Journal of Statistical Physics, Vol. 110, Nos. 1/2, January 2003 (© 2003)

Exact Results for the Universal Area Distribution of Clusters in Percolation, Ising, and Potts Models

John Cardy[1] **and Robert M. Ziff**[2]

Received May 17, 2002; accepted July 18, 2002

At the critical point in two dimensions, the number of percolation clusters of enclosed area greater than A is proportional to A^{-1}, with a proportionality constant C that is universal. We show theoretically (based upon Coulomb gas methods), and verify numerically to high precision, that $C = 1/(8\sqrt{3}\,\pi = 0.022972037\ldots)$. We also derive, and verify to varying precision, the corresponding constant for Ising spin clusters, and for Fortuin–Kasteleyn clusters of the $Q = 2$, 3 and 4-state Potts models.

KEY WORDS: Percolation; Ising model; Potts model; universality; conformal field theory; Coulomb gas methods.

1. INTRODUCTION

It is often useful to characterize critical systems by their geometric properties, for example the distribution of cluster sizes which appears to follow a power law

$$n_s \sim Bs^{-\tau} \tag{1}$$

asymptotically for large s, where n_s gives the number of clusters of s connected sites, per lattice site. The exponent τ is a universal quantity whose value is the same for all systems of a given class—for example, $\frac{187}{91}$ for all critical percolation systems in two dimensions, no matter what lattice or percolation type is considered as long as the rules are sufficiently local. The

[1] Department of Physics, Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, United Kingdom, and All Souls College, Oxford; e-mail: j.cardy1@physics.oxford.ac.uk

[2] Michigan Center for Theoretical Physics and Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2136; e-mail: rziff@umich.edu

0022-4715/03/0100-0001/0 © 2003 Plenum Publishing Corporation

coefficient or amplitude B however is non-universal, varying from lattice to lattice.

Indeed, n_s cannot have a completely universal form because it is written in terms of a lattice-level measure, the mass s. Different lattice structures have different typical site densities at the lattice level and correspondingly different values of B. In order to characterize the size distribution of clusters in a way that circumvents the site-level description, the authors of ref. 1 considered (for the case of two-dimensional percolation) the quantity $\mathrm{Nr}(\ell_m > \ell) = \tilde{N}(\ell)$ which gives the number of clusters whose maximum x- or y-dimension ℓ_m is greater or equal to a given value ℓ, divided by the total system area, $\mathscr{A} = O(L^2)$. They argued that for $L \gg \ell \gg a$ (where a is the lattice spacing), this quantity should behave as

$$\tilde{N}(\ell) \sim \frac{\tilde{C}}{\ell^2}, \tag{2}$$

with the coefficient $\tilde{C}$ being a universal quantity, identical for all 2d percolating system at the critical point. The universality of $\tilde{C}$ follows heuristically from the idea that $\tilde{N}$ represents a macroscopic measure of the large clusters of the system, and remains well defined in the limit $a \to 0$, in which the lattice disappears. The proportionality to $1/\ell^2$ is a consequence of the self-similarity of the fractal percolating system, and can also be derived by the following argument (in d dimensions): from (1) it follows that the number of clusters whose mass is greater than s scales as $s^{1-\tau}$, and because $s \sim \ell^D$, where D is the fractal dimension of the clusters, the number of clusters whose length scale is greater than ℓ scales as $\ell^{D(1-\tau)}$, or ℓ^{-d} by virtue of the hyperscaling relation $d/D = \tau - 1$. This result is valid for any critical system where the hyperscaling relation is valid. Later we shall give other, presumably equivalent, theoretical arguments.

Besides the maximum dimension ℓ_m, one can consider any other macroscopic measure of the length scale of the cluster, such as the radius of gyration or the diameter of the covering disk. For each measure, there is a corresponding value of $\tilde{C}$.

An equivalent way to write (2) is

$$N(A) \sim \frac{C}{A} \qquad (a^2 \ll A \ll L^2) \tag{3}$$

where $N(A)$ is the number of clusters (per unit area) whose area (by some measure) is greater or equal to A, and C depends upon the choice of that measure. This could be the area of the smallest disk covering the cluster,

the area enclosed by the cluster, and so on. Equation (3) is the form of the size distribution that will be considered in this paper.

We note that (3) can also be written as[2]

$$A_n \sim \frac{C}{n} \tag{4}$$

for $1 \ll n \ll 1/a^2$, where A_n represents a rank-ordering of the areas, such that A_1 is the area of the largest cluster, A_2 the area of the second-largest, etc., for a system whose total area $\mathscr{A}$ is defined as unity. Although the rank-ordering necessarily starts with clusters whose area is of the order of the area of the system, the behavior of (4) applies to clusters whose area is much smaller than 1 but larger than the lattice element area. Equation (4) gives the size distribution in a proper Zipf's-law form, in which the weight (here area) is inversely proportional to the rank. When written in terms of s, on the other hand, the behavior of the size ranking is not a simple inverse power as above (compare refs. 3 and 4) and also is not universal.

The various measures of the area of the clusters that were considered in ref. 2 included the area of the square $l_m \times l_m$, the area of a disk that just covers the cluster, the area enclosed by the external perimeter (hull) of the cluster, and the area enclosed by the Grossman–Aharony (G–A) hull of the cluster (in which fjords are excluded).[5] (Percolation hulls are fractal with dimension $D_H = 7/4$[6, 7] but enclose a non-fractal, Euclidean area.) For each of these measures, a different value of the constant C applies, and the following values were found: C(square) ≈ 0.115, C(disk) ≈ 0.104, C(G–A hull) ≈ 0.037, and C(hull) ≈ 0.024. The way these different values of C were found was that the first C(square) was measured directly on a fully populated lattice (since the measurement of the maximum x- or y-direction is an easy task), and then the rest were deduced (in an approximate way) by looking at the ratio of the area measures for individually generated clusters. It was noticed that C(disk) is close to the fractal co-dimension $d - D = 5/48 = 0.1041666\ldots$, but no exact results for any of these C were obtained.

In the present paper we report on a direct numerical and theoretical study of the constant $C = C$(hull) for the 2d measure of the area enclosed by the external perimeter or hull of percolation clusters, Ising spin clusters, and Fortuin–Kasteleyn (FK) clusters on the Potts model clusters for $Q = 2$, 3 and 4. (Of course, percolation corresponds to the Potts model for $Q = 1$ and the Ising model corresponds to $Q = 2$.)

Initially, one of us predicted that for percolation

$$C = \frac{1}{8\sqrt{3}\,\pi} = 0.022972037\ldots \tag{5}$$

Independently, the other numerically determined $C = 0.022976 \pm 0.000005$, which is completely consistent with this prediction. Additional work described below yields 0.0229723 ± 0.0000010 (one standard deviation of error). This close agreement confirms that the Coulomb gas methods that are used to derive these results are most certainly applicable to percolation and the Potts model. We also considered different lattices and types of percolation to demonstrate universality.

For Ising clusters of same-spin sites, we predict the value

$$C = \frac{1}{16\sqrt{3}\,\pi} = 0.011486019\ldots \qquad \text{(Ising spin clusters)} \tag{6}$$

exactly half the value for percolation clusters. For the Potts model with $Q = 2, 3, 4$, we also consider the areas enclosed by the FK bond clusters,[8] and find for the corresponding values of C:

$$C = \frac{1}{12\pi} = 0.026525824\ldots \qquad \text{(FK cluster, } Q = 2) \tag{7}$$

$$C = \frac{\sqrt{3}}{20\pi} = 0.027566445\ldots \qquad \text{(FK cluster, } Q = 3) \tag{8}$$

$$C = \frac{1}{4\pi^2} = 0.025330296\ldots \qquad \text{(FK cluster, } Q = 4) \tag{9}$$

The theoretical justifications of the above predictions are based upon considerations reported previously in ref. 9 and expanded upon in the second section below. In the third section we describe the numerical work we carried out to test these results; we find good numerical confirmation for all the cases. Conclusions are given in the fourth section.

2. COULOMB GAS CALCULATION OF *C*

In this section we compute the universal amplitude C in the scaling law $N(A) \sim C/A$, using Coulomb gas methods.[10] These are not rigorous, but are known to give presumably exact results for critical exponents and other universal quantities.

While the development in the Introduction emphasizes clusters and the areas enclosed by their external hulls, the focus will shift to both external and internal hulls, or loops when both are taken together. A factor of one-half will be included in the final results to compensate for this change, so that the values for C will be applicable to just external hulls, just internal hulls, or the average (but not sum) of the two.

We consider a finite but large system of linear size L, and total area $\mathscr{A} = O(L^2)$. As will become clear, the precise geometry and boundary conditions are not relevant to the calculation of C. L is considered to have dimensions of length, so that the total number of sites in the lattice is of order $(L/a)^2$, where a is the lattice spacing.

All the models we consider (percolation, Ising spin clusters, FK clusters) are special cases of either the O(n) model or the Q-state Potts model.[10]

The O(n) model is most easily considered on a honeycomb lattice, and it is equivalent to the loop gas model defined by the partition function

$$Z_{O(n)} = \sum_{\text{loop configs}} x^{\text{total length}}\, n^{\text{number of loops}} \tag{10}$$

where the sum is over all configurations of non-intersecting closed loops on the honeycomb lattice. This model has in general two critical points for each n in the interval $[-2, 2]$: $x = x_{c1}(n)$ (dilute phase) and $x = x_{c2}(n)$ (dense phase). In particular, for $n = 1$ and $x = x_{c2} = 1$ the loops form the hulls of site percolation clusters on the triangular lattice; for $n = 1$ and $x = x_{c1}$ they are the boundaries of critical Ising clusters. For $n \to 0$ we get a single self-avoiding loop, and for $n = 2$ the loops are the steps on a surface at the roughening transition.

The partition function of the Q-state Potts model is more easily considered on the square lattice, and it may be transformed into that of the random cluster model, proportional to

$$Z_Q = \sum_{\text{cluster configurations}} x^{\text{number of bonds}} Q^{\text{number of clusters}} \tag{11}$$

where $x = p/(1-p)$. The hulls of the clusters form closed loops on the medial lattice, a square lattice whose vertices lie at the midpoint of the links of the original lattice. At the critical point $x = x_c(Q) = \sqrt{Q}$, Z_Q is proportional to the partition function of a loop gas

$$Z_Q \propto \sum_{\text{loop configurations}} (\sqrt{Q})^{\text{number of loops}} \tag{12}$$

Note that both internal and external cluster hulls are counted as loops.

In both cases, then, the critical models are equivalent to loop gases with fugacity n (resp. $\sqrt{Q}$) per loop. The hulls of the FK clusters are in the same universality class as the dense phase of the O(n) model with $n = \sqrt{Q}$.

As discussed in Section 1, we are interested in the number $N(A)$ of such loops, per unit area of the lattice, whose internal area is greater than

a given A. Note that we consider A as having dimensions (length)^2. For $A \ll L^2$, we expect that $N(A)$ has a finite limit as $L \to \infty$. In order to obtain universal results, we also consider $A \gg a^2$. Our computation of the form of $N(A)$ in this regime is in two stages: first we show, from Coulomb gas arguments, that the total area contained in *all* loops behaves logarithmically, $\propto \mathscr{A} \ln(L/a)$, as $L/a \to \infty$, with a calculable coefficient; then we argue from this that in the regime of interest $N(A) \sim C/A$, with C simply related to the above coefficient.

2.1. Total Area Inside All Loops

We shall present two *a priori* independent arguments, both however based on Coulomb gas methods, for evaluating the leading behavior of the the total area in side all loops in large but finite system.

2.1.1. Wilson Loop Method

The argument of this section follows that of refs. 9 and 12, but, in order to be self-contained, we present it again, perhaps with greater clarity.

The loop gases described above may be mapped exactly onto a height model on the dual lattice, as follows. Each loop is assigned a random orientation, so that a configuration of m unoriented loops corresponds to 2^m configurations of oriented loops. There is a 1–1 mapping between the configurations of this oriented loop gas and the heights h, conventionally chosen to be integer multiples of π, as follows: assign $h = 0$ on the boundary, and increase (decrease) h by π each time a loop is crossed which goes to the left (right). The fact that the loops are closed makes this a consistent procedure.

The weights of n (resp. $\sqrt{Q}$) associated with each loop may be taken into account in the ensemble of oriented loops in (at least) two ways: the most natural would be to assign equal weights $\frac{1}{2}n$ to each orientation: we refer to this as the *real* ensemble, and denote averages with respect to this ensemble with conventional brackets $\langle \cdots \rangle$. However, these weights have the considerable disadvantage of not being local when expressed in terms of the height variables. Instead, for calculational purposes, a different weighting is usually chosen, in which the phases $e^{\pm i\alpha}$ are distributed along each loop by assigning a phase $e^{i\alpha\theta/2\pi}$ each time the loop turns leftwards through an angle θ. Each anticlockwise (clockwise) loop thus accumulates a total phase $e^{i\alpha}$ (resp. $e^{-i\alpha}$). On summing over orientations, these account for the loop weights as long as

$$n = \sqrt{Q} = 2 \cos \alpha \tag{13}$$

Clearly, these weights cannot now be interpreted as probabilities, and we refer to this as the *complex* ensemble. Averages with respect to this will be denoted by $[\cdots]$.

Within each ensemble, we may view each oriented loop as carrying a unit current in the sense of its orientation. Let $J_\mu(x, y)$ be the corresponding current *density*. For example, for a current directed along a link in the positive y-direction, located at $x = 0$, $J_x = 0$ and $J_y = \delta(x)$. This current density may be used to give a formula for the area of a single closed loop:

$$A = -\tfrac{1}{2}\iint |x-x'|\,\delta(y-y')\,J_y(x, y)\,J_y(x', y')\,dx\,dy\,dx'\,dy' \tag{14}$$

This formula is valid for any non-self-intersecting loop, and is independent of its orientation. If however we now consider the same quantity evaluated for a given configuration of a gas of many loops, and we sum over the orientation of each loop independently, $J_y(r)\,J_y(r')$ will average to zero if r and r' are on different loops. Thus the expression

$$-\tfrac{1}{2}\iint |x-x'|\,\delta(y-y')\langle J_y(x, y)\,J_y(x', y')\rangle\,d^2r\,d^2r' \tag{15}$$

where $\langle\cdots\rangle$ denotes the average over the loop gas ensemble, gives the mean total area $\langle A_{\text{tot}}\rangle$ inside *all* loops.

What is the current density $J_\mu(r)$ in the height model parametrization of the configurations? Let us imagine that the definition of the height function $h(r)$ is extended to $\mathbf{R}^2$ in such a way that it is constant within each plaquette. An obvious candidate is then $J_\mu \equiv (1/\pi)\,\epsilon_{\mu\nu}\,\partial_\nu h$. Within the real ensemble, this is clearly correct. It is easy to see that $\langle J_\mu(r)\rangle = 0$ on summing over orientations of a given loop which passes through r. However, in general $[J_\mu(r)] \neq 0$. Consider the average over orientations in the complex ensemble for a *fixed* configuration of unoriented loops:

$$[J] \propto 1\cdot e^{i\alpha} + (-1)\cdot e^{-i\alpha} \neq 0 \tag{16}$$

Instead, we have to consider a different operator as representing the current in the complex ensemble: $\tilde{J}_\mu \propto \epsilon_{\mu\nu}\,\partial_\nu e^{-2i\alpha h/\pi}$, which now gives

$$[j] \propto (e^{-2i\alpha} - 1)\,e^{i\alpha} + (e^{2i\alpha} - 1)\,e^{-i\alpha} = 0 \tag{17}$$

as required. In Coulomb gas language, j has charge $-2i\alpha/\pi$. Since there must be overall charge neutrality, this is balanced by a charge $+2i\alpha/\pi$ distributed on the boundary.

However, the orientation-averaged two-point function $\langle J(r)\, J(r')\rangle$ is *not* correctly represented by $[j(r)\, j(r')]$, since once again this has the wrong charge. Instead we should take

$$\langle J_\mu(r)\, J_\nu(r')\rangle = \lambda[J_\mu(r)\, j_\nu(r')] \tag{18}$$

Note that this does vanish when r and r' are on different loops, by virtue of (17). The constant λ is fixed by requiring that, for a fixed long loop whose sides at x and x' are parallel to the y-axis, after summing over orientations, $\langle J_y(x, y)\, J_y(x', y)\rangle = -n\delta(x)\, \delta(x')$. The factor n arises from the sum over loop orientations in the height model. This gives

$$\lambda((1-e^{-2i\alpha})\, e^{i\alpha} - (1-e^{2i\alpha})\, e^{-i\alpha}) = -n \tag{19}$$

so that $\lambda = n/(4i \sin \alpha)$.

All of this is exact, on the lattice. The weights in the height model are local but complicated, involving as they do the phase factors $e^{\pm i\alpha}$. The central assumption of the Coulomb gas approach is that, for the purposes of studying the long-distance behavior of correlation functions, they may be replaced by the continuum measure $\exp(-S)$, with $S = (g/4\pi) \int (\partial h)^2\, d^2r$. The parameter g may be determined by a number of methods[10, 11] to give $g = 1-\alpha/\pi$, so that $n = \sqrt{Q} = -2\cos(\pi g)$. The correct branches are $1 \leqslant g \leqslant 2$ for x_{c1}, and $0 \leqslant g \leqslant 1$ for x_{c2}.

Within this free field theory, it is straightforward to compute the correlation function $[J_\mu(r)\, j_\nu(r')]$ in terms of the Green function $G(r-r') = [h(r)\, h(r')] \sim -(1/g) \ln |r-r'|$. First note that

$$[h(r)\, e^{-2i\alpha h(r')/\pi}] = \sum_{p=0}^{\infty} \frac{(-2i\alpha/\pi)^p}{p!} [h(r)\, h(r')^p] \tag{20}$$

$$= \sum_{p=0}^{\infty} \frac{(-2i\alpha/\pi)^p}{p!} pG(r-r')[h(r')^{p-1}] \tag{21}$$

$$\sim (2i\alpha/\pi g) \ln |r-r'|\, [e^{-2i\alpha h(r')/\pi}] \tag{22}$$

$$= (2i\alpha/\pi g) \ln |r-r'| \tag{23}$$

The last equality follows because of the way the phase factors enter the sum over orientations, so that $[e^{-2i\alpha h(r')/\pi}] = 1$. We thus find that

$$[J_\mu(r)\, j_\nu(r')] = (2i\lambda\alpha/\pi^2 g)\, \epsilon_{\mu\kappa}\epsilon_{\nu\sigma}\partial_\kappa\partial'_\sigma \ln |r-r'| \tag{24}$$

$$= (2i\lambda\alpha/\pi^2 g)\, \epsilon_{\mu\kappa}\epsilon_{\nu\sigma} \left(\frac{2R_\kappa R_\sigma}{R^4} - \frac{\delta_{\kappa\sigma}}{R^2}\right) \tag{25}$$

where we have introduced $R = r - r'$. After a little algebra we therefore find

$$\langle J_\mu(r)\, J_\nu(r')\rangle = k(n)\,\frac{R_\mu R_\nu - \frac{1}{2}R^2\delta_{\mu\nu}}{R^4} \tag{26}$$

This form of the 2-point correlation function of a conserved current is in fact dictated by rotational invariance, but it is the coefficient $k(n)$ which is the main result: written in terms of g it is[3]

$$k(n) = \frac{n(1-g)}{\pi g \sin(\pi g)} = \frac{2(g-1)}{\pi g}\cot(\pi g) \tag{27}$$

Substituting this result into (15), we notice that the result would appear to diverge logarithmically at $r = r'$. However, this is an artifact of the continuum approximation: the result (26) is valid only for separations $|r - r'| \gg a$. Since the potential divergence is logarithmic, the amplitude of the leading term is insensitive to the precise nature of the modification at shorter distances, and therefore we may impose a simple cut-off $|r - r'| > a$ on the integral. Similarly, the precise form (26) becomes invalid for separations $O(L)$, but the short-distance leading logarithm $\ln a$ must always appear in the form $\ln(a/L)$, on dimensional grounds, independent of the precise geometry. Thus the mean total area within all loops behaves as

$$\langle A_{\text{tot}}\rangle = (k(n)/2)\,\mathscr{A}\,\ln(L/a) + O(1) \tag{28}$$

where $\mathscr{A}$ is the total area of the system.

As shown in ref. 9, in any simply connected region $\mathscr{R}$ the right-hand side of (28) is proportional to $\sum_m (1/\lambda_m)$, where the λ_m are the eigenvalues of—laplacian in $\mathscr{R}$, with Dirichlet boundary conditions. The leading term always has the universal logarithmic behavior shown above. Up to a non-universal constant which may be absorbed into the cut-off a, the $O(1)$ remainder is universal and depends only on the shape of $\mathscr{R}$. For example, for a rectangle it is related to modular forms.

2.1.2. Relation Between Total Area and the Mean Depth

In this section we show how the leading behavior of $\langle A_{\text{tot}}\rangle$ may also be found using methods of conformal field theory on a cylinder, as an extension of the results of ref. 13. Let us define, for a given configuration of unoriented loops, the "depth" $d(r)$ of a given site on the dual lattice to be the minimum number of loops which must be crossed to connect r to the

[3] There is a misprint in the corresponding equation in ref. 9.

boundary. That is, it is the number of noncontractible loops surrounding r. In the height description, it is the supremum of $h(r)/\pi$ over all possible orientations of the given set of loops. As with $h(r)$, we may extend the domain of definition of $d(r)$ to the continuum plane by assuming that it is constant over each plaquette of the dual lattice. Then a little thought shows that the total area within all loops is simply

$$A_{\text{tot}} = \int d(r)\, d^2r \tag{29}$$

so that we need to evaluate $\langle d(r)\rangle$. We do this by first evaluating $\langle d(r, r')\rangle$, where $d(r, r')$ is the minimal number of loops which separate distinct points r and r' in the infinite plane. By translational invariance this is a function of $r-r'$ only. Let us conformally map the plane into a cylinder of perimeter 2π via the usual mapping $w = \ln z$. As $|r-r'| \to \infty$ the images of these points are far apart along the cylinder, and $d(r, r')$ is therefore asymptotically the same as the number of loops which wrap around the cylinder between these points.

From the point of view of the height model with complex weights, these loops must in any case be treated separately, since the factors of $e^{\pm i\theta\alpha/2\pi}$ all cancel, so that each orientation is, a priori, counted with weight 1. As is well known, this may be compensated for by inserting operators $e^{\pm i\alpha h/\pi}$ at opposite ends of the cylinder. The free energy per unit length then is $-c/12$, where $c = 1-(6/g)(\alpha/\pi)^2$. The first term comes from the Casimir effect of the fluctuations of the h-field, while the second is the correction due to the flux between the charges at either end. This then gives the correct result for the central charge c.

Let us now count the loops which wind around the cylinder with a weight $n' = 2\cos\alpha'$, instead of weight $n = \sqrt{Q} = 2\cos\alpha$. The free energy per unit length will be simply modified by an additional term

$$\delta f = (1/2\pi^2 g)(\alpha'^2 - \alpha^2) \tag{30}$$

The mean number of loops which wrap around the cylinder, per unit length, is then found by taking $n'(\partial/\partial n')$ of this expression and setting $n' = n$. Transforming back to the plane, $\langle d(r, r')\rangle$ is given by this same coefficient, multiplying $\ln(|r-r'|/a)$. After a little algebra we then find

$$\langle d(r, r')\rangle \sim (k(n)/2)\ln(|r-r'|/a), \tag{31}$$

where $k(n)$ is given by Eq. (27). The logarithmic dependence of this function also follows from the work of ref. 14, who derived that dependence from general scaling arguments but did not find the coefficient given above.

This result, of course, applies to the number of loops separating r and r' in the infinite plane. However, notice that it has a logarithmically divergent dependence on a, which comes from loops which are much smaller in size than $|r-r'|$. This same divergence should arise if we now consider the number of loops separating a given point r from the boundary in a large but finite system, for all points whose distance from the boundary is much larger than a (but still much less than L). We conclude that

$$\int \langle d(r)\rangle \, d^2r \sim -(k(n)/2)\, \mathscr{A} \ln a \sim (k(n)/2)\, \mathscr{A} \ln(L/a), \tag{32}$$

where the last statement holds because L is the only dimensionful parameter available to compensate a. This gives a second derivation of Eq. (28).

We note in passing that $\langle d(r, r')\rangle = (1/\pi)^2 \langle (h(r)-h(r'))^2\rangle$, in the height representation. The fact that this behaves logarithmically with $|r-r'|$ is consistent with the hypothesis that, in the continuum limit, the heights are distributed according to a gaussian ensemble $\exp(-(1/2\pi k(n)) \int (\nabla h)^2 \, d^2r)$ in the *real* ensemble, even though the lattice weights are nonlocal. However, this hypothesis is incorrect: as may be shown by extending the above calculation on the cylinder to higher moments, the cumulant

$$\langle (h(r)-h(r'))^4\rangle - 3\langle (h(r)-h(r'))^2\rangle^2 \sim \text{const.} \ln(|r-r'|/a), \tag{33}$$

and does not vanish as it would in a gaussian ensemble. However, note that this cumulant decays faster than each term on the left hand side, so that asymptotically the distribution of $d(r)$ is normal, as was proved by Kesten and Zhang.[15]

2.2. Relation Between *k* and *C*

In this section we first show how the logarithmic behavior of $\langle A_{\text{tot}}\rangle$ provides further justification for the assertion that $N(A) \sim A^{-1}$ for $a^2 \ll A \ll L^2$, then show how to relate the coefficients. Recall that $N(A)$ is the number of loops with area greater than A, divided by the total area of the system. For $A \sim L^2$, this will also depend on L, so let us write it as $N(A, L)$. On dimensional grounds it has the form

$$N(A, L) = (1/A)\, F(A/a^2, L/a) \tag{34}$$

where a is the lattice spacing. For $a^2 \ll A \ll L^2$, we expect it to be independent of L, but, a priori, it could depend on a. In this regime, let us suppose it has the form

$$N(A, L) \sim 2C(1/A)(A/a^2)^{\omega} \tag{35}$$

where C is a constant and ω is some exponent. This is not of course the most general dependence which is possible, but is that which would arise if, for some reason, the area scaled non-trivially with a, that is, had a fractal structure. Such dependence, with $\omega \neq 0$, would for example occur in the distribution of masses, rather than of areas, of percolation clusters. The form (35) should of course connect smoothly onto the behavior for $A \sim a^2$, when we expect that $N \to \text{const.}$, and $A \sim L^2$, where $N \to 0$.

Now the total area A_{tot} within all loops is related to N by

$$A_{\text{tot}} = \sum_A N(A, L) \tag{36}$$

Comparing with (28), we see that the contribution to the sum from the region $a^2 \ll A \sim L^2$ will exceed $O(\ln(L/a))$ if $\omega > 0$, and similarly the contribution from $a^2 \sim A \ll L^2$ will violate this bound if $\omega < 0$. Therefore $\omega = 0$, and $N(A) \sim 2C/A$ for $a^2 \ll A \ll L^2$. Admittedly, this argument assumes the *ansatz* (35), and the reader may be more comfortable with the hyperscaling argument put forward in the Introduction. However, independently of the validity of (35), our argument shows that if $N(A) \sim C/A$, then the coefficient C is related to $k(n)$. For then the leading contribution from the region $a^2 \ll A \ll L^2$ is $2C \ln(\mathscr{A}/a^2) \sim 4C \ln(L/a)$, so that comparing once again with (28),

$$C = k(n)/8, \tag{37}$$

with $k(n)$ given by (27).

For percolation cluster hulls and FK clusters in the Q-state Potts model, we take $n = \sqrt{Q} = -2 \cos \pi g$ in the dense phase $0 \leqslant g \leqslant 1$, which yields $g = \frac{2}{3}, \frac{3}{4}, \frac{5}{6}$, 1 for $Q = 1$, 2, 3, and 4, respectively. For critical Ising spin clusters, we take $n = 1$ in the dilute phase where $1 \leqslant g \leqslant 2$, so that $g = \frac{4}{3}$. Then by (27) we find the values of C given in Section 1 and also listed in Table I (taking the limit in the case $Q = 4$). The logarithmic corrections that appear for the case $Q = 4$ are derived in the Appendix.

Table I. Predicted and Measured Values of C for Various Systems

Cluster type	C (theoretical)	C (measured)
Percolation	$1/8\sqrt{3}\,\pi = 0.022972037\ldots$	0.0229721(1)
Ising spin	$1/16\sqrt{3}\,\pi = 0.011486019\ldots$	0.01149(5)
Ising FK	$1/12\pi = 0.026525824\ldots$	0.0265
$Q = 3$ Potts FK	$\sqrt{3}/20\pi = 0.027566445\ldots$	0.0278
$Q = 4$ Potts FK	$1/4\pi^2 = 0.025330296\ldots$	0.0258

3. NUMERICAL RESULTS

To test these predictions, we carried out numerical studies of percolation on square and triangular lattices with both site and bond percolation, and the Ising/Potts models on the square lattice. For percolation we considered two ways to generate the clusters: populating the entire lattice, and individual hull generation.

3.1. Bond Percolation—Full-Lattice Population Method

In the full-lattice population method, we first assign all bonds on the lattice as occupied or vacant with probabilities p or $1-p$, respectively, and then carry out all possible hull walks around these bonds. These walks go from the center to center of each bond along the diagonals, as shown in Fig. 1, and turn by an angle $+\pi/2$ when the center of an occupied bond is encountered, and by $-\pi/2$ when the center of a vacant bond is encountered. Each walk is completed when it returns to its beginning step.

For bond percolation on the square lattice, we used a square lattice of size 512×512 with periodic boundary conditions, with p at the threshold $1/2$. We simulated 10^7 samples, amounting to a total of 5.2×10^{12} bonds occupied or not. We used the R9689 random number generator of ref. 16. In the computer program we employed an array of size 2048×2048, so that we had distinct array locations to represent the bonds and each diagonal leg of the hull walks.

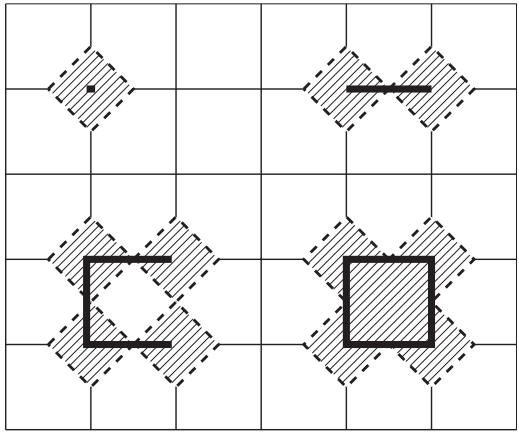

Fig. 1. Hull paths for bond percolation, with enclosed shaded areas of $\frac{1}{2}$ (top left), 1 (top right), 2 (bottom left) and $2\frac{1}{2}$ (bottom right). These are all external hulls—the last case also has an internal hull of area $\frac{1}{2}$ (not shown).

The enclosed area of a hull walk was found "on the fly" by the following method: Initially, the area is set equal to zero. When the walk steps to the right, the area is increased by one-half the y coordinate at the center of the diagonal step (where we had an array point), and decreased by one-half the y coordinate when the walk steps left. The zero point of the y coordinate is irrelevant, because its value cancels out. The factor of $1/2$ comes from the fact that each leg of the hull walk changes the x-coordinate by $\pm 1/2$; we are taking the spacing of the bond lattice to be unity. When the walk closes, this algorithm gives the area of the enclosed space, with a sign attached: positive areas correspond to external hulls that surround clusters, and negative areas corresponds to internal hulls (which are of course external to the clusters on the dual lattice).

The smallest area is $1/2$; for positive area this corresponds to the hull around an isolated site (one with no bonds attached), and negative $1/2$ corresponds to the hull inside a square of four occupied bonds, or equivalently around an isolated site on the dual lattice. The area of all the hulls are in units of $1/2$. (Alternately, one could consider the lattice spacing to be $\sqrt{2}$; then the hulls would all have integer areas, and the system area would be $2L^2$.)

Because we use periodic boundary conditions, there is the possibility that some hulls could wrap around the torus once or more before closing into a loop. The areas for such loops are undefined, unless taken in pairs, but in any case we discarded them because we are interested in clusters whose size is much smaller than the size of the system.

We found the statistics for internal and external hulls were identical (within numerical error), as one would expect for this self-dual system, and took the average of the two.

For small A we kept track of the quantity $N_A =$ the number of loops (per unit area) whose enclosed area is exactly A, where $A = \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$. According to (3), this quantity should behave as

$$N_A = N(A) - N(A+1/2) \sim \frac{C}{2A^2} \tag{38}$$

so that $2A^2 N_A \sim C$ for large A. The results are given in Table III for $A \leqslant 5$.

To check these results, we derived the exact expressions for N_A for $\frac{1}{2} \leqslant A \leqslant \frac{9}{2}$ given in Table II. These are for an arbitrary bond occupancy of p, with $q = 1-p$. For $A = \frac{1}{2} \cdots 3$ these expressions are identical to the expressions for the number of clusters (per site) containing $b = 2A-1$ bonds, which are well known.[17, 18] For larger A we had to make modifications to the bond cluster expressions to take into account graphs that contain internal open spaces with vacant bonds, which result in areas larger than

Table II. Exact Results for $N_A(p)$ for Bond Percolation on the Square Lattice at Occupancy $p = 1 - q$, for $A = \frac{1}{2} \cdots \frac{9}{2}$

A	$N_A(p)$
$\frac{1}{2}$	q^4
1	$2pq^6$
$\frac{3}{2}$	$6p^2q^8$
2	$p^3(4q^9+18q^{10})$
$\frac{5}{2}$	$p^4(q^8+32q^9+55q^{10})$
3	$p^5(8q^{10}+30q^{12}+160q^{13}+174q^{14})$
$\frac{7}{2}$	$p^6(12q^{11}+40q^{12}+332q^{14}+672q^{15}+570q^{16})$
4	$2p^6q^{11}+p^7(2q^{10}+136q^{13}+168q^{14}+336q^{15}+2030q^{16}+2712q^{17}+1908q^{18})$
$\frac{9}{2}$	$20p^7q^{13}+p^8(22q^{12}+186q^{14}+844q^{15}+868q^{16}+4064q^{17}+9972q^{18}+10880q^{19}+6473q^{20})$

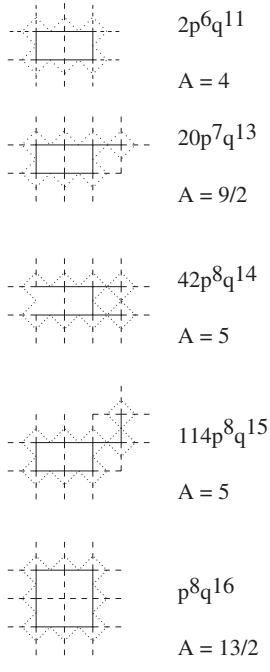

Fig. 2. Clusters contributing higher-area terms to polynomials in Table II. Solid lines represent occupied bonds, dashed lines are vacant bonds, and the dotted lines trace out the external hull. These are the graphs that have to be "moved" in the usual cluster polynomials from $A = (b+1)/2$ (where b is the number of bonds) to higher A due to the existence of enclosed open spaces.

$(b+1)/2$. We subtracted the term $2p^6q^{11}$ from $N_{\frac{7}{2}}$ and added it to N_4 to account for the area of an open 1×2 rectangle, whose external hull area is 4, not $7/2$. Likewise, the term $20p^7q^{13}$ (the 1×2 rectangle with an extra bond attached) was subtracted from N_4 and added to $N_{9/2}$. Finally, the terms $42p^8q^{14}$, $114p^8q^{15}$, and p^8q^{16}, which correspond to various graphs with area greater than $9/2$, were subtracted from $N_{9/2}$. These various diagrams are shown in Fig. 2. This shifting of terms has the effect of making N_A follow asymptotically the exponent -2 of (38) rather than the exponent $-\tau = -2.055\ldots$ followed by n_s.

Taking $p = 1/2$ and multiplying by $2A^2$, we arrive at the estimates for C listed in Table III. The agreement with our numerical results is excellent—within the small statistical error. Interestingly, the convergence of these estimates is rather quick—already, at $A = 5$, the result is within 6% of the (presumably) exact value.

To analyze the data for larger A, we considered the quantity $N(A, 2A) \equiv$ the number of clusters whose enclosed area is greater or equal to A and less than $2A$. According to (3), this quantity should behave as

$$N(A, 2A) = N(A) - N(2A) \sim \frac{C}{2A} \tag{39}$$

so that $2AN(A, 2A) \sim C$ for large A.

The measured values of $2AN(A, 2A)$ are given in Table IV. They monotonically decrease to a value 0.0229860(45) for $A = 2048$, but then slightly increase at $A = 4096$; for larger A, the increase continues, as seen in Fig. 3, (upper curve) where the data from $A = 128$ to 16384 are shown. We

Table III. Values of $2A^2N_A$ for Small A for Bond Percolation on the Square Lattice: Two Algorithms and Exact Results. Errors in Last Digit Are Given in Parentheses

A	Full Lattice	Single Hull	Exact Results
1/2	0.0312500(1)	0.031247(3)	$1/32 = 0.03125$
1	0.0312500(1)	0.031252(4)	$1/32 = 0.03125$
3/2	0.0263674(1)	0.026363(4)	$27/1024 = 0.026367188$
2	0.0253907(2)	0.025398(5)	$13/512 = 0.025390625$
5/2	0.0257491(2)	0.025744(6)	$3375/2^{17} = 0.025749207$
3	0.0254749(3)	0.025477(6)	$3339/2^{17} = 0.025474548$
7/2	0.0249188(3)	0.024917(7)	$104517/2^{22} = 0.024918795$
4	0.0249898(4)	0.025004(7)	$6551/2^{18} = 0.024990082$
9/2	0.0247714(4)	0.024778(8)	$13298985/2^{29} = 0.02477129$
5	0.0245659(5)	0.024557(8)	

Table IV. Values of $2AN(A, 2A)$ for Bond Percolation on the Square Lattice for the Two Algorithms

A	Full Lattice	Single Hull
1/2	0.0625001(1)	0.0625022(42)
1	0.0429689(1)	0.0429634(32)
2	0.0306645(2)	0.0306677(26)
4	0.0270220(2)	0.0270226(25)
8	0.0250527(3)	0.0250528(25)
16	0.0240647(4)	0.0240635(25)
32	0.0235504(6)	0.0235463(26)
64	0.0232785(8)	0.0232740(27)
128	0.0231360(11)	0.0231412(28)
256	0.0230598(16)	0.0230616(29)
512	0.0230218(22)	0.0230212(31)
1024	0.0229968(32)	0.0229948(32)
2048	0.0229860(45)	0.0229849(33)
4096	0.0229882(63)	0.0229785(35)

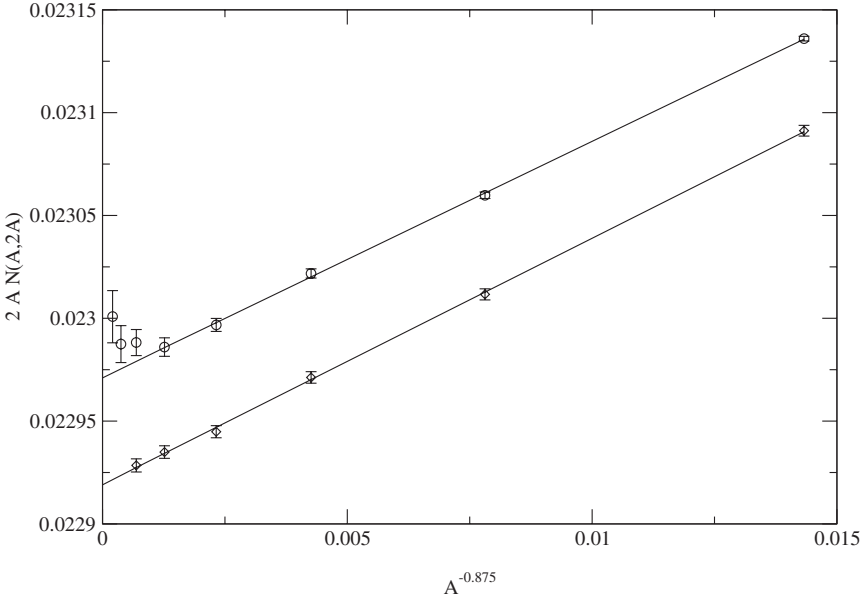

Fig. 3. Plot of $2AN(A, 2A)$ vs. $A^{-0.875}$ for bond percolation on a square lattice. Upper data points: lattice population method. Lower points (shifted down by 0.00005): single hull generation method.

attribute this increase to interference of clusters with themselves around the periodic boundary conditions, and thus ignore these data. Fitting the 5 data points from $A = 128$ to 2048 as a function of $A^{-\theta}$ to a straight line, we find a good linear fit with $\theta = 0.875$ as shown in that figure, with the equation of the line given by

$$2AN(A, 2A) = 0.0229712 + 0.01148A^{-0.875} \tag{40}$$

implying $C = 0.0229712$.

We estimate the error in the above value of C to be $\approx 10^{-6}$ from the statistical error of the data and the uncertainty in the extrapolation to infinity. The predicted value (5) falls within these error bars.

In terms of the length scale $\ell \sim A^{\frac{1}{2}}$, the above exponent corresponds to a correction of the order $\ell^{-1.75}$, which is the scaling of the hull of the cluster. Indeed, this finite-size correction can be interpreted as a surface effect,[19] reflecting the arbitrariness in locating where precisely the hull of the cluster should be placed.

As a test of our procedure, we also compared our measurement of the total number of loops (hulls of both type) per unit area ($\approx$ twice the number of clusters) with the theoretical result, which for bond percolation on the square lattice is given by the Temperley–Lieb result[20, 21]

$$\sum_A N_A \sim 3\sqrt{3} - 5 = 0.196152422\ldots \tag{41}$$

Our measured value was 0.1961572(14), larger than the above prediction by only 0.0000048(28). This difference corresponds to an excess number of 1.2 loops per lattice (found by multiplying the latter number by 512^2), which is barely discernible above the statistical error of ± 0.7. In fact, this correction can also be predicted theoretically. For a system with a rectangular boundary of aspect ratio r, the excess number of clusters is a known function $b(r)$.[21, 22] To find the excess number of loops, note that the quantity $n_c + n_{c'} - n_l$ (the number of clusters, plus the number of dual lattice clusters, minus the number of loops) equals 1 if there is a cross-configuration on the lattice or the dual lattice, and zero otherwise. Thus it follows that the excess number of loops is just $2b(r) - 2\pi_+(r)$, where $\pi_+(r)$ is the cross-configuration probability, which has been calculated by Pinson.[23] For a square system, the excess number of loops is predicted to be

$$2[b(1) - \pi_+(1)] = 2(0.883576 - 0.309526) = 1.14810 \tag{42}$$

using $b(1)$ from ref. 22 and $\pi_+(1)$ from ref. 1. This prediction happens to coincide almost exactly with the measured value (even though the error

bars of the latter are quite large). This predicted value can be tested to higher precision most easily by going to smaller lattices.

Besides the problem of clusters interfering with themselves, there is also the problem in the population method that the statistics for larger hulls are rather poor because of the relatively small number of such hulls that are generated. In the next section we consider a method that addresses both of these problems.

3.2. Bond Percolation—Single Hull Generation Method

It is well known that percolation clusters can be grown individually through a process where bonds are made occupied or not only when they are encountered (the "Leath" method). In the same way, percolation hulls can be generated individually on a blank (undetermined) lattice by a kind of growing self-avoiding walk that mimics the walk used to trace out hulls.[24] For critical bond percolation,[25] the walker moves along the edges of a square lattice (the diagonals in Fig. 1), and turns by $+\pi/2$ or $-\pi/2$ randomly at each vertex, except at vertices previously visited, where it always turns to avoid retracing itself. The walk terminates when it returns to the origin and cannot proceed further. Note that pseudo-random numbers are generated only for the bonds that are visited during the walk, making this method efficient. This walk has also been studied as a kinetic Lorentz-gas model,[26] and the results here apply to that model also.

In order for the contribution of a given hull to be the same as on the fully populated lattice, it is necessary to weight each walk by $1/t$, where t is the number of hull steps. This compensates for the fact that a hull of t steps is generated with t times the probability in the single cluster method compared to the population method, because there are t places a given walk can start from.

This weighting can also be checked as follows: The probability of generating a closed hull of at least t steps is given by[27]

$$P(t) \sim ct^{-1/7} \tag{43}$$

where c is a constant. Defining a Euclidean length scale $\ell \sim t^{1/D_H}$, it follows that the area enclosed by the walk scales as $A \sim \ell^2 \sim t^{2/D_H} = t^{8/7}$. Thus, the probability of growing a walk enclosing at least area A scales as $A^{-1/8}$, so that the probability of growing a walk of exactly area A scales as $A^{-9/8}$. When we weight a hull by the factor $1/t \sim A^{-7/8}$, we thus get the proper probability $A^{-9/8-7/8} = A^{-2}$ as given in (38).

In our simulations, we considered square lattices of size $L \times L$ with periodic b.c. This is the lattice of the hull walks, which is rotated by $\pi/4$

from the square bond lattice, and has a spacing that is $\sqrt{2}/2$ of the bond lattice spacing. Note that the square system boundary here corresponds to a diamond on the bond lattice. We stopped all walks that did not close by 65536 steps, and kept track of the areas of all the walks that closed before this cutoff, without wrapping around the periodic b.c. All walks that were stopped at the cutoff would ultimately enclose an area of at least $65536/8 = 8192$, taking into account that there are at most 4 hull steps around each wetted site, and each square on the hull-walk (rotated) lattice corresponds to an area of $1/2$.

While the statistics of walks of areas smaller than $A = 8192$ should thus be unbiased by having this cutoff, they can still be biased by the finite-size of the lattice. For runs on lattices of size 1024×1024 and smaller, we found both wraparound clusters and large deviations in the hull statistics for larger A. Even for lattices of size 2048×2048, where no wraparound occurred with this cutoff, we still found significant, obviously finite-size deviations even for $N(A, 2A)$ for A below $A = 8192$. We attribute these deviations to hulls making contact with themselves around the periodic b.c., without actually closing to wrap around. Therefore, to be absolutely certain of no finite-size effects, we went to a lattice of size 65536×65536 using the virtual lattice method of ref. 24. We checked that with the cutoff of 65536 steps, indeed no walk got anywhere near the boundary of the system.

We carried out 1.8×10^9 walks on this lattice, which, like the simulations for the 10^7 fully populated lattices, took several weeks of workstation computer time. A total of 3.2×10^{13} hull steps simulated here, compared with 1.0×10^{13} in the simulations of the populated lattices. The algorithm for the single-hull method is somewhat simpler and more efficient than that for the lattice population method.

In the single-hull method, larger hulls are generated with a higher probability than in the lattice-population method: the number generated in the interval $(A, 2A)$ (before reweighting) is proportional to $A^{-1/8}$ here, compared with A^{-1}. This is advantageous because the large hulls with their small finite-size effects are essential for finding C accurately. On the other hand, in the single-hull method a large fraction of time is spent on the walks that reach the cutoff before they close (and are discarded): Eq. (43) implies that the total number of steps for all the hulls that reach the cutoff $t_{\max}$ grows as $\sim c t_{\max}^{6/7}$, while the total number of steps for all the hulls that close before $t_{\max}$ is given by

$$\int^{t_{\max}} t \left(-\frac{dP}{dt} \right) dt \sim \frac{c}{6} t_{\max}^{6/7}. \tag{44}$$

Thus, no matter what the value of the cutoff is, a fraction $6/7 = 85.7\%$ of the work (ignoring finite-size effects) is spent generating walks that reach the cutoff without closing and are thus discarded. Still, for very large cutoffs this overhead is compensated by the increase in useful statistics for large A, making this method advantageous.

Note that, in our simulations of 3.2×10^{13} hull-walk steps, the fraction of those steps belonging to clusters that reached the cutoff $t_{\max} = 65536$ was $6.000124/7$, with the deviation from 6 in the numerator being about equal to the apparent statistical error, ≈ 0.0001. This result seems to provide a very precise confirmation that the exponent in $P(t)$ is indeed $-1/7$ (i.e., the hull fractal dimension is $D_H = 7/4$), although to quantify the precision of this result one would have to investigate different values of the cutoff $t_{\max}$ to determine the finite-size corrections.

For small A, results for $2A^2N_A$ are given in Table III agree with the exact values, confirming that the $1/t$ weighting is correct. Because the single hull method gives fewer of these small hulls than the lattice population method, these results have larger error bars. Here use used $(N_A(total))^{-1/2}$, where $N_A(total)$ is the total number of clusters of size A, to estimate the error bars.

Likewise, the results for $N(A, 2A)$ for all A, given in Table IV, are seen to agree with the lattice population results. For the largest size ranges, the single-hull method is seen to give better error bars (and are not biased by finite-size boundary effects).

A plot of the $2AN(A, 2A)$ vs. $A^{-0.875}$ for $128 \leqslant A \leqslant 4096$ is also given in Fig. 1 (shifted down by 0.00005), and the data are fit by the linear function given by

$$2AN(A, 2A) = 0.0229692 + 0.01197A^{-0.875} \tag{45}$$

which is consistent with the results of the lattice population method (40). The error bars on the intercept is about the same, 10^{-6}.

Thus, although the single-hull method is in principle advantageous, for the system size we considered we obtained C with about the same precision as the lattice population method, with about the same amount of work. However, the single-hull method allowed us to show that the curvature in the behavior of $2AN(A, 2A)$ for large A as seen in Fig. 1 was indeed due to cluster interference around the periodic boundaries.

Assuming the predicted value of C given by (5), we can also make a plot of $\log(2AN(A, 2A) - C)$ vs. $\log(A)$ (not shown); with single-hull data we find good linear behavior with a slope $-\theta = -0.88 \pm 0.01$, which is consistent with the value 0.875 that we have been using.

3.3. Site Percolation on the Square and Triangular Lattices

We also carried out simulations of site percolation on two different lattices to demonstrate the universality of the result (5) for C.

For site percolation, the logical choice for the hull walk around a cluster is to follow a path on the medial lattice whose vertices are at the center of the faces of the lattice, as shown in Fig. 4 for the square and triangular lattice. This choice allows the single isolated site to have a non-zero area, and is symmetric for internal and external hulls for the triangular lattice.

For the square lattice, we carried out 4×10^8 samples on a lattice of size 256×256, using the weighted single-hull method. (In our program we employed a computer array of size 512×512 to include the sites of the medial lattice.) With such a small lattice, finite-size effects appeared for hulls with A larger than ≈ 1024. We used occupancy probability $p = 0.592746$, which is close to the critical threshold for this system.[28] Here we generated the hulls starting from a segment between a single occupied and vacant site, which occurs in a populated system with a probability of $p(1-p)$. The latter factor was therefore included in the total weight of each hull, along with the $1/t$ weight, where here t is the number of steps along the medial lattice.

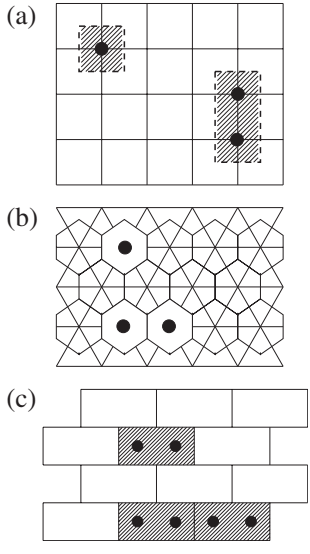

Fig. 4. Medial lattices used for hulls in site percolation: (a) square lattice, (b) triangular lattice, and (c) brick-lattice form of triangular lattice.

We found that the statistics for internal and external hulls are quite different, as one would expect by the asymmetry of this system. For example, for $A = 1$, $N_1 = p(1-p)^4 \approx 0.0163053$ for an external hull, and $N_1 = (1-p)\, p^8 \approx 0.00620604$ for an internal hull. This large difference persists as A increases, and suggests that some other definition of the hull which gives more symmetric results between external and internal hulls might be advantageous.

In Fig. 5 we show $2AN(A, 2A)$ for the two kinds of hulls, along with their average. Taking the average is the same as including both types of hulls in the area calculation (and dividing by two). Indeed, in the theoretical development in Section 2, both internal and external hulls were included in the calculation, so it is appropriate to take this average. The finite-size corrections to the average measure again followed a behavior with exponent close to -0.875, which was used in the plot in Fig. 5. The line in that figure is fit by the equation

$$2AN(A, 2A) = 0.022976 - 0.0114 A^{-0.875} \tag{46}$$

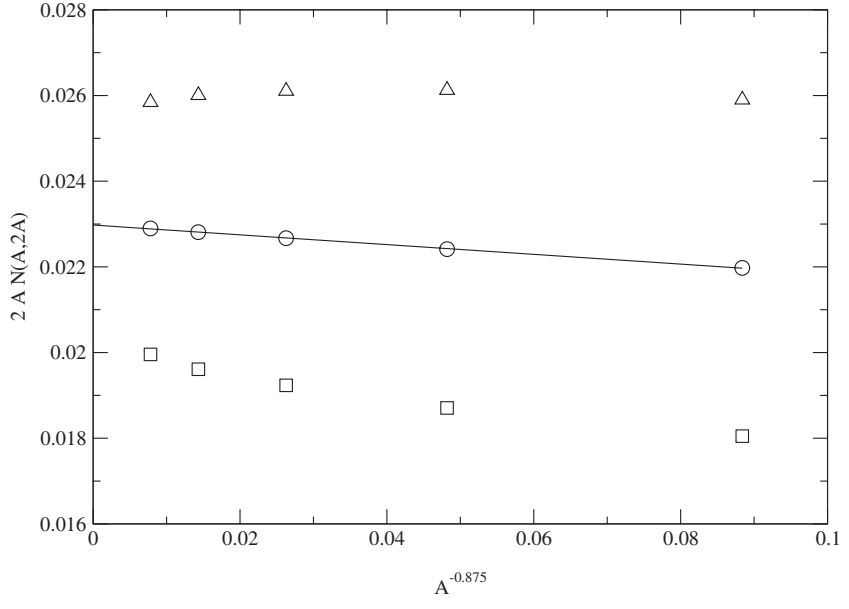

Fig. 5. Plot of $2AN(A, 2A)$ vs. $A^{-0.875}$ for site percolation clusters on a square lattice: external hulls (triangles), average (circles) and internal hulls (squares). The equation of the line fit through the average points is given in Eq. (46).

where the unit of area is one square lattice spacing on the square lattice. The average measure extrapolates (for large A) to a value ≈ 0.022976, in obvious agreement with the theoretical prediction, and making a more precise determination rather superfluous.

Note that, the coefficient to the correction term is similar in value to the coefficient for bond percolation, even though a different kind of path was used to define the hulls in the two cases. The similarity might be a coincidence, or it might reflect a fundamental equivalence of perimeter corrections for site and bond percolation on this lattice.

For site percolation on the triangular lattice the medial lattice is a honeycomb lattice with a hexagon around each vertex of the triangular lattice as shown in Fig. 4. To implement this in the computer, we used the square-lattice form of the honeycomb lattice, also shown in that figure, where the hexagons become rectangular bricks and a single site on the triangular lattice now becomes a pair of sites on the square lattice. Thus, we could use the same basic algorithm as we used for site percolation on the square lattice, with the only modification being that sites are occupied or made vacant in pairs, with a probability $1/2$.

For this case (the triangular lattice) we used the lattice-population method on an underlying square lattice of size 1024×1024 with periodic b.c. We generated 2.4×10^6 independent samples. As expected, the internal and external hulls had equal statistics, within error, reflecting the symmetry of this system.

Again, the data closely followed the $A^{-0.875}$ behavior, and we do not plot it. Fitting the data in the range $2^6 < A < 2^{12}$ (where A is measured in square-lattice units, so that the smallest hexagon corresponds to $A = 2$) we found the following behavior:

$$2AN(A, 2A) = 0.022977 - 0.0146A^{-0.875} \tag{47}$$

again agreeing with the predicted value of C. In this case, that value is approached from below for finite systems, with the definition of cluster-hull area used here.

3.4. Ising Clusters

To study the clusters of the Ising model, we considered a square lattice of size 1024×1024 with periodic b.c., and simulated the system at the critical temperature of $\exp(-J/kT) = 1 + \sqrt{2}$ (where J is the coupling constant in the Potts model formulation, $H = -J \sum \delta_{\sigma_i} \delta_{\sigma_j}$) using the Wolff variation[29] of the Swendsen–Wang method.[30] We initialized the lattice with 1000 updates, and then measured the hull area distribution treating

the system exactly as if it were one of site percolation, using the same definition of hull areas as shown in Fig. 2. This was followed by 10–100 Wolff updates, and the procedure was repeated. 140,000 realizations were generated.

As in the site percolation case, we found rather large differences between internal and external hulls, as seen in Fig. 6. Here we also found very large deviations for large A, presumably reflecting stronger correlations due to the interaction. (Indeed, runs on a smaller 256×256 lattice showed even stronger large-A deviations.) The average measure for the smaller hulls is consistent with the $A^{-0.875}$ finite-size scaling used in that figure, and a fit of the points for $A = 2$ through 16 yields the straight line as shown in that figure, given by

$$2AN(A, 2A) = 0.011487 + 0.004458A^{-0.875} \tag{48}$$

The intercept is nearly identical with the predicted value of C (6), in spite of the rather small size of clusters that were used. We estimate the error in the intercept to be ± 0.00005.

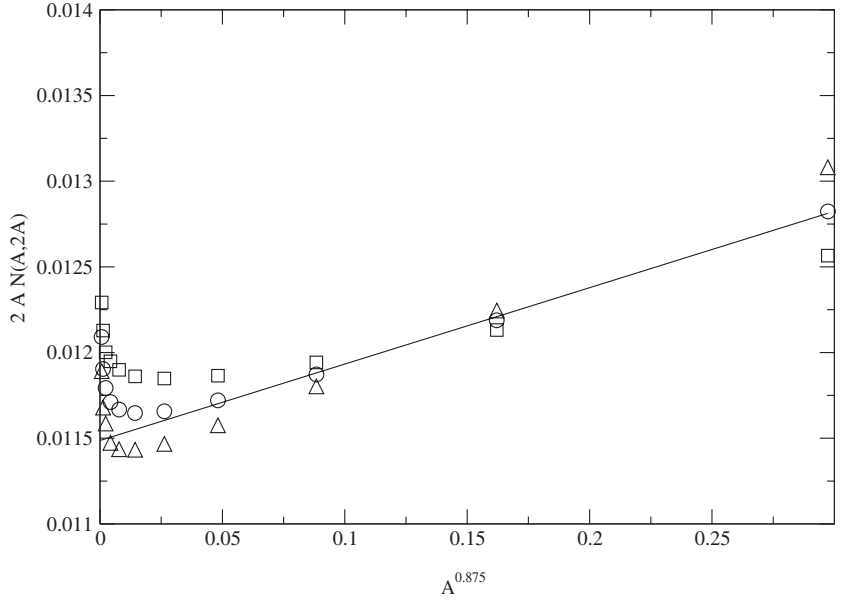

Fig. 6. Plot of $2AN(A, 2A)$ vs. $A^{-0.875}$ for Ising clusters: external hulls (triangles), average (circles) and internal hulls (squares). The line is fit through the four rightmost points of the average values.

3.5. FK Clusters on the Potts Model for $Q=2$, 3, and 4

We also studied the FK clusters on the Potts model at the critical temperature $e^{J/kT} = 1+\sqrt{Q}$. These clusters are the bond percolation clusters when bonds are drawn between neighboring identical spins with a probability $1-e^{-J/kT} = \sqrt{Q}/(1+\sqrt{Q})$. We defined the hulls exactly as in the square-lattice bond percolation case (Fig. 1) and indeed could use the same algorithm to trace out and measure the hulls after the bonds have been specified.

To thermalize the system, we used the Swendsen–Wang (SW) procedure of identifying all FK clusters on the lattice and then randomly reassigning their spins. Indeed, the FK hull measurements and the SW update method naturally go hand in had in this calculation, since the identification of the FK clusters is needed for the SW method. For $Q=2$ and $Q=3$ we used a lattice of size 512×512 and obtained the results shown in Figs. 7 and 8, where once again we find large discrepancies between internal and external clusters, and take the average of the two. That average is found to fall on a nearly straight line when plotted as a function of $A^{-\theta}$ taking $\theta = 0.875$ for $Q=2$ and $\theta = 0.7$ for $Q=3$. The extrapolated exponents are seen to approach the expected theoretical values, as shown in Table I. We simulated 82, 000 samples ($Q=2$) and 1,000,000 samples ($Q=3$).

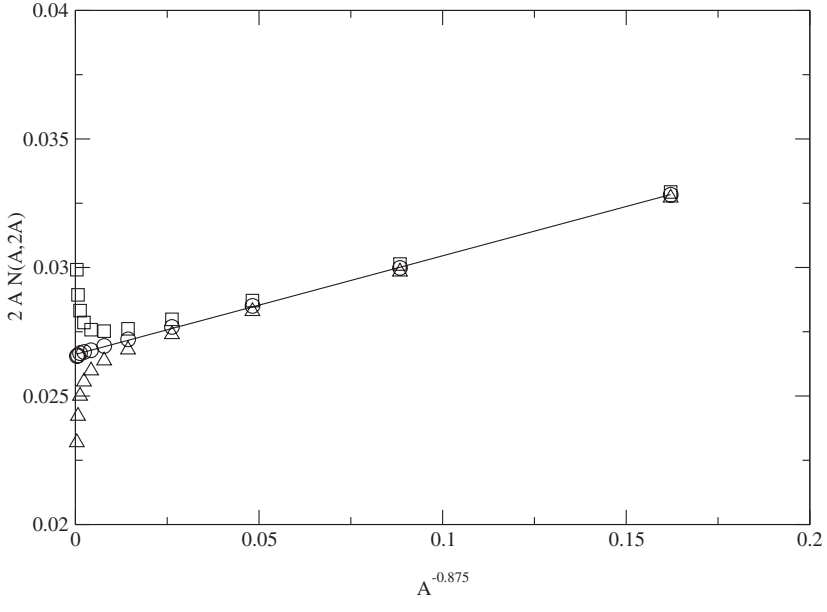

Fig. 7. Plot of $2AN(A, 2A)$ vs. $A^{-0.875}$ for FK clusters of the Ising model ($Q=2$ Potts): external hulls (triangles), average (circles) and internal hulls (squares).

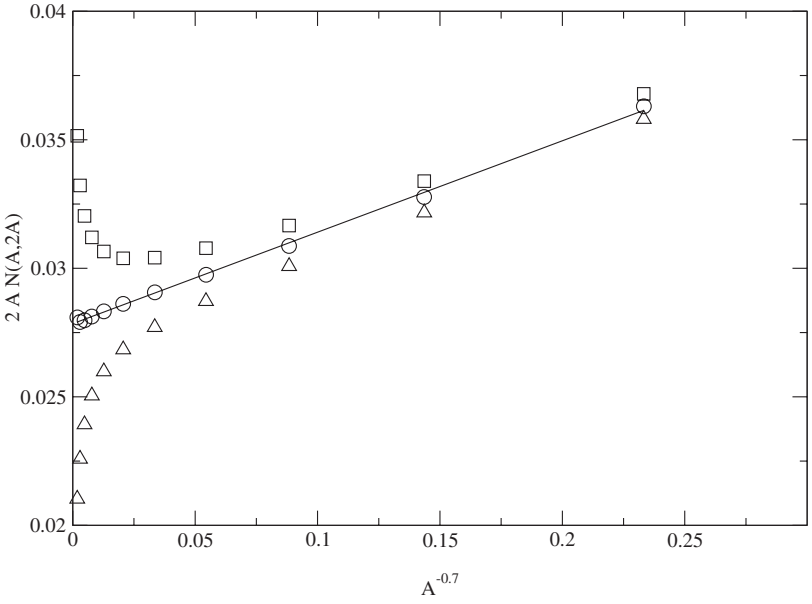

Fig. 8. Plot of $2AN(A, 2A)$ vs. $A^{-0.7}$ for FK clusters of the $Q=3$ Potts model: external hulls (triangles), average (circles) and internal hulls (squares).

Note that the discrepancy between internal and external hulls reflects an inherent asymmetry for FK clusters of the Potts model in finite periodic systems for $Q > 1$. This asymmetry is also manifested in the behavior of the fraction of bonds that are occupied, which in finite systems has a value somewhat greater than the infinite-system value of $\frac{1}{2}$.[31]

For $Q = 4$, very large differences between internal and external hulls persisted even for relatively small values of A on the 512×512 lattice, so we went to a larger lattice of size 2048×2048 (20,000 realizations) which improved the behavior somewhat. Even for this lattice, however, large finite-size effects were apparent. Similar large finite-size corrections have been seen in other Potts model studies at $Q = 4$ (e.g., refs. 31 and 32) and are generally expected to be logarithmic in character. In Appendix A we calculate these corrections analytically for this case and find

$$N(A) \sim \frac{C}{A}\left(1 - \frac{2a_2}{(\ln A)^2} + O((\ln A)^{-3})\right) \tag{49}$$

where a_2 is a constant. The above result implies that $2AN(A, 2A) = C + O((\ln A)^{-2})$. In Fig. 9 we plot our results for $2AN(A, 2A)$ as a function of $(\ln A)^{-2}$. The data fall on a straight line for large A, and the intercept

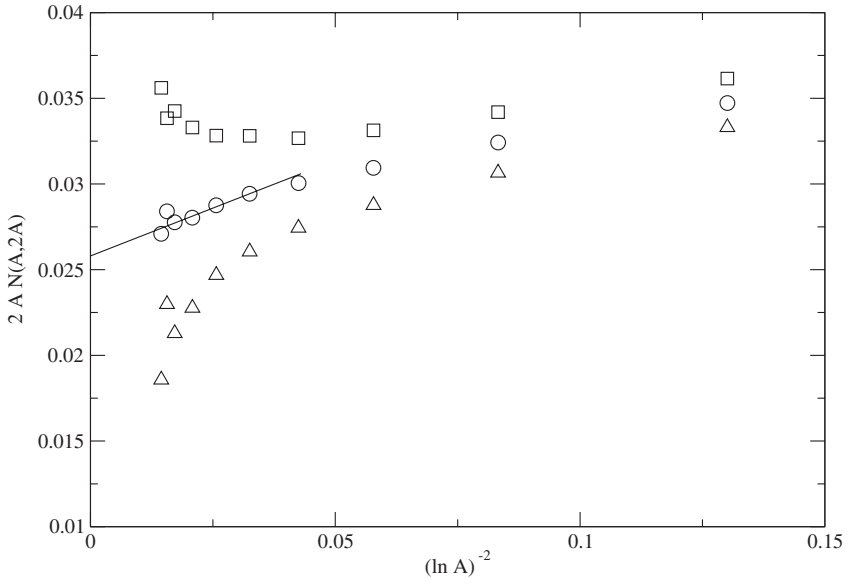

Fig. 9. Plot of $2AN(A, 2A)$ vs. $(\ln A)^{-2}$ for FK clusters of the $Q=4$ Potts model: external hulls (triangles), average (circles) and internal hulls (squares).

yields $C = 0.0258$, which is comparable to our predicted value of $1/4\pi^2 = 0.0253\ldots$.

Note that if we plot the data versus $1/\ln A$, we find about as good of a fit to linear behavior for large A, but then the intercept would 0.0231, quite a bit below the predicted value of C. Likewise, if we fit the data to a power-law as we did for other values of Q, we find fairly linear behavior with an abcissa of $A^{-0.5}$, but now the intercept is 0.0279. Thus, the data is consistent with our prediction for C combined with the predicted $1/(\ln A)^2$ finite-size scaling.

4. CONCLUSIONS

We derived and numerically confirmed predictions for the behavior of the area-size distribution of various Potts model including percolation clusters. For the latter, we also considered different lattices and percolation types (site and bond) to demonstrate universality. The theoretical ideas presented in Section 2 were well verified numerically, especially in the percolation and Ising model cases. For $Q = 4$, our results were consistent with the logarithmic finite-size behavior predicted here.

This work confirms the idea of a universal size distribution expressed by Eq. (3). An alternate way to state that result is as follows: Consider that

the unit of area is now some value A much smaller than the lattice size (which is therefore no longer of unit area). Then, (3) implies that the number of clusters whose enclosed area is greater than A, per unit area A, is a constant C, for all values of A. The lack of dependence on A is a direct consequence of the scale-free nature of this fractal system.

The arguments put forward in Section 2.1.2 also imply that the number of cluster hulls which must be crossed to connect a typical point deep inside the system to the boundary behaves as $4C \ln(L/a)$, where L is the system size, with same value of C for each universality class. So, for example, the fact that C for critical Ising spin clusters is half that for percolation clusters means, according to Zipf's law, that the nth largest cluster is the Ising case has roughly half the area of the nth largest percolation cluster. This is consistent with the fact that we have to cross one half as many cluster hulls to reach the boundary in the Ising case. It might suggest that we may go from the ensemble of percolation hulls to those of Ising clusters simply by erasing every other percolation hull, e.g., by ignoring all the internal hulls! This however is not the case, as percolation hulls have a different fractal dimension from those of Ising clusters.

The form of (3) is also consistent with the existence of the universal amplitude ratio, $R_\xi^+ = [\alpha(1-\alpha)(2-\alpha)\,\mathscr{F}_c]^{1/d}\,\xi_0$, where α is the free-energy critical exponent, $\mathscr{F}_c$ is the critical part of the free energy per unit area, and ξ_0 is the amplitude for the correlation length.[33] For any value of Q in the random cluster model, $\partial\mathscr{F}/\partial Q$ gives the mean total number of clusters per unit area $\sum_s n_s = \sum_A N_A$. At the critical point, $N_A \sim -N'(A) \sim C/A^2$ for $A \gg a^2$, and near the critical point one expects a scaling law $N_A = A^{-2}\Phi(A/\xi^2)$, where $\Phi(u)$ is some nontrivial scaling function with $\Phi(0) = C$, which decays exponentially fast as $u \to \infty$. This gives, on substitution into $\sum_A N_A \sim \int_{a^2}^{\infty} N_A\, dA$,

$$\sum_s n_s \sim \text{const.} + B\xi^{-2} \tag{50}$$

where the constant is nonuniversal, as it depends on the details of the cutoff, and

$$B = \int_0^\infty \frac{\Phi(u) - C}{u^2}\, du. \tag{51}$$

Equation (50) is of the form expected from hyperscaling,[33] with $B = (R_\xi^+)^2/]\alpha(1-\alpha)(2-\alpha)]$ directly related to the universal combination R_ξ^+ (recently given exactly for percolation by Seaton[34]). However, we see that it is given by a certain integral over a nontrivial scaling function, while C is just one limiting value of this function.

The results presented here represent the first examples where a measure of the cluster size distribution is given exactly (in the asymptotic limit), both in exponent (here, simply -1) and amplitude (the value C). The agreement between the theoretical prediction and the numerical results for percolation (to a relative accuracy of better than 10^{-4}) compares well with other precision tests of conformal field theory predictions for percolation amplitudes, for example the crossing formula[37] (where the results have been confirmed within a relative error of about 10^{-3}).[35, 36] Knowing the exact result for C at the critical point allows finite-size effects and behavior away from the critical point to be studied, without at the same time having to determine these critical parameters. In percolation especially there has been great interest in size distributions and their finite size corrections, so this result should be useful in that field.

APPENDIX A: LOGARITHMIC CORRECTIONS FOR $Q=4$

We summarize the arguments leading to Eq. (49). It has long been known that many critical quantities in the 4-state Potts model exhibit confluent logarithmic corrections. In the RG framework, this is explained by the existence of a marginally irrelevant scaling variable.[38] A general formalism for computing the form of these corrections was developed in ref. 39, was taken further in ref. 40, and recently has been applied to the fractal properties of $Q=4$ FK clusters by Aharony and Asikainen.[41] In general,[39] logarithmic corrections to susceptibilities take the form of multiplicative powers of logarithms, and are therefore numerically very significant, but in some quantities, for example the finite-size scaling of the free energy at the critical point,[42] they give only additive corrections. We shall argue that this is the case here.

Following ref. 42, suppose that the fixed-point hamiltonian is deformed by a marginal perturbation $\mathscr{H}^* \to \mathscr{H}^* + g \sum_R \Phi(R)$, where Φ is a scaling operator with scaling dimension $x_\Phi = 2$. We may develop the current-current correlation function (18) in a power series in g, the coefficient of each term being a sum over the R_j of correlation functions $\langle J_\mu(r_1)\, J_\nu(r_2)\, \Phi(R_1)\rangle$, $\langle J_\mu(r_1)\, J_\nu(r_2)\, \Phi(R_1)\, \Phi(R_2)\rangle$, and so on, each evaluated with respect to the fixed point hamiltonian. The form of the r_1 and r_2-dependence of each of these correlation functions is completely fixed by conformal invariance in two dimensions, so that they may be computed in a simple model. Choosing a gaussian theory with hamiltonian $\mathscr{H}^* = \frac{1}{2}\int (\partial\phi)^2\, d^2r$, a conserved current $J_\mu \sim \partial_\mu \phi$, and the marginal operator $\Phi \sim (\partial\phi)^2$, all the correlators may be evaluated using Wick's theorem. For the $O(g)$ correction, it turns out that the only non-zero components (in complex coordinates)

have $\mu = z$, $\nu = \bar{z}$, and *vice versa*. The form of the correlation function is

$$\langle J_z(z_1)\, J_{\bar{z}}(\bar{z}_2)\, \Phi(0)\rangle \propto 1/(z_1^2 \bar{z}_2^2) \tag{52}$$

where we have set $R_1 = 0$ for convenience.

Now the $O(g)$ correction to the total area within all loops (15) is

$$g \int \langle J_y(z_1)\, J_y(\bar{z}_2)\, \Phi(R_1)\rangle\, |x_1 - x_2|\, \delta(y_1 - y_2)\, dx_1\, dx_2\, dy_1\, dy_2\, d^2R_1 \tag{53}$$

where $J_y \propto J_z - J_{\bar{z}}$. This is to be evaluated in a large but finite region of linear size $O(L)$. As before, we shall use the infinite volume continuum limit form (52) of the correlation function, justifying this *a posteriori*. The integral in (53) is then proportional to the area $\mathscr{A}$ of the system, and we remove this factor by setting $R_1 = 0$. The remaining integral is then proportional to

$$\int_{-\infty}^{\infty} \frac{|x_1 - x_2|}{(x_1 + iy_1)^2\, (x_2 - iy_1)^2}\, dx_1\, dx_2\, dy_1 \tag{54}$$

The contour integration over y_1 vanishes unless x_1 and x_2 have the same sign: the result is then proportional to

$$\int_0^{\infty} \int_0^{\infty} \frac{|x_1 - x_2|}{(x_1 - x_2)^3}\, dx_1\, dx_2 = \int \frac{dx_1}{x_1} \sim \ln(L/a) \tag{55}$$

with an equal contribution from $x_1, x_2 < 0$. We have cut off the logarithmically divergent integral in the last step, arguing that because the divergence is only logarithmic, it was permissible to use the infinite-volume forms for the correlation function in the integrand.

The important point about this result is that it is $O(g \ln L)$, not $O(g(\ln L)^2)$, as might have been expected (recall that the leading term is $O(g^0 \ln L)$). A similar, but more tedious, calculation shows that the next term is $O(g^2 \ln L)$, and we conjecture that the nth order term is $O(g^n \ln L)$. This is consistent with the fact that, in the gaussian model, g is exactly marginal so that k, the coefficient of the $O(\ln L)$ term, depends continuously on g.

However, in the 4-state Potts model the perturbation is not exactly marginal, and instead flows logarithmically slowly to zero under the RG.

This may be taken into account[42] by replacing the bare expansion parameter g by the running coupling

$$\tilde{g}(L) = \frac{g}{1+bg\ln L} \sim (b\ln L)^{-1} + O(g^{-1}(\ln L)^{-2}) \tag{56}$$

where b is a known constant whose value is not important.

Inserting this result into the formula (15) for the total area $\langle A_{\text{tot}} \rangle \sim \mathscr{A} \sum_{A<O(L^2)} N(A)$ gives

$$\sum_{A<O(L^2)} N(A) \sim 2C\ln L \left(1 + \frac{a_1}{\ln L} + \frac{a_2}{(\ln L)^2} + O((\ln L)^{-3})\right) \tag{57}$$

where the a_j are non-universal constants. Differentiating this with respect to $L^2 \sim A$ then gives the main result quoted in (49)

$$N(A) \sim \frac{C}{A}\left(1 - \frac{2a_2}{(\ln A)^2} + O((\ln A)^{-3})\right) \tag{58}$$

The interesting feature of this result is the absence of the $O((\ln A)^{-1})$ term, proportional to a_1.

ACKNOWLEDGMENTS

J.C. thanks H. Kesten for useful comments and for drawing attention to refs. 14 and 15. He was supported in part by EPSRC Grant GR/J78327. R.Z. acknowledges a discussion with Greg Huber concerning the origin of the finite-size corrections.

REFERENCES

1. R. M. Ziff, C. D. Lorenz, and P. Kleban, *Physica A* **266**:17 (1999).
2. R. M. Ziff, C. D. Lorenz, and B. Mandelbrot, Zipf's law in percolation, to be published.
3. M. S. Watanabe, *Phys. Rev. E* **53**:4187 (1996).
4. N. Jan, D. Stauffer, and A. Aharony, *J. Statis. Phys.* **92**:325 (1998).
5. T. Grossman and A. Aharony, *J. Phys. A* **20**:L1193 (1987).
6. H. Saleur and B. Duplantier, *Phys. Rev. Lett.* **58**:2325 (1987).
7. B. Sapoval, M. Rosso, and J. F. Gouyet, *J. Physique Lett. (Paris)* **46**:L149 (1985).
8. C. M. Fortuin and P. W. Kasteleyn, *Physica* **57**:536 (1972).
9. J. L. Cardy, in *Fluctuating Geometries in Statistical Mechanics and Field Theory,* Les Houches Summer School 1994 (North-Holland, 1996); cond-mat/9409094.
10. B. Nienhuis, in *Phase Transitions and Critical Phenomena*, Vol. 11, C. Domb and J. L. Lebowitz, eds. (Academic Press, London, 1987), pp. 1–53.
11. J. Kondev, *Phys. Rev. Lett.* **78**:4329 (1997).

12. J. L. Cardy, *Phys. Rev. Lett.* **72**:1580 (1994).
13. J. L. Cardy, *Phys. Rev. Lett.* **84**:3507 (2000).
14. J. R. Chayes, L. Chayes, and R. Durrett, *J. Statis. Phys.* **45**:933 (1986).
15. H. Kesten and Y. Zhang, *Probab. Theory Relat. Fields* **107**:137 (1997).
16. R. M. Ziff, *Computers Phys.* **12**:385 (1998).
17. M. F. Sykes, D. S. Gaunt, and M. Glen, *J. Phys. A* **14**:287 (1981).
18. M. E. J. Newman, I. Jensen, and R. M. Ziff, *Phys. Rev. E* **65**:021904 (2002).
19. G. Huber, private communication.
20. H. N. V. Temperley and E. H. Lieb, *Proc. R. Soc. Lond. A.* **322**:251 (1971).
21. R. M. Ziff, S. R. Finch, and V. Adamchik, *Phys. Rev. Lett.* **79**:3447 (1997).
22. P. Kleban and R. M. Ziff, *Phys. Rev. B* **57**:R8075 (1998).
23. H. T. Pinson, *J. Statis. Phys.* **75**:1167 (1994).
24. R. M. Ziff, P. T. Cummings, and G. Stell, *J. Phys. A* **17**:3009 (1984).
25. P. Grassberger, *J. Phys. A* **19**:2675 (1986).
26. T. W. Ruijgrok and E. G. D. Cohen, *Phys. Lett. A* **133**:415 (1988).
27. R. M. Ziff, *Phys. Rev. Lett.* **56**:545 (1986).
28. M. E. J. Newman and R. M. Ziff, *Phys. Rev. Lett.* **85**:4104 (2000).
29. U. Wolff, *Phys. Rev. Lett.* **62**:361 (1989).
30. R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **58**:86 (1987).
31. C.-K. Hu, J.-A. Chen, N. Sh. Izmailian, and P. Kleban, *Phys. Rev. E* **60**:6491 (1999).
32. L.-P. Arguin, *J. Statist. Phys.* **109**:301 (2002).
33. V. Privman, P. C. Hohenberg, and A. Aharony, in *Phase Transitions and Critical Phenomena*, Vol. 14, C. Domb and J. Lebowitz, eds. (Academic Press, London, 1991), pp. 1–134.
34. K. A. Seaton, *J. Phys. A* **34**:L759 (2001).
35. R. Langlands, P. Pouliot, and Y. Saint-Aubin, *Bull. Am. Math. Soc.* **30**:1 (1994).
36. R. M. Ziff, *Phys. Rev. E* **54**:2547 (1996).
37. J. L. Cardy, *J. Phys. A* **25**:L201 (1992).
38. B. Nienhuis, A. N. Berker, E. K. Riedel, and M. Schick, *Phys. Rev. Lett.* **43**:737 (1979); B. Nienhuis, E. K. Riedel, and M. Schick, *J. Phys. A* **13**:131 (1980).
39. J. L. Cardy, M. Nauenberg, and D. J. Scalapino, *Phys. Rev. B* **22**:2560 (1980).
40. J. Salas and A. D. Sokal, *J. Statis. Phys.* **88**:567 (1998).
41. A. Aharony and J. Asikainen, preprint cond-mat/0206367, to appear in *Fractals.*
42. J. L. Cardy, *J. Phys. A Lett.* **19**:L1093 (1986).