds rely on `anc/h3/tailmem5.py` enumerating every short trail and bounding its weight, on `anc/h3/bellman.py` implementing the prefix weights (C.13) and the entries of T and T° of Definition C.3, and on the compiler and floating-point environment implementing binary64 and binary32 arithmetic with the directed rounding specified in [19]. The comparisons above provide empirical checks of the implementations.

APPENDIX D

Explicit constants

In this appendix, we make the constants in the proof of Theorem 1.1 explicit, and we derive the values $L_0 = 2\lceil \exp(10^{40}) \rceil + 5$ and $c = \exp[-\exp(\exp(10^{41}))]$ of (1.2). We follow the steps of the proof in Section 8.4. The constant c is dominated by the dimension of the quantum patch of side b, so that coarse estimates suffice in the earlier steps. In Appendix D.5, we list the constants of the paper and where they come from. The numerical inequalities below involve only rational numbers and elementary functions, and each of them can be checked by hand.

D.1 The constants of Hypothesis (P)

We first make the constants of Lemma 7.3 and Theorem 7.4 explicit.

Lemma D.1. *Hypothesis (P) holds with*

$$A = 10^{13}, \qquad \theta = 1 - 10^{-20}. \tag{D.1}$$

Proof. We follow the proofs of Lemma 7.3 and Theorem 7.4, keeping track of the constants.

A bound for the masses. Theorem 8.1 shows that Hypotheses (H1), (H2), and (H3) hold at $(c_s, c_t) = (961/1000, 187/200)$, with $\gamma_o = 2499/2500$, $C_o = 10$, and $K_t(1) \le 0.92320$ for both types. Recall the bound (7.7) from the proof of Lemma 7.3, which states that for all $n \ge 1$,

$$\frac{\mathsf{m}_t(n)}{h_t} \le 2nC_o\gamma_o^n. \tag{D.2}$$

Continuity in t. Let $\theta \in [0.9999, 1)$. The mean value theorem shows that $\theta^{-n-1} - 1 \le (n+1)(1-\theta)\theta^{-n-2}$ for $n \ge 1$. Inserting this and (D.2) into (7.5), we obtain

$$K_t(\theta) - K_t(1) = \frac{1}{h_t} \sum_{n\ge 1} \mathsf{m}_t(n)(\theta^{-n-1} - 1) \le 2C_o(1-\theta)\theta^{-2} \sum_{n\ge 1} n(n+1)\Big(\frac{\gamma_o}{\theta}\Big)^n. \tag{D.3}$$

Recall that $\sum_{n\ge 1} n(n+1)r^n = 2r/(1-r)^3$ for $0 < r < 1$. Since $\gamma_o/\theta \le 0.9996/0.9999 = 3332/3333$, we have

$$\sum_{n\ge 1} n(n+1)\Big(\frac{\gamma_o}{\theta}\Big)^n \le 2 \cdot 3333^3 < 10^{11}.$$

Using also $\theta^{-2} < 2$ and $C_o = 10$, we deduce from (D.3) that $K_t(\theta) - K_t(1) \le 4 \cdot 10^{12}(1-\theta)$. For $\theta = 1 - 10^{-20}$, this gives

$$K_\theta = \max_t K_t(\theta) \le 0.92320 + 4 \cdot 10^{-8} < 0.9233. \tag{D.4}$$

The constant A. Recall that $C_{pin} = 2C_o/(1-\gamma_o)$. The proof of Theorem 7.4 uses only that $\theta \in (\gamma_o, 1)$ and $K_\theta < 1$. For every such θ, it gives Hypothesis (P) with this θ and

$$A = \frac{h_{max}}{h_{min}} \max\Big(2, \frac{C_{pin}}{1-K_\theta}\Big). \tag{D.5}$$

Now let $\theta = 1 - 10^{-20}$, which is larger than γ_o. We have $C_{pin} = 5 \cdot 10^4$, and $1 - K_\theta > 0.0767$ by (D.4). Therefore the right side of (D.5) satisfies

$$\frac{h_{max}}{h_{min}} \max\Big(2, \frac{C_{pin}}{1-K_\theta}\Big) \le \frac{5}{4} \cdot \frac{5 \cdot 10^4}{0.0767} < 10^6.$$

Hypothesis (P) with a constant A implies (P) with every larger constant, so that (P) holds with the values (D.1). □

D.2 The window and block estimates

Throughout the rest of this appendix, we set

$$M = \lceil \exp(10^{40}) \rceil, \qquad b = 2M + 1, \tag{D.6}$$

and $\ell = M^{1/4}$, as in Section 4. We now check that this M satisfies the requirements of Theorem 5.9 and Proposition 4.5, with A and θ as in (D.1). Recall the threshold $M_g = 73^4$ of Lemma 4.3, which is also used in Section 5.5.

Lemma D.2. *Let A and θ be as in* (D.1)*. Then Hypothesis (W) holds with*

$$M_\star = 10^{180}, \qquad p_0 = \exp(-2 \cdot 10^{36}). \tag{D.7}$$

Proof. The proof of Theorem 5.9 shows that (W) holds with $p_0 = e^{-2C_{cut}}$, where C_{cut} is given by (5.31), and with $M_\star = \max(M_g, M_1)$, for every M_1 such that (5.32) holds for all $M \ge M_1$. We first show that we may take $M_1 = 10^{180}$.

We claim that

$$(2M+1)^2 \theta^{\ell/12 - 7} \le 1 \qquad \text{for all } M \ge (256 \cdot 10^{20})^8. \tag{D.8}$$

Using $2M + 1 \le 3\ell^4$, $\theta^{-7} < 8$, and $\theta^{\ell/12} \le \exp(-10^{-20}\ell/12)$, we obtain

$$(2M+1)^2 \theta^{\ell/12-7} \le 72\, \ell^8 \exp(-10^{-20}\ell/12).$$

For $\ell \geq (256 \cdot 10^{20})^2$, we have $\sqrt{\ell} \geq 256 \cdot 10^{20}$, and therefore

$$\log 72 + 8 \log \ell \leq 9\sqrt{\ell} < \frac{256}{12}\sqrt{\ell} \leq \frac{10^{-20}\ell}{12},$$

which proves (D.8). Since $(256 \cdot 10^{20})^8 = 2^{64} \cdot 10^{160} < 1.9 \cdot 10^{179}$, it follows that (5.32) holds for all $M \geq 10^{180}$. The threshold $M_{\mathrm{g}} = 73^4$ is smaller than 10^{180}, and we may therefore take $M_\star = 10^{180}$.

Next, inserting $A = 10^{13}$, $1/(1-\theta) = 10^{20}$, and $C_\star = \log 8192 < 10$ into (5.31), we obtain

$$C_{\mathrm{cut}} < 10 + 480 \cdot 10^{13}(7 + 10^{20}) + 10^{14} < 10^{36}. \tag{D.9}$$

This gives $p_0 = e^{-2C_{\mathrm{cut}}} > \exp(-2 \cdot 10^{36})$. Hypothesis (W) with a constant p_0 implies (W) with every smaller positive constant, which gives (D.7). □

The coupling. By the definition (4.2) of ε_K, we have $\varepsilon_{192} = (3072)^{-768}/2 \geq 2^{-8899}$, since $3072 < 2^{11.585}$. With p_0 as in (D.7), the success probability of one coupling trial satisfies

$$\delta = p_0 \varepsilon_{192} \geq \exp(-2 \cdot 10^{36})\, 2^{-8899} \geq \exp(-5 \cdot 10^{36}). \tag{D.10}$$

For M as in (D.6), we have $M^{1/8} \geq \exp(10^{40}/8)$ and $J = \lfloor M^{1/8} \rfloor$, so that $M^{1/8}/2 \leq J$ and $J + 5 \leq 2M^{1/8}$. The first term of Δ_M in (4.6) therefore satisfies

$$9M^{1/2}(J+5)^2 \leq 36M^{3/4} \leq \frac{M}{8},$$

where we used that $M^{1/4} \geq 288$. For the second term, we use (D.10) and $\log M \leq 10^{40} + 1$ to find that

$$\delta J \geq \frac{1}{2}\exp\Big(\frac{10^{40}}{8} - 5 \cdot 10^{36}\Big) \geq \frac{1}{2}\exp(10^{39}) \geq 4 \log M.$$

Using $b \leq 3M$, we deduce that

$$b^2(1-\delta)^J \leq b^2 e^{-\delta J} \leq b^2 M^{-4} \leq 9M^{-2} \leq \frac{M}{8}.$$

Combining the two bounds, we conclude that

$$\Delta_M \leq \frac{M}{4} < \frac{b}{8}. \tag{D.11}$$

The block dynamics. The value (D.6) satisfies $M \geq 10^{180} = \max(M_\star, M_{\mathrm{g}})$, since $\log M \geq 10^{40} > 180 \log 10$. Therefore Proposition 4.5 applies, and its bound (4.10) and (D.11) imply that

$$\mathrm{gap}(\mathcal{L}_{\mathrm{bl}}) \geq 1 - \frac{4\Delta_M}{b} \geq \frac{1}{2} \qquad \text{for all } L, N \geq b + 4. \tag{D.12}$$

D.3 The patch gap

Recall from Section 3 that the patch operator h_b is the sum of the projections P_e over the bonds e of $\mathbb{Z}^2$ that meet $\{1,\dots,b\}^2$, on the space of the patch, and that g_b is the smallest positive eigenvalue of h_b. We let

$$n_b = b^2 + 4b, \qquad m_b = 2b^2 + 2b, \qquad \dim_b = 5^{n_b} \tag{D.13}$$

denote the number of sites of the patch, the number of bonds that meet $\{1,\dots,b\}^2$, and the dimension of the space on which h_b acts, respectively. Let $\varphi_k = z_1^{4-k} z_2^k$ for $0 \le k \le 4$, so that $(\varphi_k)_{0\le k\le 4}$ is the monomial basis of V.

Lemma D.3. *For every integer $b \ge 1$, we have*

$$g_b \ge [140b(b+1)]^{-5^{b^2+4b}}. \tag{D.14}$$

Proof. We show that $70h_b$ is an integer matrix in a suitable basis, and we then use its characteristic polynomial.

The monomial basis. By the definition (3.2) of $\langle\cdot,\cdot\rangle_V$, we have

$$\langle \varphi_k, \varphi_l\rangle_V = \mathbf{1}_{k=l}\binom{4}{k}^{-1}, \qquad 0 \le k, l \le 4. \tag{D.15}$$

Indeed, the monomials are orthogonal, since a rotation about the third axis multiplies φ_k evaluated at a coherent state by a phase that depends on k. Further, if a is uniformly distributed on $\mathbb{S}^2$, then $|\zeta_1(a)|^2 = (1+a_3)/2$ is uniformly distributed on $[0,1]$, and $5\int_0^1 s^{4-k}(1-s)^k\,\mathrm{d}s = \binom{4}{k}^{-1}$.

The bond projection. For $0 \le j \le 8$, let

$$\psi_j = \sum_{k+l=j} \binom{4}{k}\binom{4}{l}\,\varphi_k \otimes \varphi_l \in V \otimes V.$$

The vector $\psi_0 = \varphi_0 \otimes \varphi_0$ is an extremal-weight vector in the subspace of total spin 4, and the operator $S^- = z_{x,2}\partial_{z_{x,1}} + z_{y,2}\partial_{z_{y,1}}$ of the bond $\{x,y\}$ satisfies $(S^-)^j\psi_0 = j!\,\psi_j$. The subspace of total spin 4 has dimension 9. It is invariant under S^- (the action of $\mathrm{SU}(2)$ by linear substitutions extends to $\mathrm{SL}(2,\mathbb{C})$, and S^- is the derivative of the substitutions $z \mapsto (z_1 + tz_2, z_2)$ at $t = 0$). The vectors $\psi_0,\dots,\psi_8$ are nonzero and have distinct weights under diagonal matrices, and they span this subspace. Using (D.15) and Vandermonde's identity, we find

$$\|\psi_j\|^2 = \sum_{k+l=j}\binom{4}{k}\binom{4}{l} = \binom{8}{j}, \qquad \langle \psi_{k+l}, \varphi_k \otimes \varphi_l\rangle = 1.$$

The vectors ψ_j are pairwise orthogonal by (D.15), since they involve disjoint sets of the monomials $\varphi_k \otimes \varphi_l$. Therefore the projection P_e onto total spin 4 of the bond $e = \{x, y\}$ satisfies

$$P_e(\varphi_k \otimes \varphi_l) = \binom{8}{k+l}^{-1} \psi_{k+l} = \sum_{i+j=k+l} \frac{\binom{4}{i}\binom{4}{j}}{\binom{8}{k+l}} \varphi_i \otimes \varphi_j. \tag{D.16}$$

Every coefficient in (D.16) has a reduced denominator that divides 70, so that $70P_e$ has integer entries in the monomial basis.

The characteristic polynomial. The operator $70h_b$ is a sum of m_b terms $70P_e$, each tensored with the identity on the other sites. In particular, $70h_b$ has integer entries in the tensor monomial basis, and its characteristic polynomial has integer coefficients. The eigenvalues of h_b lie in $[0, m_b]$, since h_b is a sum of m_b orthogonal projections.

Let $r \geq 1$ be the rank of h_b, and let $t_1 \leq \cdots \leq t_r$ denote the positive eigenvalues of $70h_b$, counted with multiplicity, so that $t_1 = 70g_b$ and $t_r \leq 70m_b$. Since h_b is self-adjoint, we have

$$\det(\lambda - 70h_b) = \lambda^{\dim_b - r} \prod_{i=1}^{r} (\lambda - t_i),$$

and the coefficient of $\lambda^{\dim_b - r}$ is $(-1)^r t_1 \cdots t_r$. This coefficient is a nonzero integer, so that $t_1 \cdots t_r \geq 1$. We conclude that

$$g_b = \frac{t_1}{70} \geq \frac{1}{70\, t_2 \cdots t_r} \geq \frac{1}{70(70m_b)^{r-1}} \geq (70m_b)^{-r} \geq (70m_b)^{-\dim_b}.$$

This is (D.14), since $70m_b = 140b(b+1)$ and $\dim_b = 5^{b^2+4b}$. □

D.4 The constants of the main theorem

We now derive (1.2). Let M and b be as in (D.6). By (D.12), this value of M may be used in the proof of Theorem 1.1, and we have $L_0 = b + 4 = 2\lceil \exp(10^{40}) \rceil + 5$. Let $L, N \geq L_0$. Combining (8.11) with Lemma D.3, we obtain

$$\operatorname{gap}(H_{L,N}) \geq \frac{g_b}{4} \geq \frac{1}{4(70m_b)^{\dim_b}}. \tag{D.17}$$

We now simplify the right side. Since $M \leq \exp(10^{40}) + 1 \leq 2\exp(10^{40})$, we have $b \leq 5\exp(10^{40}) \leq \exp(2 \cdot 10^{40})$. Therefore

$$n_b \leq 2b^2 \leq \exp(5 \cdot 10^{40}), \qquad 70m_b \leq 280b^2 \leq \exp(6 \cdot 10^{40}),$$
$$\dim_b = \exp(n_b \log 5) \leq \exp(\exp(6 \cdot 10^{40})).$$

It follows that

$$\begin{aligned} \log\bigl(4(70m_b)^{\dim_b}\bigr) &\leq \log 4 + 6 \cdot 10^{40} \exp(\exp(6 \cdot 10^{40})) \\ &\leq \exp(\exp(7 \cdot 10^{40})) \leq \exp(\exp(10^{41})). \end{aligned}$$

For the second inequality, we took logarithms and used that $\log(7 \cdot 10^{40}) + \exp(6 \cdot 10^{40}) \leq \exp(7 \cdot 10^{40})$. Inserting this bound into (D.17), we conclude that

$$\operatorname{gap}(H_{L,N}) \geq \exp[-\exp(\exp(10^{41}))] = c \qquad \text{for all } L, N \geq L_0, \tag{D.18}$$

which is (1.2).

D.5 Constants and their provenance

The following table lists the constants of the paper, their values, and where they are defined or verified.

constant	value	provenance
ε_K	$(16K)^{-4K}/2$	overlap bound for sets of size at most K, (4.2) and Lemma 4.2
d_0	24	width of the ends of a shell, (4.3)
M_{g}	73^4	threshold of the shell and cut geometry, Lemma 4.3 and Section 5.5
$C_\star$	$\log 8192 < 10$	cost of the first removal, Lemma 5.5
3/5, 1/2, 1/3	exact	conditional means and fields of corners with one present bond, Lemmas 5.2 and B.3
116/125	exact	lower bound for the pinned expectation of I_{next} in the proof of Lemma 6.8
$\bar\alpha$, $\bar\beta$	1/6, 9/25	response coefficient bounds, Lemma 6.8
c_{s}, c_{t}	961/1000, 187/200	integer certificates for all 7,962 patterns, Theorem 8.5
γ_{o}, C_{o}	2499/2500, 10	integer vector on the histories of memory 6, Proposition 8.4
h	$(1, 5/4)$	type weights, Section 7
$K_\star$	0.92320	bounds for $K_{\mathrm{coll}}(1)$ and $K_{\mathrm{diag}}(1)$, Theorem 8.1 and Proposition C.7
R, ϵ_{out}	70, at most $4.6 \cdot 10^{-6}$	box radius and constant of Lemma C.6 in the proof of Proposition C.7
θ, A	$1 - 10^{-20}$, 10^{13}	constants of Hypothesis (P), (D.1)
K_θ	at most 0.9233	(D.4)
C_{pin}	$2C_{\mathrm{o}}/(1-\gamma_{\mathrm{o}}) = 5 \cdot 10^4$	Section 7
C_{cut}	less than 10^{36}	bound for the cut cost, (5.31) and (D.9)
p_0, $M_\star$	$\exp(-2 \cdot 10^{36})$, 10^{180}	constants of Hypothesis (W), (D.7)
δ	$p_0 \varepsilon_{192} \geq \exp(-5 \cdot 10^{36})$	success probability of one coupling trial, (D.10)
J, Δ_M	$\lfloor M^{1/8} \rfloor$, at most $M/4$	number of coupling trials and discrepancy bound, (4.6) and (D.11)
M, b	$\lceil \exp(10^{40}) \rceil$, $2M+1$	(D.6)
g_b	at least $[140b(b+1)]^{-5^{b^2+4b}}$	integer characteristic polynomial of the patch, Lemma D.3
L_0, c	$2\lceil \exp(10^{40}) \rceil + 5$, $\exp[-\exp(\exp(10^{41}))]$	Theorem 1.1, (D.18)

APPENDIX E

The ancillary files

In this appendix, we describe the ancillary files of this paper and explain how to run them. They contain the programs, the data, and the recorded outputs of the computations that prove Theorem 8.1, which states that the hypotheses (H1), (H2), and (H3) hold at $(c_s, c_t) = (961/1000, 187/200)$. The mathematics behind these computations is given in Section 8 for (H2), in Appendix B for (H1), and in Appendix C for (H3), together with an account of what each program must be trusted to do. We only describe here the files, the commands, and the computing environment. The driver of Appendix E.2 repeats the computations, fully or in part, and compares their outputs with the recorded ones.

E.1 Contents of the folder

The ancillary files are in the folder `anc` of the source archive of the paper. The folder has three subfolders `anc/h1`, `anc/h2`, and `anc/h3`, one for each hypothesis, and two files at the top. The file `anc/manifest.yaml` lists every file of the folder with its role, what it checks, the command that runs it, the expected output, the time and memory needed, and the SHA-256 digest of each file other than the manifest itself. The script `anc/run_all_checks.py` runs the checks; we describe it in Appendix E.2. The folder contains 34 files.

The archive may contain, besides the source of the paper, only data files of a few formats (among them JSON, YAML, and plain logs) and scripts in a few interpreted languages, such as Python. For this reason, we store each of the three programs in C used for (H3) verbatim, as a string, inside a Python script of the same name. When it is run, the script checks the SHA-256 digest of the stored source, writes the C file to a directory, compiles it with the options of Appendix E.5, and runs the compiled program with the arguments it was given. With the option `--print-source`, it prints the C source instead. For the same reason, we store all recorded outputs with the extension `.log`.

The scripts need Python 3.8 or later, and those in `anc/h1` also need the package NumPy. The C programs need the compiler GCC with support for OpenMP. We ran the checks described below on a computer with four x86-64 cores and 15 gigabytes of memory, under Linux, with Python 3.13, NumPy 2.5, and GCC 13.3. The scripts locate their bundled data relative to their own folders. We give the commands as they are typed in the folder that contains `anc`.

The recorded outputs come from the original runs. In the archived Python scripts we made file paths relative to the script folder and added optional arguments to some commands. We also added checks for successful changes of the rounding mode in the C programs and synchronized the accesses to the proposed function in the parallel sweeps of `bellman.c`. The transition weights, enumerations, and inequalities checked by these programs are the same as in the recorded runs. The driver rejects invalid hypothesis selections. The Python checkers that require assertions refuse to run with them disabled. We list the changes in `anc/manifest.yaml`.

E.2 The driver

To repeat the computations, it suffices to run the script `anc/run_all_checks.py`, which has two modes.

```
python3 anc/run_all_checks.py --quick
python3 anc/run_all_checks.py --full
```

It prints one line for each check, with the result and the time the check took. If at least one check runs and all checks pass, the last line is `ALL CHECKS PASS`. The option `--only` takes a nonempty comma-separated list of hypotheses from `h1`, `h2`, and `h3`, as in `--only h1,h3`. Using the option `--workdir`, we can keep the compiled programs and the outputs of all runs in a given directory.

The quick mode is the default. A historical run took about six minutes on one core and at most 0.6 gigabytes of memory. This mode runs the exact check of (H2) and the checks of Appendix E.3, except that it verifies only 80 of the 7,962 certificates. Further, it compiles the three C programs, runs the independent enumeration of the short trails in the five cases of Table C.1, and runs the two programs for (H3) with small parameters, which we describe at the end of Appendix E.5. Finally, it recomputes Table C.1 from the recorded outputs.

The full mode also verifies all 7,962 certificates, in four processes by default (option `--parts`). The recorded one-process verification took about 23 minutes. The full mode also repeats the ten recorded runs of the two programs for (H3), which originally took about 70 minutes in total and up to 3.1 gigabytes of memory. In both modes, the driver leaves the files in `anc` unchanged. The magnitude guards in `anc/h1/cert.py` and the certificate checks in `anc/h2/h2_check.py` use assertions. The driver, `anc/h1/verify_all.py`, and the standalone (H2) checker require assertions to be enabled and reject `python -O`.

E.3 The local inequalities

We now turn to the folder `anc/h1`, which contains the data and programs for (H1); we refer to Appendix B.4 for their description. The file `patterns.json` lists the 7,962 patterns

of $\mathcal{P}_{\mathrm{ext}}$, up to the reflection and the duplicates of Appendix B.4. For each pattern, it records an index, the out-neighbor o, the present bonds, and the status of every ring site that is an end of a present bond. Coordinates are taken relative to the passage site, and the trail arrives from $\mathsf{S} = (0, -1)$, so that a straight pattern has $o = \mathsf{N} = (0, 1)$ and a turn has $o = \mathsf{E} = (1, 0)$. The statuses 0 and 1 are written `u` and `d`. The file `certificates.jsonl` contains one certificate for each pattern, on one line, in the format of Appendix B.4. The program `verify_all.py` checks the certificates. It uses the module `ring.py`, which reads the patterns, and the module `cert.py`, which carries out the integer computation of Appendix B.3. The program `run_prod.py` produced the certificates; it uses floating-point arithmetic only to choose them and is not needed to check them.

To check (H1), we run the following commands, where $k \in \{0, 1, 2, 3\}$ in the third line.

```
python3 anc/h1/check_certificate_file.py
python3 anc/h1/verify_all.py
python3 anc/h1/verify_all.py anc/h1/certificates.jsonl k 4
python3 anc/h1/enumerate_patterns.py
```

The first command checks the file of certificates against the list of patterns. It verifies that every pattern has exactly one certificate, that every certificate is stated for the ring functional of Appendix B (field `model` with the value `covered`) and for the tip radius $3/5$, which is also the value in `ring.py`, and that the certificate of every pattern π carries the threshold c_π, with $c_\pi = 961/1000$ for the straight patterns and $c_\pi = 187/200$ for the turns. These properties of the file are part of the proof, since `verify_all.py` checks each certificate on its own, at the threshold written in it. The command takes less than a second and ends with `CERTIFICATE FILE CHECK PASS`.

The second command is the main check, which verifies all certificates. In the recorded run, it printed `certificates 7962 verified 7962 failed []` after 1,354 seconds in one process, and this output is the file `verify_all_output.log`. The third command checks the certificates in positions k, $k + 4$, $k + 8$, and so on, so that four such commands share the work. The last command runs the independent enumeration of the patterns described in Appendix B.4. It takes about 80 seconds and reports 106,101 assignments before the choice of an out-neighbor and 7,962 classes (2,696 straight patterns and 5,266 turns), which agree with `patterns.json` class by class.

The script `closed_exact.py` contains a test in rational arithmetic for the 4,735 patterns whose ring polynomial factors once the pendant sites are removed. It is not part of the proof, since all certificates in the file are operator certificates. However, when it is run as a script, it checks these patterns a second time, in about two seconds.

E.4 The trail growth

The folder `anc/h2` contains the script `h2_check.py`, the vector `v` of Section 8.1 in the file `h2_vector.json`, and the recorded output `h2_output.log`. The check of Proposition 8.4 is the command

```
python3 anc/h2/h2_check.py
```

which took less than a second in the recorded run. It checks the inequalities in integer and rational arithmetic, then prints rounded decimal diagnostics. The driver compares this output with `h2_output.log`, whose last line reads `H2 CHECK PASS: q = 2499/2500, C_V = 10`. By default the script reads `anc/h2/h2_vector.json` and leaves it unchanged. With `--make`, it first computes a new vector by power iteration in floating-point arithmetic and writes it to that file before checking it. The option `--vector FILE` selects a different file for reading and for writing when `--make` is supplied.

E.5 The weighted trail sums

The folder `anc/h3` contains three programs. The script `tailmem5.py` contains the program `tailmem5.c`, which enumerates the trails of length at most 20 exactly and bounds their sums by rounding upward. The script `bellman.py` contains `bellman.c`, which proves the bounds for the longer trails, one case (t, e_0) at a time. The script `exact75.py` contains `exact75.c`, the independent enumeration of the short trails used in check (1) of Appendix C.6. The scripts compile the programs with the commands

```
gcc -O2 -frounding-math -ffp-contract=off tailmem5.c -lm -o tailmem5
gcc -O2 -fopenmp -frounding-math -ffp-contract=off bellman.c -lm -o bellman
gcc -O2 -frounding-math -ffp-contract=off exact75.c -lm -o exact75
```

The programs use the C library function `fesetround` to set the rounding mode and check that each call succeeds. The certification assumes that the compiler and the floating-point environment honor the selected rounding mode [19]. With GCC we use `-frounding-math` to request this behavior. GCC documents the option as experimental.[1] The option `-ffp-contract=off` disables contraction of a product and a sum into a single operation.

The file `jobs.yaml` lists the command lines of the fifteen recorded runs and of the short runs of the quick mode. For example, the recorded runs for the collinear type and the first direction `N` are the following.

```
python3 anc/h3/tailmem5.py B 10 20 28 109 109 22 0.961 0.935 \
```

[1] `https://gcc.gnu.org/onlinedocs/gcc/Optimize-Options.html`

```
        0.16666666666666667 0.36 1 1.25 1 > short_trails_B_1.log 2>&1
python3 anc/h3/bellman.py B 1 70 600 > bellman_B_1.log
python3 anc/h3/exact75.py B 20 0.961 0.935 0.16666666666666667 0.36 \
        1 1.25 1 20 > exact75_B_1.log
```

We store the recorded outputs in the files `short_trails_X_d.log`, `bellman_X_d.log`, and `exact75_X_d.log`. Here `X` $=$ `B` stands for the collinear type coll and `X` $=$ `C` for the diagonal type diag, and `d` $= 1, 2, 3$ stands for the first direction $\mathsf{e}_0 = \mathsf{N}, \mathsf{W}, \mathsf{S}$. There are five cases (the first direction E leads from u into c, and the case $(\mathrm{coll}, \mathsf{S})$ follows from $(\mathrm{coll}, \mathsf{N})$ by the symmetry used in the proof of Proposition C.7). The labels of the phases in the outputs are given in Appendix C.6. Recall that a run of `bellman.py` is valid only if it reports `violated rows = 0` and ends with `PASS`.

We now explain how the entries of Table C.1 are read off from these files. Each line `E n value` of `short_trails_X_d.log` reports an upward-rounded upper bound for the sum over trails of length n. The line `# EXACT` reports the aggregate bound rounded separately. The driver computes the first row by adding the decimals from the `E` lines as exact rationals. For coll, it counts the lines of `short_trails_B_1.log` twice and those of `short_trails_B_2.log` once. For diag, it counts the lines of the three files `short_trails_C_d.log` once each. The entry of the case $(\mathsf{t}, \mathsf{e}_0)$ in the second, third, and fourth rows comes from the bound in the last line of `bellman_X_d.log`, with reflection for $(\mathrm{coll}, \mathsf{S})$. The driver rounds each result upward to six decimal places. The printed decimals bound the binary64 numbers from above because the GNU C library converts to decimals in the current rounding direction. For instance, the `# tailmem5` line prints the binary64 upper bound on $\bar{\beta}$ as 0.360001. The lines `D n value`, for $22 \le n \le 108$, come from the periodic-quotient calculation described in Appendix C.6. In the empirical comparison for the five cases, the ratio of their sum to the corresponding entry of the table ranged from 0.948 to 0.951.

With the recorded arguments, `tailmem5.c` needs 3.1 gigabytes of memory, almost all of it for the computation of the lines `D`, and about three minutes on one core. The program `bellman.c` needs about 1.3 gigabytes and about ten minutes on four cores for each case; these are the times of the recorded runs. The program `exact75.c` needs a few seconds. In earlier checks with these compiler options, the outputs of `exact75.py` in the five cases and of `tailmem5.py` in the case $(\mathrm{coll}, \mathsf{W})$ agreed with the recorded files byte for byte. The archive preserves the outputs of the original five runs of `bellman.py` with $R = 70$. The original parallel sweeps accessed the shared array for U without synchronization. The current code synchronizes those accesses. The sweeps propose U, and their results can depend on the order of updates. The final verification runs in one thread and checks the inequalities for the resulting U. The full mode of the driver requires a passing final verification, a bound at most the corresponding entry of Table C.1, and free-vector and prefix diagnostics that match the recorded run.

The quick mode runs the two programs with small parameters. It runs `tailmem5.py` with histories of 4 steps and a quotient of side 7, instead of 10 and 28. The depth-first search that computes the lines `E` is the same for all values of these two parameters. In earlier checks, these runs took less than 20 seconds for each case, and the lines `E` agreed with the recorded ones for all five cases. They also agreed with the last column of the output of `exact75.py`, up to a relative error of at most $8.3 \cdot 10^{-13}$. Next, the quick mode runs `bellman.py` with the box of radius $R = 23$ and without sweeps, so that `U` keeps its initial values in the box: the prescribed values at c and v, and `Z` stored in binary32 elsewhere. In earlier checks, each run took about 25 seconds and used 0.6 gigabytes of memory. These runs reproduced the number of prefixes and the slack of `Z` of the recorded runs and passed the final verification with bounds much larger than the table entries. With the last argument `1`, `bellman.py` also prints the value for $n = 22$ of check (2) of Appendix C.6. In the earlier runs this value agreed with the line `D 22` of the corresponding file `short_trails_X_d.log` up to a relative error between $1.7 \cdot 10^{-6}$ and $2.5 \cdot 10^{-6}$. This option requires $R \geq 23$.

APPENDIX F

Uniqueness of the infinite-volume frustration-free state

The coupling of Section 4.4 also controls the dependence of the spins in a fixed set on the boundary configuration of a growing box. We use this fact and the polynomial description of the ground space to prove uniqueness on the infinite square lattice.

For a finite set $A \subset \mathbb{Z}^2$, let $\mathcal{H}_A = \bigotimes_{x \in A} V_x$, with the spin-2 spaces of Section 3.1. We regard an operator on $\mathcal{H}_A$ as an operator on $\mathcal{H}_{A'}$ for $A \subset A'$ by tensoring it with the identity. An infinite-volume state is a state on the norm closure of the union of these local matrix algebras. We call the state ω *frustration-free* if $\omega(P_e) = 0$ for every bond e of $\mathbb{Z}^2$.

Theorem F.1. *The spin-2 AKLT model on $\mathbb{Z}^2$ has a unique infinite-volume frustration-free state.*

F.1 The central marginal of the classical law

Let $A \subset \mathbb{Z}^2$ be finite and nonempty, and set $R = \max_{x \in A} |x|_\infty$. Recall the box $\mathcal{C}_M$, the length $\ell = M^{1/4}$, and the number of trials $J = \lfloor M^{1/8} \rfloor$. By Theorems 5.9, 7.4 and 8.1, Hypothesis (W) holds, so we may use the coupling of Section 4.4 with $\delta = p_0 \varepsilon_{192} > 0$. For all sufficiently large M, we have $A \subset \mathcal{C}_M$ and $M + 1 - R > (J+5)\ell$, since $(J+5)\ell / M \to 0$.

Suppose that two boundary configurations of $\mathcal{C}_M$ differ at just one site y adjacent to $\mathcal{C}_M$. In the coupling, the probability that none of the trials succeeds is at most $(1-\delta)^J$, by (4.8). If some trial at i succeeds, then the final configurations agree on F_{i+1}, by the agreement argument in the proof of Proposition 4.4. Since $|y|_\infty = M + 1$, for all $x \in A$ we have

$$|x - y|_\infty \geq M + 1 - R > (J+5)\ell \geq i\ell,$$

where we used (4.7). Thus $A \subset F_{i+1}$, and the two marginals on A differ in total variation by at most $(1-\delta)^J$.

The law $\mu^{\eta}_{\mathcal{C}_M}$ depends on η only at the $4(2M+1)$ exterior sites adjacent to $\mathcal{C}_M$. For two arbitrary boundary configurations η and η', change these spins from η to η' one at a time. By the triangle inequality for total variation,

$$\sup_{\eta, \eta'} \left\| \mu^{\eta}_{\mathcal{C}_M, A} - \mu^{\eta'}_{\mathcal{C}_M, A} \right\|_{\mathrm{TV}} \leq 4(2M+1)(1-\delta)^J \longrightarrow 0 \qquad \text{as } M \to \infty. \tag{F.1}$$

Here the total variation distance of two probability measures is the supremum, over measurable events, of the absolute difference of their probabilities. For each such event, (F.1) bounds the variation of its probability over all boundary configurations. The same bound thus holds between any two mixtures of these marginals.

F.2 Coherent-state laws of open-box ground states

For a finite set A, let L_s be the evaluation functional $\varphi \mapsto \varphi(\zeta(s))$ on $\mathcal{H}_A$, for $s \in (\mathbb{S}^2)^A$, and let $E_A(s) = L_s^* L_s$. By (3.2) and the tensor product construction of the inner product,

$$\int E_A(s) \prod_{x \in A} 5 \, \mathrm{d}s_x = \mathrm{Id}_{\mathcal{H}_A}.$$

For a density matrix ρ on $\mathcal{H}_A$, the *coherent-state law* of ρ is the probability law with density $\mathrm{Tr}(\rho E_A(s))$ with respect to $\prod_{x \in A} 5 \, \mathrm{d}s_x$. In particular, for a unit vector φ, the coherent-state law of its rank-one projection has density $|\varphi(\zeta(s))|^2$. For $A \subset A'$, the operator $E_{A'}(s)$ is the tensor product of $E_A(s_A)$ and $E_{A' \setminus A}(s_{A' \setminus A})$. Given a density matrix on $\mathcal{H}_{A'}$, integrating out $s_{A' \setminus A}$ in its coherent-state law therefore gives the coherent-state law of its restriction to A.

The open box we use is $\mathcal{C}_{M+1}$. Let $\mathcal{F}_M$ be the set of all bonds with both sites in $\mathcal{C}_{M+1}$, and set $H_M^{\mathrm{open}} = \sum_{e \in \mathcal{F}_M} P_e$. If φ is a unit vector in its kernel, then $P_e \varphi = 0$ for every $e \in \mathcal{F}_M$ because the terms are nonnegative. The divisibility argument of Lemmas 3.1 and 3.2 applies to these bonds. Indeed, the proof of Lemma 3.1 only uses the tensor factors at the ends of a bond, and the polynomials q_e for distinct bonds are pairwise non-associated irreducibles, as in the proof of Lemma 3.2. Thus

$$\varphi = q_{\mathcal{F}_M} \chi,$$

where χ is a polynomial in the variables at sites of $\mathcal{C}_{M+1} \setminus \mathcal{C}_M$. To see the last assertion, every site of $\mathcal{C}_M$ belongs to four bonds of $\mathcal{F}_M$, so the product $q_{\mathcal{F}_M}$ uses all four degrees there.

By (3.6), the coherent-state density of this vector is proportional to

$$|\chi(\zeta(s))|^2 \prod_{\{x,y\} \in \mathcal{F}_M} (1 - s_x \cdot s_y).$$

Make the change of variables $n_x = (-1)^{x_1 + x_2} s_x$ on $\mathcal{C}_{M+1}$. Each bond factor then becomes $1 + n_x \cdot n_y$. Conditional on the spins of the outer layer $\mathcal{C}_{M+1} \setminus \mathcal{C}_M$, the factor from χ and all bond factors within that layer are constant. Every bond of $\mathbb{Z}^2$ that meets $\mathcal{C}_M$ belongs to $\mathcal{F}_M$. A version of the conditional law of $n_{\mathcal{C}_M}$ is therefore $\mu^{\eta}_{\mathcal{C}_M}$ of (3.17), where η agrees with the transformed outer-layer spins on the sites adjacent to $\mathcal{C}_M$ and is arbitrary on the

other exterior sites. Its density is obtained by normalizing the remaining bond factors, whose integral is positive for every η by Section 3.5.

It follows that the law of n_A for $A \subset \mathcal{C}_M$ is a mixture of $\mu^{\eta}_{\mathcal{C}_M,A}$. The same is true of any density matrix supported in $\ker H_M^{\mathrm{open}}$, by its decomposition into a convex combination of rank-one projections in the kernel. Thus (F.1) applies to any two such density matrices. Taking the inverse change of signs on A preserves total variation distance, so the bound applies also to the coherent-state laws of their restrictions to A.

F.3 Determination of the state

For each finite A, the coherent-state law determines the density matrix on $\mathcal{H}_A$. To see this, let B be the difference of two density matrices with the same coherent-state law. Then $\operatorname{Tr}(BE_A(s)) = 0$ almost everywhere. Choose the orthonormal basis of V formed by $\binom{4}{k}^{1/2} z_1^{4-k} z_2^k$ for $0 \le k \le 4$, using (D.15), and its tensor product on A. On the chart in which the first spinor component is nonzero at every site, use the representatives $z_x = (1, t_x)/\sqrt{1+|t_x|^2}$ with $t_x \in \mathbb{C}$. The choice of phases of these representatives does not change $E_A(s)$, since all polynomials have the same degree in each z_x. On this chart, $\operatorname{Tr}(BE_A(s))$ is continuous, and the measure $\prod_{x\in A} \mathrm{d}s_x$ has full support. Hence it vanishes throughout the chart. Multiplying by $\prod_{x\in A}(1+|t_x|^2)^4$ yields a polynomial in the monomials

$$\prod_{x\in A} t_x^{k_x} \overline{t_x}^{l_x}, \qquad 0 \le k_x, l_x \le 4.$$

Their coefficients are the entries of B in the tensor basis, each multiplied by a nonzero factor. These monomials are linearly independent as functions of $(t_x)_{x\in A}$, since the formal variables t_x and $\overline{t_x}$ are obtained by an invertible complex linear change of variables from $\operatorname{Re} t_x$ and $\operatorname{Im} t_x$. Thus $B = 0$.

Proof of Theorem F.1. Let ω be an infinite-volume frustration-free state. Its restriction to $\mathcal{C}_{M+1}$ has a density matrix ρ_M with $\operatorname{Tr}(\rho_M H_M^{\mathrm{open}}) = 0$. Since both operators in this trace are nonnegative, ρ_M is supported in $\ker H_M^{\mathrm{open}}$. Indeed, each eigenvector of ρ_M with positive eigenvalue has zero expectation of H_M^{open} and lies in its kernel.

If ω' is another such state, apply (F.1) and the argument of Appendix F.2 to the restrictions of ω and ω' to $\mathcal{C}_{M+1}$. For every fixed finite A, their coherent-state laws on A are independent of M as soon as $A \subset \mathcal{C}_M$. The bound in (F.1) tends to zero, so these two laws are equal. The density matrices on A are equal by the preceding argument. The local matrix algebras have dense union, and we conclude that $\omega = \omega'$.

For existence, the open-box kernel is nonzero for every M. For $x \in \mathcal{C}_{M+1}$, let d_x be the number of bonds of $\mathcal{F}_M$ at x. The nonzero polynomial $q_{\mathcal{F}_M} \prod_{x\in\mathcal{C}_{M+1}} z_{x,1}^{4-d_x}$ belongs to $\mathcal{H}_{\mathcal{C}_{M+1}}$ and lies in $\ker H_M^{\mathrm{open}}$ by Lemma 3.1. Normalize it and take its vector state. Along

a diagonal subsequence of growing boxes, the restrictions of these states to each fixed finite box converge, by compactness of the finite-dimensional sets of density matrices. The limits are consistent under restriction, so they give a state on the local algebras and hence on their norm closure. For each bond e, the finite-volume expectations of P_e are zero once both sites belong to the box. The limit state is therefore frustration-free. □

References

[1] Houssam Abdul-Rahman, Marius Lemm, Angelo Lucia, Bruno Nachtergaele, and Amanda Young. A class of two-dimensional AKLT models with a gap. In Houssam Abdul-Rahman, Robert Sims, and Amanda Young, editors, *Analytic Trends in Mathematical Physics*, volume 741 of *Contemporary Mathematics*, pages 1–21. American Mathematical Society, Providence, RI, 2020. doi: 10.1090/conm/741/14917.

[2] Ian Affleck, Tom Kennedy, Elliott H. Lieb, and Hal Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Physical Review Letters*, 59(7):799–802, 1987. doi: 10.1103/PhysRevLett.59.799.

[3] Ian Affleck, Tom Kennedy, Elliott H. Lieb, and Hal Tasaki. Valence bond ground states in isotropic quantum antiferromagnets. *Communications in Mathematical Physics*, 115(3):477–528, 1988. doi: 10.1007/BF01218021.

[4] Sven Erick Alm. Upper bounds for the connective constant of self-avoiding walks. *Combinatorics, Probability and Computing*, 2(2):115–136, 1993. doi: 10.1017/S0963548300000547.

[5] Daniel P. Arovas, Assa Auerbach, and F. D. M. Haldane. Extended Heisenberg models of antiferromagnetism: Analogies to the fractional quantum Hall effect. *Physical Review Letters*, 60(6):531–534, 1988. doi: 10.1103/PhysRevLett.60.531.

[6] Erwin H. Bareiss. Sylvester's identity and multistep integer-preserving Gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968. doi: 10.2307/2004533.

[7] Claudio Castelnovo, Claudio Chamon, Christopher Mudry, and Pierre Pujol. From quantum mechanics to classical statistical physics: Generalized Rokhsar–Kivelson Hamiltonians and the "Stochastic Matrix Form" decomposition. *Annals of Physics*, 318(2):316–344, 2005. doi: 10.1016/j.aop.2005.01.006.

[8] Mu-Fa Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In *Probability Towards 2000 (New York, 1995)*, volume 128 of *Lecture Notes in Statistics*, pages 123–136. Springer, New York, 1998. doi: 10.1007/978-1-4612-2224-8_7.

[9] Jean-Dominique Deuschel and Agoston Pisztora. Surface order large deviations for high-density percolation. *Probability Theory and Related Fields*, 104(4):467–482, 1996. doi: 10.1007/BF01198162.

[10] R. L. Dobrushin and S. B. Shlosman. Completely analytical Gibbs fields. In J. Fritz, A. Jaffe, and D. Szász, editors, *Statistical Physics and Dynamical Systems (Kőszeg, 1984)*,

pages 371–403. Birkhäuser, Boston, MA, 1985. doi: 10.1007/978-1-4899-6653-7_21.

[11] R. L. Dobrushin and S. B. Shlosman. Completely analytical interactions: Constructive description. *Journal of Statistical Physics*, 46(5-6):983–1014, 1987. doi: 10.1007/BF01011153.

[12] Artur Garcia-Saez, Valentin Murg, and Tzu-Chieh Wei. Spectral gaps of Affleck–Kennedy–Lieb–Tasaki Hamiltonians using tensor network methods. *Physical Review B*, 88(24):245118, 2013. doi: 10.1103/PhysRevB.88.245118.

[13] David Gosset and Evgeny Mozgunov. Local gap threshold for frustration-free spin systems. *Journal of Mathematical Physics*, 57(9):091901, 2016. doi: 10.1063/1.4962337.

[14] Wenhan Guo, Nicholas Pomata, and Tzu-Chieh Wei. Nonzero spectral gap in several uniformly spin-2 and hybrid spin-1 and spin-2 AKLT models. *Physical Review Research*, 3(1):013255, 2021. doi: 10.1103/PhysRevResearch.3.013255.

[15] F. D. M. Haldane. Continuum dynamics of the 1-D Heisenberg antiferromagnet: Identification with the O(3) nonlinear sigma model. *Physics Letters A*, 93(9):464–468, 1983. doi: 10.1016/0375-9601(83)90631-X.

[16] F. D. M. Haldane. Nonlinear field theory of large-spin Heisenberg antiferromagnets: Semiclassically quantized solitons of the one-dimensional easy-axis Néel state. *Physical Review Letters*, 50(15):1153–1156, 1983. doi: 10.1103/PhysRevLett.50.1153.

[17] C. L. Henley. From classical to quantum dynamics at Rokhsar–Kivelson points. *Journal of Physics: Condensed Matter*, 16(11):S891–S898, 2004. doi: 10.1088/0953-8984/16/11/045.

[18] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, second edition, 2013. doi: 10.1017/CBO9781139020411.

[19] *IEEE Standard for Floating-Point Arithmetic*. IEEE, New York, 2019. IEEE Std 754-2019 (Revision of IEEE 754-2008).

[20] Thomas Jackson, Bruno Nachtergaele, and Amanda Young. Local topological quantum order and spectral gap stability for the AKLT models on the hexagonal and Lieb lattices. *Communications in Mathematical Physics*, 407(12):261, 2026. doi: 10.1007/s00220-026-05752-9.

[21] Michael J. Kastoryano. Spectral gap of the hexagonal AKLT model via boundary-state factorization. arXiv:2609.27993 [quant-ph], 2026. Preprint.

[22] Michael J. Kastoryano, Angelo Lucia, and David Pérez-García. Locality at the boundary implies gap in the bulk for 2D PEPS. *Communications in Mathematical Physics*, 366(3):895–926, 2019. doi: 10.1007/s00220-019-03404-9.

[23] Tom Kennedy, Elliott H. Lieb, and Hal Tasaki. A two-dimensional isotropic quantum antiferromagnet with unique disordered ground state. *Journal of Statistical Physics*, 53(1-2):383–415, 1988. doi: 10.1007/BF01011563.

[24] Harry Kesten. Aspects of first passage percolation. In *École d'Été de Probabilités de Saint-Flour XIV—1984*, volume 1180 of *Lecture Notes in Mathematics*, pages 125–264. Springer, Berlin, 1986. doi: 10.1007/BFb0074919.

[25] A. N. Kirillov and V. E. Korepin. A resonance valence bond in quasicrystals. *Leningrad Mathematical Journal*, 1(2):343–377, 1990. Russian original: Algebra i Analiz **1** (1989), no. 2, 47–76. English text also available as arXiv:0909.2211 under the title "The valence bond solid in quasicrystals".

[26] Stefan Knabe. Energy gaps and elementary excitations for certain VBS-quantum antiferromagnets. *Journal of Statistical Physics*, 52(3-4):627–638, 1988. doi: 10.1007/BF01019721.

[27] Marius Lemm. Finite-size criteria for spectral gaps in D-dimensional quantum spin systems. In Houssam Abdul-Rahman, Robert Sims, and Amanda Young, editors, *Analytic Trends in Mathematical Physics*, volume 741 of *Contemporary Mathematics*, pages 121–132. American Mathematical Society, Providence, RI, 2020. doi: 10.1090/conm/741/14923.

[28] Marius Lemm and Evgeny Mozgunov. Spectral gaps of frustration-free spin systems with boundary. *Journal of Mathematical Physics*, 60(5):051901, 2019. doi: 10.1063/1.5089773.

[29] Marius Lemm, Anders W. Sandvik, and Ling Wang. Existence of a spectral gap in the Affleck–Kennedy–Lieb–Tasaki model on the hexagonal lattice. *Physical Review Letters*, 124(17):177204, 2020. doi: 10.1103/PhysRevLett.124.177204.

[30] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, second edition, 2017. doi: 10.1090/mbk/107. With contributions by Elizabeth L. Wilmer.

[31] Elliott H. Lieb. The classical limit of quantum spin systems. *Communications in Mathematical Physics*, 31(4):327–340, 1973. doi: 10.1007/BF01646493.

[32] Angelo Lucia and Amanda Young. A nonvanishing spectral gap for AKLT models on generalized decorated graphs. *Journal of Mathematical Physics*, 64(4):041902, 2023. doi: 10.1063/5.0139706.

[33] Angelo Lucia, Alvin Moon, and Amanda Young. Stability of the spectral gap and ground state indistinguishability for a decorated AKLT model. *Annales Henri Poincaré*, 25(8):3603–3648, 2024. doi: 10.1007/s00023-023-01398-8.

[34] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Communications in Mathematical Physics*, 161(3):447–486, 1994. doi: 10.1007/BF02101929.

[35] F. Martinelli, E. Olivieri, and R. H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics*, 165(1):33–47, 1994. doi: 10.1007/BF02099735.

[36] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In Pierre Bernard, editor, *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Mathematics*, pages 93–191. Springer, Berlin, 1999. doi: 10.1007/978-3-540-48115-7_2.

[37] John Martyn, Kohtaro Kato, and Angelo Lucia. Deformations of the boundary theory of the square-lattice AKLT model. *Physical Review B*, 102(3):035121, 2020. doi: 10.1103/PhysRevB.102.035121.

[38] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009. doi: 10.1016/j.jfa.2008.11.001.

[39] A. M. Perelomov. Coherent states for arbitrary Lie group. *Communications in Mathematical Physics*, 26(3):222–236, 1972. doi: 10.1007/BF01645091.

[40] Nicholas Pomata and Tzu-Chieh Wei. AKLT models on decorated square lattices are gapped. *Physical Review B*, 100(9):094429, 2019. doi: 10.1103/PhysRevB.100.094429.

[41] Nicholas Pomata and Tzu-Chieh Wei. Demonstrating the Affleck–Kennedy–Lieb–Tasaki spectral gap on 2D degree-3 lattices. *Physical Review Letters*, 124(17):177203, 2020. doi: 10.1103/PhysRevLett.124.177203.

[42] André Pönitz and Peter Tittmann. Improved upper bounds for self-avoiding walks in $\mathbf{Z}^d$. *Electronic Journal of Combinatorics*, 7:R21, 2000. doi: 10.37236/1499.

[43] J. M. Radcliffe. Some properties of coherent spin states. *Journal of Physics A: General Physics*, 4(3):313–323, 1971. doi: 10.1088/0305-4470/4/3/009.

[44] Ádám Timár. Boundary-connectivity via graph theory. *Proceedings of the American Mathematical Society*, 141(2):475–480, 2013. doi: 10.1090/S0002-9939-2012-11333-4.

[45] Warwick Tucker. *Validated Numerics: A Short Introduction to Rigorous Computations*. Princeton University Press, Princeton, NJ, 2011. doi: 10.1515/9781400838974.

[46] F. Verstraete, M. M. Wolf, D. Pérez-García, and J. I. Cirac. Criticality, the area law, and the computational power of projected entangled pair states. *Physical Review Letters*, 96(22):220601, 2006. doi: 10.1103/PhysRevLett.96.220601.

[47] Tzu-Chieh Wei, Ian Affleck, and Robert Raussendorf. Affleck–Kennedy–Lieb–Tasaki state on a honeycomb lattice is a universal quantum computational resource. *Physical Review Letters*, 106(7):070501, 2011. doi: 10.1103/PhysRevLett.106.070501.