Journal of Statistical Physics, Vol. 82, Nos. 5/6, 1996

The Covariance Matrix of the Potts Model: A Random Cluster Analysis

C. Borgs[1] and J. T. Chayes[2]

Received September 14, 1994; final August 10, 1995

We consider the covariance matrix, $G^{mn} = q^2\langle\delta(\sigma_x, m); \delta(\sigma_y, n)\rangle$, of the d-dimensional q-states Potts model, rewriting it in the random cluster representation of Fortuin and Kasteleyn. In any of the q ordered phases, we identify the eigenvalues of this matrix both in terms of representations of the unbroken symmetry group of the model and in terms of random cluster connectivities and covariances, thereby attributing algebraic significance to these stochastic geometric quantities. We also show that the correlation length corresponding to the decay rate of one of the eigenvalues is the same as the inverse decay rate of the diameter of finite clusers. For dimension $d = 2$, we show that this correlation length and the correlation length of the two-point function with free boundary conditions at the corresponding dual temperature are equal up to a factor of two. For systems with first-order transitions, this relation helps to resolve certain inconsistencies between recent exact and numerical work on correlation lengths at the self-dual point β_o. For systems with second order transitions, this relation implies the equality of the correlation length exponents from above and below threshold, as well as an amplitude ratio of two. In the course of proving the above results, we establish several properties of independent interest, including left continuity of the inverse correlation length with free boundary conditions and upper semicontinuity of the decay rate for finite clusters in all dimensions, and left continuity of the two-dimensional free boundary condition percolation probability at β_o. We also introduce DLR equations for the random cluster model and use them to establish ergodicity of the free measure. In order to prove these results, we introduce a new class of events which we call decoupling events and two inequalities for these events. The first is similar to the FKG inequality, but holds for events which are neither increasing nor decreasing; the second is similar to the van den Berg–Kesten inequality in standard percolation. Both inequalities hold for an arbitrary FKG measure.

[1] Center for Theoretical Study, Charles University, Prague, and Institut für Theoretische Physik, Freie Universität Berlin, D-14195 Berlin; Permanent address: Institut für Theoretische Physik, Universität Leipzig, Augustusplatz 10/11, 04109 Leipzig; E-mail: borgs@tph100.physik.uni-leipzig.de.

[2] Department of Mathematics, UCLA, Los Angeles, California 90095; E-mail: jchayes@math.ucla.edu.

0022-4715/96/0300-1235$09.50/0

KEY WORDS: Amplitude ratios; correlation inequalities; correlation length; covariance matrix; Fortuin–Kasteleyn representation; Potts model; random cluster model; Widom scaling.

1. INTRODUCTION: BACKGROUND AND DISCUSSION OF RESULTS

The q-state Potts model has been the subject of increasing interest in recent years. On the one hand, it has been studied by probabilists and statistical mechanicists due to its relationship to the random cluster model,[13, 1] where many of the known results for percolation are open and interesting problems. On the other hand, the phase transitions in the Potts model provide a paradigm for testing numerical methods developed for more complex transitions, such as deconfinement in lattice QCD: The Potts model is relatively easy to simulate with efficient algorithms (see, e.g., ref. 37), it can be tuned from a second-order through a weakly first-order to a strongly first-order transition by varying the number of states q, and many quantities of interest are explicitly known for dimension $d=2$, thus allowing for a direct test of numerical methods. Finally, many of the exact results on the Potts model have recently been shown to have fascinating algebraic interpretations (see, e.g., Section VII.B in the review of ref. 39).

Motivated by discrepancies between recent exact and numerical results on the correlation length of the Potts model, we have undertaken to identify and study the relevant length scales in the problem. We relate these scales both to the algebraic structure of the unbroken symmetry group and to stochastic geometric quantities in the random cluster representation of the Potts model. In the process, we show that the some of natural stochastic geometric quantities one defines in the random cluster representation—e.g., the finite-cluster connectivity—have independent algebraic significance. In two dimensions, we prove a relation between various scales which is an extension of known relations for percolation and the Ising magnet, and which establishes a strong form of Widom scaling for Potts models with continuous transitions. We also prove an analog of this relation for two-dimensional Potts models with discontinuous transitions; this analog helps to explain the apparent discrepancy between the exact and numerical results.

Adopting a field-theoretic perspective, we identify the relevant lengths in the model by studying the eigenvalues of the covariance matrix

$$G_b^{mn}(x-y)=\langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_b \tag{1.1}$$

Here $\sigma_x \in \{0, ..., q-1\}$ are the usual spins of the Potts model, $\delta(\cdot, \cdot)$ is the Kronecker delta, $\langle \cdot \rangle_b$ is the expectation with respect to the infinite-volume state obtained from finite-volume states with "b" boundary conditions, and $\langle A; B \rangle_b \equiv \langle AB \rangle_b - \langle A \rangle_b \langle B \rangle_b$ is the truncated expectation of the functions A and B.

In the disordered phase, we consider the covariance matrix with free boundary conditions, $G^{mn}_{\text{free}}(x-y)$. We find that this is proportional to the standard two-point function, which in turn is equal to the connectivity function in the random cluster representation:

$$\begin{aligned} G^{mn}_{\text{free}}(x-y) &= (q\delta(m,n)-1) \left\langle \frac{1}{q-1} (q\delta(\sigma_x, \sigma_y) - 1) \right\rangle_{\text{free}} \\ &= (q\delta(m,n)-1)\, \tau_{\text{free}}(x-y) \end{aligned} \tag{1.2}$$

Here the connectivity, $\tau_{\text{free}}(x-y)$, is the probabililty with respect to the free-boundary-condition random cluster measure that x and y lie in the same component. That $\tau_{\text{free}}(x-y)$ is equal to the two-point function in finite volume is well known both to probabilists and to numerical physicists, the latter of whom use this equivalence to measure the two-point function according to the "improved estimators" approach. Our only contribution here is to verify the equivalence in infinite volume. We note that, in the disordered phase, the covariance matrix contains no more information than the standard two-point function, or equivalently, the connectivity function.

The problem is more subtle in the ordered phase, where we consider the matrix $G^{mn}_c(x-y)$ with fixed constant boundary conditions, $c \in S = \{0, ..., q-1\}$. Defining the finite-cluster connectivity $\tau^{\text{fin}}_{\text{wir}}(x-y)$ to be the probability, in the so-called wired random cluster measure, that x find y lie in the same finite component, and the infinite-cluster covariance $C_{\text{wir}}(x-y)$ to be the covariance, again in the wired measure, of the events that x and y lie in the infinite component, we prove that the matrix elements $G^{mn}_c(x-y)$ are linear combinations of $\tau^{\text{fin}}_{\text{wir}}(x-y)$ and $C_{\text{wir}}(x-y)$, namely

$$\begin{aligned} G^{mn}_c(x-y) = (q\delta(m,n)-1)\, \tau^{\text{fin}}_{\text{wir}}(x-y) &+ (q\delta(m,c)-1) \\ &\times (q\delta(n,c)-1)\, C_{\text{wir}}(x-y) \end{aligned} \tag{1.3}$$

We remark that while the finite-volume analog of (1.3) is a straightforward consequence of the random cluster representation, the proof of the infinite-volume limit involves some subtleties related to how the infinite cluster emerges from large finite clusters in the wired problem (for more details, see the remark following Proposition 3.4).

Percolation analogs of $\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)$ and $C_{\mathrm{wir}}(x-y)$—in the absence of boundary conditions—have arisen previously in ref. 7, where they appeared as a natural decomposition of the truncated percolation connectivity in the ordered phase. There, however, they did not have independent signficance, appearing only as a sum. The question naturally arises whether they have independent significance here. Obviously, this is not the case for $q=2$, for which (1.3) can be rewritten as

$$G_c^{mn}(x-y)=(2\delta(m,n)-1)(\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)+C_{\mathrm{wir}}(x-y))$$

involving again only the sum $\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)+C_{\mathrm{wir}}(x-y)$.

For $q\geqslant 3$, however, the fixed-boundary-condition covariance matrix $G_c^{mn}(x-y)$ has a richer structure. We prove that it has a simple eigenvalue zero and a nontrivial simple eigenvalue

$$G_{\mathrm{wir}}^{(1)}(x-y)=q\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)+q(q-1)\,C_{\mathrm{wir}}(x-y) \tag{1.4}$$

both corresponding to the trival representation of the unbroken subgroup S_{q-1} of permutations of $S\backslash\{c\}$, as well as one $(q-2)$-fold degenerate eigenvalue

$$G_{\mathrm{wir}}^{(2)}(x-y)=q\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y) \tag{1.5}$$

corresponding to the remaining orthogonal subspace.[3] Thus we see that for $q\geqslant 3$, the finite-cluster cluster connectivity, $\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)$, has independent algebraic significance as an eigenvalue of the covariance matrix, and hence also physical significance in terms of the associated one-particle spectrum. As for the infinite cluster covariance $C_{\mathrm{wir}}(x-y)$, we will show in Theorem 4.3 that its *decay rate* is equal to the *decay rate* of the eigenvalue $G_{\mathrm{wir}}^{(1)}$ whenever the magnetization is positive. Thus although $C_{\mathrm{wir}}(x-y)$ does not have independent algebraic significance, its decay rate does.

For completeness, we note that the free boundary condition matrix, $G_{\mathrm{free}}^{mn}(x-y)$, can be diagonalized as well, yielding a simple eigenvalue zero and a $(q-1)$-fold degenerate eigenvalue

$$G_{\mathrm{free}}(x-y)=q\tau_{\mathrm{free}}(x-y) \tag{1.6}$$

Given the eigenvalues (1.4)–(1.6), one naturally defines the inverse correlation lengths:

[3] For the Ising model $(q=2)$, $G_c^{mn}(x-y)$ has only the trivial eigenvalue zero and the eigenvalue $G_{\mathrm{wir}}^{(1)}(x-y)$.

$$\frac{1}{\xi_{\text{free}}(\beta)} = -\lim_{|x| \to \infty} \frac{1}{|x|} \log G_{\text{free}}(x) \tag{1.7}$$

$$\frac{1}{\xi_{\text{wir}}^{(1)}(\beta)} = -\lim_{|x| \to \infty} \frac{1}{|x|} \log G_{\text{wir}}^{(1)}(x) \tag{1.8}$$

and

$$\frac{1}{\xi_{\text{wir}}^{(2)}(\beta)} = -\lim_{|x| \to \infty} \frac{1}{|x|} \log G_{\text{wir}}^{(2)}(x) \tag{1.9}$$

Here, as usual, β is the inverse temperature of the model. In all cases, the limits are taken so that x lies along a coordinate axis. In order to establish the existence of the limits, we return to the spin representation and use reflection positivity. We also give alternative subadditive proofs of the existence of the limits (1.7) and (1.9), which, though more complicated than the reflection positivity arguments, have the advantage that they hold for non-integer $q \geqslant 1$ and can be used to establish additional properties. In particular, we use subadditivity to show left continuity of the inverse correlation length $1/\xi_{\text{free}}(\beta)$ and upper semicontinuity of the inverse correlation length $1/\xi_{\text{wir}}^{(2)}(\beta)$. We also use subadditivity arguments to prove that $\xi_{\text{wir}}^{(2)}(\beta)$ is equal to several other geometrical correlation lengths in the problem, one of which is the decay raty of the diameter of finite clusters in the wired measure—a quantity which should be easily accessible to numerical measurement.

All three correlation lengths coincide in the high-temperature regime, where their common value is often denoted by $\xi_{\text{dis}}(\beta)$. In the low-temperature regime, we expect $\xi_{\text{free}}(\beta) \equiv \infty$. The nontrivial correlation length in this regime is often denoted by $\xi_{\text{ord}}(\beta)$. Here, however, we see that for $q \geqslant 3$, there are two *a priori* different nontrivial lengths, $\xi_{\text{wir}}^{(1)}(\beta)$ and $\xi_{\text{wir}}^{(2)}(\beta)$. Equations (1.4) and (1.5) immediately imply that

$$\xi_{\text{wir}}^{(1)}(\beta) \geqslant \xi_{\text{wir}}^{(2)}(\beta) \tag{1.10}$$

so that the correlation length $\xi_{\text{wir}}^{(1)}$ of the symmetric state (i.e., symmetric with respect to S_{q-1}) is not smaller than those of the unsymmetric states. An interesting open question is whether or not the inequality is strict. It is worth noting that in percolation, analogs of $C_{\text{wir}}(x-y)$ and $\tau_{\text{wir}}^{\text{fin}}(x-y)$ in the absence of boundary conditions have equal exponential decay rates[7], which here would imply equality of $\xi_{\text{wir}}^{(1)}(\beta)$ and $\xi_{\text{wir}}^{(2)}(\beta)$. However, it is not at all clear whether the Potts models for $q \geqslant 3$ should have analogous behavior.

We return finally to our original question, namely the discrepancy between the exact and numerical correlation lengths of two-dimensional Potts models with discontinuous transitions. Explicit calculations based on a mapping to the six-vertex model yielded a correlation length $\xi_{\text{dis}}(\beta_o)$ of the disordered phase at the self-dual point[6, 26, 24] β_0 which disagreed with previous numerical measurements[33, 17] of the ordered correlation length at the transition point $\xi_{\text{ord}}(\beta_o)$ by roughly a factor of 2, suggesting the possible relation $\xi_{\text{ord}}(\beta_o) = \frac{1}{2}\xi_{\text{dis}}(\beta_o)$.[2] A continuous transition analog of this relation is already known for both two-dimensional bond percolation, where $\xi(p) = \frac{1}{2}\xi(1-p)$ has been rigorously established for all $p > p_c$,[7] and the two-dimensional Ising magnet, where $\xi(\beta) = \frac{1}{2}\xi(\beta^*)$ has been established via exact solution for all $\beta > \beta_o$ (ref. 31; see also ref. 10). Here, as usual, β^* is the dual inverse temperature. However, from our results discussed above, we now know that the situation is more complicated in the q-state Potts model, $q \geqslant 3$, than it is in either percolation or the Ising magnet, since in the ordered phase the Potts model has two *a priori* different correlation lengths. One of our principal results is a relation of the conjectured form in terms of the smaller ordered correlation length, $\xi^{(2)}_{\text{wir}}$.

Our result follows from a dichotomy which we prove for all two-dimensional random cluster models with $q \geqslant 1$. In addition to the conjectured relation, the dichotomy implies Widom scaling for Potts models with continuous transitions. Let $P^{\text{free}}_{\infty}(\beta)$ be the percolation probabilty in the free-boundary-condition random cluster measure. Our dichotomy is: If $P^{\text{free}}_{\infty}(\beta^*) = 0$, then

$$\xi^{(2)}_{\text{wir}}(\beta) = \tfrac{1}{2}\xi_{\text{free}}(\beta^*) \tag{1.11}$$

whereas if $P^{\text{free}}_{\infty}(\beta^*) > 0$, then

$$\xi_{\text{free}}(\beta) = \xi^{(1)}_{\text{wir}}(\beta) = \xi^{(2)}_{\text{wir}}(\beta) \tag{1.12}$$

In order to interpret the dichotomy, we supplement it with the two-dimensional relation

$$P^{\text{wir}}_{\infty}(\beta)\, P^{\text{free}}_{\infty}(\beta^*) = 0 \tag{1.13}$$

where $P^{\text{wir}}_{\infty}(\beta)$ is the percolation probability in the wired measure, which is of course equal to the spontaneous magnetization $M(\beta)$. Note that (1.13) shows that $P^{\text{free}}_{\infty}(\beta^*) > 0$ implies $M(\beta) = 0$, so that (1.12) is simply the equality of the three correlation lengths in the high-temperature regime, as mentioned earlier.

Our more interesting corollaries follow from the first branch of the dichotomy, i.e., the duality relation (1.11). In order to see this, we combine (1.13) with the obvious bound $P_\infty^{\text{wir}}(\beta) \geqslant P_\infty^{\text{free}}(\beta)$ to obtain $P_\infty^{\text{free}}(\beta) P_\infty^{\text{free}}(\beta^*) = 0$, so that $P_\infty^{\text{free}}(\beta_o) = 0$. Since $P_\infty^{\text{free}}(\beta^*)$ is an increasing function of β^*, this in turn implies

$$P_\infty^{\text{free}}(\beta^*) = 0 \qquad \text{for all} \quad \beta \geqslant \beta_o \tag{1.14}$$

Equation (1.14) implies in particular that $P_\infty^{\text{free}}(\beta)$ is left continuous at the self-dual point β_o. Moreover, it means that that the first branch of the dichotomy [i.e., Eq. (1.11)] holds throughout the low-temperature phase $\beta \geqslant \beta_o$. For systems with first-order transitions, this implies the conjectured relation at β_o:

$$\xi_{\text{wir}}^{(2)}(\beta_o) = \tfrac{1}{2} \xi_{\text{free}}(\beta_o) \tag{1.15}$$

For systems with second-order transitions, (1.11) is a generalization of the aforementioned results on two-dimensional percolation[7] and the Ising magnet.[31] In particular, it gives a strong form of Widom scaling as $\beta \to \beta_o$: If $\xi_{\text{free}}(\beta^*)$ diverges with critical exponent ν, $\xi_{\text{free}}(\beta^*) \sim |\beta^* - \beta_o|^{-\nu}$ as $\beta^* \uparrow \beta_o$, (1.11) implies that $\xi_{\text{wir}}^{(2)}(\beta)$ diverges with the same exponent: $\xi_{\text{wir}}^{(2)}(\beta) \sim |\beta - \beta_t|^{-\tilde{\nu}}$ as $\beta \downarrow \beta_o$ with $\tilde{\nu} = \nu$.

As noted above, the interpretation (and in fact the proof) of the dichotomy (1.11) and (1.12) requires the relation (1.13), which we obtain as a special case of a general two-dimensional result of Gandolfi, Keane, and Russo (GKR).[18] However, in order to apply the GKR theorem, we need to know that the free random measure is ergodic, a result which we establish in all dimensions. We prove ergodicity by introducing suitable DLR equations[8, 29] for the random cluster problem. Here the justification of the DLR equations is much more delicate than in standard spin systems due to the nonlocal nature of the random cluster weights: because of this nonlocality, the specification used to construct the DLR states is not quasilocal, and thus standard theorems do not apply.

Before reviewing the organization of the paper, let us briefly discuss our methods. These methods are necessarily quite different from those used in the analysis of the Bernoulli percolation model, since the random cluster model lacks several properties which are used extensively in percolation —namely, independence of events occurring on fixed disjoint sets and the van den Berg–Kesten (BK)[3] inequality for events occurring on random disjoint sets. Moreover, the random cluster model has an additional feature—boundary conditions—which significantly complicates its analysis relative to the independent model. However, it is by actually focusing on the boundary conditions that we are able to circumvent the other difficulties and in fact derive two correlation inequalities which we expect will be useful in many other contexts. We do this by noting that in many cases, the events of interest carry with them boundary conditions which *decouple* them from other events and thus effectively overcome the coupling of the random cluster weights. This idea is formalized by introducing the notion of *decoupling events.* We use our decoupling events in formulating and proving two sets of inequalities which effectively replace independence and the BK inequality. The independence is replaced by a relation which resembles the FKG inequality, but contains two decoupling events and holds for a much larger class of events than the original FKG inequality—in particular, for events which are neither increasing nor decreasing. The BK inequality is replaced by a relation which resembles the independent BK inequality but contains a decoupling event. Both inequalities hold for *any* FKG measure, and thus in particular for the free and wired random cluster measures with $q \geqslant 1$.

The organization of this paper is as follows. In first two parts of Section 2, we review the necessary properties of the standard spin and random cluster representations of the Potts model. The third part of Section 2 contains our inequalities for decoupling events. In the last part of the section, we derive the DLR equation and establish ergodicity of the free measure. Section 3 is concerned with the covariance matrix. In the first part of the section, we derive the finite- and infinite-volume representations of the matrix with free and constant boundary conditions, in particular establishing the infinite-volume limits of $\tau_{\text{free}}(x-y)$, $\tau_{\text{wir}}^{\text{fin}}(x-y)$, and $C_{\text{wir}}(x-y)$ from their finite-volume analogs. The matrix is diagonalized in the second part of Section 3. Section 4 concerns the correlation lengths ξ_{free}, $\xi_{\text{wir}}^{(1)}$, and $\xi_{\text{wir}}^{(2)}$. In the first part of the section, we establish existence of the lengths using reflection positivity, as reviewed in the appendix. The second part of the section concerns alternative characterizations of ξ_{free}, $\xi_{\text{wir}}^{(1)}$, and $\xi_{\text{wir}}^{(2)}$, proved via subadditivity arguments and our inequalities for decoupling events. In particular, we show that $1/\xi_{\text{free}}$ is left continuous and $1/\xi_{\text{wir}}^{(2)}$ is upper semicontinuous; we prove that $\xi_{\text{wir}}^{(1)}$ is the decay rate of C_{wir} whenever the magnetization is positive; and we establish that $\xi_{\text{wir}}^{(2)}$ coincides with the decay rate of the diameter of finite clusters, as well as with the limiting decay rate of connectivity functions for clusters in boxes. Section 5 contains our proof of the two-dimensional dichotomy (1.11) and (1.12), as well as derivations of a few results on two-dimensional percolation probabilities. The first part of this section contains a discussion of the heuristics of the duality relation (1.11) in terms of the behavior of interfaces in.the system. In the second and third parts of the section, we prove upper and lower bounds of the form needed for the duality relation (1.11). Finally, in the fourth part of the section, we combine these upper and lower bounds with several results from Section 4 and the relation (1.13) to obtain our dichotomy.

Note Added. After submission of this Paper, we learned that G. Grimmett has simultaneously obtained results[21] which parallel some of those in our Section 2.4. In particular, Grimmett also introduces DLR equations and gives a very nice proof that both the free and the wired measure are Gibbs states. He then draws some of the same conclusions as we do in our corollaries to Theorem 5.5. The emphasis and the main results of the two papers are, however, very different: while Grimmett focuses on states of the random cluster model, we focus on the decay of correlations in the random cluster model, and correlation inequalities for general FKG measures.

2. PRELIMINARIES

2.1. Definition of the Spin Model

We consider the q-states Potts ferromagnet, a model with spins σ_x in the set $S = \{0, 1, ..., q-1\}$, $q \geqslant 1$. In a finite volume $\Lambda \subset \mathbb{Z}^d$, the Hamiltonian with free boundary conditions is

$$H_{\text{free}}(\sigma_\Lambda) = - \sum_{\langle x, y\rangle \in B(\Lambda)} (\delta(\sigma_x, \sigma_y) - 1) \tag{2.1}$$

where the sum goes over the set $B(\Lambda)$ of all nearest neighbor pairs $\langle x, y\rangle$ for which both x and y lie in Λ. The Hamiltonian with c-boundary conditions, $c \in S = \{0, 1, ..., q-1\}$, is

$$H_c(\sigma_\Lambda) = H_{\text{free}}(\sigma_\Lambda) - \sum_{x \in \Lambda,\, y \in \partial\Lambda} (\delta(\sigma_x, \sigma_y) - 1 \tag{2.2}$$

where $\partial\Lambda = \{x \notin \Lambda \mid \text{dist}(x, \Lambda) = 1\}$ is the (outer) boundary of Λ. Using the label b for "free" or $c \in S$, one introduces the partition function with boundary condition b as

$$Z_b(\Lambda) = \sum_{\sigma_\Lambda} e^{-\beta H_b(\sigma_\Lambda)} \tag{2.3}$$

where the sum runs over all configurations σ_Λ: $\Lambda \to S$, $x \mapsto \sigma_x$, and β is the inverse temperatures $\beta = 1/k_B T$.

As usual, an observable A with support $\operatorname{supp} A$ is a function A: $\sigma_\Lambda \to \mathbb{C}$ which does not depend on the spin variables σ_x for $x \notin \operatorname{supp} A$. A local observable is an observable with a support $\operatorname{supp} A$ not depending on Λ. Expectation values of a local observable A are defined as

$$\langle A\rangle_{b,\Lambda} = \frac{1}{Z_b(\Lambda)} \sum_{\sigma_\Lambda} A(\sigma_\Lambda)\, e^{-\beta H_b(\sigma_\Lambda)} \tag{2.4}$$

If A and $\tilde{A}$ are two local observables, one also considers the truncated expectation value defined by

$$\langle A;\tilde{A}\rangle_{b,\Lambda}=\langle A\tilde{A}\rangle_{b,\Lambda}-\langle A\rangle_{b,\Lambda}\langle \tilde{A}\rangle_{b,\Lambda} \tag{2.5}$$

For observables of the form

$$A_p=\exp(i(\sigma,p))=\exp\left(i\sum_{x\in\operatorname{supp}A_p}\sigma_x p_x\right)$$

where p is a function of finite support from $\mathbb{Z}^d$ into $\hat{S}=\{0, 2\pi/q, ..., 2\pi(q-1)/q\}$, the expectation values $\langle A_p\rangle_{\text{free},\Lambda}$ are monotone increasing (i.e., nondecreasing) in Λ, while the expectation values $\langle A_p\rangle_{0,\Lambda}$ are monotone decreasing in Λ by Griffiths' second inequality,[19] as generalized by Ginibre.[16] As a consequence, for $b=$ "free" or $b=0$, the thermodynamic limit

$$\langle A_p\rangle_b=\lim_{\Lambda\to\mathbb{Z}^d}\langle A_p\rangle_{b,\Lambda} \tag{2.6}$$

exists for all local observables of the form $A_p=e^{i(\sigma,p)}$ and hence for all local observables A. In (2.6), the limit may be taken through any increasing sequence of sets. Using the permutation symmetry of the Hamitonian (2.2), one concludes that the limit (2.6) exists for all boundary conditions b considered here (i.e., free or any constant boundary conditions). Also by Griffiths' second inequality,[19,16] the limit (2.6) is translation invariant.

The order parameter of the Potts model is the magnetization

$$M(\beta)=\frac{1}{q-1}\langle q\delta(\sigma_{x_0},0)-1\rangle_0 \tag{2.7}$$

where x_0 is an arbitrary point in $\mathbb{Z}^d$ (recall that the infinite-volume states $\langle\cdot\rangle_b$ are translation invariant). It is known that $M(\beta)$ is increasing in β,[16,1] decreasing in q,[1] and that the infinite-volume states $\langle\cdot\rangle_c$, $c\in S$, are equal to $\langle\cdot\rangle_{\text{free}}$ if and only if $M(\beta)=0$.[1],[4] Defining the transition point

$$\beta_t=\inf\{\beta\mid M(\beta)>0\} \tag{2.8}$$

we remark that it is believed that $M(\beta_t)$ is increasing in q, and that

$$q_c=\max\{q\in\mathbb{N}\mid M(\beta_t)=0\} \tag{2.9}$$

[4] Actually, ref. 1 proved that *all* infinite-volume Gibbs states are equal to $\langle\cdot\rangle_{\text{free}}$ if and only if $M(\beta)=0$.

is 4 for $d=2$ and 2 for $d>2$. The fact that $M(\beta_t)>0$, i.e., the existence of a first-order phase transition, has been rigorously established for all $d\geqslant 2$ provided q is sufficiently large (ref. 25; see also refs. 27 and 28).

2.2. The Random Cluster Representation: Review of Basic Properties

It is often useful to reexpress the q-state Potts model as an integer value of a two-parameter interacting percolation model, the so-called random cluster model of Fortuin and Kasteleyn.[13] In order to set our notation and state the results we will use in the rest of this paper, we briefly review the derivation and some basic properties of the FK representation. The representation is defined in terms of configurations $\omega\in\Omega\equiv\{0,1\}^{\mathbb{B}_d}$, where $\mathbb{B}_d=\{\langle x,y\rangle \mid x,y\in\mathbb{Z}^d\}$ is the nearest-neighbor bond lattice. For subsets $B\subset\mathbb{B}_d$, the configuration space is denoted by $\Omega_B\equiv\{0,1\}^B$.

Let us start with the finite-volume partition funtion with free boundary conditions. We write the Gibbs factor $\exp[-\beta H_{\text{free}}(\sigma_\Lambda)]$ as

$$\prod_{\langle x,y\rangle\in B(\Lambda)} e^{\beta(\delta(\sigma_x,\sigma_y)-1}$$

and expand the product with the help of the identity

$$e^{\beta(\delta(\sigma_x,\sigma_y)-1}=(1-p)+p\delta(\sigma_x,\sigma_y),\qquad \text{where}\quad p=1-e^{-\beta} \tag{2.10}$$

We identify each term of this expansion with a configuration $\omega\in\Omega_{B(\Lambda)}$; ω is chosen so that it is zero on those bonds for which the factor in the product is $1-p$, and one on those bonds for which the factor is $p\delta(\sigma_x,\sigma_y)$. Geometrically, we think of the bonds $b=\langle x,y\rangle$ for which $\omega(b)=1$ as occupied or ordered, and those for which $\omega(b)=0$ as vacant or disordered. With a slight abuse of notation, we sometimes use the symbol ω to denote the set of occupied bonds in $B(\Lambda)$, and ω^c to denote the set of empty bonds in $B(\Lambda)$; see, e.g., (2.11) below.

Rewriting the Gibbs factor in expanded form, we obtain

$$Z_{\text{free}}(\Lambda)=\sum_{\omega\in\Omega_{B(\Lambda)}}\sum_{\sigma_\Lambda}(1-p)^{|\omega^c|}p^{|\omega|}\prod_{\langle x,y\rangle\in\omega}\delta(\sigma_x,\sigma_y) \tag{2.11}$$

Evaluating the sum over σ_Λ, we pick up a factor q for each connected component of the graph (Λ,ω) (regarding isolated points as separate clusters). Denoting the number of clusters in this graph by $\#(\omega)$, we find

$$Z_{\text{free}}(\Lambda)=\sum_{\omega\in\Omega_{B(\Lambda)}}(1-p)^{|\omega^c|}p^{|\omega|}q^{\#(\omega)} \tag{2.12}$$

It is an easy exercise to generalize (2.12) to the expectations of local observables $A = A(\sigma)$. One obtains

$$\langle A \rangle_{\text{free}, \Lambda} = \sum_{\omega \in \Omega_{B(\Lambda)}} G_{\text{free}, \Lambda}(\omega)\, E_{\text{free}}(A \mid \omega) \tag{2.13}$$

where

$$G_{\text{free}, \Lambda}(\omega) = \frac{1}{Z_{\text{free}}(\Lambda)} (1-p)^{|\omega^c|}\, p^{|\omega|}\, q^{\#(\omega)} \tag{2.14}$$

is the weight of the configuration ω, while $E_{\text{free}}(\cdot \mid \omega)$ is an average over spins, with the spins constrained to be constant on each connected cluster of ω and with values for different clusters being chosen uniformly from $\{0, 1, ..., q-1\}$. We remark that for the purposes of interpreting expectations of this sort, it is often convenient to consider the joint distribution on the spin and bond variables with weights given by the terms in (2.11), as introduced implicitly in ref. 37 and explicitly in ref. 9. In terms of this distribution, the expectation $E_{\text{free}}(\cdot \mid \omega)$ is an average over the conditional distribution of spins, given the bond variables.

For constant boundary conditions, one obtains a similar representation, with the following differences (as noted in ref. 1):

(a) The set $B(\Lambda)$ is replaced by the set $B^+(\Lambda)$ of all nearest neighbor pairs $\langle x, y \rangle$ for which at least one of the two points x and y lies in Λ.

(b) The points of the boundary $\partial\Lambda$ are regarded as *preconnected* or *wired*, in the sense that these points are taken to be lying in one cluster. This of course modifies the value of $\#(\omega)$.

(c) The expectation $E_{\text{free}}(A|\omega)$ in (2.14) is replaced by $E_c(A|\omega)$, where the average is computed with the additional constraint that spins in clusters connected to the boundary now only assume the value $\sigma_x = c$.

We have

$$\langle A \rangle_{c, \Lambda} = \sum_{\omega \in \Omega_{B^+(\Lambda)}} G_{\text{wir}, \Lambda}(\omega)\, E_c(A \mid \omega) \tag{2.15}$$

where

$$G_{\text{wir}, \Lambda}(\omega) = \frac{1}{Z_{\text{wir}}(\Lambda)} (1-p)^{|\omega^c|}\, p^{|\omega|}\, q^{\#(\omega)} \tag{2.16}$$

and $Z_{\text{wir}} = \sum_{c \in S} Z_c = q Z_0$.

We denote by $\mu_{\text{free}, \Lambda}(\cdot)$ and $\mu_{\text{wir}, \Lambda}(\cdot)$ the finite-volume measures defined by the weights (2.14) and (2.16), respectively.

Remark. The measures $\mu_{\text{free}, \Lambda}(\cdot)$ and $\mu_{\text{wir}, \Lambda}(\cdot)$ are defined on the probability spaces $(\Omega_{B(\Lambda)}, \mathscr{F}_{B,(\Lambda)})$, and $(\Omega_{B^+(\Lambda)}, \mathscr{F}_{B^+(\Lambda)})$ respectively. (In general, we use $\mathscr{F}_B$ to denote the σ-algebra generated by cylinder events $A \subset \Omega_B$.) It is sometimes convenient to extend these to measures on the full space $(\Omega, \mathscr{F})$ by declaring all bonds in $\mathbb{B}_d \backslash B(\Lambda)$ to be vacant for $\mu_{\text{free}, \Lambda}(\cdot)$, and all bonds in $\mathbb{B}_d \backslash B^+(\Lambda)$ to be occupied for $\mu_{\text{wir}, \Lambda}(\cdot)$.

An important property of the FK representation is that it obeys the Harris–FKG inequality. This inequality, first proved for percolation in ref. 22 and proved for a large class of models in ref. 14, was established for the $q \geqslant 1$ random cluster representation in ref. 11 (see also ref. 1). We begin with the standard:

Definition 2.1. Consider the natural partial order on bond configurations $\omega \in \Omega_B$, $B \subset \mathbb{B}_d$, namely $\omega \prec \omega'$ if $\omega(b) = 1 \Rightarrow \omega'(b) = 1$. A function $f\colon \Omega_B \to \mathbb{R}$ is said to be *increasing* if it is nondecreasing with respect to this partial order, i.e., $f(\omega) \leqslant f(\omega')$ for all $\omega \prec \omega'$. An event is said to be increasing if its indicator is an increasing function. Similarly, a function f is *decreasing* if the function $-f$ is increasing, and an event is decreasing if its complement is increasing.

A measure μ on $(\Omega_B, \mathscr{F}_B)$ is said to be an *FKG measure* if it obeys the so-called Harris–FKG inequality

$$\mu(A_1 \cap A_2) \geqslant \mu(A_1)\, \mu(A_2) \tag{2.17}$$

for all pairs of increasing events $A_1, A_2 \in \mathscr{F}_B$. It is said to be a *strong FKG measure* if for each cylinder event $C \in \mathscr{F}_B$, the conditional measure $\mu(\cdot \mid C)$ is an FKG measure. Finally, a measure μ on $(\Omega_B, \mathscr{F}_B)$ is said to *FKG dominate* a measure ν on $(\Omega_B, \mathscr{F}_B)$, denoted by

$$\nu \underset{\text{FKG}}{\leqslant} \mu$$

if $\nu(A) \leqslant \mu(A)$ for all increasing events $A \in \mathscr{F}_B$.

Proposition 2.2.[11, 1] Let $q \geqslant 1$. Then the finite-volume free and wired FK measures $\mu_{\text{free}, \Lambda}$ and $\mu_{\text{wir}, \Lambda}$ are strong FKG measures.

Consequences (see, e.g., ref. 1)

1. The finite-volume measures are monotonic in the volume:

$$\mu_{\text{free}, \Lambda} \underset{\text{FKG}}{\leqslant} \mu_{\text{free}, \Lambda'} \qquad \text{if} \quad \Lambda \subset \Lambda' \tag{2.18}$$

and

$$\mu_{\mathrm{wir},\Lambda} \underset{\mathrm{FKG}}{\geqslant} \mu_{\mathrm{wir},\Lambda'} \qquad \text{if} \quad \Lambda \subset \Lambda' \tag{2.19}$$

from which it follows that the infinite-volume measures

$$\mu_{\mathrm{free}}(\,\cdot\,) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\mathrm{free},\Lambda}(\,\cdot\,) \tag{2.20}$$

and

$$\mu_{\mathrm{wir}}(\,\cdot\,) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\mathrm{wir},\Lambda}(\,\cdot\,) \tag{2.21}$$

exist for all monotone local functions, and hence for all local functions. Furthermore, these infinite-volume measures are translation invariant and inherit the strong FKG property.

2. The wired measures FKG dominate the free measures, i.e.,

$$\mu_{\mathrm{free},\Lambda} \underset{\mathrm{FKG}}{\leqslant} \mu_{\mathrm{wir},\Lambda} \tag{2.22}$$

and

$$\mu_{\mathrm{free}} \underset{\mathrm{FKG}}{\leqslant} \mu_{\mathrm{wir}} \tag{2.23}$$

Another useful property of these measures is that they have *finite energy*, a notion introduced by Newman and Schulman.[32]

Definition 2.3. Let $B \subset \mathbb{B}_d$, $|B| < \infty$, and $\phi \in \Omega_B$ a configuration on B. If $\omega \in \Omega$ is a configuration on the full space, let $\phi(\omega)$ be the configuration which agrees with ϕ on B and with ω on B^c:

$$\phi(\omega)(b) = \begin{cases} \phi(b), & b \in B \\ \omega(b), & b \in B^c \end{cases}$$

Finally, if $A \subset \Omega$ is an event, let $\phi(A) = \{\phi(\omega) \mid \omega \in A\}$. The measure μ on Ω is said to have *finite energy* if for every finite $B \subset \mathbb{B}_d$ and for every $\phi \in \Omega_B$,

$$\mu(A) > 0 \Rightarrow \mu(\phi(A)) > 0$$

It is easy to see that finite energy is equivalent to the statement: For each bond b, the conditional probability of the event that b is occupied, given the configuration on all the other bonds, is nontrivial:

$$0 < \mu(\omega(b) = 1 \mid \omega(\tilde{b}), \tilde{b} \neq b) < 1$$

For the free and wired measures, it was observed in ref. 1 that this probability can be explicitly calculated:

$$\mu(\omega(b)=1 \,|\omega(\tilde{b}), \tilde{b}\neq b) = \begin{cases} p & \text{if the endpoints of } b \text{ are connected, regardless of } \omega(\dot{b}) \\ \dfrac{p}{p+q(1-p)} & \text{otherwise} \end{cases} \tag{2.24}$$

where $\mu=\mu_{\text{free}}$ or μ_{wir}. Thus for all $q\geqslant 1$ and all $p\neq 0, 1$, the random cluster measures μ_{free} and μ_{wir} have finite energy. Note that this is not true in all random cluster measures: Boundary conditions can impose constraints which exclude certain configurations.

Given stationarity and finite energy, it follows immediately from a general result of Burton and Keane[5] that the infinite cluster is unique:

Proposition 2.4. For any $q\geqslant 1$ and any $p\in(0, 1)$, the free and wired random cluster states have at most one infinite cluster with probability one.

Since the Burton and Keane theorem requires only stationarity, it applies also to nonextremal states, and therefore allows the possibility of a convex combination of states with zero and one infinite cluster. If, in addition, the measures are ergodic, then at any given value of p, there is *either* zero *or* one infinite cluster with probability one. This is presumably the case for both the free and wired measures, although we only prove it for the free state (see Section 2.4). Of course, ergodicity does not exclude the possibility that, for a fixed value of p, the wired state has an infinite cluster and the free state does not—indeed, for q large enough, this is exactly what happens at the transition point.

2.3. Two Useful Inequalities

There are three main technical tools for factoring intersections of events in standard Bernoulli percolation: the FKG inequality for monotone events, independence for events which occur on nonrandom disjoint sets, and the van den Berg–Kesten[3] inequality for events which occur on random disjoint sets. As discussed in the last section, the free and wired random cluster measures obey an FKG inequality. However, due to the nonlocality of the weights (2.14) and (2.16), they satisfy neither an independence condition nor a BK inequality. Indeed, it is clear from (2.24) that the probability of even a simple bond occupation event can be enhanced by the occurrence of some other event at an arbitrarily long distance from the bond in question. In this subsection, we provide alternatives to independence and the BK inequality for many events of interest in a general setting.

As a substitute for independence of events occurring on nonrandom disjoint sets, we might try to use the FKG inequality as a bound, provided that the desired events are monotone. However, many of the events we care about—especially in the low-temperature phase—are not monotone. For example, the probability of a connection via finite clusters is the *intersection* of an increasing and a decreasing event. The presence of boundary conditions, which very often complicates proofs in the random cluster model, can be used to our advantage here. Certain boundary conditions *decouple* a set from its exterior. Many events of interest carry with them decoupling boundary conditions for the (random) sets on which they occur. We make this notion precise by introducing the definition of a *decoupling event* below. It turns out that, given this definition, it is possible to prove a general inequality which is similar to the FKG inequality and which replaces independence for events whose random boundaries occur within disjoint nonrandom sets. Our inequality holds for any FKG measure and for events which are intersections of *arbitrary* events with monotone decoupling events.

As explained above, the BK inequality is certainly not true in general for the random cluster model—there are numerous examples in which the occurrence of one event enhances the occurrence of another. However, this enhancement cannot take place if the two events are *decoupled* from one another, in a sense to be made precise in the definition below. Thus we prove a second inequality, which replaces the BK inequality of Bernoulli percolation, and which holds for the intersection of an arbitrary event, an increasing event, and a decreasing decoupling event.

In Proposition 2.6 below, we actually present two versions of each of our inequalities: one which is easy to formulate (but not that useful), and a more involved one which is of the form needed for our applications. All of these inequalities hold for general FKG measures. We also give a useful corollary that concerns monotonicity in the volume and FKG domination in the random cluster model. We begin with the definition of a decoupling event.

Definition 2.5. Given a probability space $(\Omega, \mathscr{F}, \mu)$ and events A_1, A_2, $D \in \mathscr{F}$, we say that D is a *decoupling event* for A_1 and A_2, if

$$\mu(A_1 \cap A_2 | D) = \mu(A_1 | D)\, \mu(A_2 | D) \tag{2.25}$$

For brevity, we will sometimes say D decouples A_1 from A_2.

While this definition makes sense in any probability space, it may be useful to illustrate it with a typical example from the random cluster model. Consider a set $B \subset \mathbb{B}_d$ such that $\mathbb{B}_d \backslash B = B_1 \cup B_2$, $B_1 \cap B_2 = \emptyset$. The event that the bonds of B are vacant then decouples any event $A_1 \in \mathscr{F}_{B_1 \cup B}$ from any event $A_2 \in \mathscr{F}_{B_2 \cup B}$. In this paper, such decoupling events typically occur when B is the boundary of a finite occupied cluster. Returning to the general context of Definition 2.5, we have:

Proposition 2.6. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space with Ω partially ordered and μ an FKG measure with respect to this order. Then the following inequalities hold.

1. The First Inequality:

(i) Consider two arbitrary events $A_1, A_2 \in \mathscr{F}$, and two increasing (or two decreasing) events $D_1, D_2 \in \mathscr{F}$ such that D_1 decouples A_1 from D_2 while D_2 decouples A_2 from $A_1 \cap D_1$. Then $E_1 = A_1 \cap D_1$ and $E_2 = A_2 \cap D_2$ obey the inequality

$$\mu(E_1 \cap E_2) \geqslant \mu(E_1)\, \mu(E_2) \tag{2.26}$$

(ii) More generally, let E_i, $i = 1, 2$, be disjoint unions of the form

$$E_i = \bigcup_{k \in K_i} A_{i,k} \cap D_{i,k} \tag{2.27}$$

where K_i are countable index sets, $A_{i,k} \in \mathscr{F}$ are arbitrary events, $D_{i,k} \in \mathscr{F}$ are all increasing (or all decreasing) events, and $D_{1,k}$ decouples $A_{1,k}$ from $D_{2,k'}$ while $D_{2,k'}$ decouples $A_{2,k'}$ from $A_{1,k} \cap D_{1,k}$ for all $k \in K_1$ and $k' \in K_2$. Then E_1 and E_2 obey the inequality (2.26).

2. The Second Inequality:

(i) Let $A_1 \in \mathscr{F}$ be an increasing event, $A_2 \in \mathscr{F}$ be arbitrary, and $D \in \mathscr{F}$ be a decreasing event which decouples A_1 from A_2. Then

$$\mu(A_1 \cap D \cap A_2) \leqslant \mu(A_1)\, \mu(D \cap A_2) \leqslant \mu(A_1)\, \mu(A_2) \tag{2.28}$$

(ii) More generally let $A_1 \in \mathscr{F}$ be an increasing event, and let $A_2 \in \mathscr{F}$ and $D \in \mathscr{F}$ be events for which $D \cap A_2$ can be rewritten as a disjoint union of the form (2.27), with $D_{2,k}$ decreasing events that decouple A_1 from $A_{2,k}$ for all $k \in K_2$. Then the bound (2.28) remains valid.

Proof. Rewriting the left hand side of (2.26) as

$$\mu(D_2)\, \mu(A_1 \cap D_1 \cap A_2 \,|\, D_2)$$

and using the fact that D_2 decouples A_2 from $A_1 \cap D_1$, we obtain

$$\mu(A_1 \cap D_1 \cap A_2 \cap D_2) = \mu(A_1 \cap D_1 \cap D_2)\, \mu(A_2 \mid D_2)$$

Applying the same procedure to the term $\mu(A_1 \cap D_1 \cap D_2)$ and using the decoupling event D_1, we get

$$\mu(A_1 \cap D_1 \cap A_2 \cap D_2) = \mu(D_1 \cap D_2)\, \mu(A_1 \mid D_1)\, \mu(A_2 \mid D_2)$$

which by the FKG inequality (2.17) implies (2.26). Part 1(ii) of the proposition then follows from the countable additivity of the measure μ and the fact that the events E_1 and E_2 are disjoint unions of events for which (2.26) is valid.

In order to prove 2(i), we observe that

$$\mu(A_1 \cap D \cap A_2) = \mu(D)\, \mu(A_1 \mid D)\, \mu(A_2 \mid D)$$

by the definition of conditional expectations and (2.25). Using the FKG inequality (2.17) to bound $\mu(A_1 \mid D)$ by $\mu(A_1)$, we find that the bound (2.28) now follows. Again, 2(ii) follows from 2(i) and the countable additivity of the measure. ∎

Remark. It is clear from the above proof that the inequality (2.26) is reversed if one of the two decoupling events D_1 and D_2 is increasing and the other is decreasing. Similarly, the first inequality in (2.28) is reversed if A_1 and D are both decreasing or both increasing.

Corollary. Let $q \geqslant 1$, $\Lambda \subset \mathbb{Z}^d$, and $b = \text{wir}$ or free. Consider the random cluster measure $\mu_{b,\Lambda}$ and the corresponding probability space $(\Omega_{B_b}, \mathscr{F}_{B_b}, \mu_{b,\Lambda})$, where $B_b = B^+(\Lambda)$ if $b = \text{wir}$ and $B_b = B(\Lambda)$ if $b = \text{free}$. Let $B \subset B_b$, and let E be an event of the form (2.27), where the index set is the set of all subsets of B, i.e.,

$$E = \bigcup_{S \subset B} A_S \cap D_S$$

with $A_S \in \mathscr{F}_{B_b}$ arbitrary events, and $D_S \in \mathscr{F}_{B_b}$ decreasing events that decouple A_S from all events in $\mathscr{F}_{B_b \setminus B}$. If the events D_S are decoupling events with respect to the measure $\mu_{\text{free},\Lambda}$, then

$$\mu_{\text{free},\Lambda'}(E) \geqslant \mu_{\text{free},\Lambda}(E) \qquad \text{provided} \quad \Lambda' \subset \Lambda \quad \text{and} \quad B \subset B(\Lambda') \tag{2.29}$$

If the events D_S are decoupling events with respect to the measure $\mu_{\text{wir},\Lambda}$, then

$$\mu_{\text{wir},\Lambda'}(E) \leqslant \mu_{\text{wir},\Lambda}(E) \qquad \text{provided} \quad \Lambda' \subset \Lambda \quad \text{and} \quad B \subset B^+(\Lambda') \tag{2.30}$$

and

$$\mu_{\text{wir}, \Lambda}(E) \leqslant \mu_{\text{free}, \Lambda}(E) \qquad \text{provided} \quad B \subset B(\Lambda) \tag{2.31}$$

Proof. Let A_1 be the event that all bonds in $B(\Lambda)\backslash B(\Lambda')$ are vacant. Then $A_1 \in \mathscr{F}_{B(\Lambda)\backslash B(\Lambda')} \subset \mathscr{F}_{B(\Lambda)\backslash B}$ is decoupled from A_S by the event D_S. Since both A_1 and D_S are decreasing events,

$$\mu_{\text{free}, \Lambda}(E \mid A_1) \geqslant \mu_{\text{free}, \Lambda}(E)$$

by the remark following the proof of Proposition 2.6. Observing that $\mu_{\text{free}, \Lambda'}(E) = \mu_{\text{free}, \Lambda}(E \mid A_1)$, this proves (2.29). Defining A_1 as the event that all bonds in $B^+(\Lambda)\backslash B^+(\Lambda')$ are occupied [all bonds in $B(\Lambda)^+\backslash B(\Lambda)$ are vacant], we prove the remaining two inequalities in the same way. ∎

In order to illustrate the utility of Proposition 2.6, we conclude this subsection with applications of each of the two inequalities. These applications will be needed in our subsequent analysis and may be of independent interest. As usual, we denote by $C(x) = C(x; \omega)$ the set of occupied bonds connected to x in the configuration ω, and define $\{x \leftrightarrow y\}$ as the event that x is connected to y by a finite path of occupied bonds. We also define, for each finite set $\Lambda \subset \mathbb{Z}^d$ and any two points $x, y \in \Lambda$, the event $R^{\text{fin}}_{x,y}(\Lambda)$ that x and y are connected by a cluster $C(x) \subset B(\Lambda)$.

Proposition 2.7. Let $q \geqslant 1$, $\Lambda \subset \mathbb{Z}^d$ be finite or infinite, and let $\mu = \mu_{\text{wir}, \Lambda}$ or $\mu_{\text{free}, \Lambda}$. Then for all finite $\Lambda_1, \Lambda_2 \subset \Lambda$ with $B^+(\Lambda_1) \cap B(\Lambda_2) = B(\Lambda_1) \cap B^+(\Lambda_2) = \varnothing$, and all $x, y \in \Lambda_1$, $z, w \in \Lambda_2$,

$$\mu(R^{\text{fin}}_{x,y}(\Lambda_1) \cap R^{\text{fin}}_{z,w}(\Lambda_2)) \geqslant \mu(R^{\text{fin}}_{x,y}(\Lambda_1))\, \mu(R^{\text{fin}}_{z,w}(\Lambda_2))$$

Proposition 2.8. Let $q \geqslant 1$, $\Lambda \subset \mathbb{Z}^d$ be finite, and let $\mu = \mu_{\text{wir}, \Lambda}$ or $\mu_{\text{free}, \Lambda}$. Let $x, y, z, w \in \Lambda$, and let D be the event $\{x \not\leftrightarrow z\} \cap \{y \not\leftrightarrow w\}$. Then

$$\mu(\{x \leftrightarrow y\} \cap D \cap \{z \leftrightarrow w\}) \leqslant \mu(x \leftrightarrow y)\, \mu(z \leftrightarrow w)$$

Clearly, Proposition 2.7 is an application of the first inequality in Proposition 2.6—the connections in question occur on fixed disjoint sets, $B(\Lambda_1)$ and $B(\Lambda_2)$, and due to the finiteness of the clusters, each connection carries its own decoupling event. Note that if $B^+(\Lambda_1) \cap B^+(\Lambda_2) = \varnothing$, then in percolation, the probability of the intersection of the events in Proposition 2.7 would factor exactly. Here our first inequality replaces this independence. In fact, given that the decoupling events can overlap, Proposition 2.7 gives a new result even in the case of percolation. Proposition 2.8 is an application of the second inequality in Proposition 2.6—the connections in question occur on random disjoint sets separated by the decoupling event D. This obviously replaces the BK inequality. Note that, in marked contrast to percolation, the inequality would fail to hold if we removed the decoupling event.

Proof of Proposition 2.7. Introducing $\mathscr{B}_1$ as the family of all sets $B \subset B(\Lambda_1)$ such that B connects x to y, and $\mathscr{B}_2$ as the family of all $B \subset B(\Lambda_2)$ connecting z to w, we decompose $R^{\text{fin}}_{x,y}(\Lambda_1)$ and $R^{\text{fin}}_{z,w}(\Lambda_2)$ as

$$R^{\text{fin}}_{x,y}(\Lambda_1) = \bigcup_{B \in \mathscr{B}_1} \{C(x) = B\} = \bigcup_{B \in \mathscr{B}_1} \{\omega_B = 1\} \cap \{\omega_{\partial^* B} = 0\}$$

and

$$R^{\text{fin}}_{z,w}(\Lambda_2) = \bigcup_{B \in \mathscr{B}_2} \{C(z) = B\} = \bigcup_{B \in \mathscr{B}_2} \{\omega_B = 1\} \cap \{\omega_{\partial^* B} = 0\}$$

Here ω_B is the configuration ω restricted to the set B and $\partial^* B$ is the set of all bonds in $\mathbb{B}_d \backslash B$ which are connected to B. Observing that for all $B \in \mathscr{B}_1$ and all $\tilde{B} \in \mathscr{B}_2$, $D_{1,B} = \{\omega_{\partial^* B} = 0\}$ decouples $A_{1,B} = \{\omega_B = 1\}$ from all events in $\mathscr{F}_{B^c} \supset \mathscr{F}_{B^+(\Lambda_2)}$, while $D_{2,\tilde{B}} = \{\omega_{\partial^* \tilde{B}} = 0\}$ decouples $A_{2,\tilde{B}} = \{\omega_{\tilde{B}} = 1\}$ from all events in $\mathscr{F}_{\tilde{B}^c} \supset \mathscr{F}_{B^+(\Lambda_1)}$, one easily verifies that $R^{\text{fin}}_{x,y}(\Lambda_1)$ and $R^{\text{fin}}_{z,w}(\Lambda_2)$ are events of the form considered in part 1(ii) of Proposition 2.6.

Proof of Proposition 2.8. Defining $A_1 = \{x \leftrightarrow y\}$ and $A_2 = \{z \leftrightarrow w\}$, we rewrite A_1 as the disjoint union $A_1^{\text{fin}} \cup A_1^{\text{inf}}$, with

$$A_1^{\text{fin}} = A_1 \cap \{x \not\leftrightarrow \partial\Lambda\}$$

and

$$A_1^{\text{inf}} = A_1 \cap \{x \leftrightarrow \partial\Lambda\}$$

Notice that $\mu_{\text{free},\Lambda}(A_1^{\text{inf}}) = 0$, since with free boundary conditions, x cannot be connected to the outer boundary $\partial\Lambda = \{x \notin \Lambda \mid \text{dist}(x, \Lambda) = 1\}$. Introducing the family $\mathscr{B}_1$ of sets $B \subset B(\Lambda)$ that connect x to y but do not connect x to z or y to w, we then decompose $A_1^{\text{fin}} \cap D$ as

$$A_1^{\text{fin}} \cap D = \bigcup_{B \in \mathscr{B}_1} \{\omega_B = 1\} \cap \{\omega_{\partial^* B} = 0\}$$

Observing that for all $B \in \mathscr{B}_1$, the event $\{\omega_{\partial^* B} = 0\}$ decouples A_2 from the event $\{\omega_B = 1\}$, we obtain

$$\mu(A_1^{\text{fin}} \cap D \cap A_2) \leqslant \mu(A_1^{\text{fin}} \cap D)\, \mu(A_2) \leqslant \mu(A_1^{\text{fin}})\, \mu(A_2)$$

where we have used the second inequality of Proposition 2.6 in the first step. This completes the proof for the free measure.

In order to complete the proof for the wired measure, we will show

$$\mu(A_1^{\mathrm{inf}} \cap D \cap A_2) \leqslant \mu(A_1^{\mathrm{inf}})\, \mu(A_2)$$

To this end, we define

$$A_2^{\mathrm{fin}} = A_2 \cap \{z \not\leftrightarrow \partial\Lambda\}$$

Since the wiring would connect two points if they were both connected to the boundary, we have

$$A_1^{\mathrm{inf}} \cap D \cap A_2 = A_1^{\mathrm{inf}} \cap D \cap A_2^{\mathrm{fin}}$$

with probability one with respect to the wired measure. Applying the same strategy as before, we then obtain

$$\mu(A_1^{\mathrm{inf}} \cap D \cap A_2) = \mu(A_1^{\mathrm{inf}} \cap D \cap A_2^{\mathrm{fin}}) \leqslant \mu(A_1^{\mathrm{inf}})\, \mu(A_2^{\mathrm{fin}}) \leqslant \mu(A_1^{\mathrm{inf}})\, \mu(A_2)$$

as claimed. ▌

Remarks. 1. As can be seen from the above proof, the finite-volume free measure actually obeys the stronger inequality

$$\mu_{\mathrm{free},\Lambda}(\{x \leftrightarrow y\} \cap D \cap \{z \leftrightarrow w\}) \leqslant \mu_{\mathrm{free},\Lambda}(\{x \leftrightarrow y\} \cap D)\, \mu_{\mathrm{free},\Lambda}(z \leftrightarrow w)$$

2. Using uniqueness of the infinite cluster[5] (see also Proposition 2.4 above), it follows immediately that $A_1^{\mathrm{inf}} \cap D \cap A_2 = A_1^{\mathrm{inf}} \cap D \cap A_2^{\mathrm{fin}}$ with probability one with respect to the infinite-volume measures μ_{free} and μ_{wir}. Hence Proposition 2.8 holds for these measures as well.

2.4. DLR Equations and States of the Random Cluster Model

In this subsection, we introduce the notion of (unconstrained) DLR states for the random cluster model, prove that the free measure is such a state, and use this to show that it is ergodic—a property we will need in our subsequent analysis. It is usually straightforward to establish such results by invoking the general theory of Gibbs states (see, e.g., refs. 34 and 15). However, the general theory requires that the finite-volume expectations used to construct the DLR states are *quasilocal* functions of the boundary conditions, a property which fails to hold here due to the nonlocality of the random cluster weights. Thus the DLR equation has to be established explicitly.

We start by defining finite-volume measures with general unconstrained boundary conditions—conditions which permit any component to be connected to any other component. The set of states generated by all such boundary conditions is quite natural in the random cluster model. A larger class including constrained states will be discussed briefly at the end of this subsection. Each measure is defined on an arbitrary finite set of bonds $B \subset \mathbb{B}_d$ with boundary

$$\partial B = \{x \in \mathbb{Z}^d \mid \exists y, z \in \mathbb{Z}^d \text{ with } \langle x, y \rangle \in B, \langle x, z \rangle \in B^c\}$$

We specify the boundary condition by introducing a *wiring diagram* W, which is a disjoint partition of ∂B into $n_W = 1, ..., |\partial B|$ components:

$$W = \{W_1, ..., W_{n_W}\} \qquad \text{with} \quad \partial B = \bigcup_{i=1}^{n_W} W_i, \qquad W_i \cap W_j = \varnothing \qquad \text{if} \quad i \neq j$$

We denote by $\mathscr{W}(\partial B)$ the set of all such wiring diagrams, i.e., the set of all disjoint partitions of ∂B. Each component W_i of the wiring diagram W is considered to be preconnected or wired, so that all bonds $b \in B$ connected to points of W_i are regarded as being connected to each other. The number of components $\#(\omega)$ is then computed as usual. The random cluster weight

$$G_{W,B}(\omega) = \frac{1}{Z_W(B)} (1-p)^{|\omega^c|} p^{|\omega|} q^{\#(\omega)} \tag{2.32}$$

defines the finite-volume measure $\mu_{W,B}(\cdot)$. Denoting by W_{free} the partition with $n_W = |\partial B|$ components and by W_{wir} the partition with only a single component, we see that

$$\mu_{\text{free}, \Lambda}(\cdot) = \mu_{W_{\text{free}}, B(\Lambda)}(\cdot), \qquad \mu_{\text{wir}, \Lambda}(\cdot) = \mu_{W_{\text{wir}}, B^+(\Lambda)}(\cdot)$$

so that the free and (fully) wired measures are just special cases of $\mu_{W,B}(\cdot)$. Note that among the measures $\mu_{W,B}(\cdot)$ are some that cannot be obtained as transforms of any finite-volume states in the spin system, namely those in which W has more than q components W_i with $|W_i| \geq 2$.

There is a natural partial order on the set $\mathscr{W}(\partial B)$. If $W, W' \in \mathscr{W}(\partial B)$, we say that W' is *coarser* than W, denoted by $W' \succ W$, if for each $W'_i \in W'$ there exist $W_{i_1}, W_{i_2}, ..., W_{i_m} \in W$ such that $W'_i = \bigcup_{j=1}^{m} W_{i_j}$. Notice that W_{free} is the least coarse and W_{wir} is the most coarse of all wiring diagrams. Moreover, if $W' \succ W$ then $\mu_{W',B}$ dominates $\mu_{W,B}$ in the sense of FKG (see Definition 2.1 above).

Each configuration $\omega \in \Omega$ induces a wiring diagram on each finite set $B \subset \mathbb{B}_d$. The *induced wiring diagram* $W(B, \omega)$ is a partition into components of ∂B, each of which is connected using occupied bonds in ω_{B^c}. Thus each $\omega \in \Omega$ gives rise to a sequence of induced finite-volume measures $\mu_{W(B,\omega),B}$ for any increasing sequence of sets $B \subset \mathbb{B}_d$. Henceforth we will extend the induced finite-volume measure $\mu_{W(B,\omega),B}$ to a measure on the full space $(\Omega, \mathscr{F})$ by declaring all bonds in B^c to have the configuration specified by ω. [Compare this to our extensions of of $\mu_{\text{free},\Lambda}$ and $\mu_{\text{wir},\Lambda}$ discussed in the remark following Eq. (2.16).] Using the form (2.32) of the weights $G_{W,B}$ and our definition of induced wiring diagrams, it is straightforward to check that the (extended) induced finite-volume measures obey the *consistency condition*

$$\mu_{W(B,\omega),B}(A) = \int \mu_{W(B,\omega),B}(d\tilde{\omega})\, \mu_{W(\tilde{B},\tilde{\omega}),\tilde{B}}(A) \tag{2.33}$$

for all local events $A \in \mathscr{F}$, any finite set B, and all $\tilde{B} \subset B$.

For each finite B, we may define the function $\pi_B\colon (\mathscr{F}, \Omega) \to \mathbb{R}$ by $\pi_B(A|\omega) = \mu_{W(B,\omega),B}(A)$. Since the family $\gamma = \{\pi_B | B \subset \mathbb{B}_d, |B| < \infty\}$ is a set of proper probability kernels obeying the consistency condition (2.33), γ is a *specification* in the sense of ref. 34.

A DLR equation[8,29] is just an infinite-volume analog of a consistency condition like (2.33). Thus we introduce the (unconstrained) DLR equation for an infinite-volume random cluster state μ:

$$\mu(A) = \int \mu(d\omega)\, \mu_{W(B,\omega),B}(A) \tag{2.34}$$

where $A \in \mathscr{F}$ is any local observable and $B \subset \mathbb{B}_d$ is any finite set. As usual, the DLR equation (2.34)—if it holds—allows us to write the infinite-volume expectation of A as an average over finite-volume expectations. It is closed in the sense that the average is computed with respect to the given measure μ. Note that this is different from the equation for states given in ref. 1, where a random cluster measure was obtained as a transform of a measure obeying the DLR equation in the spin system. On the other hand, a DLR equation was implicit in the discussion of states in ref. 20; there, however, the question of existence of solutions to the equation was not addressed.

Let us denote the set of states obeying (2.34) by $\mathscr{G} = \mathscr{G}(\gamma)$, where as above γ denotes the specification. States $\mu \in \mathscr{G}$ will be called *DLR states* or *Gibbs states. A priori* it is not clear whether $\mathscr{G}$ is nonempty, i.e., whether there exists any μ satsifying (2.34). One might try to construct such a μ as a subsequential limit of finite-volume measures $\mu_{W,B}$—which clearly exists by compactness—but the question of whether such a limit obeys (2.34) involves a delicate interchange of limits. The theory of Gibbs states[34,15] provides general conditions under which (2.34) is satisfied, one of which is quasilocality of the specification.

A function f is *quasilocal* if it can be approximated in the supremum norm by local functions, a property which is equivalent (ref. 15, Remark 2.21) to the statement

$$\sup_{\omega, \eta : \omega_B = \eta_B} |f(\omega) - f(\eta)| \to 0 \qquad \text{as} \quad B \to \mathbb{B}_d$$

A specification $\{\pi_B\}$ is quasilocal if the functions $\pi_B(A, \cdot)$ are quasilocal for all finite $B \subset \mathbb{B}_d$ and all local events $A \in \mathscr{F}$.

Unfortunately, due to nonlocality of the weights $G_{W, B}$, our specification is not quasilocal. For example, the probability of the simple event $\{\omega(b) = 1\}$, conditioned on the bonds in $\mathbb{B}_d \backslash \{b\}$, changes discontinuously depending on whether or not the endpoints of b are connected by a path (of any length) in $\mathbb{B}_d \backslash \{b\}$ [see eq. (2.24)]. The general theory of Gibbs states therefore cannot be applied here. However, we can verify the DLR equation (2.34) explicitly in the case of the free measure:

Proposition 2.9. For all $q \geqslant 1$ and $0 \leqslant \beta \leqslant \infty$, $\mu_{\text{free}} \in \mathscr{G}$.

Proof. Let $B \subset \mathbb{B}_d$ be a finite set and $A \in \mathscr{F}$ a local event. We wish to show

$$\mu_{\text{free}}(A) = \int \mu_{\text{free}}(d\omega)\, \mu_{W(B, \omega), B}(A) \tag{2.35}$$

By the finite-volume consistency condition (2.33) and convergence of the finite-volume measures (2.20), it suffices to prove

$$\lim_{\Lambda \to \mathbb{Z}_d} \int \mu_{\text{free}, \Lambda}(d\omega)\, \mu_{W(B, \omega), B}(A) = \int \mu_{\text{free}}(d\omega)\, \mu_{W(B, \omega), B}(A) \tag{2.36}$$

Inserting the partition of unity $\sum_{W \in \mathscr{W}(\partial B)} \mathbb{1}_{\{W(B, \omega) = W\}} = 1$ into (2.36) and noting that $\mu_{W, B}(A)$ is independent of Λ, we see that it is enough to prove

$$\lim_{\Lambda \to \mathbb{Z}_d} \mu_{\text{free}, \Lambda}(\{W(B, \omega) = W\}) = \mu_{\text{free}}(\{W(B, \omega) = W\}) \tag{2.37}$$

i.e., that the probability of a given wiring diagram converges.

Let $R_{W_i}(B^c)$ denote the event that all sites within the set W_i are connected to each other via bonds in B^c, let $S_W(B^c) = \bigcap_{W_i \in W} R_{W_i}(B^c)$, and let $N_{W_i, W_j}(B^c)$ denote the event that none of the sites in W_i is connected to any of the sites in W_j via bonds in B^c. Then

$$\{W(B, \omega) = W\} = S_W(B^c) \cap \bigcap_{\substack{W_i, W_j \in W \\ i \neq j}} N_{W_i, W_j}(B^c) \tag{2.38}$$

By inclusion-exclusion, it is not hard to show that if $B(\Lambda) \supset B$, then

$$\mu_{\text{free}, \Lambda}(\{W(B, \omega) = W\}) = \sum_{\substack{\tilde{W} \in \mathcal{W}(\partial B): \\ \tilde{W} \succ W}} k_{\tilde{W}}(W)\, \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B^c)) \tag{2.39}$$

where the sum is over $\tilde{W}$ coarser than W [see the definition a paragraph below (2.32)], and $k_{\tilde{W}}(W) \in \mathbb{Z}$ are computable coefficients with $k_W(W) = 1$. Thus by (2.37)–(2.39), we only need to show that for all $\tilde{W} \in \mathcal{W}(\partial B)$

$$\lim_{\Lambda \to \mathbb{Z}_d} \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B^c)) = \mu_{\text{free}}(S_{\tilde{W}}(B^c)) \tag{2.40}$$

Let $\tilde{W} \in \mathcal{W}(\partial B)$ and choose Λ such that $B(\Lambda) \supset B$. Due to the free boundary conditions on $\partial\Lambda$, the argument of the left-hand side of (2.40) can be rewritten as

$$\mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B^c)) = \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda)\backslash B)) \tag{2.41}$$

Approximating the wiring event $S_{\tilde{W}}(B^c)$ by local events, we see that the right hand side of (2.40) is actually a double limit:

$$\begin{aligned} \mu_{\text{free}}(S_{\tilde{W}}(B^c)) &= \lim_{\Lambda' \to \mathbb{Z}_d} \mu_{\text{free}}(S_{\tilde{W}}(B(\Lambda')\backslash B)) \\ &= \lim_{\Lambda' \to \mathbb{Z}_d} \lim_{\Lambda \to \mathbb{Z}_d} \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda')\backslash B)) \end{aligned} \tag{2.42}$$

Thus by (2.40)–(2.42), we must show

$$\lim_{\Lambda \to \mathbb{Z}_d} \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda)\backslash B)) = \lim_{\Lambda' \to \mathbb{Z}_d} \lim_{\Lambda \to \mathbb{Z}_d} \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda')\backslash B)) \tag{2.43}$$

In order to prove this, we note that for all $\Lambda' \subset \Lambda$

$$\mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda')\backslash B)) \leqslant \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda')\backslash B)) \leqslant \mu_{\text{free}, \Lambda}(S_{\tilde{W}}(B(\Lambda)\backslash B)) \tag{2.44}$$

where the first inequality follows from the monotonicity property (2.18) and the second is just a consequence of $S_{\tilde{W}}(B(\Lambda')\backslash B) \subset S_{\tilde{W}}(B(\Lambda)\backslash B)$ if $\Lambda' \subset \Lambda$. Taking the limits $\Lambda \to \mathbb{Z}_d$ and $\Lambda' \to \mathbb{Z}_d$, we find that Eq. (2.44) yields (2.43) and hence (2.35). ∎

Remarks. 1. The only property of the free measure that was used to reduce the proposition to Eq. (2.40) was convergence of the finite-volume measures. Thus the wired analog of (2.40)—i.e. convergence of the probability of the wiring events $S_{\tilde{W}}(B^c)$ with respect to the finite-volume wired measures—is sufficient to prove $\mu_{\text{wir}} \in \mathcal{G}$. Unfortunately, however, μ_{wir} does not have nice monotonicity properties like those in Eq (2.44).

2. Equation (2.43) of the proof is our first example of the problem of interchange of limits which arises again in Proposition 3.4 and in many theorems in Section 4. Whenever we deal with the infinite-volume limit of an event which is not confined to a finite volume, we encounter a double limit—one for the construction of the infinite-volume measure and the other for the approximation of the given event by local events. Hence the problem of interchange of limits. This problem does not arise in percolation because the measure is defined directly in the infinite-volume limit. Here, when we can deal with interchange, it is usually accomplished via either simple FKG monotonicity [Eqs. (2.18) and (2.19)] or our monotonicity involving decoupling events (corollary to Proposition 2.6).

It is now straightforward to show that μ_{free} is ergodic. We have:

Theorem 2.10. Let H be any nontrivial subgroup of the translation group and let $\mathcal{G}_o \subset \mathcal{G}$ be the set of all H-invariant DLR states. Then for all $q \geqslant 1$, μ_{free} is extremal in $\mathcal{G}_o$ and hence is H-ergodic.

Proof. As noted earlier, W_{free} is the least coarse of all wiring diagrams, so that

$$\mu_{W_{\text{free}}, B} \underset{\text{FKG}}{\leqslant} \mu_{W, B} \qquad \text{for all} \quad W \in \mathcal{W}(\partial B) \tag{2.45}$$

and thus by convergence of the measure (2.20)

$$\mu_{\text{free}} \underset{\text{FKG}}{\leqslant} \mu \qquad \text{for all} \quad \mu \in \mathcal{G} \tag{2.46}$$

Given that $\mu_{\text{free}} \in \mathcal{G}$ (Proposition 2.9), it follows immediately from (2.46) that μ_{free} is extremal in $\mathcal{G}$ and hence also in $\mathcal{G}_o$ (since μ_{free} is of course H-invariant). Ergodicity then follows from the fact that all extremal measures in $\mathcal{G}_o$ are H-ergodic (ref. 34, Theorem 4.1). ∎

Remarks. 1. *FKG Ordering of States*: Using the fact that the wired state is the coarsest of all states, we have analogs of (2.45) and (2.46) for the wired measure, and thus

$$\mu_{\text{free}} \underset{\text{FKG}}{\leqslant} \mu \underset{\text{FKG}}{\leqslant} \mu_{\text{wir}} \qquad \text{for all} \quad \mu \in \mathcal{G} \tag{2.47}$$

Note of course that this does not imply $\mu_{\text{wir}} \in \mathcal{G}$.

2. *The Size of $\mathcal{G}$*: By Proposition 2.9, $\mu_{\text{free}} \in \mathcal{G}$, so that $|\mathcal{G}| \geqslant 1$ for all $q \geqslant 1$ and all inverse temperatures β. Let $P_\infty^{\text{wir}}(\beta)$ denote the percolation probability in the wired measure, which of course coincides with the magnetization for integer q. According to a result of ref. 1 (Theorem A.2),

whenever $P_{\infty}^{\text{wir}}(\beta)=0$ (i.e., $\beta \leqslant \beta_t$ for systems with second-order transitions and $\beta < \beta_t$ for those with first-order transitions) $\mu_{\text{free}} = \mu_{\text{wir}}$, so that by (2.47) and Proposition 2.9, $|\mathscr{G}|=1$. It is expected that $|\mathscr{G}|=1$ also for $\beta > \beta_t$, but there are only incomplete results for $d=2$: The two-dimensional dual of the result of ref. 1 says $P_{\infty}^{\text{wir}}(\beta^*)=0$ implies $|\mathscr{G}|=1$, i.e. there is one state for $\beta > \beta_t^*$, which presumably coincides with β_t (see also ref. 20). However, one expects more states at the transition point in systems with first-order transitions. For q large enough and $d=2$, convergent expansions[25, 27] can be used to show that there are $q+1$ distinct translation-invariant spin states (which transform into two distinct translation-invariant random cluster states—the free and the wired). There are presumably no non-translation-invariant states. Thus we expect $|\mathscr{G}|=2$ for $\beta=\beta_t$ and q large enough in $d=2$. In three dimensions, convergent expansions[30] can be used to show that for q large enough, in addition to the translation-invariant states discussed above, there are infinitely many non-translation-invariant "Dobrushin-type" states corresponding here to states constructed from wiring diagrams which coincide with W_{wir} above a certain hyperplane and with W_{free} below that plane. We expect that these expansions can also be used to show that these non-translation-invariant states satisfy our DLR equation (2.34), so that at $\beta=\beta_t$, $|\mathscr{G}|=\infty$ for q large enough in $d \geqslant 3$, in contrast to the conjecture of ref. 20.

3. *States with Constraints*: In the remark above, we mentioned "Dobrushin-type" states which we expect to be in $\mathscr{G}$; these states were constructed from a combination of wired and free boundary conditions. There are, however, many Dobrushin-type states in the spin system whose transforms are not in $\mathscr{G}$—namely, mixed states in which various components of the boundary have different values of the spin. In the random cluster model, these correspond to states with constraints—certain components cannot be connected to other components. Therefore, in order to formulate DLR equations for these states, one has to supplement our wiring diagrams with some notion of constraints. While this is possible for individual finite-volume states, it is not clear how constraints should be induced by a given configuration $\omega \in \Omega$, nor whether the resulting measures would obey even finite-volume consistency conditions.

3. THE COVARIANCE MATRIX

3.1. The Random Cluster Representation of the Covariance Matrix

In this section, we rewrite the covariance matrices with free and constant boundary conditions,

$$G_{\text{free}}^{nm}(x-y)=\langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_{\text{free}} \tag{3.1}$$

and

$$G_c^{mn}(x-y)=\langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_c \tag{3.1}$$

in terms of the random cluster representation of Fortuin and Kasteleyn.[13] We do this by first deriving finite-volume expressions and then taking infinite-volume limits.

3.1.1. The Covariance Matrix in Finite Volume. Before deriving our representation for the covariance matrix, we recall the corresponding result for the (finite-volume) magnetization

$$M_x(\beta, \Lambda)=\frac{1}{q-1}\langle q\delta(\sigma_x, 0)-1\rangle_{0, \Lambda} \tag{3.3}$$

Using the symbol $X \leftrightarrow Y$ for the event that the set X is connected to the set Y by a finite path of occupied bonds, we see that the expression (2.15) almost immediately gives

$$M_x(\beta, \Lambda)=\mu_{\text{wir}, \Lambda}(x \leftrightarrow \partial\Lambda) \tag{3.4}$$

For future reference, we note that this can be easily generalized to the expectation of $e^{ip\sigma_x}$ with $p \in \hat{S}\backslash\{0\}=\{2\pi/q, ..., 2\pi(q-1)/q\}$. We obtain

$$\langle e^{ip\sigma_x}\rangle_{0, \Lambda}=\mu_{\text{wir}, \Lambda}(x \leftrightarrow \partial\Lambda) \qquad \text{if} \quad p \in \hat{S}\backslash\{0\} \tag{3.5}$$

We begin by considering the finite-volume two-point function with free boundary conditions,

$$G_{\text{free}, \Lambda}^{mn}(x, y)=\langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_{\text{free}, \Lambda} \tag{3.6}$$

Using the fact that $\langle q\delta(\sigma_x, m)\rangle_{\text{free}, \Lambda}=1$ for all m and all $x \in \Lambda$, we first rewrite $G_{\text{free}, \Lambda}^{nm}(x, y)$ as an untruncated expectation value

$$G_{\text{free}, \Lambda}^{mn}(x, y)=\langle (q\delta(\sigma_x, m)-1)(q\delta(\sigma_y, n)-1)\rangle_{\text{free}, \Lambda} \tag{3.7}$$

Now observe that

$$E_{\text{free}}((q\delta(\sigma_x, m)-1)(q\delta(\sigma_y, n)-1) \mid \omega)=0$$

if x and y are not connected in the configuration ω, while

$$E_{\text{free}}((q\delta(\sigma_x, m)-1)(q\delta(\sigma_y, n)-1) \mid \omega)=q\delta(m, n)-1$$

if x and y are connected. Thus, defining the connectivity in the FK representation

$$\tau_{\text{free}, \Lambda}(x, y) = \mu_{\text{free}, \Lambda}(x \leftrightarrow y) \tag{3.8}$$

we obtain the following:

Lemma 3.1. The finite-volume covariance matrix with free boundary conditions has the representation

$$G^{mn}_{\text{free}, \Lambda}(x, y) = (q\delta(m, n) - 1)\, \tau_{\text{free}, \Lambda}(x, y) \tag{3.9}$$

Remark. The result (2.9) in ref. 1 for the usual two-point function,

$$\frac{1}{q-1} \langle q\delta(\sigma_x, \sigma_y) - 1 \rangle_{\text{free}, \Lambda} = \tau_{\text{free}, \Lambda}(x, y)$$

is proportional to the trace of our expression (3.9).

Next, we rewrite the finite-volume covariance matrix with constant boundary conditions,

$$G^{mn}_{c, \Lambda}(x - y) = \langle q\delta(\sigma_x, m); q\delta(\sigma_y, n) \rangle_{c, \Lambda} \tag{3.10}$$

To this end, we define the finite-cluster connectivity

$$\tau^{\text{fin}}_{\text{wir}, \Lambda}(x, y) = \mu_{\text{wir}, \Lambda}(\{x \leftrightarrow y\} \cap \{x \not\leftrightarrow \partial\Lambda\}) \tag{3.11}$$

and the covariance of the events that x and y are connected to the boundary $\partial\Lambda$

$$\begin{aligned} C_{\text{wir}, \Lambda}(x, y) = {} & \mu_{\text{wir}, \Lambda}(\{x \leftrightarrow \partial\Lambda\} \cap \{y \leftrightarrow \partial\Lambda)\} \\ & - \mu_{\text{wir}, \Lambda}(x \leftrightarrow \partial\Lambda)\, \mu_{\text{wir}, \Lambda}(y \leftrightarrow \partial\Lambda) \end{aligned} \tag{3.12}$$

We have:

Lemma 3.2. The finite-volume covarlance matrix with constant boundary conditions has the representation

$$\begin{aligned} G^{mn}_{c, \Lambda}(x, y) = {} & (q\delta(m, n) - 1)\, \tau^{\text{fin}}_{\text{wir}, \Lambda}(x, y) \\ & + (q\delta(m, c) - 1)(q\delta(n, c) - 1)\, C_{\text{wir}, \Lambda}(x, y) \end{aligned} \tag{3.13}$$

Proof. By the symmetry of the model, it is enough to establish the lemma for $c=0$. In a first step, we prove a similar relation for $\langle e^{ip\sigma_x}; e^{-ip'\sigma_y}\rangle_{0,\Lambda}$ namely

$$\langle e^{ip\sigma_x}; e^{-ip'\sigma_y}\rangle_{0,\Lambda} = (1-\delta(p,0))(1-\delta(p',0))(C_{\mathrm{wir},\Lambda}(x,y)+\delta(p,p')\,\tau^{\mathrm{fin}}_{\mathrm{wir},\Lambda}(x,y)) \tag{3.14}$$

Assume w.l.o.g. that $p\neq 0$ and $p'\neq 0$, since otherwise $\langle e^{ip\sigma_x}; e^{-ip'\sigma_y}\rangle_{0,\Lambda}=0$. Then recalling the definition of truncated expectation values and observing that by (3.5)

$$\langle e^{ip\sigma_x}\rangle_{0,\Lambda}\langle e^{-ip'\sigma_y}\rangle_{0,\Lambda}=\mu_{\mathrm{wir},\Lambda}(x\leftrightarrow\partial\Lambda)\,\mu_{\mathrm{wir},\Lambda}(y\leftrightarrow\partial\Lambda)$$

the proof of (3.14) reduces to showing that

$$\langle e^{ip\sigma_x}e^{-ip'\sigma_y}\rangle_{0,\Lambda}=\mu_{\mathrm{wir},\Lambda}(\{x\leftrightarrow\partial\Lambda\}\cap\{y\leftrightarrow\partial\Lambda\})+\delta(p,p')\,\tau^{\mathrm{fin}}_{\mathrm{wir},\Lambda}(x,y) \tag{3.15}$$

We consider the cases $p=p'$ and $p\neq p'$ separately: If $p\neq p'$, the expectation $E_0(\exp(ip\sigma_x)\exp(-ip'\sigma_y)\,|\,\omega)$ is zero unless both x and y are connected to the boundary, in which case $E_0(\exp(ip\sigma_x)\exp(-ip\sigma_y)\,|\,\omega)=1$. As a consequence,

$$\langle e^{ip\sigma_x}e^{-ip'\sigma_y}\rangle_{0,\Lambda}=\mu_{\mathrm{wir},\Lambda}(\{x\leftrightarrow\partial\Lambda\}\cap\{y\leftrightarrow\partial\Lambda\})\qquad\text{if}\quad p\neq p' \tag{3.16}$$

If $p=p'$, we consider two cases: either $x\leftrightarrow\partial\Lambda$ in the configuration ω or $x\not\leftrightarrow\partial\Lambda$. In the first case, $E_0(\exp(ip\sigma_x)\exp(-ip'\sigma_y)\,|\,\omega)=1$ if $y\leftrightarrow\partial\Lambda$ as well, and $E_0(\exp(ip\sigma_x)\exp(-ip'\sigma_y)\,|\,\omega)=0$ if $y\not\leftrightarrow\partial\Lambda$, yielding a contribution of $\mu_{\mathrm{wir},\Lambda}(x\leftrightarrow\partial\Lambda$ and $y\leftrightarrow\partial\Lambda)$. In the second case, $E_0(\exp(ip\sigma_x)\exp(-ip'\sigma_y)\,|\,\omega)=1$ if $x\leftrightarrow y$ and $E_0(\exp(ip\sigma_x)\exp(-ip'\sigma_y)\,|\,\omega)=0$ if $x\not\leftrightarrow y$, yielding $\tau^{\mathrm{fin}}_{\mathrm{wir},\Lambda}(x,y)$. Thus

$$\langle e^{ip\sigma_x}e^{-ip'\sigma_y}\rangle_{0,\Lambda}=\mu_{\mathrm{wir},\Lambda}(\{x\leftrightarrow\partial\Lambda\}\cap\{y\leftrightarrow\partial\Lambda\}) + \tau^{\mathrm{fin}}_{\mathrm{wir},\Lambda}(x,y)\qquad\text{if}\quad p=p' \tag{3.17}$$

Equations (3.16) and (3.17) establish (3.15) and hence (3.14).

Given (3.14), the proof of the lemma is an easy exercise: observing that the delta functions $q\delta(\sigma_x,m)$ and $q\delta(\sigma_y,n)$ can be rewritten as

$$\sum_{p\in\mathcal{S}} e^{ip(\sigma_x-m)}\qquad\text{and}\qquad\sum_{p'\in\mathcal{S}} e^{-ip'(\sigma_y-n)}$$

respectively, we multiply both sides of (3.14) by $\exp[i(p'n-pm)]$ and sum over p and p' to obtain (3.13) for $c=0$. ∎

3.1.2. The Covariance Matrix in Infinite Volume. In this subsection, we extend our representations of the covariance matrix with free and wired boundary conditions to the infinite volume. To this end, we again denote by $C(x) = C(x; \omega)$ the set of occupied bonds connected to x in the configuration ω, and define the (translation-invariant) analogs of expression (3.8) for the connectivity,

$$\tau_{\text{free}}(x-y) = \mu_{\text{free}}(x \leftrightarrow y) \tag{3.18}$$

expression (3.11) for the finite-cluster connectivity,

$$\tau_{\text{wir}}^{\text{fin}}(x-y) = \mu_{\text{wir}}(x \leftrightarrow y \text{ and } |C(x)| < \infty) \tag{3.19}$$

and expression (3.12) for the covariance,

$$C_{\text{wir}}(x-y) = \text{Cov}_{\mu_{\text{wir}}}(|C(x)| = \infty, |C(y)| = \infty) \tag{3.20}$$

where in general $\text{Cov}_\mu(A, B) = \mu(A \cap B) - \mu(A)\mu(B)$ is the covariance of events A and B with respect to a measure μ. We call the function $C_{\text{wir}}(x-y)$ defined in (3.20) the *infinite-cluster covariance*. Our infinite-volume representation is contained in:

Theorem 3.3. The covariance matrices $G_{\text{free}}^{mn}(x-y)$ and $G_c^{mn}(x-y)$ can be expressed as

$$G_{\text{free}}^{mn}(x-y) = (q\delta(m, n) - 1)\, \tau(x, y) \tag{3.21}$$

and

$$\begin{aligned} G_c^{mn}(x-y) = (q\delta(m, n) - 1)\, \tau_{\text{wir}}^{\text{fin}}(x-y) + (q\delta(m, c) - 1) \\ \times (q\delta(n, c) - 1)\, C_{\text{wir}}(x-y) \end{aligned} \tag{3.22}$$

Proof. Given the corresponding finite-volume statements in Lemmas 3.1 and 3.2, the theorem is an immediate consequence of Proposition 3.4 below.

Remark. • For the diagonal elements of the covariance matrix, an analog of Eq. (3.22) was already stated in Proposition 1.1 of ref. 23 with a proof referring to techniques developed in ref. 1. However, we do not see how these techniques, originally developed to treat the increasing events defining the magnetization, apply to two-point functions, in particular how they can be used to establish the infinite-volume limit for $\tau_{\text{wir}}^{\text{fin}}$.

Proposition 3.4. Let $q \geqslant 1$ be real, let $\tau_{\text{free}, \Lambda}(x, y)$, $\tau^{\text{fin}}_{\text{wir}, \Lambda}(x, y)$, and $C_{\text{wir}, \Lambda}(x, y)$ be the quantities defined in Eqs. (3.8), (3.11), and (3.12), and let $\tau_{\text{free}}(x-y)$, $\tau^{\text{fin}}(x-y)$, and $C_{\text{wir}}(x-y)$ be the corresponding infinite-volume quantities, defined in Eqs. (3.18)–(3.20). Then the infinite-volume limits of $\tau_{\text{free}, \Lambda}(x, y)$, $\tau^{\text{fin}}_{\text{wir}, \Lambda}(x, y)$, and $C_{\text{wir}, \Lambda}(x, y)$ exist, and

$$\tau_{\text{free}}(x-y) = \lim_{\Lambda \to \mathbb{Z}^d} \tau_{\text{free}, \Lambda}(x, y) \tag{3.23}$$

$$\tau^{\text{fin}}_{\text{wir}}(x-y) = \lim_{\Lambda \to \mathbb{Z}^d} \tau^{\text{fin}}_{\text{wir}, \Lambda}(x, y) \tag{3.24}$$

and

$$C_{\text{wir}}(x-y) = \lim_{\Lambda \to \mathbb{Z}^d} C_{\text{wir}, \Lambda}(x, y) \tag{3.25}$$

Remark. For local observables, the existence of the thermodynamic limit follows immediately from the FKG monotonicity properties (2.18) and (2.19)—see Eqs. (2.20) and (2.21). This, however, does not imply the relations (3.23)–(3.25), since the events in question are nonlocal; the relations can only be established after an interchange of limits. In the ordered phase, this interchange is not merely technical—it is related to the question of how the infinite cluster emerges from large clusters in a finite volume. Thus it depends sensitively on boundary conditions. For example, for a *free* boundary condition analog of the finite-cluster connectivity (3.19), an infinite-volume statement like (3.24) is actually false.

Proof. Introducing the event $R_{x,y}(\Lambda)$ that x and y are connected in $B(\Lambda)$, we obtain for the right hand side of (3.23)

$$\lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{free}, \Lambda}(x \leftrightarrow y) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda))$$

while the left hand side is

$$\mu_{\text{free}}(x \leftrightarrow y) = \lim_{\Lambda' \to \mathbb{Z}^d} \mu_{\text{free}}(R_{x,y}(\Lambda')) = \lim_{\Lambda' \to \mathbb{Z}^d} \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda'))$$

We therefore have to show that

$$\lim_{\Lambda' \to \mathbb{Z}^d} \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda')) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda)) \tag{3.26}$$

In order to prove (3.26), we combine the monotonicity property (2.18) with the fact that $R_{x,y}(\Lambda') \subset R_{x,y}(\Lambda)$ if $\Lambda' \subset \Lambda$ to get

$$\mu_{\text{free}, \Lambda'}(R_{x,y}(\Lambda')) \leqslant \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda')) \leqslant \mu_{\text{free}, \Lambda}(R_{x,y}(\Lambda)) \qquad \text{if} \quad \Lambda' \subset \Lambda \tag{3.27}$$

Taking the limits $\Lambda \to \mathbb{Z}^d$ and $\Lambda' \to \mathbb{Z}^d$, we find that the inequality (3.27) implies Eq. (3.26).

In order to prove (3.24), we consider the event $R^{\text{fin}}_{x,y}(\Lambda)$ that x and y are connected by a cluster $C(x) \subset B(\Lambda)$, as introduced in Proposition 2.7. Recalling that $\partial\Lambda \equiv \{x \notin \Lambda \mid \text{dist}(x, \Lambda) = 1\}$, we see that $R^{\text{fin}}_{x,y}(\Lambda)$ is the intersection of the event $R_{x,y}(\Lambda)$ with the event that x is not connected to $\partial\Lambda$. We claim that

$$\mu_{\text{wir},\Lambda'}(R^{\text{fin}}_{x,y}(\Lambda')) \leqslant \mu_{\text{wir},\Lambda}(R^{\text{fin}}_{x,y}(\Lambda')) \leqslant \mu_{\text{wir},\Lambda}(R^{\text{fin}}_{x,y}(\Lambda)) \qquad \text{if} \quad \Lambda' \subset \Lambda \tag{3.28}$$

As before, the second inequality follows from the fact that $R^{\text{fin}}_{x,y}(\Lambda') \subset R^{\text{fin}}_{x,y}(\Lambda)$ if $\Lambda' \subset \Lambda$, which implies that $\mu_{\text{wir},\Lambda}(R^{\text{fin}}_{x,y}(\Lambda'))$ is monotone increasing in Λ'. However, the monotonicity of $\mu_{\text{wir},\Lambda}(R^{\text{fin}}_{x,y}(\Lambda'))$ in Λ is less obvious because $R^{\text{fin}}_{x,y}(\Lambda')$ is neither an increasing nor a decreasing event. It is, however, an event of the form (2.27) considered in Proposition 2.6 and its corollary. Namely,

$$R^{\text{fin}}_{x,y}(\Lambda') = \bigcup_B \{C(x) = B\} = \bigcup_B \{\omega_B = 1\} \cap \{\omega_{\partial^* B} = 0\} \tag{3.29}$$

where the union goes over all connected sets $B \subset B(\Lambda')$ that join x to y, ω_B is the configuration ω restricted to the set B, and $\partial^* B$ is the set of all bonds in $\mathbb{B}_d \backslash B$ which are connected to B, as in the proof of Proposition 2.7. Thus by the corollary to Proposition 2.6, $\mu_{\text{wir},\Lambda}(R^{\text{fin}}_{x,y}(\Lambda'))$ is an increasing function of $\Lambda' \subset \Lambda$, which is actually stronger than the first inequality of (3.28). This completes the proof of (3.24).

In order to prove (3.25), we remark that it has been already shown in ref. 1 (Theorem 2.3c) that

$$\mu_{\text{wir}}(|C(x)| = \infty) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{wir},\Lambda}(x \leftrightarrow \partial\Lambda)$$

The proof of (3.25) therefore reduces to showing

$$\mu_{\text{wir}}(|C(x)| = \infty \text{ and } |C(y)| = \infty)) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{wir},\Lambda}(x \leftrightarrow \partial\Lambda \text{ and } y \leftrightarrow \partial\Lambda) \tag{3.30}$$

Proceeding as before, we now introduce $R^{\partial\Lambda}_{x,y}$ as the event that both x and y are connected to $\partial\Lambda$. With this notation, Eq. (3.30) can be rewritten as

$$\lim_{\Lambda' \to \mathbb{Z}^d} \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{wir},\Lambda}(R^{\partial\Lambda'}_{x,y}) = \lim_{\Lambda \to \mathbb{Z}^d} \mu_{\text{wir},\Lambda}(R^{\partial\Lambda}_{x,y}) \tag{3.31}$$

Using (2.19) instead of (2.18), and observing that $R^{\partial\Lambda'}_{x,y} \supset R^{\partial\Lambda}_{x,y}$ if $\Lambda' \subset \Lambda$, we obtain

$$\mu_{\mathrm{wir},\Lambda'}(R^{\partial\Lambda'}_{x,y}) \geqslant \mu_{\mathrm{wir},\Lambda}(R^{\partial\Lambda'}_{x,y}) \geqslant \mu_{\mathrm{wir},\Lambda}(R^{\partial\Lambda}_{x,y}) \qquad \text{if} \quad \Lambda' \subset \Lambda \tag{3.32}$$

As before, the proof is completed by taking the limits $\Lambda \to \mathbb{Z}^d$ and $\Lambda' \to \mathbb{Z}^d$. ∎

3.2. The Covariance Matrix and Its Eigenvalues

Here we analyze the structure of the covariance matrix

$$G^{mn}_b(x-y) = \langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_b \tag{3.33}$$

with free and constant boundary conditions, summarizing our results in Theorem 3.5 at the end of the section. Before discussing particular boundary conditions, we note that in general

$$\sum_m G^{mn}_b(x-y) = \sum_n G^{mn}_b(x-y) = 0 \tag{3.34}$$

which follows from the fact that any truncated expectation $\langle A; B\rangle_b$ vanishes if either A or B is constant, and from the obvious relation $\sum_{m\in S} \delta(\sigma_x, m) = 1$. In particular, this implies that, independent of boundary conditions, G^{mn}_b always has a trivial eigenvalue 0, corresponding to an eigenvector $\vec{v}_0 = (1, ..., 1) \in \mathbb{R}^q$.

Now consider the matrix with free boundary conditions

$$G^{mn}_{\mathrm{free}}(x-y) = \langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_{\mathrm{free}} \tag{3.35}$$

Due to the permutation symmetry of the Hamilton function (2.1) and the symmetry of the boundary conditions, all diagonal elements are equal, as are all off-diagonal elements. Combining this with the observation (3.34), we conclude that

$$G^{mn}_{\mathrm{free}}(x-y) = \left(\delta(m,n) - (1-\delta(m,n))\frac{1}{q-1}\right) G^{00}_{\mathrm{free}}(x-y) \tag{3.36}$$

Given (3.36), the matrix $G^{mn}_{\mathrm{free}}(x-y)$ is easily diagonalized. We find one trivial eigenvalue 0, corresponding to an eigenvector $\vec{v}_0 = (1, ..., 1)$, and one $(q-1)$-fold degenerate eigenvalue

$$G_{\mathrm{free}}(x-y) = \frac{q}{q-1} G^{00}_{\mathrm{free}}(x-y) = \frac{q}{q-1} \langle q\delta(\sigma_x, \sigma_y) - 1\rangle_{\mathrm{free}} \tag{3.37}$$

corresponding to the $(q-1)$-dimensional eigenspace orthogonal to $\vec{v}_0$. In the second equality in (3.37), we have reexpressed $G^{00}_{\text{free}}(x-y)$ as the usual two-point function.

Remark. The above results imply that, in the free boundary condition case, the covariance matrix of the q-state Potts model does not contain more information than the standard two-point function. As we will see below [and as should be clear from the fact that $G^{mn}(x-y)$ always has one trivial eigenvalue], the same is true of the covariance matrix of the Ising model $(q=2)$ with constant boundary conditions. This may explain why the covariance matrix has not been more widely studied previously. However, as we shall see below and in subsequent sections, the $q>3$-state matrix with constant boundary conditions does have additional content, and this content has a clear stochastic geometric interpretation.

Next we analyze the covariance matrix with constant boundary conditions,

$$G^{mn}_c(x-y) = \langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_c \tag{3.38}$$

Starting with the special case $q=2$, we use (3.34) to conclude that

$$G^{00}_c(x-y) = G^{11}_c(x-y) = -G^{01}_c(x-y) = -G^{10}_c(x-y)$$

Combined with the fact that $G^{00}_0(x-y) = G^{11}_1(x-y)$ by the symmetry of the model, we obtain

$$G^{nm}_c(x-y) = (\delta(m,n) - (1-\delta(n,m))\, G^{00}_0(x-y) \qquad \text{for} \quad q=2 \tag{3.39}$$

Observing that the matrix structure of (3.39) is identical to that of (3.36) with $q=2$, we see that we again obtain a trivial eigenvalue of 0 and an eigenvalue

$$G^{(1)}_{\text{wir}}(x-y) = G^{00}_0(x-y) = \langle s_x; s_y\rangle_0 \qquad \text{for} \quad q=2 \tag{3.40}$$

Here we have rewritten $G^{00}_0(x-y)$ in terms of standard Ising spins $s_x = 2\delta(\sigma_x, 0) - 1$.

For $q \neq 2$, the matrix structure of $G^{mn}_c(x-y)$ is less trivial. Using relation (3.34) and the fact that constant boundary conditions $c \in S$ leave the symmetry of permutations among elements of $S\backslash\{c\}$ unbroken, it is easy to show that there are only two independent matrix elements. Taking these to be $G^{00}_0(x-y)$ and $G^{11}_0(x-y)$, we obtain

$$
\begin{aligned}
&G_c^{cc}(x-y)=G_0^{00}(x-y) \qquad \text{if} \quad c\in S\\
&G_c^{nn}(x-y)=G_0^{11}(x-y) \qquad \text{if} \quad n\neq c\\
&G_c^{cn}(x-y)=G_c^{nc}(x-y)=-\frac{1}{q-1}G_0^{00}(x-y) \qquad \text{if} \quad n\neq c \qquad (3.41)\\
&G_c^{mn}(x-y)=\frac{1}{(q-1)(q-2)}G_0^{00}(x-y)-\frac{1}{q-2}G_0^{11}(x-y)\\
&\qquad\qquad \text{if} \quad n\neq m, \quad n,m\neq c
\end{aligned}
$$

In order to diagonalize $G_c^{mn}(x-y)$, we begin by observing that the expectation $\langle\,\cdot\,\rangle_c$ is invariant under the group S_{q-1} of permutations of $S\backslash\{c\}$. Diagonalizing G_c^{mn} on the Hilbert space corresponding to the trivial representation of S_{q-1}, we identify two eigenvectors: $\vec{v}_0=(1,\ldots,1)$, corresponding to the simple eigenvalue zero, and $\vec{v}_1$, with components $(v_1)_m=q\delta(m,c)-1$, corresponding to the nontrivial simple eigenvalue

$$
G_{\mathrm{wir}}^{(1)}(x-y)=\frac{q}{q-1}G_0^{00}(x-y)=\frac{q}{q-1}\langle q\delta(\sigma_x,0);q\delta(\sigma_y,0)\rangle_0 \qquad (3.42)
$$

On the remaining $(q-2)$-dimensional subspace orthogonal to $\vec{v}_0$ and $\vec{v}_1$, we finally obtain the $(q-2)$-fold degenerate eigenvalue

$$
\begin{aligned}
G_{\mathrm{wir}}^{(2)}(x-y)&=\frac{q-1}{q-2}G_0^{11}(x-y)-\frac{1}{(q-1)(q-2)}G_0^{00}(x-y)\\
&=G_0^{11}(x-y)-G_0^{12}(x-y) \qquad (3.43)
\end{aligned}
$$

It is interesting to note that, as in (3.37), it is possible to express the eigenvalue $G_{\mathrm{wir}}^{(2)}(x-y)$ in terms of an *un*truncated expectation. Indeed, we may simply rewrite the second line in (3.43) as

$$
\begin{aligned}
G_{\mathrm{wir}}^{(2)}(x-y)&=\tfrac{1}{2}\langle (q\delta(\sigma_x,1)-q\delta(\sigma_x,2));(q\delta(\sigma_y,1)-q\delta(\sigma_y,2))\rangle_0\\
&=\tfrac{1}{2}\langle (q\delta(\sigma_x,1)-q\delta(\sigma_x,2))(q\delta(\sigma_y,1)-q\delta(\sigma_y,2))\rangle_0 \qquad (3.44)
\end{aligned}
$$

We both summarize the results of this section and establish their stochastic geometric significance in the following:

Theorem 3.5. Consider the q-state Potts model with $q\geqslant 2$. Then the free-boundary-condition covariance matrix $G_{\mathrm{free}}^{mn}(x-y)$ has the simple eigenvalue zero corresponding to the eigenvector $\vec{v}_0=(1,\ldots,1)$, and a $(q-1)$-fold degenerate eigenvalue

$$
G_{\mathrm{free}}(x-y)=\frac{q}{q-1}\langle q\delta(\sigma_x,\sigma_y)-1\rangle_{\mathrm{free}} \qquad (3.45)
$$

corresponding to the subspace orthogonal to $\vec{v}_0$. For $q \geqslant 3$ and all $c \in S$, the constant-boundary-condition covariance matrix $G_c^{mm'}(x-y)$ has the simple eigenvalue zero corresponding to the eigenvector $\vec{v}_0 = (1, ..., 1)$, a nontrivial simple eigenvalue

$$G_{\text{wir}}^{(1)}(x-y) = \frac{q}{q-1} \langle q\delta(\sigma_x, 0); q\delta(\sigma_y, 0)\rangle_0 \tag{3.46}$$

corresponding to the eigenvector $\vec{v}_1$, with components $(v_1)_m = q\delta(m, c) - 1$, where $\vec{v}_0$ and $\vec{v}_1$ belong to the trivial representation of the unbroken subgroup S_{q-1}, and a $(q-2)$-fold degenerate eigenvalue

$$G_{\text{wir}}^{(2)}(x-y) = \tfrac{1}{2}\langle (q\delta(\sigma_x, 1) - q\delta(\sigma_x, 2))(q\delta(\sigma_y, 1) - q\delta(\sigma_y, 2))\rangle_0 \tag{3.47}$$

corresponding to the subspace orthogonal to $\vec{v}_0$ and $\vec{v}_1$. For $q = 2$, the matrix $G_c^{mm'}(x-y)$ has only the trivial eigenvalue zero and the eigenvalue $G_{\text{wir}}^{(1)}(x-y)$.

Moreover, the eigenvalues $G_{\text{free}}(x-y)$, $G_{\text{wir}}^{(1)}(x-y)$, and $G_{\text{wir}}^{(2)}(x-y)$ can be expressed in the random cluster representation as

$$G_{\text{free}}(x-y) = q\tau_{\text{free}}(x-y) \tag{3.48}$$

$$G_{\text{wir}}^{(1)}(x-y) = q\tau_{\text{wir}}^{\text{fin}}(x-y) + q(q-1)\, C_{\text{wir}}(x-y) \tag{3.49}$$

and

$$G_{\text{wir}}^{(2)}(x-y) = q\tau_{\text{wir}}^{\text{fin}}(x-y) \tag{3.50}$$

Proof. It only remains to establish the random cluster representations (3.48)–(3.50) of the eigenvalues. But these follow immediately from expressions (3.37), (3.42), and (3.43) for the eigenvalues in terms of the matrix elements, and expressions (3.21) and (3.22) relating the matrix elements to the random cluster connectivities and cluster covariance. ∎

4. THE CORRELATION LENGTHS

4.1. Existence of the Lengths ξ_{free}, $\xi_{\text{wir}}^{(1)}$, and $\xi_{\text{wir}}^{(2)}$

In this subsection, we establish the existence of the limits, (1.7)–(1.9), using standard reflection positivity arguments. Namely, introducing the unit lattice vector $\hat{e}_1 = (1, 0, ..., 0) \in \mathbb{Z}^d$, we prove the following:

Theorem 4.1. Let $q \geqslant 2$ be an integer, and let $G(t)$ denote $G_{\text{free}}(t\hat{e}_1)$, $G_{\text{wir}}^{(1)}(t\hat{e}_1)$, or (for $q \geqslant 3$) $G_{\text{wir}}^{(2)}(t\hat{e}_1)$; see Theorem 3.5. Then $G(t)$ is a nonnegative, monotone decreasing, and log convex function of t, so that $(G(t)/G(0))^{1/t}$ is monotone increasing in t, the limit

$$\frac{1}{\xi} \equiv -\lim_{t \to \infty} \frac{\log G(t)}{t} \tag{4.1}$$

exists, and the function $G(t)$ obeys the *a priori* bound

$$G(t) \leqslant G(0)\, e^{-t/\xi} \tag{4.2}$$

Furthermore, denoting the correlation lengths defined in (4.1) by ξ_{free}, $\xi^{(1)}_{\text{wir}}$, and $\xi^{(2)}_{\text{wir}}$, we have

$$\xi^{(1)}_{\text{wir}} \geqslant \xi^{(2)}_{\text{wir}} \tag{4.3}$$

Proof. We start with the observation that each $G(t)$ can be written as a truncated (infinite-volume) expectation of the form

$$G(t) = \langle A; T^t A \rangle_b \tag{4.4}$$

where b denotes either free or constant boundary conditions, $A = A(\sigma_x)$ is an observable which depends on the spin variable σ_x of a single point $x \in \mathbb{Z}^d$, and $T^t A$ is the translation of the observable A by $t\hat{e}_1$. Equation (4.4) follows from (3.37) and (3.35) for G_{free}, from (3.42) for $G^{(1)}_{\text{wir}}$, and from (3.44) for $G^{(2)}_{\text{wir}}$.

Due to the reflection positivity of the model (see Appendix A for a review of the basic ideas), T can be represented as a non-negative contraction $(0 \leqslant T \leqslant 1)$ on a Hilbert space $\mathscr{H}$, and

$$G(t) = (\psi, T^t \psi)$$

for a suitable vector $\psi \in \mathscr{H}$. Obviously, this implies that $G(t)$ is a monotone decreasing, nonnegative function of t. By the Cauchy-Schwarz inequality,

$$G(\tfrac{1}{2}(t_1 + t_2)) = (T^{t_1/2}\psi, T^{t_2/2}\psi) \leqslant [G(t_1)\, G(t_2)]^{1/2}$$

so that $G(t)$ is log convex. Noting $0 < G(0) < \infty$, this implies that $(G(t)/G(0))1/t$ is monotone increasing in t, which in turn immediately implies existence of the limit and the *a priori* bound. Finally, (4.3) follows immediately from the representation in Theorem 3.5 and existence of the limits. ▌

Remark. For G_{free} and $G^{(2)}_{\text{wir}}$, the existence of the corresponding correlation lengths can also be established by subadditivity arguments (see Section 4.2 below). While these arguments are more involved than the reflection positivity proof presented above, they have the advantage that they give existence of limits analogous to (1.7) and (1.9) for noninteger values of q, defined directly in terms of $\tau_{\rm free}$ and $\tau_{\rm wir}^{\rm fin}$ [cf. Eqs. (3.48) and (3.50)]. Moreover, a slight variation of these arguments can be used to establish left continuity of the inverse correlation length $1/\xi_{\rm free}(\beta)$ (Theorem 4.2) and upper semicontinuity of $1/\xi_{\rm wir}^{(2)}$ (Theorem 4.3). On the other hand, subadditivity does not establish log convexity of $G(t)$ and hence monotonicity of the full sequence $(G(t)/G(0))^{1/t}$, as we have from the above theorem for integer q. Furthermore, we are not aware of any proof of the existence of $\xi_{\rm wir}^{(1)}$ which does not involve reflection positivity.

4.2. Equivalent Characterizations of $\xi_{\rm free}$, $\xi_{\rm wir}^{(1)}$, and $\xi_{\rm wir}^{(2)}$

We already have stochastic geometric representations for the correlation lengths $\xi_{\rm free}$ and $\xi_{\rm wir}^{(2)}$ as the decay rates of of $\tau_{\rm free}$ and $\tau_{\rm wir}^{\rm fin}$—see Theorem 3.5. In this subsection, we provide a stochastic geometric representation for $\xi_{\rm wir}^{(1)}$ (Theorem 4.4) and give alternative representations for $\xi_{\rm free}$ and $\xi_{\rm wir}^{(2)}$ (Theorem 4.3, Lemmas 4.6–4.8). On the one hand, these alternative representations allow us to prove several results on the behavior of $\xi_{\rm free}$ and $\xi_{\rm wir}^{(2)}$, in particular left continuity of $1/\xi_{\rm free}(\beta)$ (Theorem 4.2), upper semicontinuity of $1/\xi_{\rm wir}^{(2)}$ (Theorem 4.3), and the two-dimensional dichotomy (1.11) and (1.12) involving $\xi_{\rm free}$ and $\xi_{\rm wir}^{(2)}$ discussed in the introduction. On the other hand, the alternative representations may be of interest for numerical determinations—in particular the representation of $\xi_{\rm wir}^{(2)}$ in terms of the probability $\tau_{\rm wir}^{\rm diam}(n)$ that the diameter of the cluster $C(0)$ is n (Theorem 4.3). The representation of $\xi_{\rm wir}^{(1)}$ as the decay rate of the covariance $C_{\rm wir}$, provided the magnetization $M(\beta)>0$, may be of interest both to mathematicians and numerical physicists. It is worth noting that many of the results of this subjection are generalizations of corresponding percolation results of ref. 7 to $q \geqslant 1$, but the proofs are quite different due to the lack of independence, the lack of a BK inequality, and the presence of boundary conditions.

Some of the results in this section (and most of the proofs) are of a rather technical nature. In particular, we introduce many connectivity functions and ultimately show that they have only a few independent decay rates. However, in the process, the notation and the arguments become rather cumbersome. In order to simplify matters, we first introduce only a few "physical" connectivity functions and summarize the results of independent interest on $\xi_{\rm free}$, $\xi_{\rm wir}^{(2)}$ and $\xi_{\rm wir}^{(1)}$ in Theorems 4.2–4.4, respectively. The remainder of the subsection is devoted both to the proof of these results and to the statement and proof of several more technical results which we will need for our proof of the two-dimensional dichotomy in Section 5.

We start with a few definitions. For $b=\text{wir}$ or free, we introduce the on-axis connectivity function

$$\tau_b(L\hat{e}_1)=\mu_b(0\leftrightarrow L\hat{e}_1) \tag{4.5}$$

the on-axis finite-cluster connectivity

$$\tau_b^{\text{fin}}(L\hat{e}_1)=\mu_b(\{0\leftrightarrow L\hat{e}_1\}\cap\{|C(0)|<\infty\}) \tag{4.6}$$

the diameter function

$$\tau_b^{\text{diam}}(L)=\mu_b(\text{diam } C(0)=L) \tag{4.7}$$

where diam $C(0)$ denotes the diameter of the cluster $C(0)$ in the ℓ_∞ norm, i. e., the maximum diameter in any of the d coordinate directions, and the covariance

$$C_b(x-y)=\mu_b(\{|C(x)|=\infty\}\cap\{|C(y)|=\infty\})-P_\infty^b(\beta)^2 \tag{4.8}$$

where

$$P_\infty^b(\beta)=\mu_b(|C(0)|=\infty) \tag{4.9}$$

We denote the corresponding correlation lengths—whenever they exist—by ξ_b, ξ_b^{fin}, ξ_b^{diam}, and ξ_b^C.

Theorem 4.2. Let $0\leqslant\beta\leqslant\infty$ and $q\geqslant 1$. Then the correlation lenths ξ_{wir} and ξ_{free} exist, $\xi_{\text{free}}\leqslant\xi_{\text{wir}}$, and $1/\xi_{\text{free}}$ is a left continuous function of β.

Theorem 4.3. Let $0<\beta<\infty$ and $q\geqslant 1$. Then for $b=$ wir or free, the correlation lengths ξ_b^{fin} and ξ_b^{diam} exist and are equal: $\xi_b^{\text{fin}}=\xi_b^{\text{diam}}$. Also $\xi_{\text{wir}}^{\text{fin}}\leqslant\xi_{\text{free}}^{\text{fin}}$, and $1/\xi_{\text{wir}}^{\text{fin}}$ is an upper semicontinuous function of β. If $q\geqslant 3$ is an integer, then in addition

$$\xi_{\text{wir}}^{(2)}=\xi_{\text{wir}}^{\text{fin}}=\xi_{\text{wir}}^{\text{diam}} \tag{4.10}$$

Remark. Combined with the obvious inequality $\xi_{\text{free}}^{\text{fin}}\leqslant\xi_{\text{free}}$, the inequalities from Theorem 4.2 and Theorem 4.3 give

$$\xi_{\text{wir}}^{\text{fin}}\leqslant\xi_{\text{free}}^{\text{fin}}\leqslant\xi_{\text{free}}\leqslant\xi_{\text{wir}}$$

provided $0<\beta<\infty$ and $q\geqslant 1$.

Theorem 4.4. Let $0 \leqslant \beta \leqslant \infty$ and $q \geqslant 2$ be an integer. Then the correlation length $\xi^C_{\rm wir}$ exists and

$$\xi^C_{\rm wir} = \xi^{(1)}_{\rm wir} \qquad \text{if} \quad M(\beta) > 0 \tag{4.11}$$

while

$$\xi^C_{\rm wir} = 0 \quad \text{and} \quad \xi^{(1)}_{\rm wir} = \xi^{(2)}_{\rm wir} = \xi_{\rm free} \qquad \text{if} \quad M(\beta) = 0 \tag{4.12}$$

In order to prove Theorem 4.4, we use a proposition which may be of independent interest and is stated next.

Proposition 4.5. Let $0 \leqslant \beta \leqslant \infty$ and $q \geqslant 1$. Then

$$C_b(x-y) \geqslant \tau^{\rm fin}_b(x-y)\, P^b_\infty(\beta) \tag{4.13}$$

Proof of Proposition 4.5. The proof of this proposition is an easy generalization of the proof of the corresponding statement in ref. 7. For a set $B \subset \mathbb{B}_d$, let $P(B)$ be the set of points x such that $x \in \partial b$ for some bond $b \in B$. Denoting by $\mathscr{B}_x$ the family of finite connected subsets $B \subset \mathbb{B}_d$ for which $x \in P(B)$, we have

$$\begin{aligned}
C_b(x-y) &= \mu_b(|C(x)| < \infty)\, \mu_b(|C(y)| = \infty) \\
&\quad - \mu_b(\{|C(x)| < \infty)\} \cap \{|C(y)| = \infty\}) \\
&= \sum_{B \in \mathscr{B}_x} (\mu_b(C(x) = B)\, \mu_b(|C(y)| = \infty) \\
&\quad - \mu_b(\{C(x) = B\} \cap \{|C(y)| = \infty\})) \\
&= \sum_{B \in \mathscr{B}_x} (\mu_b(C(x) = B)(\mu_b(|C(y)| = \infty) \\
&\quad - \mu_b(|C(y)| = \infty \mid C(x) = B)) \\
&\geqslant \sum_{\substack{B \in \mathscr{B}_x: \\ y \in P(B)}} (\mu_b(C(x) = B)(\mu_b(|C(y)| = \infty) \\
&\quad - \mu_b(|C(y)| = \infty \mid C(x) = B)) \\
&= \sum_{\substack{B \in \mathscr{B}_x: \\ y \in P(B)}} \mu_b(C(x) = B)\, \mu_b(|C(y)| = \infty) = \tau^{\rm fin}_b(x-y)\, P^b_\infty(\beta)
\end{aligned}$$

where in the fourth step we have used that for all $B \in \mathscr{B}_x$

$$\{C(x) = B\} = \{\omega_B = 1\} \cap \{\omega_{\partial^* B} = 0\} \equiv A_2 \cap D$$

is an event of the form considered in the second inequality [part 2(i)] of Proposition 2.6, and hence

$$\mu_b(|C(y)|=\infty)-\mu_b(|C(y)|=\infty \mid C(x)=B)\geqslant 0$$

Here, as in the proof of Proposition 2.7, ω_B is the configuration ω restricted to B, and the boundary $\partial^* B$ of B is the set of all bonds in $\mathbb{B}_d \backslash B$ which are connected to B. ∎

Proof of Theorem 4.4. Since $M(\beta)=P_\infty^{\mathrm{wir}}(\beta)$, we have that for $M(\beta)>0$

$$\begin{aligned}
&(q-1)\, C_{\mathrm{wir}}(x-y)+\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)\\
&\quad\leqslant\left((q-1)+\frac{1}{M(\beta)}\right) C_{\mathrm{wir}}(x-y)\\
&\quad\leqslant\left(1+\frac{1}{(q-1)M(\beta)}\right)((q-1)\, C_{\mathrm{wir}}(x-y)+\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y))
\end{aligned}$$

by Proposition 4.5 and the fact that $\tau_{\mathrm{wir}}^{\mathrm{fin}}\geqslant 0$. Combined with Theorem 4.1., which guarantees the existence of the inverse correlation length

$$\frac{1}{\xi_{\mathrm{wir}}^{(1)}}=-\lim_{L\to\infty}\frac{\log((q-1)\, C_{\mathrm{wir}}(L\hat{e}_1)+\tau_{\mathrm{wir}}^{\mathrm{fin}}(L\hat{e}_1))}{L}$$

provided $q\geqslant 2$ is an integer, we obtain the statement of Theorem 4.4 for $M(\beta)>0$. On the other hand, if $M(\beta)=P_\infty^{\mathrm{wir}}(\beta)=0$, then $\tau_{\mathrm{wir}}^{\mathrm{fin}}=\tau_{\mathrm{wir}}$ and $C_{\mathrm{wir}}(x-y)=0$, which implies $\xi_{\mathrm{wir}}^{(1)}=\xi_{\mathrm{wir}}^{(2)}$ and $\xi_{\mathrm{wir}}^{C}=0$. Finally, $M(\beta)=0$ implies $\mu_{\mathrm{wir}}=\mu_{\mathrm{free}}$ (see Ref. 1), and hence $\xi_{\mathrm{wir}}^{(2)}=\xi_{\mathrm{free}}$. ∎

In order to prove Theorems 4.2 and 4.3, we will need several approximations to the connectivity functions $\tau_b(L\hat{e}_1)$ and $\tau_b^{\mathrm{fin}}(L\hat{e}_1)$. Additional approximations will be needed to prove the dichotomy (1.11) and (1.12) discussed in the introduction. Rather than introducing them as they arise, we define all of them here, so that the reader may more easily refer back to the definitions. We will consider several subsets of $\mathbb{Z}^d$, namely the the "cylinder"

$$H(L)=\{x\in\mathbb{Z}^d \mid 0\leqslant x_1\leqslant L\} \tag{4.14}$$

the "tunnel"

$$T(M)=\{x\in\mathbb{Z}^d \mid -M/2\leqslant x_i\leqslant (M+1)/2,\, i=2,\ldots,d\} \tag{4.15}$$

and the "box"

$$\Lambda(L, M) = T(M) \cap H(L) \tag{4.16}$$

We then consider the following approximations to $\tau_b(L\hat{e}_1)$ in the cylinder, tunnel, and box:

$$\tau_b^{\text{cyl}}(L) = \mu_{b,\, H(L)}(0 \leftrightarrow L\hat{e}_1) \tag{4.17}$$

$$\tau_b^{\text{tun}}(L, M) = \mu_{b,\, T(M)}(0 \leftrightarrow L\hat{e}_1) \tag{4.18}$$

and

$$\tau_b^{\text{box}}(L, M) = \mu_{b,\, \Lambda(L, M)}(0 \leftrightarrow L\hat{e}_1) \tag{4.19}$$

In all cases $b = \text{wir}$ or free. Assuming that they exist, we denote the corresponding correlation lengths by ξ_b^{cyl}, $\xi_b^{\text{tun}}(M)$, and $\xi_b^{\text{box}}(M)$. We also consider the several approximations to $\tau_b^{\text{fin}}(L\hat{e}_1)$, namely

$$\begin{aligned}\hat{\tau}_b^{\text{cyl}}(L) = \mu_{b,\, H(L)}(\{0 \leftrightarrow L\hat{e}_1\} \cap \{|C(0)| < \infty\}\\ \cap \{0 \not\leftrightarrow \partial H(L)\})\end{aligned} \tag{4.20}$$

$$\hat{\tau}_b^{\text{box}}(L, M) = \mu_{b,\, \Lambda(L, M)}(\{0 \leftrightarrow L\hat{e}_1\} \cap \{0 \not\leftrightarrow \partial \Lambda(L, M)\}) \tag{4.21}$$

$$\begin{aligned}\tilde{\tau}_b^{\text{cyl}}(L) = \mu_b(\{0 \leftrightarrow L\hat{e}_1\} \cap \{|C(0)| < \infty\}\\ \cap \{C(0) \subset B(H(L))\})\end{aligned} \tag{4.22}$$

and

$$\tilde{\tau}_b^{\text{box}}(L, M) = \mu_b(\{0 \leftrightarrow L\hat{e}_1\} \cap \{C(0) \subset B(\Lambda(L, M))\}) \tag{4.23}$$

and we denote the corresponding correlation lengths—whenever they exist—by $\hat{\xi}_b^{\text{cyl}}$, $\hat{\xi}_b^{\text{box}}(M)$, $\tilde{\xi}_b^{\text{cyl}}$, and $\tilde{\xi}_b^{\text{box}}(M)$.

We note that the distinction between (4.17)–(4.21) and (4.22), (4.23) is that in the former quantities the probabilities are computed with respect to measures that live on the relevant sets Λ, while in the latter the probabilites are computed with respect to the full measures μ_b, but the events in question occur in the relevant sets $B(\Lambda)$.

Our first lemma gives the equivalence of several definitions of the correlation length ξ_b and will be used at the end of this section to prove Theorem 4.2.

Lemma 4.6. Let $0 \leqslant \beta \leqslant \infty$ and $q \geqslant 1$. Let $\tau(L)$ denote $\tau_{\text{wir}}(L\hat{e}_1)$, $\tau_{\text{free}}(L\hat{e}_1)$, $\tau_{\text{free}}^{\text{tun}}(L, M)$, $\tau_{\text{free}}^{\text{cyl}}(L)$, or $\tau_{\text{free}}^{\text{box}}(L, M)$. Then the correlation length ξ

corresponding to $\tau(L)$ exists, and $\tau(L) \leqslant e^{-L/\xi}$. Furthermore, the correlation lengths $\xi^{\text{tun}}_{\text{free}}(M)$ and $\xi^{\text{box}}_{\text{free}}(M)$ are monotone decreasing in M, $\xi^{\text{tun}}_{\text{free}}(M) = \xi^{\text{box}}_{\text{free}}(M)$, and

$$\xi_{\text{free}} = \xi^{\text{cyl}}_{\text{free}} = \xi^{\text{tun}}_{\text{free}} = \xi^{\text{box}}_{\text{free}} \tag{4.24}$$

where $\xi^{\text{tun}}_{\text{free}} = \lim_{M \to \infty} \xi^{\text{tun}}_{\text{free}}(M)$ and $\xi^{\text{box}}_{\text{free}} = \lim_{M \to \infty} \xi^{\text{box}}_{\text{free}}(M)$.

Proof. Considering an arbitrary subset $\Lambda \subset \mathbb{Z}^d$ and two points x and y in Λ, we note that by the FKG inequality (2.17)

$$\mu_{b,\Lambda}(x \leftrightarrow y) \geqslant \mu_{b,\Lambda}(x \leftrightarrow z)\, \mu_{b,\Lambda}(z \leftrightarrow y) \tag{4.25}$$

for all $z \in \Lambda$; furthermore, by the FKG monotonicity (2.18)

$$\mu_{\text{free},\Lambda}(x \leftrightarrow y) \geqslant \mu_{\text{free},\Lambda'}(x \leftrightarrow y) \tag{4.26}$$

for all $\Lambda' \subset \Lambda$ containing x and y. Using these inequalities, one obtains subadditivity, and hence existence of the corresponding correlation length ξ, together with the *a priori* bound $\tau(L) \leqslant e^{-L/\xi}$ for all five connectivity functions $\tau(L)$ considered in the theorem. Observing that the monotonicity (4.26) implies the monotonicity of $\tau^{\text{box}}_{\text{free}}(L, M)$ and $\tau^{\text{tun}}_{\text{free}}(L, M)$ in M, one obtains the monotonicity of $\xi^{\text{tun}}_{\text{free}}(M)$ and $\xi^{\text{box}}_{\text{free}}(M)$ in M, as well as the justification of the interchange of limits

$$\lim_{M \to \infty} \lim_{L \to \infty} \frac{\log \tau^{\text{tun}}_{\text{free}}(L, M)}{L} = \lim_{L \to \infty} \lim_{M \to \in \infty} \frac{\log \tau^{\text{tun}}_{\text{free}}(L, M)}{L}$$

and similarly for $\tau^{\text{box}}_{\text{free}}$. The only additional ingredient needed in the proof of the equalities

$$\xi_{\text{free}} = \lim_{M \to \infty} \xi^{\text{tun}}_{\text{free}}(M) \qquad \text{and} \qquad \xi^{\text{cyl}}_{\text{free}} = \lim_{M \to \infty} \xi^{\text{box}}_{\text{free}}(M)$$

is that $\mu_{\text{free},T(M)}(x \leftrightarrow y)$ converges to $\mu_{\text{free}}(x \leftrightarrow y)$ (and similarly for $\tau^{\text{box}}_{\text{free}}$), which is established in the same way as (3.23).

We are left with the proof of the equalities $\xi_{\text{free}} = \xi^{\text{cyl}}_{\text{free}}$ and $\xi^{\text{tun}}_{\text{free}}(M) = \xi^{\text{box}}_{\text{free}}(M)$. To this end, we use (4.25) and (4.26) to get the bound

$$\begin{aligned} \mu_{\text{free}}(o \leftrightarrow nL\hat{e}_1) &\geqslant \mu_{\text{free},H(nL)}(0 \leftrightarrow nL\hat{e}_1) \\ &\geqslant \prod_{i=0}^{n-1} \mu_{\text{free},H(nL)}(iL\hat{e}_1 \leftrightarrow (i+1)\, L\hat{e}_1) \end{aligned} \tag{4.27}$$

Taking the limit $n \to \infty$, and noting that all but, say, $\sqrt{n}$ terms on the right-hand side have arguments which are sufficiently far from the boundary, we obtain

$$e^{-L/\xi_{\text{free}}} \geqslant e^{-L/\xi_{\text{free}}^{\text{cyl}}} \geqslant \tau_{\text{free}}(L\hat{e}_1)$$

which, in the limit $L \to \infty$, implies that $\xi_{\text{free}} = \xi_{\text{free}}^{\text{cyl}}$. The equality of $\xi_{\text{free}}^{\text{tun}}(M)$ and $\xi_{\text{free}}^{\text{box}}(M)$ is proved in the same way. ∎

The next two lemmas give several useful relations between the correlation lengths corresponding to (4.20)–(4.23), and are important ingredients for the proof of Theorem 4.3 (see below) and for the proof of the dichotomy (1.11) and (1.12) (see Section 5). In order to state the first of these two lemmas, we introduce for each $x_\perp \in \mathbb{Z}^{d-1} \cap [-M, M]^{d-1}$ the off-axis connectivity function in the box

$$\tilde{\tau}_b^{\text{box}}(L, M; x_\perp) = \mu_b(\{0 \leftrightarrow (L, x_\perp)\} \cap \{C(0) \subset B(\Lambda(L, M))\}) \quad (4.28)$$

and for each $x_\perp \in \mathbb{Z}^{d-1}$ the off-axis connectivity function in the cylinder

$$\tilde{\tau}_b^{\text{cyl}}(L; x_\perp) = \mu_b(\{0 \leftrightarrow (L, x_\perp)\} \cap \{|C(0)| < \infty\} \cap \{C(0) \subset B(H(L))\}) \quad (4.29)$$

We note that $A_M = \{C(0) \subset B(\Lambda(L, M))\}$ is an increasing sequence of events which converges to the event $\{|C(0)| < \infty\} \cap \{C(0) \subset B(H(L))\}$. As a consequence,

$$\tilde{\tau}_b^{\text{box}}(L, M; x_\perp) \nearrow \tilde{\tau}_b^{\text{cyl}}(L; x_\perp) \qquad \text{as} \quad M \nearrow \infty \quad (4.30)$$

Lemma 4.7. Let $0 < \beta < \infty$ and $q \geqslant 1$. Then for $b =$ wir or free, the correlation lengths $\tilde{\xi}_b^{\text{cyl}}$ and $\tilde{\xi}_b^{\text{box}}(M)$ corresponding to the connectivity functions (4.22) and (4.23), as well as the limit $\tilde{\xi}_b^{\text{box}} = \lim_{M \to \infty} \tilde{\xi}_b^{\text{box}}(M)$, exist and

$$\tilde{\xi}_b^{\text{cyl}} = \tilde{\xi}_b^{\text{box}} \quad (4.31)$$

In addition,

$$\tilde{\tau}_b^{\text{box}}(L, M; x_\perp) \leqslant C(\beta, q) \exp[-L/\tilde{\xi}_b^{\text{box}}(M)] \leqslant C(\beta, q) \exp(-L/\tilde{\xi}_b^{\text{box}}) \quad (4.32)$$

and

$$\tilde{\tau}_b^{\text{cyl}}(L; x_\perp) \leqslant C(\beta, q) \exp(-L/\tilde{\xi}_b^{\text{cyl}}) \quad (4.33)$$

where $C(\beta, q) < \infty$ is continuons as a function of β and independent of L and M.

Remark. We will later show that $\tilde{\xi}_b^{\text{box}} = \xi_b^{\text{fin}}$ [see Eq. (4.53)] so that Lemma 4.7, together with Theorem 4.3, gives us yet another characterization of $\xi_{\text{wir}}^{(2)}$.

Proof. In order to prove the lemma, we first establish the existence of a constant $C(\beta, q) < \infty$ such that for all $x_\perp \in \mathbb{Z}^{d-1} \cap [-M, M]^{d-1}$,

$$\tilde{\tau}_b^{\text{box}}(L_1, M; x_\perp)\, \tilde{\tau}_b^{\text{box}}(L_2, M; x_\perp) \leqslant C(\beta, q)\, \tilde{\tau}_b^{\text{box}}(L_1 + L_2 + 1, M; 0) \tag{4.34}$$

To this end, we first rewrite the left-hand side of (4.34) in a form which allows us to apply Proposition 2.7. We consider the boxes

$$\begin{aligned} \Lambda_1 &= \{x \in T(M) \mid 0 \leqslant x_1 \leqslant L_1\} \\ \Lambda_2 &= \{x \in T(M) \mid L_1 + 1 \leqslant x_1 \leqslant L_1 + L_2 + 1\} \\ \Lambda &= \Lambda(L_1 + L_2 + 1, M) \end{aligned}$$

and the points $x_- = (L_1, x_\perp)$, $x_+ = (L_1 + 1, x_\perp)$, and $y = (L_1 + L_2 + 1, 0)$. Using the translation and reflection invariance of the measures μ_b, we then rewrite

$$\tilde{\tau}_b^{\text{box}}(L_1, M; x_\perp)\, \tilde{\tau}_b^{\text{box}}(L_2, M; x_\perp) = \mu_b(R_{0, x_-}^{\text{fin}}(\Lambda_1))\, \mu_b(R_{x_+, y}^{\text{fin}}(\Lambda_2)) \tag{4.35}$$

where, as in Proposition 2.7, $R_{0, x_-}^{\text{fin}}(\Lambda_1)$ is the event that 0 and x_- are connected by a cluster $C(0) \subset B(\Lambda_1)$, and similarly for $R_{x_+, y}^{\text{fin}}(\Lambda_2)$. Observing that $B^+(\Lambda_1) \cap B(\Lambda_2) = B(\Lambda_1) \cap B^+(\Lambda_2) = \varnothing$, we apply Proposition 2.7 to obtain

$$\mu_b(R_{0, x_-}^{\text{fin}}(\Lambda_1))\, \mu_b(R_{x_+, y}^{\text{fin}}(\Lambda_2)) \leqslant \mu_b(R_{0, x_-}^{\text{fin}}(\Lambda_1) \cap R_{x_+, y}^{\text{fin}}(\Lambda_2)) \tag{4.36}$$

Next we note that all configurations $\omega \in R_{0, x_-}^{\text{fin}}(\Lambda_1) \cap R_{x_+, y}^{\text{fin}}(\Lambda_2)$ would contribute to the event $R_{0, y}^{\text{fin}}(\Lambda)$ if the vacant bond $\langle x_-, x_+ \rangle$ were occupied. Using finite energy in the form (2.24), we therefore conclude that

$$\mu_b(R_{0, x_-}^{\text{fin}}(\Lambda_1) \cap R_{x_+, y}^{\text{fin}}(\Lambda_2)) \leqslant C(\beta, q)\, \mu_b(R_{0, y}^{\text{fin}}(\Lambda)) \tag{4.37}$$

for a suitable constant $C(\beta, q) < \infty$. Observing that

$$\mu_b(R_{0, y}^{\text{fin}}(\Lambda)) \equiv \tilde{\tau}_b^{\text{box}}(L_1 + L_2 + 1, M; 0)$$

we obtain the subadditivity bound (4.34). By standard arguments, the bound (4.34), together with the monotone convergence (4.30), implies the existence of the correlation lengths $\tilde{\xi}_b^{\text{cyl}}$ and $\tilde{\xi}_b^{\text{box}}(M)$, and the limit $\tilde{\xi}_b^{\text{box}} = \lim_{L \to \infty} \tilde{\xi}_b^{\text{box}}(M)$, the *a priori* bounds (4.32) and (4.33), and the equality of $\tilde{\xi}_b^{\text{cyl}}$ and $\tilde{\xi}_b^{\text{box}}$. Finally, we note that by the finite energy relation (2.24), without loss of generality $C(\beta, q)$ may be chosen to be a continuous function of β. ∎

Lemma 4.8. Let $0<\beta<\infty$ and $q\geqslant 1$. Then the correlation lengths $\hat{\xi}^{\mathrm{cyl}}_{\mathrm{wir}}$ and $\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}(M)$ corresponding to the connectivity functions (4.20) and (4.21), as well as the limit $\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}=\lim_{M\to\infty}\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}(M)$, exist and

$$\hat{\xi}^{\mathrm{cyl}}_{\mathrm{wir}}=\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}=\tilde{\xi}^{\mathrm{box}}_{\mathrm{wir}} \tag{4.38}$$

In addition,

$$\hat{\tau}^{\mathrm{box}}_{\mathrm{wir}}(L,M)\leqslant C(\beta,q)\exp[-L/\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}(M)]\leqslant C(\beta,q)\exp(-L/\hat{\xi}^{\mathrm{box}}_{\mathrm{wir}}) \tag{4.39}$$

and

$$\hat{\tau}^{\mathrm{cyl}}_{\mathrm{wir}}(L)\leqslant C(\beta,q)\exp(-L/\hat{\xi}^{\mathrm{cyl}}_{\mathrm{wir}}) \tag{4.40}$$

where $C(\beta,q)<\infty$ is continuous as a function of β and independent of L and M.

Remark. While Lemma 4.7 gives us alternative representations of ξ^{fin}_{b} in terms of decay rates of infinite-volume connectivities, this lemma —together with Eq. (4.53) and Theorem 4.3—gives us representations of $\xi^{(2)}_{\mathrm{wir}}$ in terms of finite-volume quantities.

Proof. Following the proof of Lemma 4.6, we first establish two inequalities analogous to (4.25) and (4.26). In order to state them, we introduce the cylinders

$$H(L_1,L_2)=\{x\in\mathbb{Z}^d\mid L_1\leqslant x_1\leqslant L_2\} \tag{4.41}$$

the boxes

$$\Lambda(L_1,L_2,M)=H(L_1,L_2)\cap T(M) \tag{4.42}$$

[with $T(M)$ as defined in (4.15)], the events

$$R^{\mathrm{fin}}_{L_1,L_2}(\Lambda)=\{L_1\hat{e}_1\leftrightarrow L_2\hat{e}_1\}\cup\{C(0)\subset B(\Lambda)\} \tag{4.43}$$

and in particular

$$\tilde{R}^{\mathrm{fin}}_{L_1,L_2}(M)=R^{\mathrm{fin}}_{L_1,L_2}(\Lambda(L_1,L_2,M)) \tag{4.44}$$

We then claim that for a suitable constant $C(\beta,q)<\infty$,

$$\mu_{\mathrm{wir},\Lambda}(\tilde{R}^{\mathrm{fin}\,\bullet}_{L_1,L_2}(M))\,\mu_{\mathrm{wir},\Lambda}(\tilde{R}^{\mathrm{fin}}_{L_2+1,L_3}(M))\leqslant C(\beta,q)\,\mu_{\mathrm{wir},\Lambda}(\tilde{R}^{\mathrm{fin}}_{L_1,L_3}(M)) \tag{4.45}$$

if $\Lambda\supset\Lambda(L_1,L_2,M)$,

$$\mu_{\mathrm{wir},\Lambda'}(R^{\mathrm{fin}}_{L_1,L_2}(\Lambda))\leqslant\mu_{\mathrm{wir},\Lambda''}(R^{\mathrm{fin}}_{L_1,L_2}(\Lambda)) \tag{4.46}$$

if $\Lambda'' \supset \Lambda' \supset \Lambda$, and

$$\mu_{\text{wir}, \Lambda}(R^{\text{fin}}_{L_1, L_2}(\Lambda')) \leqslant \mu_{\text{wir}, \Lambda}(R^{\text{fin}}_{L_1, L_2}(\Lambda'')) \tag{4.47}$$

if $\Lambda' \subset \Lambda'' \subset \Lambda$. While the first of these three inequalities is proved in exactly the same way as (4.34) using finite energy and Proposition 2.6, the second follows from Proposition 2.6 (and its corollary) alone since $R^{\text{fin}}_{L_1, L_2}(\Lambda)$ is an event of the form (2.27); see proof of Proposition 2.7 and Eq. (3.29). The last of the inequalities follows from the fact that $R^{\text{fin}}_{L_1, L_2}(\Lambda') \subset R^{\text{fin}}_{L_1, L_2}(\Lambda'')$ if $\Lambda' \subset \Lambda''$.

Given (4.45)–(4.47), the proof of Lemma 4.8 is analogous to that of Lemma 4.6, with

$$\begin{aligned} C(\beta, q)^{n-1} & \mu_{\text{wir}}(\tilde{R}_{0, (nL+n-1)}(M)) \\ & \geqslant C(\beta, q)^{n-1} \mu_{\text{wir}, \Lambda(0, nL+n-1, M)}(\tilde{R}_{0, (nL+n-1)}(M)) \\ & \geqslant \prod_{i=0}^{n-1} \mu_{\text{wir}, \Lambda(0, nL+n-1, M)}(\tilde{R}_{i(L+1), L+i(L+1)}(M)) \end{aligned} \tag{4.48}$$

replacing the inequality (4.27). ∎

We finally turn to the proofs of Theorems 4.2 and 4.3.

Proof of Theorem 4.2. The existence of the correlation lengths ξ_{free} and ξ_{wir} has already been established in Lemma 4.6, and the inequality $\xi_{\text{free}} \leqslant \xi_{\text{wir}}$ follows immediately from the FKG ordering (2.23), so all that remains to show is left continuity of $1/\xi_{\text{free}}(\beta)$. Due to Eq. (4.24), $1/\xi_{\text{free}}(\beta)$ is a limit of finite-volume (and hence continuous) quantities, namely

$$\frac{1}{\xi_{\text{free}}(\beta)} = -\lim_{M \to \infty} \lim_{L \to \infty} \frac{\log \tau^{\text{box}}_{\text{free}}(L, M)}{L} \tag{4.49}$$

As shown in the proof of Lemma 4.6, the finite-volume connectivity $\tau^{\text{box}}_{\text{free}}(L, M)$ is subadditive in L and monotone increasing in M. It is also monotone increasing and continuous in β. Choosing suitable subsequences, e.g., $L = 2^n$, and noting the minus sign in (4.49), this gives $1/\xi_{\text{free}}(\beta)$ as the limit of a decreasing sequence of continuous decreasing functions, and hence establishes the desired left continuity. ∎

Proof of Theorem 4.3. We start with the obvious bounds

$$\tilde{\tau}^{\text{box}}_b(L, M) \leqslant \tilde{\tau}^{\text{cyl}}_b(L) \leqslant \tau^{\text{fin}}_b(L) \leqslant \sum_{n \geqslant L} \tau^{\text{diam}}_b(n) \tag{4.50}$$

and

$$\tilde{\tau}^{\text{box}}_b(L, M) \leqslant \tau^{\text{diam}}_b(L) \qquad \text{for all} \quad M \leqslant L \tag{4.51}$$

Now consider the event $\{C(0)=B\}$, where B is some given set of diameter L. Using suitable rotations and translations by vectors $t\in\mathbb{Z}^d$, $|t|\leqslant L$, each such cluster B can be transformed into a cluster $\tilde{B}\subset\Lambda(L,2L)$ connecting the origin to a point $x=(l,x_\perp)$ in the boundary of $\Lambda(L,2L)$. As a consequence,

$$\begin{aligned}\tau_b^{\mathrm{diam}}(L)&\leqslant d(2L+1)^d\sum_{\substack{x\in\mathbb{Z}^{d-1}\\|x_\perp|\leqslant L}}\tilde{\tau}_b^{\mathrm{box}}(L,2L,x_\perp)\\&\leqslant d(2L+1)^{2d-1}\exp(-L/\tilde{\xi}_b^{\mathrm{box}})\end{aligned}\tag{4.52}$$

where we have used the *a priori* bound (4.32) of Lemma 4.7 in the last step.

Combining the bounds (4.51) and (4.52) with Lemma 4.7, we immediately obtain the existence of the correlation length ξ_b^{diam} and the equality of ξ_b^{diam} and $\tilde{\xi}_b^{\mathrm{box}}$. Combining the bounds (4.50) and (4.52) gives the existence of ξ_b^{fin} and the equality of ξ_b^{fin} and $\tilde{\xi}_b^{\mathrm{box}}$, provided $\tilde{\xi}_b^{\mathrm{box}}<\infty$. If, on the other hand, $\tilde{\xi}_b^{\mathrm{box}}=\tilde{\xi}_b^{\mathrm{cyl}}=\infty$, we use the bound $\tilde{\tau}_b^{\mathrm{cyl}}(L)\leqslant\tau_b^{\mathrm{fin}}(L)\leqslant 1$ to prove that the inverse correlation length exists and is equal to zero. Thus we have the existence of the correlation lengths ξ_b^{fin} and ξ_b^{diam} and the equality

$$\tilde{\xi}_b^{\mathrm{box}}=\xi_b^{\mathrm{fin}}=\xi_b^{\mathrm{diam}}\tag{4.53}$$

The final equivalence of Theorem 4.3, namely $\xi_{\mathrm{wir}}^{(2)}=\xi_{\mathrm{wir}}^{\mathrm{fin}}$ for integer $q\geqslant 3$, follows immediately from relation (3.50) of Theorem 3.5.

We are therefore left with the proofs of the inequality $\xi_{\mathrm{wir}}^{\mathrm{fin}}\leqslant\xi_{\mathrm{free}}^{\mathrm{fin}}$ and the upper semicontinuity of $1/\xi_{\mathrm{wir}}^{\mathrm{fin}}$. Noting that τ_b^{fin} is the probability of an event of the form considered in Proposition 2.6, the inequality follows immediately from the infinite-volume limit of (2.31). To prove the upper semicontinuity, we note that by Lemma 4.8 and Eq. (4.53), $1/\xi_{\mathrm{wir}}^{\mathrm{fin}}$ can be written as a limit of finite-volume quantities, namely

$$\frac{1}{\xi_{\mathrm{wir}}^{\mathrm{fin}}}=\frac{1}{\tilde{\xi}_{\mathrm{wir}}^{\mathrm{box}}}=\frac{1}{\hat{\xi}_{\mathrm{wir}}^{\mathrm{box}}}=-\lim_{M\to\infty}\lim_{L\to\infty}\frac{\log\hat{\tau}_{\mathrm{wir}}^{\mathrm{box}}(L,M)}{L}\tag{4.54}$$

Combined with the *a priori* bound (4.39) of Lemma 4.8, this implies

$$e^{-1/\xi_{\mathrm{wir}}^{\mathrm{fin}}}=\sup_{L,M}\left\{\left(\frac{\hat{\tau}_{\mathrm{wir}}^{\mathrm{box}}(L,M)}{C(\beta,q)}\right)^{1/L}\right\}$$

and hence

$$\frac{1}{\xi_{\mathrm{wir}}^{\mathrm{fin}}}=\inf_{L,M}\left\{\frac{\log C(\beta,q)-\log\hat{\tau}_{\mathrm{wir}}^{\mathrm{box}}(L,M)}{L}\right\}\tag{4.55}$$

Since both $\log C(\beta, q)$ and $-\log \hat{\tau}^{\text{box}}_{\text{wir}}(L, M)$ are continuous and hence upper semicontinuous functions β, and since the infimum of upper semicontinuous functions is upper semicontinuous, this establishes the upper semicontinuity of $1/\xi^{\text{fin}}_{\text{wir}}$. ∎

In the above proof, we actually obtained one additional equivalence which is not stated in Theorem 4.3, but which will be necessary in the proof of our dichotomy. Namely, by Eq. (4.53), we have:

Corollary. Let $0<\beta<\infty$ and $q \geqslant 1$. Then $\tilde{\xi}^{\text{box}}_b = \xi^{\text{fin}}_b$.

5. THE TWO-DIMENSIONAL DICHOTOMY AND RELATED RESULTS

5.1. Heuristics and Preliminaries

The goal of this section is a proof of the two-dimensional dichotomy, the principal part of which is the duality relation (1.11) for all β in the low-temperature regime. In this subsection, we discuss the heuristics of the relation, state our results, and briefly review two-dimensional duality in the random cluster model. In the next two subsections, we derive upper and lower bounds on $\tau^{\text{fin}}_{\text{wir}}$ and its approximations in terms of τ_{free} and approximations to $\tau^{\text{fin}}_{\text{free}}$. Finally, in the fourth subsection, we put these bounds together with the equivalence lemmas of Section 4 and the ergodicity theorem of Section 2 to prove the dichotomy.

In order to explain the heuristics of the duality relation (1.11), let us consider the representation of the random cluster model in terms of the order–disorder contours introduced in ref. 28 (see also ref. 4). In this representation, contours are defined as (the connected components of) the boundaries between regions of occupied bonds, regarded as ordered regions, and those of vacant bonds, regarded as disordered regions. Notice that in the wired measure, any finite cluster of occupied bonds must be separated from the infinite occupied cluster by a (disordered) region of vacant bonds. Thus all configurations contributing to $\tau^{\text{fin}}_{\text{wir}}(x-y)$ have at least two contours surrounding the points x and y—one being the boundary between the cluster connecting x and y and the disordered region, and the second being the boundary between the disordered region and the infinite cluster.

Let us begin by considering systems with first-order transitions at the transition point β_o. Since both the ordered and disordered phases are stable at β_0, the two contours need not remain near each other. Indeed, under similar circumstances, it is proved in ref. 30 that two such order–disorder interfaces tend to behave like independent interfaces, leading to a surface tension σ_{oo} between two ordered phases which is exactly twice the surface tension σ_{od} between an ordered and a disordered phase. Now, in our case, the minimal combined area of the two interfaces is $4\,|x-y|$. Moreover, we expect that large interfaces are suppressed at a first-order transition. Thus we expect $\tau_{\text{wir}}^{\text{fin}}(x-y)$ to decay exponentially with a rate $4\sigma_{od}=2\sigma_{oo}$, which would imply

$$1/\xi_{\text{wir}}^{\text{fin}}=2\sigma_{oo} \tag{5.1}$$

Obviously, this relation should also be satisfied trivially at β_o for systems with second-order transitions—both sides should vanish.

Now consider the regime $\beta>\beta_o$ in a system with either a first- or second-order transition. In this regime, the disordered phase is unstable, so that large regions of vacant bonds are suppressed. Thus the two contours surrounding x and y tend to bind together, leading to a single order–order interface surrounding the points of minimal area $2|x-y|$. This leads to a exponential decay with a rate $2\sigma_{oo}$ and hence again the relation (5.1).

Note that due to the duality relation $\sigma_{oo}(\beta)=1/\xi_{\text{free}}(\beta^*)$, Eq. (5.1) is equivalent to the desired relation (1.11)

It would be interesting to make the above heuristic arguments rigorous. While this could presumably be done for sufficiently large q, a direct translation of these heursitics into a proof for arbitrary q seems much more difficult. We therefore follow a different route, based on our inequalities involving decoupling events (Proposition 2.6) and the equivalences established in Section 4.

Before stating our main result, let us recall that the dual inverse temperature β^* is defined by

$$(e^{\beta}-1)(e^{\beta^*}-1)=q \tag{5.2}$$

Our main result is:

Theorem 5.1. Let $d=2$, $q\geqslant 1$ real, and $0<\beta<\infty$. Then either

$$P_{\infty}^{\text{free}}(\beta^*)=0 \qquad \text{and} \qquad \xi_{\text{wir}}^{\text{fin}}(\beta)=\tfrac{1}{2}\xi_{\text{free}}(\beta^*) \tag{5.3}$$

or

$$P_{\infty}^{\text{free}}(\beta^*)>0 \qquad \text{and} \qquad \xi_{\text{wir}}^{\text{fin}}(\beta)=\xi_{\text{free}}(\beta) \tag{5.4}$$

Remarks 1. If, in addition, $q\geqslant 3$ is an integer, it follows easily from the results of the last section and the duality relation (1.13) (a proof of which is given in Section 5.4) that (5.3) and (5.4) may be replaced by the dichotomy: Either

$$P_\infty^{\text{free}}(\beta^*) = 0 \qquad \text{and} \qquad \xi_{\text{wir}}^{(1)}(\beta) \geqslant \xi_{\text{wir}}^{(2)}(\beta) = \tfrac{1}{2}\xi_{\text{free}}(\beta^*) \tag{5.3$'$}$$

or

$$P_\infty^{\text{free}}(\beta^*) > 0 \qquad \text{and} \qquad \xi_{\text{wir}}^{(1)}(\beta) \geqslant \xi_{\text{wir}}^{(2)}(\beta) = \xi_{\text{free}}(\beta) \tag{5.4$'$}$$

2. It follows from the duality relation (1.13) and the monotonicity of P_∞^b [b = free or wir] as a function of β that the first branch of the dichotomy [i.e., (5.3) or (5.3′)] occurs when $\beta \geqslant \beta_o$, the self-dual point, and that the second branch [i.e., (5.4) or (5 4′)] occurs whenever $\beta < \beta_t = \inf\{\beta | M(\beta) > 0\}$. It is presumably the case that $\beta_o = \beta_t$, but rigorously this is only known for sufficiently large q (see, e.g., ref. 27, where this is shown for $q > 25$).

We close this subsection with a few remarks on duality. As usual, the dual site lattice $(\mathbb{Z}^*)^2$ is the set of points $x^* = (x_1^*, x_2^*) \in (\mathbb{Z} + \frac{1}{2})^2$ with half-integer coordinates, and the dual bond lattice $\mathbb{B}_2^*$ is the set of nearest neighbor bonds in $(\mathbb{Z}^*)^2$. To each bond $b \in \mathbb{B}_2$, there corresponds a dual bond $b^* \in \mathbb{B}_2^*$ which has the same midpoint as b. Similarly, to each configuration ω on $B \subset \mathbb{B}_2$, there corresponds a dual configuration ω^* on $B^* = \{b^* | b \in B\}$, given by

$$\omega^*(b^*) = \begin{cases} 0 & \text{if} \quad \omega(b) = 1 \\ 1 & \text{if} \quad \omega(b) = 0 \end{cases} \tag{5.5}$$

We will sometimes refer to the bonds $b^* \in \mathbb{B}_2^*$ for which $\omega^*(b^*) = 1$ as occupied dual bonds. Given a finite box

$$\Lambda = \{x \in \mathbb{Z}^2 \mid 0 \leqslant x_1 \leqslant L, \ -M/2 \leqslant x_2 \leqslant (M+1)/2\}$$

and the corresponding set of bonds $B^+(\Lambda)$, one defines the dual of Λ as

$$\Lambda^* = \{x \in (\mathbb{Z}^*)^2 \mid \exists y \in (\mathbb{Z}^*)^2 \text{ with } \langle x, y \rangle \in (B^+(\Lambda))^*\} \tag{5.6}$$

Note that in general Λ and Λ^* are not of the same cardinality. Using the appropriate Euler relation to relate $\#(\omega)$ to $\#(\omega^*)$, it is straightforward to check that for a given configuration ω on $B^+(\Lambda)$ and its dual ω^* on $B(\Lambda^*)$,

$$G_{\text{wir}, \beta, \Lambda}(\omega) = G_{\text{free}, \beta^*, \Lambda^*}(\omega^*) \tag{5.7}$$

where we have explicitly indicated the temperature dependence of the weights (2.14) and (2.16). Thus for each $A \in \mathscr{F}_{B^+(\Lambda)}$,

$$\mu_{\mathrm{wir},\,\beta,\,\Lambda}(A) = \mu_{\mathrm{free},\,\beta^*,\,\Lambda^*}(A^*) \tag{5.8}$$

where A^* is the event $A^* = \{\omega^* \mid \omega \in A\}$. In the next two subsections, we will often characterize events $A \in \mathscr{F}_{B^+(\Lambda)}$ in terms of the corresponding dual events. A typical example is the event that the cluster of the origin does not touch the boundary $\partial\Lambda$, which is equivalent to the existence of a dual cluster in $B(\Lambda^*)$ containing a closed loop which surrounds the origin.

5.2. The Upper Bound

Our upper bound is stated in terms of the finite-volume approximation $\hat{\tau}^{\mathrm{box}}_{\mathrm{wir}}$ to $\tau^{\mathrm{fin}}_{\mathrm{wir}}$; see Eq. (4.21). As in the last subsection, we will often explicitly indicate the β dependence of the relevant quantities.

Theorem 5.2. Let $d = 2$, $0 < \beta < \infty$, and $q \geqslant 1$. Then there exists a constant $C_1(\beta, q) < \infty$ such that

$$\hat{\tau}^{\mathrm{box}}_{\mathrm{wir},\,\beta}(L, M) \leqslant C_1(\beta, q)(\tau_{\mathrm{free},\,\beta^*}((L-1)\,\hat{e}_1))^2 \tag{5.9}$$

Proof. Let Λ denote the box $\Lambda(L, M)$, see Eq. (4.16). By its definition (4.21), the connectivity function $\hat{\tau}^{\mathrm{box}}_{\mathrm{wir},\,\beta}(L, M)$ is the probability, in the measure $\mu_{\mathrm{wir},\,\beta,\,\Lambda}$, of the event $R^{\mathrm{fin}}_{0,\,L}(\Lambda) = \{0 \leftrightarrow L\hat{e}_1\} \cap \{0 \not\leftrightarrow \partial\Lambda\}$. Equivalently, $R^{\mathrm{fin}}_{0,\,L}(\Lambda)$ can be defined as the intersection of the event $\{0 \leftrightarrow L\hat{e}_1\}$ and the event that there is a closed loop γ^* of occupied dual bonds surrounding the points 0 and $L\hat{e}_1$. Consider the points

$$x^*_{\pm} = (-1/2,\ \pm 1/2) \qquad \text{and} \qquad y^*_{\pm} = (L + 1/2,\ \pm 1/2)$$

in Λ^*. Given the fact that the connection from 0 to $L\hat{e}_1$ must occur without touching $\partial\Lambda$, it is clear that the dual loop γ^* must consist of four pieces: the bond $\langle x^*_-, x^*_+ \rangle$, a path γ^*_+ connecting the point x^*_+ to the point y^*_+, the bond $\langle y^*_-, y^*_+ \rangle$, and a path γ^*_- connecting the point y^*_- to the point x^*_-. Moreover, the two paths $\gamma^*_{\pm} : x^*_{\pm} \to y^*_{\pm}$ must occur in $B(\Lambda^*) \backslash \{\langle x^*_-, x^*_+ \rangle, \langle y^*_-, y^*_+ \rangle\}$. Let us denote by R^*_L, R^*_R, R^*_-, and R^*_+ the four events described above, namely (dual) occupation of the bond $\langle x^*_-, x^*_+ \rangle$, the bond $\langle y^*_-, y^*_+ \rangle$ and some paths $\gamma^*_{\pm} : x^*_{\pm} \to y^*_{\pm}$ in $B(\Lambda^*) \backslash \{\langle x^*_-, x^*_+ \rangle, \langle y^*_-, y^*_+ \rangle\}$, respectively. Then $R^{\mathrm{fin}}_{0,\,L}(\Lambda) = \{0 \leftrightarrow L\hat{e}_1\} \cap R^*_L \cap R^*_R \cap R^*_- \cap R^*_+$.

Consider now a configuration $\omega \in R^*_L \cap R^*_R \cap R^*_- \cap R^*_+$. It is an easily verified geometrical fact that $\omega \in R^{\mathrm{fin}}_{0,\,L}(\Lambda)$ if and only if the dual cluster joining x_+^* to y_+^* and the dual cluster joining x_-^* to y_-^* are connected only via the two bonds $\langle x_-^*, x_+^* \rangle$ and $\langle y_-^*, y_+^* \rangle$, i.e., if and only if there is no dual connection between x_-^* and x_+^* in the set $B(\Lambda^*)\backslash\{\langle x_-^*, x_+^* \rangle, \langle y_-^*, y_+^* \rangle\}$. Using finite energy in the form (2.24) to convert the two occupation events R_L^* and R_R^* to the events that the bonds $\langle x_-^*, x_+^* \rangle$ and $\langle y_-^*, y_+^* \rangle$ are *vacant*, and the duality relation (5.8) to transform the measure $\mu_{\text{wir}, \beta, \Lambda}$ into the free measure $\mu_{\text{free}, \beta^*, \Lambda^*}$, we therefore obtain

$$\hat{\tau}_{\text{wir}, \beta}^{\text{box}}(L, M) \leqslant C_1(\beta, q)\, \mu_{\text{free}, \beta^*, \Lambda^*}(\{x_-^* \leftrightarrow y_-^*\} \cap \{x_+^* \leftrightarrow y_+^*\} \cap \{x_-^* \not\leftrightarrow x_+^*\}) \tag{5.10}$$

with $C_1(\beta, q) < \infty$ if $0 < \beta < \infty$.

Next, we note that by Proposition 2.8,

$$\begin{aligned}\mu_{\text{free}, \beta^*, \Lambda}&(\{x_-^* \leftrightarrow y_-^*\} \cap \{x_+^* \leftrightarrow y_+^*\} \cap \{x_-^* \not\leftrightarrow x_+^*\}) \\ &\leqslant \mu_{\text{free}, \beta^*, \Lambda^*}(x_+^* \leftrightarrow y_+^*)\, \mu_{\text{free}, \beta^*, \Lambda^*}(x_-^* \leftrightarrow y_-^*)\end{aligned} \tag{5.11}$$

Using the monotonicity (2.18), we obtain

$$\mu_{\text{free}, \beta^*, \Lambda^*}(x_\pm^* \leftrightarrow y_\pm^*) \leqslant \mu_{\text{free}, \beta^*}(x_\pm^* \leftrightarrow y_\pm^*) = \tau_{\text{free}, \beta^*}((L+1)\hat{e}_1) \tag{5.12}$$

which, combined with (5.10) and (5.11), proves the theorem.

5.3. The Lower Bound Our lower bound is given in terms of the approximation $\tilde{\tau}_{\text{free}}^{\text{box}}$ to $\tau_{\text{fin}}^{\text{free}}$; see Eq. (4.23).

Theorem 5.3. Let $d=2$, $0<\beta<\infty$, and $q \geqslant 1$. Then there exists a constant $C_2(\beta, q) > 0$ such that for all positive integers M and L

$$\tau_{\text{wir}, \beta}^{\text{fin}}(L\hat{e}_1) \geqslant C_2(\beta, q)^M (\tilde{\tau}_{\text{wir}, \beta^*}^{\text{box}}(L-1, M))^2 \tag{5.13}$$

Proof. In order to prove the theorem, we introduce several sets in both $\mathbb{Z}^2$ and in its dual $(\mathbb{Z}^*)^2$. Consider the dual boxes

$$\Lambda_+^* = \{x^* \in (\mathbb{Z}^*)^2 \mid \tfrac{1}{2} \leqslant x_1^* \leqslant L - \tfrac{1}{2}, \tfrac{1}{2} \leqslant x_2^* \leqslant 2M + \tfrac{1}{2}\}$$

$$\Lambda_-^* = \{x^* \in (\mathbb{Z}^*)^2 \mid \tfrac{1}{2} \leqslant x_1^* \leqslant L - \tfrac{1}{2}, -(2M + \tfrac{1}{2}) \leqslant x_2^* \leqslant -\tfrac{1}{2}\}$$

dual points

$$x_\pm^* = \tfrac{1}{2}\hat{e}_1 \pm (M + \tfrac{1}{2})\hat{e}_2, \qquad x_\pm^* = (L - \tfrac{1}{2})\hat{e}_1 \pm (M + \tfrac{1}{2})\hat{e}_2$$

dual bonds

$$b^*_{\pm,l} = \langle x^*_\pm, x^*_\pm - \hat{e}_1 \rangle, \qquad b^*_{\pm,r} = \langle y^*_\pm, y^*_\pm + \hat{e}_1 \rangle$$

and dual vertical lines γ^*_l and γ^*_r joining the point $x^*_- - \hat{e}_1$ to the point $x^*_+ - \hat{e}_1$, and the point $y^*_- + \hat{e}_1$ to the point $y^*_+ + \hat{e}_1$, respectively. In addition to these sets in $(\mathbb{Z}^*)^2$ and $\mathbb{B}^*_2$, consider the points

$$\tilde{x}_\pm = \pm M\hat{e}_2 \qquad \text{and} \qquad \tilde{y}_\pm = L\hat{e}_1 \pm M\hat{e}_2$$

in $\mathbb{Z}^2$ and the vertical lines $\tilde{\gamma}_l$ and $\tilde{\gamma}_r$ in $\mathbb{B}_2$ that join the point $\tilde{x}_-$ to the point $\tilde{x}_+$, and the point $\tilde{y}_-$ to the point $\tilde{y}_+$, respectively.

Consider now the events $R^*_\pm$ that the points $x^*_\pm$ and $y^*_\pm$ are connected by a path of dual occupied bonds in $\Lambda^*_\pm$, and the events R^*_α that all bonds in $\gamma^*_\alpha \cup b^*_{+,\alpha} \cup b^*_{-,\alpha}$ ($\alpha = l, r$) are occupied. The event $R^*_+ \cap R^*_- \cap R^*_l \cap R^*_r$ then clearly implies the existence of an occupied dual path surrounding the points 0 and $L\hat{e}_1$, so that the event $\{0 \leftrightarrow L\hat{e}_1\} \cap R^*_+ \cap R^*_- \cap R^*_l \cap R^*_r$ is contained in the desired event $R^{\text{fin}}_{0,L} = \{0 \leftrightarrow L\hat{e}_1\} \cap \{|C(0)| < \infty\}$. As a consequence

$$\begin{aligned} \tau^{\text{fin}}_{\text{wir},\beta}(L\hat{e}_1) &= \mu_{\text{wir},\beta}(R^{\text{fin}}_{0,L}) \\ &\geq \mu_{\text{wir},\beta}(\{0 \leftrightarrow L\hat{e}_1\} \cap R^*_+ \cap R^*_- \cap R^*_l \cap R^*_r) \end{aligned} \tag{5.14}$$

Our goal is to modify the event in the argument of (5.14) so that (1) the event $\{0 \leftrightarrow L\hat{e}_1\}$ is guaranteed to occur, and (2) the two dual paths across Λ^*_+ and Λ^*_- carry decoupling events that allow us to apply Proposition 2.6 to factor their probabilities. We begin by degrading our estimate (5.14) by constructing vertical lines of occupied (direct) bonds:

$$\tau^{\text{fin}}_{\text{wir},\beta}(L\hat{e}_1) \geq \mu_{\text{wir},\beta}(\{\omega_{\tilde{\gamma}_l \cup \tilde{\gamma}_r} = 1\} \cap \{0 \leftrightarrow L\hat{e}_1\} \cap R^*_+ \cap R^*_- \cap R^*_l \cap R^*_r)$$

Here, as usual, ω_B denotes the configuration restricted to B. Using finite energy in the form (2.24) to flip the $4M+6$ bonds in $\gamma^*_l \cup b^*_{+,l} \cup b^*_{-,l} \cup \gamma^*_r \cup b^*_{+,r} \cup b^*_{-,r}$, we then obtain

$$\tau^{\text{fin}}_{\text{wir},\beta}(L\hat{e}_1) \geq C(\beta, q)^{4M+6} \mu_{\text{wir},\beta}(\{\omega_{\tilde{\gamma}_l \cup \tilde{\gamma}_r} = 1\} \cap \{0 \leftrightarrow L\hat{e}_1\} \cap R^*_+ \cap R^*_-)$$

with a suitable constant $C(\beta, q) > 0$.

Consider now the events $(R^*_\pm)^{\text{fin}}$ that $x^*_\pm$ and $y^*_\pm$ are connected by dual clusters $C^*(x^*_\pm) \subset B(\Lambda^*_\pm)$, i.e., by clusters that lie entirely within $B(\Lambda^*_\pm)$, and hence are surrounded by decoupling circuits of occupied (direct) bonds in $(B^+(\Lambda^*_\pm))^*$. Clearly $R^*_\pm \supset (R^*_\pm)^{\text{fin}}$ and thus

$$\begin{aligned} &\tau^{\text{fin}}_{\text{wir},\beta}(L\hat{e}_1) \\ &\quad \geq C(\beta, q)^{4M+6} \mu_{\text{wir},\beta}(\{\omega_{\tilde{\gamma}_l \cup \tilde{\gamma}_r} = 1\} \cap \{0 \leftrightarrow L\hat{e}_1\} \cap (R^*_+)^{\text{fin}} \cap (R^*_-)^{\text{fin}}) \end{aligned} \tag{5.15}$$

We now claim that

$$\{\omega_{\tilde{\gamma}_l \cup \tilde{\gamma}_r} = 1\} \cap \{0 \leftrightarrow L\hat{e}_1\} \cap (R_+^*)^{\text{fin}} \cap (R_-^*)^{\text{fin}}$$
$$= \{\omega_{\tilde{\gamma}_l \cup \tilde{\gamma}_r} = 1\} \cap (R_+^*)^{\text{fin}} \cap (R_-^*)^{\text{fin}} \tag{5.16}$$

In order to see this, let ω be a configuration in $(R_+^*)^{\text{fin}}$. As noted above, the condition $C^*(x_\pm^*) \subset B(\Lambda^*)$ implies the existence of a closed path of occupied bonds in $(B^+(\Lambda_+^*))^*$ surrounding x_+^* and y_+^*. Given $\omega \in (R_+^*)^{\text{fin}}$, let γ be the innermost such path. Since γ surrounds x_+^* and y_+^*, but lies within $(B^+(\Lambda_+^*))^*$, it must visit the points $\tilde{x}_+$ and $\tilde{y}_+$; thus it provides a connection between $\tilde{x}_+$ and $\tilde{y}_+$ by a path of occupied bonds. Observing that the vertical paths $\tilde{\gamma}_l$ and $\tilde{\gamma}_r$ connect the point 0 to the point $\tilde{x}_+$ and the point $L\hat{e}_1$ to the point $\tilde{y}_+$, we see that there is automatically an occupied path from 0 to $L\hat{e}_1$, which completes the proof of (5.16).

Using once more finite energy, the relations (5.15) and (5.16) together with the duality relation (5.8) now imply that

$$\begin{aligned}\tau^{\text{fin}}_{\text{wir},\beta}(L\hat{e}_1) &\geq C(\beta,q)^{4M+6}\,\mu_{\text{wir},\beta}(\{\omega_{\tilde{\gamma}_l} \cup \tilde{\gamma}_r = 1\} \cap (R_+^*)^{\text{fin}} \cap (R_-^*)^{\text{fin}}) \\ &\geq C(\beta,q)^{8M+6}\,\mu_{\text{wir},\beta}((R_+^*)^{\text{fin}} \cap (R_-^*)^{\text{fin}}) \\ &= C(\beta,q)^{8M+6}\,\mu_{\text{free},\beta^*}(R_+^{\text{fin}} \cap R_-^{\text{fin}})\end{aligned} \tag{5.17}$$

where $R_\pm^{\text{fin}}$ are the events dual to $(R_\pm^*)^{\text{fin}}$.

Finally, we use the fact that R_+^{fin} and R_-^{fin} are events of the form considered in Proposition 2.7, so that

$$\mu_{\text{free},\beta^*}(R_+^{\text{fin}} \cap R_-^{\text{fin}}) \geq \mu_{\text{free},\beta^*}(R_+^{\text{fin}})\,\mu_{\text{free},\beta^*}(R_-^{\text{fin}}) = (\tilde{\tau}^{\text{box}}_{\text{free}}(L, M))^2 \tag{5.18}$$

This completes the proof of Theorem 5.3. ∎

Notice that, in contrast to the proof of Theorem 5.2, the above proof does not invoke monotonicity properties which depend on boundary conditions. Thus it can be used equally well to give a lower bound on $\tau^{\text{fin}}_{\text{free}}$, namely:

Corollary. Let $d=2$, $0<\beta<\infty$, and $q \geq 1$. Then there exists a constant $C_2(\beta, q) > 0$ such that for all positive integers M and L

$$\tau^{\text{fin}}_{\text{free},\beta}(L\hat{e}_1) \geq C_2(\beta,q)^M\,(\tilde{\tau}^{\text{box}}_{\text{wir},\beta^*}(L-1, M))^2 \tag{5.19}$$

5.4. The Dichotomy and Percolation Probabilities

In this subsection we prove Theorem 5.1. We start with a proposition which is essentially a corollary to the upper and lower bounds of Theorems 5.2 and 5.3.

Proposition 5.4. Let $d=2$, $0<\beta<\infty$, and $q\geqslant 1$. Then

$$\xi_{\rm wir}^{\rm fin}(\beta)=\tfrac{1}{2}\xi_{\rm free}(\beta^*) \qquad \text{if} \quad P_\infty^{\rm free}(\beta^*)=0 \tag{5.20}$$

and

$$\xi_{\rm wir}^{\rm fin}(\beta)=\xi_{\rm free}(\beta) \qquad \text{if} \quad P_\infty^{\rm wir}(\beta)=0 \tag{5.21}$$

Proof. Introducing the notation $\hat{\xi}_{\rm wir}^{\rm box}(M;\beta)$ and $\tilde{\xi}_{\rm free}^{\rm box}(M;\beta)$ to indicate the β dependence of the correlation lengths $\hat{\xi}_{\rm wir}^{\rm box}(M)$ and $\tilde{\xi}_{\rm free}^{\rm box}(M)$ corresponding to $\hat{\tau}_{\rm wir}^{\rm box}$ and $\tilde{\tau}_{\rm free}^{\rm box}$, the upper and lower bounds of Theorem 5.2 and 5.3 imply that

$$\hat{\xi}_{\rm wir}^{\rm box}(M;\beta)\leqslant\tfrac{1}{2}\xi_{\rm free}(\beta^*) \tag{5.22}$$

and

$$\tfrac{1}{2}\tilde{\xi}_{\rm free}^{\rm box}(M;\beta^*)\leqslant\xi_{\rm wir}^{\rm fin}(\beta) \tag{5.23}$$

Taking the limit $M\to\infty$, and observing that the left-hand side of (5.22) goes to $\xi_{\rm wir}^{\rm fin}(\beta)$ by Lemma 4.8 and the corollary at the end of Section 4, while the left-hand side of (5.23) goes to $\xi_{\rm free}^{\rm fin}(\beta^*)$ by the same corollary and Lemma 4.7, we conclude that

$$\tfrac{1}{2}\xi_{\rm free}^{\rm fin}(\beta^*)\leqslant\xi_{\rm wir}^{\rm fin}(\beta)\leqslant\tfrac{1}{2}\xi_{\rm free}(\beta^*) \tag{5.24}$$

Since $\xi_{\rm free}^{\rm fin}(\beta^*)=\xi_{\rm free}(\beta^*)$ if $P_\infty^{\rm free}(\beta^*)=0$, this implies the first part of the proposition. If $P_\infty^{\rm wir}(\beta)=0$, then $\xi_{\rm wir}^{\rm fin}(\beta)=\xi_{\rm wir}(\beta)$. In addition, by the results of ref. 1, $\mu_{{\rm wir},\beta}=\mu_{{\rm free},\beta}$ whenever $P_\infty^{\rm wir}(\beta)=0$ and hence $\xi_{\rm free}(\beta)=\xi_{\rm wir}(\beta)$. This implies the second part of the proposition. ∎

In order to complete the proof of Theorem 5.1, we use the fact, proven in Section 2.4, that the free measure $\mu_{\rm free}$ is ergodic under any nontrivial subgroup of the translation group (Theorem 2.10). Since, in addition, $\mu_{\rm free}$ is an FKG measure which is invariant under horizontal and vertical translations and axis reflections, a bond percolation analog of the theorem of ref. 18 applies, leading to the following result.

Theorem 5.5. Let $d=2$, $0<\beta<\infty$, and $q\geqslant 1$, and assume that $P_\infty^{\rm free}(\beta)>0$. Then, with probability one with respect to the free measure $\mu_{{\rm free},\beta}$, any finite set of sites in $\mathbb{Z}^2$ is surrounded by a circuit of occupied bonds.

Corollaries. Let $d=2$, $0<\beta<\infty$, and $q\geqslant 1$. Then

(1) $$P_\infty^{\rm free}(\beta)\,P_\infty^{\rm wir}(\beta^*)=0 \tag{5.25}$$

(2) $P_\infty^{\rm free}(\beta)=0$ for all $\beta \leqslant \beta_o$. In particular, $P_\infty^{\rm free}(\beta)$ is left continuous at β_o.

(3) If $P_\infty^{\rm wir}(\beta)=0$ or $P_\infty^{\rm free}(\beta)>0$, then $\mu_{{\rm free},\beta}(\,\cdot\,)=\mu_{{\rm wir},\beta}(\,\cdot\,)$, and, in particular, $P_\infty^{\rm free}(\beta)=P_\infty^{\rm wir}(\beta)$.

Proof. As noted above, the theorem follows from (a bond percolation analog of) ref. 18 and Theorem 2.10. Corollary 1 then follows immediately, and Corollary 2 also follows easily—see Eq. (1.14) and the paragraph preceding it. The first part of Corollary 3, namely that $P_\infty^{\rm wir}(\beta)=0$ implies $\mu_{{\rm free},\beta}(\,\cdot\,)=\mu_{{\rm wir},\beta}(\,\cdot\,)$, is a result of ref. 1. That equality of the measures is also implied by $P_\infty^{\rm free}(\beta)>0$ follows from (5.25), ref. 1 and the self-duality of the model.

Remarks. 1. Theorem 5.1 (the dichotomy) now follows immediately from Proposition 5.4 and Eq. (5.25).

2 It turns out that, although not explicitly stated, Corollary 2 has already been established by Welsh[38] in the course of the proof of his Theorem 7.3. We note that Welsh's proof does not require ergodicity, but only stationarity of the measure $\mu_{\rm free}$. Instead it invokes uniqueness of the infinite cluster and an unpublished argument of Zhang.

3 We expect that left continuity of $P_\infty^{\rm free}(\beta)$ at the transition point holds in all dimensions provided $q \geqslant 1$. However, we do *not* expect $P_\infty^{\rm free}(\beta)$ to be right continuous at the transition point if the system has a first-order transition; indeed, for q sufficiently large, this can be established using Pirogov–Sinai theory, as used, e.g., in ref. 28. This is to be contrasted with the behavior of $P_\infty^{\rm wir}(\beta)$. By standard percolation arguments,[35] namely expressing $P_\infty^{\rm wir}(\beta)$ as the decreasing limit of the finite-volume quantities (3.4) (which are continuous and nondecreasing in β), $P_\infty^{\rm wir}(\beta)$ is right continuous for all β and all $q \geqslant 1$ in dimension d >1. However, in dimension $d \geqslant 2$, convergent expansions have been used to show $P_\infty^{\rm wir}(\beta)$ is not left continuous at the transition point provided q is sufficiently large[25] (see also refs. 27 and 28).

4. Corollary 3 implies that in two dimensions the Gibbs state is unique at all β except those for which $P_\infty^{\rm free}(\beta)=0$ while $P_\infty^{\rm wir}(\beta)>0$. Presumably, this never occurs for systems with second-order transitions ($q \leqslant 4$ in $d=2$), and occurs only at a single point—the transition point—for systems with first-order transitions ($q>4$ in $d=2$). Again, this can be proven via expansion methods in $d \geqslant 2$ for q sufficiently large.

We conclude this section with a little result which is an easy consequence of Proposition 5.4. The result shows that continuity of the magnetization at β_o ensures criticality of the transition, i.e., divergence of the correlation length(s).

Proposition 5.6. Let $d=2$ and $q \geqslant 1$. Then $M(\beta_o) \equiv P_\infty^{\text{wir}}(\beta_o)=0$ implies $\xi_{\text{wir}}^{\text{fin}}(\beta_o)=\infty$ and hence also $\xi_{\text{free}}^{\text{fin}}(\beta_o)=\xi_{\text{free}}(\beta_o)=\xi_{\text{wir}}(\beta_o)=\infty$.

Proof. By the assumption $P_\infty^{\text{wir}}(\beta_o)=0$, the FKG ordering of states (2.23) and the definition of β_o, we have $0=P_\infty^{\text{free}}(\beta_o)=P_\infty^{\text{free}}(\beta_o^*)$, and hence by the first branch (5.20) of Proposition 5.4

$$\xi_{\text{wir}}^{\text{fin}}(\beta_o)=\tfrac{1}{2}\xi_{\text{free}}(\beta_o^*)=\tfrac{1}{2}\xi_{\text{free}}(\beta_o) \tag{5.26}$$

On the other hand, again by the assumption $P_\infty^{\text{wir}}(\beta_o)=0$, we have the second branch (5.21) of Proposition 5.4, namely

$$\xi_{\text{wir}}^{\text{fin}}(\beta_o)=\xi_{\text{free}}(\beta_o) \tag{5.27}$$

From (5.26) and (5.27), we conclude that either $\xi_{\text{free}}(\beta_o)=0$ or $\xi_{\text{free}}(\beta_o)=\infty$. The first case is easily ruled out by considering, e.g., $\tau_{\text{free}}(\hat{e}_1)$. That the other correlation lengths also diverge is an immediate consequence of the remark following Theorem 4.3. ∎

APPENDIX. REFLECTION POSITIVITY AND THE TRANSFER MATRIX

The concept of reflection positivity and its consequences are well-known tools in the context of field theory. For the convenience of the reader we give a brief review in this appendix.

We consider a (finite or infinite) lattice $\Lambda \subset \mathbb{Z}^d$ which is invariant under reflections at a plane Σ. Here Σ is either a lattice plane or a plane which lies halfway between two lattice planes. Denoting the reflection at Σ by r, we then decompose Λ as $\Lambda=\Lambda_+ \cup \Lambda_-$, where Λ_+ are the points on one side of Σ, $\Lambda_-=r(\Lambda_+)$ are the points on the other side of Σ, and $\Lambda_- \cap \Lambda_+=\Sigma \cap \Lambda$ (which is of course empty if Σ lies between two lattice planes).

For a local observable A with support $\operatorname{supp} A \subset \Lambda_+$, one introduces the reflected observable $r(A)$ as

$$(r(A))(\sigma)=A(r(\sigma)) \tag{A.1}$$

where $r(\sigma)_x=\sigma_{r(x)}$. Reflection positivity of the Potts model is the statement that

$$\langle \overline{r(A)}\, A\rangle_{b,\Lambda} \geqslant 0 \tag{A.2}$$

The proof of (A.2) is standard; for the strategy, see, e.g., refs. 12 and 36.

The inequality (A.2) has several important consequences. Here we are mainly interested in the representation of truncated expectation values as matrix elements of a suitably defined transfer matrix T. In order to define the transfer matrix T in the setting considered here, we need, in addition to the reflection invariance of Λ, that Λ is invariant under translations perpendicular to Σ. We therefore assume that Λ is of the form

$$\Lambda = \mathbb{Z} \times \Lambda_1$$

where Λ_1 is a (finite or infinite) sublattice of $\mathbb{Z}^{d-1}$. We then consider the algebra $\mathscr{A}_+$ of local observables with support in Λ_+, where from now on

$$\Lambda_+ = \{(x, \vec{x}) \in \mathbb{Z}^d \mid x \geqslant 0, \vec{x} \in \Lambda_1\}$$

while $\Sigma = \{(x, \vec{x}) \in \mathbb{Z}^d | x = 0, \vec{x} \in \Lambda_1\}$. Due to (A.2), the equation

$$\langle A, B \rangle := \langle \overline{r(A)}\, B \rangle_{b, \Lambda} \tag{A.3}$$

defines a positive semi-definite scalar product over $\mathscr{A}_+$. Dividing out the corresponding null space $\mathscr{N}$ and completing the resulting space in the usual way, this leads to the definition of a Hilbert space $\mathscr{H} = \overline{\mathscr{A}_+ / \mathscr{N}}$.

Next, we introduce, for each local observable $A \in \mathscr{A}_+$, the observable TA which is obtained from A by translation by one lattice unit in the positive direction perpendicular to Σ. It is an easy consequence of the Cauchy–Schwarz inequality for the scalar product (A.3) (see ref. 36 for details) that T obeys the inequalities

$$0 \leqslant \langle A, TA \rangle \leqslant \langle A, A \rangle \tag{A.4}$$

The operator T therefore defines a positive transfer matrix, which obeys the inequalities

$$0 \leqslant T \leqslant 1$$

as an operator on $\mathscr{H}$. Observing that the vector Ω corresponding to the constant function $1 \in \mathscr{A}_+$ is an eigenvector of T with eigenvalue 1, we note that the norm of T is one.

We finally consider the interpretation of truncated expectation values in the above Hilbert space representation. Since

$$\langle \overline{r(A)} \rangle_{b, \Lambda} = \langle \overline{r(A)} \cdot 1 \rangle_{b, \Lambda} = \langle A, \Omega \rangle \tag{A.5a}$$

while

$$\langle T^n A \rangle_{b, \Lambda} = \langle \Omega, T^n A \rangle = \langle T^n \Omega, A \rangle = \langle \Omega, A \rangle \tag{A.5b}$$

one immediately obtains

$$\langle \overline{r(A)};\, T^n A\rangle_{b,\Lambda} = \langle A, T^n A\rangle - \langle A, \Omega\rangle\langle \Omega, A\rangle \tag{A.6}$$

Introducing the projection operator $P_\perp$ onto the Hilbert space orthogonal to Ω, we find that Eq. (A.6) becomes

$$\langle \overline{r(A)};\, T^n A\rangle_{b,\Lambda} = \langle A_\perp, T^n A_\perp\rangle \tag{A.7}$$

where

$$A_\perp = P_\perp A = A - \langle \Omega, A\rangle \Omega \tag{A.8}$$

If the support of A is a subset of the lattice plane Σ, $r(A) = A$, and Eq. (A.7) reduces to

$$\langle \bar{A};\, T^n A\rangle_{b,\Lambda} = \langle A_\perp, T^n A_\perp\rangle \tag{A.9}$$

Equation (A.9) is an important technical tool in the proof of the existence of the correlation lengths $\xi^{(1)}_{\text{wir}}$ and $\xi^{(2)}_{\text{wir}}$.

Remark: In the context of Euclidean field theory, the direction perpendicular to Σ is often interpreted as the Euclidean time. The Hilbert space a $\mathscr{H} = \overline{\mathscr{A}_+/\mathscr{N}}$ is then nothing but the quantum mechanical Hilbert space of the considered model, and T is the generator of the Euclidean time translations, i.e. $T = e^{-\varepsilon}H$, where ε is the lattice spacing and H is the Hamilton operator of the theory.

However, $\mathscr{H}$ and T have no such interpretation for the classical Potts model. This is due to the fact that here Λ is the lattice of a *classical* system, and not a lattice approximation to Euclidean space-time.

ACKNOWLEDGMENTS

C.B. is a DFG Heisenberg Fellow. J.T.C. is paritally supported by NSF grant DMS-940382.

REFERENCES

1. M. Aizenman, J. T. Chayes, L. Chayes and C. M. Newman, Discontinuity of the magnetization in the one-dimensional $1/|x-y|^2$ Ising and Potts models, *J. Stat. Phys.* **50**:1–40 (1988).
2. C. Borgs and W. Janke, An explicit formula for the interface tension of the 2D Potts model, *J. Phys. I France* **2**:2011–2018 (1992).

3. J. van den Berg and H. Kesten, Inequalities with applications to percolation and reliability, *J. Appl. Prob.* **22**:556–569 (1985).
4. C. Borgs, R. Kotecký, and S. Miracle-Sole, Finite-size scaling for the Potts model, *J. Stat. Phys.* **62**:529–551 (1991).
5. R. M. Burton and M. Keane, Density and uniqueness in percolation, *Commun. Math. Phys.* **121**:501–505 (1989).
6. E. Buffenoir and S. Wallon, The correlation length of the Potts model at the first-order transition point, *J. Phys. A: Math. Gen.* **26**:3045–3062 (1993).
7. J. T. Chayes, L. Chayes, G. R. Grimmett, H. Kesten, and R. H. Schonmann, The correlation length for the high-density phase of Bernoulli percolation, *Ann. Prob.* **17**:1277–1302 (1989).
8. R. L. Dobrushin, The description of a random field by means of conditional probabilities and conditions of its regularity, *Theory Prob. Appl.* **13**:197–224 (1968).
9. R. G. Edwards and A. D. Sokal, Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm, *Phys. Rev. D* **38**:2009–2012 (1988).
10. M. E. Fisher, Aspects of equilibrium critical phenomena, *J. Phys. Soc. Japan* **26**:87–93 (1969).
11. C. M. Fortuin, On the random cluster model. III. The simple random-cluster process, *Physica* **59**:545–570 (1972).
12. J. Fröhlich, R. Israel, E. H. Lieb, and B. Simon, Phase transitions and reflection positivity. I. General theory and long range lattice models, *Commun. Math. Phys.* **62**:1–34 (1978).
13. C. M. Fortuin and P. W. Kasteleyn, On the random cluster model. I. Introduction and relation to other models, *Physica* **57**:536–564 (1972).
14. C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets, *Commun. Math. Phys.* **22**:89–103 (1971).
15. H.-O. Georgii, *Gibbs Measures and Phase Transitions* (de Gruyter, Berlin, 1988).
16. J. Ginibre, General formulation of Griffths' inequalities, *Commun. Math. Phys.* **16**:310–328 (1970).
17. S. Gupta and A. Irbäck, Physics beyond instantons: Measuring the physical correlation length, *Phys. Lett. B* **286**:112–117 (1992).
18. A. Gandolfi, M. Keane, and L. Russo, On the uniqueness of the infinite occupied cluster in dependent two-dimensional site percolation, *Ann. Prob.* **16**:1147–1157 (1988).
19. R. B. Griffiths, Correlations in Ising ferromagnets. II, *J. Math. Phys.* **8**:484–489 (1964).
20. G. Grimmett, Percolative problems, in *Probability and Phase Transition*, G. Grimmett, ed. (Kluwer, Dordrecht, 1993), pp. 69–86.
21. G. Grimmett, The stochastic random-cluster process and the uniqueness of random-cluster measures, Preprint (1994).
22. T. E. Harris, A lower bound for the critical probability in a certain percolation process, *Proc. Camb. Philos Soc.* **56**:13–20 (1960).
23. J. Imbrie and C. M. Newman, An intermediate phase with slow decay of correlations in one dimensional $1/|x-y|^2$ percolation, Ising and Potts models, *Commun. Math. Phys.* **118**:303–336 (1988).
24. A. Klümper, Investigation of excitation spectra of exactly solvable models using inversion relations, *Int. J. Mod. Phys. B* **4**:871–893 (1990).
25. R. Kotecký and S. Shlosman, First-order phase transitions in large entropy lattice systems, *Commun. Math. Phys.* **83**:493–550 (1982).
26. A. Klümper, A. Schadschneider, and J. Zittartz, Inversion relations, phase transitions and transfer matrix excitations for special spin models in two dimensions, *Z. Phys. B* **76**:247–258 (1989).
27. L. Laanait, A. Messager, and J. Ruiz, Phase coexistence and surface tensions for the Potts model, *Commun. Math. Phys.* **105**:527–545 (1986).

28. L. Laanait, A. Messaged S. Miracle-Sole, J. Ruiz, and S. Shlosman, Interfaces in the Potts model I: Pirogov–Sinai theory of the Fortuin–Kasteleyn representation, *Commun. Math. Phys.* **140**:81–91 (1991).
29. O. E. Lanford and D. Ruelle, Observables at infinity and states with short range correlations in statistical mechanics, *Commun. Math Phys.* **13**:194–215 (1969).
30. A. Messager, S. Miracle-Sole, J. Ruiz, and S. Shlosman, Interfaces in the Potts model II: Antonov's rule and rigidity of the order disorder interface, *Commun. Math. Phys.* **140**:275–290 (1991).
31. B. McCoy and T. T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, Cambridge, Massachusetts, 1973).
32. C. M. Newman and L. S. Schulman, Infinite clusters in percolation models, *J. Stat. Phys.* **26**:613–628 (1981).
33. P. Peczak and D. P. Landau, Monte Carlo study of finite-size effects at a weakly first-order phase transition, *Phys. Rev. B* **39**:11932–11942 (1989).
34. C. Preston, *Random Fields* (Springer-Verlag, Berlin, 1976).
35. L. Russo, A note on percolation, *Z. Wahrsch. Geb.* **43**:39–48 (1978).
36. E. Seiler, *Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics* (Springer Verlag, Berlin, 1982).
37. R. H. Swendsen and J. S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
38. D. J. A. Welsh, Percolation in the random cluster process and Q-state Potts model, *J. Phys. A: Math. Gen.* **26**:2471–2483 (1993).
39. F. Y. Wu, Knot theory and statistical mechanics, *Rev. Mod. Phys.* **64**:1099–1131 (1992).