Probab. Theory Relat. Fields 91, 507–532 (1992)

Probability Theory and Related Fields

Stability of the Wulff minimum and fluctuations in shape for large finite clusters in two-dimensional percolation

Kenneth S. Alexander*

Department of Mathematics DRB 155, University of Southern California, Los Angeles, CA 90089-1113, USA

Received March 28, 1990; in revised form October 14, 1991

AMS 1980 Subject Classifications: 60K35, 52A40, 82A43

Summary. For two-dimensional Bernoulli percolation at density p above the critical point, there exists a natural norm g determined by the rate of decay of the connectivity function in every direction. If W is the region of unit area with boundary of minimum possible g-length, then it is known [4] that as $N \to \infty$, with probability approaching 1, conditionally on $N \leqq |C(0)| < \infty$, the cluster $C(0)$ of the origin approximates W in shape to within a factor of $1 \pm \eta(N)$ for some $\eta(N) \to 0$. Here a bound is established for the size $\eta(N)$ of the fluctuations. Other types of conditioning which result in the formation of a shape approximating W are also considered.

This is related to the quadratic stability of the variational minimum achieved by the Wulff curve ∂W: for some $k > 0$, if γ is a curve enclosing a region of unit area such that the Hausdorff distance $d_H(\gamma + v, dW) \geqq \delta$ for every translate $\gamma + v$, then the g-length $g(\gamma) \geqq g(\partial W) + k\delta^2$, at least for δ small.

I Introduction

Let us consider Bernoulli bond percolation on the square lattice. The sites are elements of $\mathbb{Z}^2$; bonds (i.e. pairs of adjacent sites) are independently occupied with probability p and vacant with probability $1-p$. $p \in [0, 1]$ is called the density. The cluster $C(x)$ of site x consists of those sites y such that x is connected to y by a path of occupied bonds, an event denoted by "$x \leftrightarrow y$". As is standard, we let $|A|$ denote the number of sites in a subset A of $\mathbb{Z}^2$, and let P_p denote probability when the density is p.

When $p > p_c = 1/2$, it is known that $P_\infty(p) := P_p[|C(0)| = \infty] > 0$, and in fact there exists (a.s.) a unique infinite cluster C_∞. Conversely for $p \leqq 1/2$, all clusters are finite a.s. These results can be found in [2], [10], and [11].

Our interest here is in the percolating regime $p > 1/2$. This is the analog of the low temperature phase in other statistical mechanics models. Large finite

* Research supported by NSF grant number DMS-8702906

clusters can be thought of as (contained in) "bubbles" in C_∞, separated from C_∞ by a skin of vacant bonds. The analogous feature for the Ising and other models is a droplet of one phase immersed in another, which arises in a particular microcanonical ensemble. The bubble or droplet tends to take on a certain characteristic shape when it is large, to minimize, subject to a fixed droplet volume, the surface energy of the interface between phases. See [1], [4], and [8] for more on the Ising-percolation analogy, and [7] for results on droplet shape for the Ising model. Our main result here is an in-probability bound, conditional on $N \leqq |C(0)| \leqq \infty$, on the fluctuations (in Hausdorff distance) of $C(0)$ from this characteristic shape.

For each bond B, thought of now as a unit-length line segment in $\mathbb{R}^2$, there exists a *dual bond* B^*, the unit-length perpendicular bisector of B. The dual sites are the corresponding endpoints $x^* := (x_1 + 1/2, x_2 + 1/2)$ for $x = (x_1, x_2) \in \mathbb{Z}^2$. B^* is defined to be vacant precisely when B is occupied. The dual lattice is isomorphic to the original lattice; when the density for the lattice is p, the dual lattice has density $1-p$. We let $[x^* \leftrightarrow y^*]$ denote the event that x^* is connected to y^* by a path of occupied dual bonds. $(\mathbb{Z}^2)^*$ denotes the set of all dual sites.

Let $p > p_c$; there exists ([9]) a constant $0 < \sigma = \sigma(p) < \infty$ such that

$$P_p[0^* \leftrightarrow (ne_1)^*] \approx e^{-\sigma n}$$

where $e_1 := (1, 0)$ and $a_n \approx b_n$ means the ratio of the logarithms converges to 1. In fact there is a norm $g = g_p$ on $\mathbb{R}^2$ such that

$$\begin{aligned} &\|x\|_\infty \leqq g(x) \leqq \|x\|_1, \quad g(e_1) = 1, \\ &P_p[0^* \leftrightarrow x^*] \leqq e^{-\sigma g(x)} \qquad \text{for all } x \in \mathbb{Z}^2, \\ &P_p[0^* \leftrightarrow (nx)^*] \approx e^{-\sigma g(x) n} \qquad \text{for all } x \in \mathbb{Q}^2, \end{aligned} \tag{1.1}$$

where $n \to \infty$ through values such that $nx \in \mathbb{Z}^2$. For these and further properties of g, see [4].

Let $\gamma: [0, t] \to \mathbb{R}^2$ be a curve; when confusion is unlikely we will also let γ denote the image of this curve. Let $g(\gamma)$ denote the g-length of γ and $R(\gamma)$ the closed region enclosed by γ (i.e. the complement of the unbounded component of $\mathbb{R}^2 \backslash \gamma$.) Let $\mathscr{K}$ be the set of all rectifiable closed curves (not necessarily self-avoiding) in $\mathbb{R}^2$. Let $|A|$ denote the Euclidean area of a region A in $\mathbb{R}^2$. If ∂A is a rectifiable closed curve, we let γ_A denote this curve. The problem

$$\text{minimize } g(\gamma) \text{ subject to } |R(\gamma)| \geqq 1 \quad \text{and} \quad \gamma \in \mathscr{K} \tag{1.2}$$

has a unique minimizer γ_W where W is given by the Wulff construction [W]: for $e_\theta := (\cos\theta, \sin\theta)$,

$$\begin{aligned} &\tilde{W} := \tilde{W}(p) := \{x \in \mathbb{R}^2 : x \cdot e_\theta \leqq g(e_\theta) \quad \text{for all } \theta \in [0, 2\pi]\} \\ &W := W(p) := \tilde{W} / |\tilde{W}|^{1/2}. \end{aligned} \tag{1.3}$$

This was proved by Taylor ([13], [14]) for the minimum over piecewise C^1 curves; see [4] for the elementary extension to $\mathscr{K}$. In convex-analysis terms $\tilde{W}$ is the polar set of the unit ball of g. The set $G := \{g(e_\theta) e_\theta : \theta \in [0, 2\pi]\}$ is the polar plot of g; if for each point $x \in G$ we draw a line through x perpendicular to $[0, x]$ and discard from $R(G)$ any portion outside the line, the remainder is $\tilde{W}$. Let

$$\mathscr{W} := \mathscr{W}(p) := g(\gamma_W)$$

denote the minimum in (1.2).

As in [4] define the metric ρ on subsets of $\mathbb{R}^2$ by

$$\rho(A, B) := \inf_{v \in \mathscr{R}^2} d_H(A+v, B)$$

where $A+v$ denotes the translate of A by v and d_H denotes Hausdorff distance:

$$d_H(A, B) = \max\{\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\} \tag{1.4}$$

with d denoting Euclidean distance. ρ measures differences in shape, independent of location. Given a bounded set $A \subset \mathbb{Z}^2$, the set of bonds

$$\partial_e A := \{\{x, y\} : x \text{ adjacent to } y, x \in A, \text{ and there exists a lattice path from } y \text{ to } \infty \text{ outside } A\}$$

is called the *external boundary* of A; the corresponding dual bonds $\{B^* : B \in \partial_e A\}$ form a circuit (i.e. a closed lattice path, self-avoiding except where the endpoints join) around A. Thus a large finite value of $|C(0)|$ implies the existence of a large circuit of occupied dual bonds enclosing the origin.

Conversely, working only heuristically, suppose there exists a large occupied dual circuit γ enclosing a region containing n sites including the origin. Bonds in $R(\gamma)$ are essentially unconditioned by the existence of γ, so $R(\gamma)$ should look like a broken-off piece of a typical configuration at density p. In particular, the "bubble" $R(\gamma)$ should contain a single large cluster consisting of roughly a fraction $P_\infty(p)$ of the sites in $R(\gamma)$, i.e. a cluster of nP_∞ sites. A cluster of size N then occurs when $|R(\gamma)|$ is of order N/P_∞, which implies $g(\tilde{\gamma}) \geqq (N/P_\infty)^{1/2} g(\gamma_W)$ where $\tilde{\gamma}$ is a "smoothed" version of γ. If $\tilde{\gamma}$ is not shaped like γ_W then this inequality becomes strict. From the definition of g, the probability of occurrence of a given smoothed occupied dual circuit $\tilde{\gamma}$ is of order $\exp(-\sigma g(\tilde{\gamma}))$. These ideas underlie the following result from [4].

Theorem 1.1 *For Bernoulli bond percolation on the square lattice at each fixed density* $p \in (p_c, 1)$, *there exists* $\eta(N) = \eta(N, p) \to 0$ *such that conditionally on* $N \leqq |C(0)| < \infty$, *with probability approaching* 1 *as* $N \to \infty$, *there exists an occupied dual loop* γ *enclosing* 0 *with*

$$\rho(\gamma_W, (P_\infty/N)^{1/2} \gamma) \leqq \eta(N).$$

Further, for some $\varepsilon(N) = \varepsilon(N, p)$ *with* $\varepsilon(N) \to 0$, *and some* $\varepsilon'(N) = O(|\varepsilon(N)|^{1/2})$,

$$P_p[N \leqq |C(0)| < \infty] = \exp(-(1+\varepsilon(N))\, \sigma(p)\, \mathscr{W}(p)\, P_\infty(p)^{-1/2} N^{1/2}) \tag{1.5}$$

and

$$P_p[|C(0)| = N] = \exp(-(1+\varepsilon'(N))\, \sigma(p)\, \mathscr{W}(p)\, P_\infty(p)^{-1/2} N^{1/2}). \tag{1.6}$$

The analog of g-length, in the Ising model as considered in [7], is the integral over the boundary of the surface tension in the normal direction. Of course the g-length of the boundary is the integral of the g-norm of the tangent vector, not the normal, but the analogy holds because g is invariant under rotation by 90 degrees. See [4] for further discussion.

Another context in which such bubbles in C_∞ arise is as follows. Let $\Lambda_L := [-L/2, L/2]^2$. As $L \to \infty$, $|C_\infty \cap \Lambda_L|/|\Lambda_L|$ converges to its expected value P_∞ a.s. For $0 < \lambda < 1$ and L large we can consider the (rare) large-deviation event

$$F_L(\lambda) := [|C_\infty \cap \Lambda_L|/|\Lambda_L| \leqq (1-\lambda) P_\infty]. \tag{1.7}$$

One way for $F_L(\lambda)$ to occur is for an occupied dual circuit γ enclosing a fraction λ of Λ_L to occur, while in $\Lambda_L \backslash R(\gamma)$ the fraction of sites in C_∞ is near its typical value P_∞. As in Theorem 1.1 this is the overwhelmingly most common method of occurrence of $F_L(\lambda)$, and circuits γ shaped like γ_W predominate, giving the following result from [4].

Theorem 1.2 *For Bernoulli bond percolation on the square lattice, at each fixed density $p > p_c$ and each $0 < \lambda < (\operatorname{diam} \gamma_W)^{-2}$, there exist $\zeta(L) = \zeta(L, p) \to 0$ such that conditionally on $F_L(\lambda)$, with probability approaching 1 as $L \to \infty$, there exists an occupied dual circuit γ in Λ_L satisfying*

$$\rho(\gamma_W, \gamma/L\lambda^{1/2}) \leqq \zeta(L). \tag{1.8}$$

Further, for some $\psi(L) = \psi(L, p) \to 0$,

$$P_p(F_L(\lambda)) = \exp(-(1+\psi(L))\, \lambda^{1/2} \sigma(p)\, \mathscr{W}(p)\, L). \tag{1.9}$$

Note that in (1.9), $\lambda^{1/2} \mathscr{W}(p) L$ is the g-length of the boundary of a Wulff shape covering a fraction λ of Λ_L.

For the Ising model in Λ_L at low temperatures with "plus" boundary conditions, the analog of $F_L(\lambda)$ is the event $F_L^I(\lambda)$ that the surplus fraction of "plus" in Λ_L (i.e. the fraction of "plus" sites minus the fraction of "minus" sites) is $(1-2\lambda)m$, where m is the magnetization (i.e. the expected surplus fraction). Conditioning on this event produces what is called a microcanonical ensemble. Dobrushin, Kotecky, and Schlosman showed in [7] that at each fixed very low temperature, conditionally on $F_L^I(\lambda)$, with probability approaching 1 as $L \to \infty$, there is a "droplet" of minus phase covering a fraction λ of Λ_L, with boundary γ closely approximating a fixed curve γ_{W_I} in shape.

It is natural to ask how closely γ approximates γ_W in shape in Theorems 1.1 and 1.2. In [7] it is shown that for the Ising model, the error in shape analogous to $\zeta(L)$ is $O(L^{-\alpha})$ for some $\alpha > 0$. Here we will establish the following.

Theorem 1.3 *In Theorems* 1.1 *and* 1.2, *we have*

$$|\eta(N)| = O(N^{-1/6} (\log N)^{1/3}) \tag{1.10}$$

$$|\varepsilon(N)| = O(N^{-1/3} (\log N)^{2/3}) \tag{1.11}$$

$$|\zeta(L)| = O(L^{-1/3} (\log L)^{1/3}) \tag{1.12}$$

$$|\psi(L)| = O(L^{-2/3} (\log L)^{2/3}). \tag{1.13}$$

Of course the rates in (1.10) and (1.11) differ from those in (1.12) and (1.13) because the scale of the Wulff shape in Theorem 1.1 is $N^{1/2}$ while in Theorem 1.2 it is L.

We suspect that the rates in Theorem 1.3 are not optimal. In fact one might speculate that the fluctuations of γ about γ_W are Gaussian, meaning $\zeta(L)$ should be of order $L^{-1/2}$.

In very broad outline our proofs do not differ greatly from the proofs in [4] of Theorems 1.1 and 1.2 above. Of course a number of new elements, along with greater care in the details, are needed to obtain the rates in Theorem 1.3. Of particular note are (1) the use of Theorem 2.1, Lemma 2.5, Proposition 4.2, and Lemma 4.3, which in effect give rates where in [4] only the fact of convergence was used; (2) the improved method, following Theorem 3.2, for approximating a dual circuit by a polygonal path; (3) the sorting by size of dual circuits, and families of circuits, in the proofs of (1.11) and (1.13) – see the remarks after (5.3).

II Stability of the Wulff minimum

It was shown in [4] that the probability of occurrence of a given large occupied dual circuit (after appropriate smoothing) is essentially a function of its g-length. Among circuits enclosing a given area, then, those not shaped like γ_W have lower probability due to greater g-length. For our analysis we need to quantify the relationship between deviation from γ_W and increase in g-length.

Theorem 2.1 *Let φ be a norm on $\mathbb{R}^2$, and let Y be the corresponding unit-area Wulff shape. There exists a constant $k=k(\varphi)>0$ such that if $\gamma\in\mathscr{K}$, $|\mathbb{R}(\gamma)|\geqq 1$ and $\rho(\gamma,\gamma_Y)=\delta>0$ then $\varphi(\gamma)\geqq\varphi(\gamma_Y)+k(\delta\wedge 1)^2$.*

A similar result was stated without proof in the research announcement [12].

Throughout this section φ is a norm on $\mathbb{R}^2$, and $\tilde{Y}$ and Y denote the corresponding Wulff set and unit-area Wulff shape (cf. (1.3).) $\|\cdot\|$ denotes the Euclidean norm, ψ_Y denotes the norm with unit ball Y, d_φ denotes the distance associated to φ, and $k_1, k_2, \ldots$ are constants which depend only on φ.

For the Euclidean norm on $\mathbb{R}^2$, and γ the boundary of a convex set, Theorem 2.1 follows from an inequality of Bonnesen ([6], Sect. 38). For general norms, when γ traces the boundary of a convex set, we will make use of a special case of a theorem of Wallen ([15]) which extends the inequality of Bonnesen. Let $A+B:=\{x+y: x\in A, y\in B\}$ denote the Minkowski sum of subsets A, B of $\mathbb{R}^2$. Given two convex sets A and B, there exists a constant S_{AB}, the *mixed volume* of A and B, such that

$$|xA+yB|=x^2|A|+2xyS_{AB}+y^2|B| \qquad \text{for all } x,y>0;$$

see [6] for details. Note that $S_{AA}=|A|$. The following is standard – see [13].

Lemma 2.2 *Let C be a bounded convex set in $\mathbb{R}^2$. Then*

$$\varphi(\gamma_C)=\lim_{\varepsilon\downarrow 0}(|C+\varepsilon\tilde{Y}|-|C|)/\varepsilon=2S_{C\tilde{Y}}.$$

In particular,

$$\varphi(\gamma_Y)=2|\tilde{Y}|^{1/2}.$$

Given a convex set C in $\mathbb{R}^2$, let

$$R_C=R_C(\varphi):=\inf_{v\in\mathbb{R}^2}\sup_{x\in C+v}\psi_Y(x)$$

denote the radius of the smallest ψ_Y-ball (centered anywhere) which contains C, and let

$$r_C=r_C(\varphi):=\sup_{v\in\mathbb{R}^2}\inf_{x\notin C+v}\psi_Y(x)$$

be the radius of the largest ψ_Y-ball (centered anywhere) contained in C. The first half of the following is a special case of the theorem of Wallen [15]; the second half follows from Lemma 2.2.

Lemma 2.3 *Let C be a bounded convex set in $\mathbb{R}^2$ and S the mixed volume of C and $\tilde{Y}$. Then*

$$S^2-|C||\tilde{Y}|\geqq|\tilde{Y}|(R_C-r_C)^2/4.$$

Consequently, if $|C|\geqq 1$, then

$$\varphi(\gamma_C)\geqq\varphi(\gamma_Y)(1+(R_C-r_C)^2/4)^{1/2}.$$

Let d_Y denote the distance associated to the norm ψ_Y, let d_H^Y be the corresponding Hausdorff distance (cf. (1.4)), and let ρ_Y denote the corresponding translated Hausdorff distance:

$$\rho_Y(A,B):=\inf_{v\in\mathbb{R}^2} d_H^Y(A+v,B).$$

There exists a constant $k_1=k_1(\varphi)>0$ such that

$$k_1^{-1}\,d_Y\leqq d\leqq k_1\,d_Y$$

so that also

$$k_1^{-1}\,\rho_Y\leqq\rho\leqq k_1\,\rho_Y. \tag{2.1}$$

Lemma 2.4 *There exists a constant $k_2=k_2(\varphi)>0$ such that if C is a convex set in $\mathbb{R}^2$ with $|C|=1$ then $R_C-r_C\geqq k_2\,\rho(\gamma_C,\gamma_Y)$.*

Proof. We may assume C is closed. Note $r_C\leqq 1\leqq R_C$. Let $v\in\mathbb{R}^2$ be such that $C+v\subset R_C\,Y$. Let $x\in C+v$ be such that $x+r_C\,Y\subset C+v$. Then $\psi_Y(x)\leqq R_C-r_C$, and γ_{C+v} lies between $x+r_C\,Y$ and $\gamma_{R_C Y}$, so

$$\begin{aligned}\rho_Y(\gamma_C,\gamma_Y)&\leqq d_H^Y(\gamma_{C+v},\gamma_Y)\\&\leqq d_H^Y(\gamma_{C+v},\gamma_{R_C Y})+d_H^Y(\gamma_{R_C Y},\gamma_Y)\\&\leqq d_H^Y(x+r_C\,Y,\gamma_{R_C Y})+R_C-1\\&\leqq\psi_Y(x)+d_H^Y(r_C\,Y,\gamma_{R_C Y})+R_C-r_C\\&\leqq 3(R_C-r_C).\end{aligned}$$

The lemma now follows from (2.1). □

A point x in A or B which achieves the maximum in the definition (1.4) of $d_H(A, B)$ will be called a *most remote point* for A and B.

With Lemmas 2.3 and 2.4 it is easy to prove Theorem 2.1 when γ is the boundary of a convex set. For general curves γ we use the next lemma. Let $C(\gamma)$ denote the convex hull of $R(\gamma)$, let E_γ denote the set of extreme points of $C(\gamma)$, and let γ^* denote $\gamma_{C(\gamma)}$, which we assume traces $\partial C(\gamma)$ in the direction of positive orientation. Note $C(\gamma)=R(\gamma^*)$, and

$$\varphi(\gamma)\geqq\varphi(\gamma^*) \tag{2.2}$$

(see the Appendix of [4].) Let $k_3=k_3(\varphi)\leqq 1$ be such that

$$d_\varphi\geqq k_3 d.$$

Let $\mathscr{I}$ denote the set of all self-avoiding curves in $\mathscr{K}$.

Lemma 2.5 *Let $\gamma\in\mathscr{K}$ with $|C(\gamma)|=1$ and $d_H(\gamma,\gamma^*)=\delta>0$. Then either*

$$|R(\gamma)|\leqq 1-\pi k_3^2\,\delta^2/288 \tag{2.3}$$

or

(2.4) there exist $u, v\in E_\gamma$ and $x, y, z\in\gamma$ such that

$$[u, v]\subset\gamma^* \tag{2.4a}$$

$$d(y,\gamma^*)=\delta \tag{2.4b}$$

(2.4c) γ passes through u, x, y, z and v in that order

$$\varphi(v-z)+\varphi(z-y)+\varphi(y-x)+\varphi(x-u)\geqq\varphi(u-v)+k_3\,\delta. \tag{2.4d}$$

Further, there exists a constant $k_4=k_4(\varphi)>0$ such that for $\gamma\in\mathscr{K}$ with $|C(\gamma)|=1$, if $d_H(\gamma,\gamma^*)=\delta>0$ then

$$\varphi(\gamma)/|R(\gamma)|^{1/2}\geqq\varphi(\gamma_Y)+k_4(\delta\wedge 1)^2. \tag{2.5}$$

Proof. The idea is as follows: γ traces a "dent" of depth δ in $C(\gamma)$. Since we do not assume φ is strictly convex, we cannot be sure that this dent makes γ longer than γ^*. However, if the area of the dent is at least $\pi k_3^2\,\delta^2/288$ then (2.3) holds. Alternatively if the dent area is smaller than $\pi k_3^2\,\delta^2/288$ (so the dent is essentially a thin crevice), y is the bottom of the crevice, and x and z are the rim points of the crevice, then the path formed from γ^* by replacing the segment from u to v across the top of the dent with the "crevice path" $u\to x\to y\to z\to v$ has φ-length at least $k_3\,\delta$ greater than that of γ^*.

Let y be a most remote point for γ and γ^*. Suppose first that $y\in\gamma^*$, so $d(y,\gamma)=\delta$. Then half of the open ball $B(y,\delta)$ is in $R(\gamma^*)\backslash R(\gamma)$, so

$$|R(\gamma)|\leqq|R(\gamma^*)|-\pi\delta^2/2 \tag{2.6}$$

and (2.3) follows.

Suppose then that $y\in\gamma$. Let u and v be the points of $\gamma\cap\gamma^*$ most closely preceding and following y in γ, so that $[u, v]\subset\gamma^*$. Now

$$\varphi(v-y)+\varphi(y-u)\geqq 2k_3\,\delta.$$

If $\varphi(u-v) \leqq k_3 \delta$ then (2.4) follows, with $x=u$ and $z=v$. Thus we suppose that $\varphi(u-v) > k_3 \delta$. Let

$$m := [6\varphi(u-v)/k_3 \delta] + 1$$

and let $u=u_1, u_2, \ldots, u_m=v$ be m evenly spaced points in the line $[u, v]$, so that

$$\varphi(u_{i+1}-u_i) = \varphi(u-v)/(m-1) \geqq k_3 \delta/6.$$

Let B_i be the open φ-ball of radius $k_3 \delta/12$ centered at u_i, and let γ_{uv} be the section of γ from u to v. Let H_{uv} be the closed half space with u, $v \in \partial H_{uv}$ and $R(\gamma^*) \subset H_{uv}$.

Suppose first that $\gamma_{uv} \cap B_i = \phi$ for some i (necessarily $1 < i < m$.) Then similarly to (2.6),

$$|R(\gamma)| \leqq |R(\gamma^*)| - |B_i|/2 \leqq 1 - \pi k_3^2 \delta^2/288$$

and (2.3) follows.

Suppose alternatively that for all $i \leqq m$, $\gamma_{uv} \cap B_i \neq \phi$. Let j be the largest index $i < m$ such that γ_{uv} visits B_i before passing through y. Let x be the last point of γ_{uv} in the closure $\bar{B}_j$ before y, and let z be the first point of γ_{uv} in $\bar{B}_{j+1}$ after y. Then

$$\begin{aligned}
&\varphi(v-z) + \varphi(z-y) + \varphi(y-x) + \varphi(x-u) \\
&\quad \geqq d_\varphi(v, \bar{B}_{j+1}) + d_\varphi(y, \bar{B}_{j+1}) + d_\varphi(y, \bar{B}_j) + d_\varphi(u, \bar{B}_j) \\
&\quad \geqq (m-j-1)\, \varphi(u-v)/(m-1) - k_3 \delta/12 + 2k_3(\delta - \delta/12) \\
&\qquad + (j-1)\, \varphi(u-v)/(m-1) - k_3 \delta/12 \\
&\quad = \varphi(u-v) - \varphi(u-v)/(m-1) + 5k_3 \delta/3 \\
&\quad \geqq \varphi(u-v) + k_3 \delta
\end{aligned}$$

so (2.4) holds.

Turning to (2.5), suppose first that $\gamma \in \mathscr{I}$. Under (2.3), (2.5) follows from

$$\varphi(\gamma) \geqq \varphi(\gamma^*) \geqq \varphi(\gamma_Y).$$

Under (2.4), (2.5) follows from $|R(\gamma)| \leqq 1$ and

$$\varphi(\gamma) \geqq \varphi(\gamma^*) + k_3 \delta \geqq \varphi(\gamma_Y) + k_3 (\delta \wedge 1)^2.$$

For general $\gamma \in \mathscr{K}$ there exists a sequence of polygonal paths γ_n (which may self-intersect) converging uniformly to γ, with $\varphi(\gamma_n) \to \varphi(\gamma)$. For each n there is another polygonal path α_n which traces $\partial R(\gamma_n)$ in the direction of positive orientation, so that $\varphi(\alpha_n) \leqq \varphi(\gamma_n)$. Now α_n may also have points of self-intersection, but no transversal ones, so by perturbing α_n by at most, say, $1/n$ we can obtain a polygonal path $\beta_n \in \mathscr{I}$, with $\varphi(\beta_n) \leqq \varphi(\alpha_n) + 1/n$. Then $\beta_n \to \gamma$ uniformly and $|C(\beta_n)| \to 1$. Finally let $\zeta_n := \beta_n / |C(\beta_n)|^{1/2}$. Then (2.5) holds for ζ_n, and

$$\limsup \varphi(\zeta_n) \leqq \varphi(\gamma), \qquad \limsup |R(\zeta_n)| \geqq |R(\gamma)|,$$

and

$$d_H(\zeta_n, \zeta_n^*) \to \delta$$

which shows that (2.5) holds for γ. □

Proof of Theorem 2.1 Let k_5 be such that the Euclidean norm $\|\cdot\| \leqq k_5 \varphi$ and let $0 < k_6 < \min(1, (2k_5^{-1} \varphi(\gamma_Y))^{-1})$. Let $\beta := \gamma/|C(\gamma)|^{1/2}$, so $|R(\beta^*)| = 1$.

If $|C(\gamma)| \geqq 1 + k_6(\delta \wedge 1)^2$ then using (2.2), for some $k_7 > 0$,

$$\begin{aligned}\varphi(\gamma) &\geqq \varphi(\beta)(1 + k_6(\delta \wedge 1)^2)^{1/2} \\ &\geqq \varphi(\beta^*)(1 + k_7(\delta \wedge 1)^2) \\ &\geqq \varphi(\gamma_Y)(1 + k_7(\delta \wedge 1)^2).\end{aligned}$$

If $|C(\gamma)| < 1 + k_6(\delta \wedge 1)^2$ and $d_H(\gamma, \gamma^*) \geqq \delta/2$ then $d_H(\beta, \beta^*) \geqq \delta/4$ so by Lemma 2.5 applied to β,

$$\varphi(\gamma) = |R(\gamma)|^{1/2} \varphi(\beta)/|R(\beta)|^{1/2} \geqq \gamma(\gamma_Y) + k_4((\delta/4) \wedge 1)^2.$$

Suppose then that $|C(\gamma)| < 1 + k_6(\delta \wedge 1)^2$ and $d_H(\gamma, \gamma^*) < \delta/2$. We may assume that $\varphi(\gamma) < 2\varphi(\gamma_Y)$. Then $\rho(\gamma^*, \gamma_Y) > \delta/2$ while

$$\rho(\gamma^*, \beta^*) \leqq (|C(\gamma)|^{1/2} - 1) \operatorname{diam}(\beta^*) \leqq (k_6 \delta/2)(k_5^{-1} \varphi(\beta^*)/2) \leqq \delta/4.$$

Therefore $\rho(\beta^*, \gamma_Y) > \delta/4$. By Lemmas 2.4 and 2.3,

$$\varphi(\gamma) \geqq \varphi(\beta^*) \geqq \varphi(\gamma_Y)(1 + (k_2 \delta/4)^2/4)^{1/2} \geqq \varphi(\gamma_Y) + k_8(\delta \wedge 1)^2$$

for some $k_8 > 0$.

In all cases the theorem follows. □

III Upper bounds on the probabilities of dual circuits

When a bond configuration gives rise to a large cluster, the bonds dual to the external boundary of the cluster form an occupied dual circuit enclosing the cluster. The results in [4] (cf. Theorem 1.1 above) show that the probability that the cluster $C(0)$ is large and finite is essentially just the probability that an occupied dual circuit enclosing sufficient area surrounds the origin. For a cluster of size N, the area enclosed should be roughly N/P_∞, since the cluster density is roughly P_∞. Thus we need some estimates for the probabilities of large occupied dual circuits.

Throughout this section, $c_1, c_2, \ldots$ will denote constants which depend only on p. As p is fixed but arbitrary in each result, our notation will sometimes suppress the dependence on p of various constants. When circuits γ are viewed as curves, we always assume they are traced in the direction of positive orientation and parametrized by $[0, 1]$, with $\gamma(0)$ the leftmost point among those with minimal y-coordinate in γ. We will begin with upper bonds on circuit probabilities.

Theorem 3.1 *For each $p \in (p_c, 1)$ there exists a constant $c_1 = c_1(p) > 0$ such that for all $A \geqq 2$,*

$$P_p[\text{there exists an occupied dual circuit } \gamma \text{ enclosing } 0 \text{ with } |\mathbb{R}(\gamma^*)| \geqq A] \leqq \exp(-\sigma \mathscr{W} A^{1/2} + c_1 A^{1/6} (\log A)^{2/3}).$$

Theorem 3.2 *For each $p\in(p_c, 1)$ there exist constants $c_i = c_i(p) > 0$ $(i=2,3,4)$ and $A_0(p) > 0$ such that if $A \geqq A_0$ and*

$$c_2 \geqq \delta \geqq c_3 A^{-1/6} (\log A)^{1/3} \tag{3.1}$$

then

(3.2) P_p [there exists an occupied dual circuit γ enclosing 0 with
$|\mathbb{R}(\gamma)| \geqq A$ and $\rho(\gamma_W, \gamma/|R(\gamma)|^{1/2}) \geqq \delta] \leqq \exp(-\sigma \mathscr{W} A^{1/2}(1+c_4\delta^2))$.

The lower bound on δ in (3.1) ensures that the increased length that γ must have because it is not shaped like γ_W is not counterbalanced by other factors, including the reduction in length when γ is replaced by a smoothed approximation.

A result similar to Theorem 3.1 was proved in [4] (Lemma 4.1), but with a larger error term. The proof used the notion of the m-skeleton of a dual circuit γ, defined (for γ enclosing 0) as follows. Let s_0 be the lowest dual site of γ on the positive (dual) y axis. Inductively define s_{n+1} to be the first site in γ after s_n for which $g(s_{n+1}-s_n) \geqq m$. If J is the largest value of n for which s_n can be so defined, then the sequence $(s_0, s_1, \ldots, s_J, s_0)$, abbreviated (s_i), is called the *m-skeleton* of γ. Corresponding to this m-skeleton is a polygonal path, which must enclose almost as much area as γ does; the maximum error in area is of order Jm^2. Therefore the g-length of the polygonal path is at least $\mathscr{W}(|\mathbb{R}(\gamma)| - O(Jm^2))^{1/2}$, a fact which can be used to bound the probability in Theorem 3.1.

Here we introduce a modified type of skeleton which reduces the error in area to order m^2. The basic idea is to construct the skeleton from the boundary of the convex hull of $R(\gamma)$, rather than from γ itself. Let γ^* be a curve which traces the boundary of the convex hull of $R(\gamma)$, and let E_γ denote the set of extreme points of $R(\gamma^*)$. Then

$$E_\gamma \subset \gamma^* \cap \gamma \cap (\mathbb{Z}^2)^*.$$

Let $t_0 := 0$ and $s_0 := \gamma^*(0) = \gamma(0)$. Note that $s_0 \in E_\gamma$. Define inductively

$$\begin{aligned}
t'_{n+1} &:= \inf\{t > t_n : g(\gamma*(t) - \gamma^*(t_n)) \geqq m \text{ or } t = 1\}\\
t''_{n+1} &:= \sup\{t \leqq t'_{n+1} : \gamma^*(t) \in E_\gamma\}\\
t'''_{n+1} &:= \inf\{t > t_n : g(\gamma*(t) - \gamma^*(t_n)) \geqq m, \gamma^*(t) \in E_\gamma\}\\
t_{n+1} &:= \begin{cases} t''_n & \text{if } t''_{n+1} \neq t_n \\ t'''_{n+1} & \text{if } t''_{n+1} = t_n \end{cases}\\
s_{n+1} &:= \gamma^*(t_{n+1}),
\end{aligned}$$

stopping when we reach a value J for which $t_{J+1} = 1$, so that $s_{J+1} = s_0$. In words, we go forward from s_n along γ^* until we reach the boundary of the g-ball of radius m about s. We then backtrack along γ^* to find a point of E_γ. If this does not require going all the way back to s_n, then this point of E_γ is labeled s_{n+1}. If the backtracking does take us all the way back to s_n, we then go forward along γ^* outside the radius-m g-ball, necessarily in a straight line from s_n, until we find a point of E_γ, which becomes s_{n+1}. We continue until we return to s_0. The sequence $(s_0, s_1, \ldots, s_J, s_0)$, abbreviated (s_i), is then called the *m-hull skeleton* of γ. The corresponding polygonal path is called the *m-hull skeletal path* of γ and will be denoted γ_m. Some important observations: first, s_{n+2} is always outside the radius-m g-ball about s_n, i.e.

$$g(s_{n+2}-s_n)>m \quad \text{for all } 0\leqq n\leqq J-2. \tag{3.3}$$

Second,

$$\text{if } g(s_{n+1}-s_n)>m \quad \text{then } [s_n, s_{n+1}]\subset\gamma^* \quad \text{and } \gamma \text{ does not cross } [s_n, s_{n+1}] \tag{3.4}$$

And third, since the vertices of γ_m are a subsequence of the cyclically ordered vertices of the convex polygon γ^*,

$$R(\gamma_m) \text{ is convex.} \tag{3.5}$$

Lemma 3.3 *There exists a constant $c_5=c_5(p)>0$ such that for every dual circuit γ and every $m\geqq 1$, $|R(\gamma_m)|\geqq|R(\gamma^*)|-c_5 m^2$.*

Proof. Let $\beta_n\in[0,\pi]$ be the angle between the left and right derivatives of γ_m at s_n. Then

$$\sum_{n=0}^{J}\beta_n=2\pi. \tag{3.6}$$

Each point of $R(\gamma^*)\setminus R(\gamma_m)$ lies, for some n, in the region Q_n between $[s_n, s_{n+1}]$ and the section of γ^* from s_n to s_{n+1}. If $\max(\beta_n, \beta_{n+1})\leqq\pi/4$ then by (3.4) either $|Q_n|=0$ or $|Q_n|\leqq c_6\|s_{n+1}-s_n\|^2(\beta_n+\beta_{n+1})\leqq c_7 m^2(\beta_n+\beta_{n+1})$. By (3.6) there are at most 16 values of n for which $\max(\beta_n, \beta_{n+1})>\pi/4$; for these n we have by (3.4) that either $|Q_n|=0$ or $|Q_n|\leqq c_8 m^2$. Thus by (3.6)

$$|R(\gamma^*)\setminus R(\gamma_m)|\leqq 16c_8 m^2+\sum_{n=0}^{J} c_7 m^2(\beta_n+\beta_{n+1})\leqq c_9 m^2. \quad \square$$

The following will be used several times, so we will isolate it as a lemma.

Lemma 3.4 *For all $N\geqq 2$, $n\geqq 1$, and $l>0$,*

$$P_p\Big[\text{there exist disjoint occupied dual paths } v_0\leftrightarrow v_1\leftrightarrow\ldots\leftrightarrow v_j \text{ for some}$$
$$v_0, \ldots, v_j\in\Lambda_N\cap(\mathbb{Z}^2)^* \text{ with } j\leqq n \text{ and } \sum_{i=0}^{j-1} g(v_{i+1}-v_i)\geqq l\Big]$$
$$\leqq\exp(-\sigma l+15n\log N).$$

Proof. The number of possible choices for $v_0, \ldots, v_j$ is at most $((N+1)^2+1)^{n+1}$. For each such choice, the van den Berg-Kesten inequality ([5]) and (1.1) tell us that

$$\begin{aligned}
&P_p[\text{there exist disjoint occupied dual paths } v_0\leftrightarrow v_1\leftrightarrow\ldots\leftrightarrow v_j]\\
&\quad\leqq\prod_{i=0}^{j-1} P_p[v_i\leftrightarrow v_{i+1}]\\
&\quad\leqq\exp\left(-\sum_{i=0}^{j-1}\sigma g(v_{i+1}-v_i)\right)\\
&\quad\leqq\exp(-\sigma l)
\end{aligned}$$

and the lemma follows. $\square$

Proof of Theorem 3.1 Let $\mathscr{R}_0(A)$ denote the event in the statement of the theorem. Let m satisfy

(3.7) $$c_5 m^2 < A/2;$$

m will be specified more precisely later. When $\mathscr{R}_0(A)$ occurs we have

$$|R(\gamma_m)| \geqq |R(\gamma^*)| - c_5 m^2$$

by Lemma 3.3, so

(3.8) $$g(\gamma_m) \geqq \mathscr{W}(A - c_5 m^2)^{1/2}.$$

Let $n := 2[2\mathscr{W} A^{1/2}/m]$. The dual circuit γ passes through the vertices of γ_m in the order given by the skeleton $(s_0, s_1, \ldots, s_J, s_0)$. There therefore exist disjoint occupied dual paths $s_0 \leftrightarrow s_1 \leftrightarrow \ldots \leftrightarrow s_J \leftrightarrow s_0$ and by (3.3)

$$L := \sum_{i=0}^{J \wedge n} g(s_{i+1} - s_i) \geqq (J \wedge n)\, m/2.$$

If $J > n$ then this shows

(3.9) $$L \geqq nm/2 \geqq \mathscr{W} A^{1/2}.$$

If $J \leqq n$ then by (3.8)

(3.10) $$L = g(\gamma_m) \geqq \mathscr{W} A^{1/2}(1 - c_5 m^2/A)^{1/2} \geqq \mathscr{W} A^{1/2} - c_{10} m^2/A^{1/2}.$$

If γ is not contained in Λ_{4A} then the event in Lemma 3.4 occurs with $n=2$, $N=5A$, and $l=4A$. Thus from that lemma,

(3.11) $$\begin{aligned} &P_p[\mathscr{R}_0(A) \text{ occurs with } \gamma \not\subset \Lambda_{4A}] \\ &\quad \leqq \exp(-4\sigma A + 30 \log 4A) \\ &\quad \leqq \exp(-\sigma \mathscr{W} A^{1/2} + 30 \log 4A). \end{aligned}$$

(This uses the fact that $\mathscr{W} \leqq 4$, which follows from (1.1) and comparison of W to a unit square.) If γ is contained in Λ_{4A} then the event in Lemma 3.4 occurs with $N=4A$ and (by (3.9) and (3.10)) $l = \mathscr{W} A^{1/2} - c_{10} m^2/A^{1/2}$. Thus again from that lemma,

(3.12) $$\begin{aligned} &P_p[\mathscr{R}_0(A) \text{ occurs with } \gamma \subset \Lambda_{4A}] \\ &\quad \leqq \exp(-\sigma \mathscr{W} A^{1/2} + c_{11} m^2/A^{1/2} + c_{12} A^{1/2}(\log A)/m). \end{aligned}$$

Observe that the first error term $c_{11} m^2/A^{1/2}$ in (3.12) comes essentially from the error in area between γ and γ_m, and the second error term $c_{12} A^{1/2}(\log A)/m$ from the number of possible m-hull skeletons in Λ_{4A}. For optimal tradeoff, the sum of the two error terms in the exponent in (3.12) is minimized, up to a constant, by taking $m = (A \log A)^{1/3}$. Then together, (3.11) and (3.12) prove the theorem. □

Proof of Theorem 3.2 Let $\mathscr{R}'_0(A,\delta)$ denote the event in the statement of the theorem. Let us first handle the following three special cases:

Case 1: $|R(\gamma^*)| \geqq A(1+c_{13}\,\delta^2)$.
Case 2: $g(\gamma_m) \geqq 2\mathscr{W} A^{1/2}$ (but not Case 1).
Case 3: $d_H(\gamma,\gamma^*) \geqq A^{1/2}\,\delta/4$ (but not Case 1 or 2).

Here m and $c_{13} \leqq 1$ are constants (depending on p) to be specified later. Briefly, in Case 1, γ^* encloses more area than necessary, hence is longer than necessary. In Case 2, γ is far longer than it need be to enclose area A. In Case 3, γ traces a deep dent in $R(\gamma^*)$ hence again is longer than it need be to enclose area A. Outside of these three cases, when $\mathscr{R}'_0(A,\delta)$ occurs we will show that γ_m is not shaped like γ_W, so Theorem 2.1 can be applied.

For Case 1 we can apply Theorem 3.1: if $c_3(p)$ is large enough and $c_2(p)$ is small enough (depending on our choice of c_{13} – cf. Case 3a below) then

$$
\begin{aligned}
(3.13)\qquad & P[\mathscr{R}'_0(A,\delta) \text{ occurs under Case 1}]\\
&\leqq \exp(-\sigma\mathscr{W} A^{1/2} - c_{13}\,\sigma\mathscr{W} A^{1/2}\,\delta^2/3 + 2c_1\,A^{1/6}(\log A)^{2/3})\\
&\leqq \exp(\sigma\mathscr{W} A^{1/2} - A^{1/2}\,\delta^2(c_{13}\,\sigma\mathscr{W}/3 - 2c_1/c_3^2))\\
&\leqq \exp(-\sigma\mathscr{W} A^{1/2}(1+c_{13}\,\delta^2/4)).
\end{aligned}
$$

Consider then Case 2: $g(\gamma_m) \geqq 2\mathscr{W} A^{1/2}$, but not Case 1. Let $(s_0, s_1, \ldots, s_J, s_0)$ be the vertices of γ_m and let s_M be the first vertex such that the g-length of the polygonal path $s_0 \to s_1 \to \ldots \to s_M$ is at least $2\mathscr{W} A^{1/2}$. Then by (3.3),

$$
m(M-2)/2 \leqq \sum_{j=0}^{M-2} g(s_{j+1}-s_j) < 2\mathscr{W} A^{1/2}
$$

so that

$$
M < 4\mathscr{W} A^{1/2}/m + 2 \leqq 8\mathscr{W} A^{1/2}/m
$$

provided $m < A^{1/2}$. Thus the event in Lemma 3.4 occurs with $n = 8\mathscr{W} A^{1/2}/m$, $l = 2\mathscr{W} A^{1/2}$, and $N = 2A$, the latter because we are not in Case 1. Thus from that Lemma,

$$
\begin{aligned}
(3.14)\qquad & P[\mathscr{R}'_0(A,\delta) \text{ occurs under Case 2}]\\
&\leqq \exp(-2\sigma\mathscr{W} A^{1/2} + 120\mathscr{W} A^{1/2}(\log 2A)/m)\\
&\leqq \exp(-\sigma\mathscr{W} A^{1/2}(2 - 240(\log A)/\sigma m))\\
&\leqq \exp(-\sigma\mathscr{W} A^{1/2}(1+\delta^2))
\end{aligned}
$$

provided c_2 (and hence δ) is small enough and $m \geqq c_{14} \log A$.

Turning to Case 3, $d_H(\gamma,\gamma^*) \geqq A^{1/2}\,\delta/4$ but not Case 1 or 2, after rescaling γ and γ^* by a factor of $|R(\gamma^*)|^{1/2}$, we get two subcases, according to whether (2.3) or (2.4) holds for the curve $\gamma/|R(\gamma^*)|^{1/2}$ in Lemma 2.5. Suppose first that we have

Case 3a: $|R(\gamma)|/|R(\gamma^*)| \leqq 1 - (\pi k_3^2/288)\,A(\delta/4)^2/|R(\gamma^*)|$.

Then

$$|R(\gamma^*)| \geqq |R(\gamma)| + (\pi k_3^2/288)\, A(\delta/4)^2 > A(1+c_{13}\,\delta^2)$$

provided c_{13} is chosen sufficiently small. But this violates the assumption that we are not in Case 1, so Case 3a never occurs. This leaves us with the following.

Case 3*b*: For some u, v, x, y, z, (2.4) holds.

Since each vertex s_n is in the set E_γ of extreme points, from (2.4a) we see that u and v are between the same pair s_n, s_{n+1} of adjacent vertices of γ_m. Consider the polygonal path β which runs

$$s_0 \to s_1 \to \dots \to s_n \to u \to x \to y \to z \to v \to s_{n+1} \to \dots \to s_J \to s_0.$$

By Lemma 3.3, $|R(\gamma_m)| \geqq A - c_5\, m^2$, so from (2.4d),

$$\begin{aligned} g(\beta) &\geqq g(\gamma_m) + c_{15}\, A^{1/2}\,\delta \\ &\geqq \mathscr{W}(A - c_5\, m^2)^{1/2} + c_{15}\, A^{1/2}\,\delta \\ &\geqq \mathscr{W} A^{1/2} + c_{16}\, A^{1/2}\,\delta \end{aligned}$$

provided $m \leqq c_{17}\,\delta^{1/2} A^{1/2}$. Analogously to the reasoning in Case 2, since we are not in Case 1 or 2 the number of vertices in γ_m is at most $8\mathscr{W} A^{1/2}/m$, provided c_2 is small enough and $m < A^{1/2}$. Thus the event in Lemma 3.4 occurs with $n = 8\mathscr{W} A^{1/2}/m$, $l = \mathscr{W} A^{1/2} + c_{16}\, A^{1/2}\,\delta$, and $N = 2A$. Therefore

$$\begin{aligned} (3.15) \qquad & P[\mathscr{R}'_0(A,\delta) \text{ occurs under Case 3b}] \\ &\leqq \exp(-\sigma \mathscr{W} A^{1/2} - c_{16}\,\sigma A^{1/2}\,\delta + 120 \mathscr{W} A^{1/2}(\log 2A)/m) \\ &\leqq \exp(-\sigma \mathscr{W} A^{1/2}(1 + c_{18}\,\delta)) \\ &\leqq \exp(-\sigma \mathscr{W} A^{1/2}(1+\delta^2)) \end{aligned}$$

provided that c_2 is small enough and $m \geqq c_{19}\,\delta^{-1} \log A$.

Now suppose that $\mathscr{R}'_0(A,\delta)$ occurs but not under Case 1, 2, or 3. We wish to show that, like γ, γ_m is not shaped like γ_W. From (1.1),

$$\operatorname{diam}(\gamma_m) \leqq g(\gamma_m) \leqq 2\mathscr{W} A^{1/2}.$$

Further, both $|R(\gamma)|$ and $|R(\gamma_m)|$ are between $A - c_5\, m^2$ and $A(1 + c_{13}\,\delta^2)$. Therefore

$$\begin{aligned} (3.16) \qquad & \rho(\gamma_m/|R(\gamma)|^{1/2}, \gamma_m/|R(\gamma_m)|^{1/2}) \\ &\leqq \big|\,|R(\gamma)|^{-1/2} - |R(\gamma_m)|^{-1/2}\big| \operatorname{diam}(\gamma_m) \\ &\leqq 2\mathscr{W}((1 - c_5\, m^2/A)^{-1/2} - (1 + c_{13}\,\delta^2)^{-1/2}) \\ &\leqq c_{20}\, m^2/A + c_{21}\,\delta^2 \\ &\leqq \delta/4 \end{aligned}$$

provided again that c_2 is small enough and $m \leqq c_{22}\,\delta^{1/2} A^{1/2}$. In addition, as we are not in Case 3,

$$\begin{aligned} (3.17) \qquad \rho(\gamma/|R(\gamma)|^{1/2}, \gamma_m/|R(\gamma)|^{1/2}) &\leqq A^{-1/2}(\rho(\gamma,\gamma^*) + \rho(\gamma^*, \gamma_m)) \\ &\leqq \delta/4 + A^{-1/2} m \leqq \delta/2, \end{aligned}$$

provided $m \leqq \delta A^{1/2}/4$. Combining (3.16) and (3.17) with the definition of $\mathscr{R}'_0(A, \delta)$ we see that, as desired,

$$\rho(\gamma_W, \gamma_m/|R(\gamma_m)|^{1/2}) \geqq \delta/4.$$

From Theorem 2.1, then,

$$\begin{aligned} g(\gamma_m) &\geqq |R(\gamma_m)|^{1/2}(\mathscr{W} + k\delta^2/16) \\ &\geqq (A - c_5 m^2)^{1/2}(\mathscr{W} + k\delta^2/16) \\ &\geqq \mathscr{W} A^{1/2}(1 + c_{23}\delta^2) \end{aligned}$$

provided that $m \leqq c_{24}\delta A^{1/2}$ and, as usual, that c_2 is sufficiently small. As in Case 2, this implies that the event in Lemma 3.4 occurs with $n = 8\mathscr{W} A^{1/2}/m$, $l = \mathscr{W} A^{1/2}(1 + c_{23}\delta^2)$, and $N = 2A$, so that

$$\begin{aligned} (3.18) \qquad & P[\mathscr{R}'_0(A, \delta) \text{ occurs but not under Case 1, 2, or 3}] \\ &\leqq \exp(-\sigma\mathscr{W} A^{1/2}(1 + c_{23}\delta^2) + 120\mathscr{W} A^{1/2}(\log 2A)/m) \\ &\leqq \exp(-\sigma\mathscr{W} A^{1/2}(1 + c_{25}\delta^2)) \end{aligned}$$

provided that $m \geqq c_{26}\delta^{-2}\log 2A$.

All the provisions we have made on m can be simultaneously satisfied (if A_0 and c_3 are sufficiently large) by taking $m = (A\log A)^{1/3}$. Thus (3.13), (3.14), (3.15), and (3.18) prove the theorem. □

IV Lower bounds for the probabilities of dual circuits

The main result of this section is the following counterpart of Theorem 3.1.

Theorem 4.1 *For each $p \in (p_c, 1)$ and $A \geqq 2$ there exists a convex polygon $Q = Q(A, p)$ containing 0, and a constant $c_{27} = c_{27}(p) > 0$, such that $|Q| = A$ and*

$$\begin{aligned} & P_p[\text{there exists an occupied dual circuit enclosing } Q] \\ &\geqq \exp(-\sigma\mathscr{W} A^{1/2} - c_{27} A^{1/6}(\log A)^{2/3}). \end{aligned}$$

Further, as $A \to \infty$, $\rho(\gamma_W, A^{-1/2}\gamma_Q) \to 0$.

The idea, similar to Theorem 1.B of [4], is to construct a polygon which approximates $A^{1/2}R(\gamma_W)$ from inside, rescale it slightly so that it has area A, and calculate a lower bound for the probability that each adjacent pair of vertices of this polygon is connected by a path of occupied dual bonds, entirely outside the polygon. We will need the following result from [3].

Proposition 4.2 *There exists $r > 0$ such that for each $p \in (p_c, 1)$ there is a constant $c_{28} = c_{28}(p) > 0$ with the property that for every $x, y \in (\mathbb{Z}^2)^*$,*

$$\begin{aligned} & P_p[x \leftrightarrow y \text{ by a path of occupied dual bonds}] \\ &\geqq c_{28}\|y - x\|^{-r} e^{-\sigma g(y-x)}. \end{aligned}$$

To keep our dual circuit outside the polygon, we will need to refine Proposition 4.2. First some definitions: for $C \subset \mathbb{R}^2$ let

$$\partial^* C := \{x \in (\mathbb{Z}^2)^* \setminus C : x \text{ is adjacent to } (\mathbb{Z}^2)^* \cap C\}.$$

Let l_{xy} denote the line through x and y, and H_{xy} the closed halfspace to the right of l_{xy} as one moves from x to y. Let u_{xy} be the inward unit normal to H_{xy}. Let U_g denote the unit ball of g.

Lemma 4.3 *For each $p\in(p_c, 1)$ there exists a constant $c_{29}=c_{29}(p)>0$ such that for every $x\neq y\in(\mathbb{Z}^2)^*$,*

$$P_p[x\leftrightarrow y \text{ in } H_{xy}]\geqq c_{29}\|x-y\|^{-2}P_p[x\leftrightarrow y].$$

Proof. It is sufficient to prove the result for $\|x-y\|$ large. Suppose $x\leftrightarrow y$ via a self-avoiding occupied dual path α. Let V be a vertex of α which minimizes the Euclidean inner product $\langle\cdot, u_{xy}\rangle$ over α. Then $H_{V,V+y-x}$ is a parallel translate of H_{xy} with V in its boundary, so that α breaks down into two segments, $x\leftrightarrow V$ and $V\leftrightarrow y$ both in $H_{V,V+y-x}$. We wish to interchange these two segments. More precisely, given any $v\in(\mathbb{Z}^2)^*$, we have by the Harris-FKG inequality ([10]):

$$\begin{aligned}&P_p[x\leftrightarrow v \text{ in } H_{v,v+y-x}]\,P_p[v\leftrightarrow y \text{ in } H_{v,v+y-x}]\\&\quad=P_p[v\leftrightarrow y \text{ in } H_{v,v+y-x}]\,P_p[y\leftrightarrow y+v-x \text{ in } H_{v,v+y-x}]\\&\quad\leqq P_p[v\leftrightarrow v+y-x \text{ in } H_{v,v+y-x}]\\&\quad=P_p[x\leftrightarrow y \text{ in } H_{xy}].\end{aligned}$$

From this, the van den Berg-Kesten inequality ([5]), Proposition 4.2, and (1.1), if $\|x-y\|$ is large,

$$\begin{aligned}P_p[x\leftrightarrow y]&\leqq P_p[x\leftrightarrow y,\ V\in x+2g(y-x)\,U_g]\\&\quad+P_p[x\leftrightarrow z \text{ for some } z\in\partial^*(x+2g(y-x)\,U_g)]\\&\leqq\sum_{v\in x+2g(y-x)U_g}P_p[\exists \text{ disjoint paths } x\leftrightarrow v \text{ and } v\leftrightarrow y \text{ both in } H_{v,v+y-x}]\\&\quad+|\partial^*(x+2g(y-x)\,U_g)|\,e^{-2\sigma g(y-x)}\\&\leqq c_{30}\|y-x\|^2\sup_{v\in(\mathbb{Z}^2)^*}P_p[x\leftrightarrow v \text{ in } H_{v,v+y-x}]\\&\quad\cdot P_p[v\leftrightarrow y \text{ in } H_{v,v+y-x}]+c_{31}\|y-x\|\,e^{-2\sigma g(y-x)}\\&\leqq c_{30}\|y-x\|^2P_p[x\leftrightarrow y \text{ in } H_{xy}]+P_p[x\leftrightarrow y]/2\end{aligned}$$

and the lemma follows easily. □

Proof of Theorem 4.1 It is sufficient to prove the result for large A. Let $m:=(A\log A)^{1/3}$, let β_m denote the m-hull skeletal path of $A^{1/2}\gamma_W$, and let

$$Q:=(A/|R(\beta_m)|)^{1/2}R(\beta_m).$$

Let $q_0, q_1, \ldots, q_J, q_{J+1}=q_0$ be the vertices of Q, cyclically ordered in the direction of positive orientation around the boundary of Q. It is easy to see that there exist $s_0, \ldots, s_{J+1}\in(\mathbb{Z}^2)^*$ such that

$$g(s_i-q_i)\leqq 2$$

and such that the polygonal path $s_0\to\ldots\to s_{J+1}=s_0$ lies outside Q. Note that by (3.3), $(J-1)\,m/2\leqq g(\beta_m)\leqq\mathscr{W}A^{1/2}$, so that

$$(J+1)\log A\leqq 3\mathscr{W}A^{1/6}(\log A)^{2/3}. \tag{4.1}$$

From the Harris-FKG inequality ([10]), Lemma 3.3, Proposition 4.2, Lemma 4.3, (4.1), and Lemma 3.3, then, for r as in Proposition 4.2 and A large,

$$\begin{aligned}
&P_p[\text{there exists an occupied dual circuit enclosing } Q] \\
&\quad \geqq P_p[s_i \leftrightarrow s_{i+1} \text{ in } H_{s_i s_{i+1}} \text{ for all } i \leqq J] \\
&\quad \geqq \prod_{i=0}^{J} P_p[s_i \leftrightarrow s_{i+1} \text{ in } H_{s_i s_{i+1}}] \\
&\quad \geqq \prod_{i=0}^{J} c_{32} \|s_{i+1}-s_i\|^{-(r+2)} e^{-\sigma g(s_{i+1}-s_i)} \\
&\quad \geqq A^{-(J+1)(r+2)} \exp\left(-\sigma \sum_{i=0}^{J} g(q_{i+1}-q_i)\right) \\
&\quad \geqq \exp(-c_{33} A^{1/6} (\log A)^{2/3}) \exp(-\sigma g((A/|R(\beta_m)|)^{1/2} \beta_m)) \\
&\quad \geqq \exp(-\sigma \mathscr{W} A^{1/2} - c_{27} A^{1/6} (\log A)^{2/3}). \quad \square
\end{aligned}$$

V Error terms for large finite cluster probabilities

In this section we will prove the error estimates (1.10) and (1.11) for the probabilities in Theorem 1.1. Without explicitly saying so each time, we will make statements which are actually only valid for sufficiently large N. Given a region, or set of sites, A, let

$$f_{\leqq n}(A) := |\{x \in A \cap \mathbb{Z}^2 : |C(x)| \leqq n\}| / |A \cap \mathbb{Z}^2|.$$

We will need the following result from [4].

Lemma 5.1 *For each $n \geqq 2$, each finite $A \subset \mathbb{Z}^2$, each $p \in (0, 1)$ and each $\varepsilon \in (0, 1)$,*

$$P_p[|f_{\leqq n}(A) - E f_{\leqq n}(A)| > \varepsilon] \leqq 18 \exp(-\varepsilon^2 |A| / 324 n^2).$$

Proof of (1.11). Suppose $N \leqq |C(0)| < \infty$ and let γ denote the outermost occupied dual circuit surrounding 0. Let κ_N satisfy

$$N^{-5/12} (\log N)^{7/3} \ll \kappa_N \ll N^{-1/3} (\log N)^{2/3}.$$

From Theorem 3.1 and the upper bound on κ_N,

$$\begin{aligned}
(5.1) \qquad &P_p[N \leqq |C(0)| < \infty, |R(\gamma)| \geqq (1-\kappa_N) N/P_\infty] \\
&\quad \leqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} + c_{34} N^{1/6} (\log N)^{2/3}).
\end{aligned}$$

Suppose then that $|R(\gamma)| < (1-\kappa_N) N/P_\infty$. Let $\mathscr{C}_{Nj}$ denote the class of all dual circuits α satisfying

$$0 \in R(\alpha) \quad \text{and} \quad (1 - 2^{j+1} \kappa_N) N/P_\infty \leqq |R(\alpha)| < (1 - 2^j \kappa_N) N/P_\infty.$$

For some $j \geqq 0$ and some $\alpha \in \mathscr{C}_{Nj}$, the following two events occur:

(i) $\gamma = \alpha$

(ii) the bond configuration inside $R(\alpha)$ includes a cluster of size N or more.

The key is that these two events are independent, since (i) depends only on bonds outside $R(\alpha)$. For small values of j, neither of the events (i) (summed over $\mathscr{C}_{Nj}$) or (ii) is by itself unlikely enough to provide the bound we need, but we will show the product of the two probabilities is small enough.

Fix $j \geqq 0$, let $n := [(\log N)^2/\sigma^2 W^2]$, define $\Delta_n := P_p[n < |C(0)| < \infty]$, and define $\kappa \in (2^j \kappa_N, 2^{j+1} \kappa_N]$ by

$$|R(\alpha)| = (1-\kappa) N/P_\infty .$$

By (1.5),

$$\Delta_n < \kappa_N/2(1-\kappa_N) \leqq \kappa/2(1-\kappa).$$

Therefore by the Corollary to Lemma 4.2 of [4], the probability, denoted $P_{\geqq N|R(\alpha)}$, of (ii) satisfies

$$\begin{aligned} P_{\geqq N|R(\alpha)} &\leqq 18 \exp(-c_{35}\, n^{-2} \kappa^2 N) \\ &\leqq \exp(-c_{36}\, 2^{2j} \kappa_N^2\, N/(\log N)^4). \end{aligned} \tag{5.2}$$

On the other hand, by Theorem 3.1,

$$\begin{aligned} P_p[\gamma \in \mathscr{C}_{Nj}] \leqq \exp(&-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} + c_{37}\, 2^j \kappa_N N^{1/2} \\ &+ c_{38} N^{1/6} (\log N)^{2/3}). \end{aligned} \tag{5.3}$$

Notice that shrinking $|R(\gamma)|$ by a factor of roughly $1-2^j\kappa_N$ from its "natural" size N/P_∞ (i.e. assuming $\gamma \in \mathscr{C}_{Nj}$) shortens γ and thereby increases the probability that γ forms. Hence the second term in the exponent in (5.3), which is of the same order as the reduction in length. But (due to the lower bound on κ_N) this and the third term in the exponent in (5.3) are more than compensated for by the reduced probability in (5.2) that a size-N cluster forms, which has roughly the square of the length reduction in the exponent. This compensation occurs for each fixed size range (i.e. each j) but not when all sizes are lumped together, which is what requires that we sort the possible curves γ by size.

Thus we obtain

$$\begin{aligned} &P_p[N \leqq |C(0)| < \infty, |R(\gamma)| < (1-\kappa_N) N/P_\infty] \\ &\leqq \sum_{j=0}^{\infty} \sum_{\alpha \in \mathscr{C}_{Nj}} P_p[\gamma = \alpha]\, P_{\geqq N|R(\alpha)} \\ &\leqq \sum_{j=0}^{\infty} P_p[\gamma \in \mathscr{C}_{Nj}] \sup_{\alpha \in \mathscr{C}_{Nj}} P_{\geqq N|R(\alpha)} \\ &\leqq \sum_{j=0}^{\infty} \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} - c_{39}\, 2^{2j} \kappa_N^2\, N/(\log N)^4) \\ &\leqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} - c_{40}\, \kappa_N^2\, N/(\log N)^4). \end{aligned} \tag{5.4}$$

Together (5.1) and (5.4) provide the desired upper bound:

$$P_p[N \leqq |C(0)| < \infty] \leqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} + c_{41} N^{1/6} (\log N)^{2/3}). \tag{5.5}$$

We turn next to a lower bound on this same probability. Let θ_N satisfy

$$N^{-1/2}(\log N)^2 \ll \theta_N \ll N^{-1/3}(\log N)^{2/3}.$$

Define polygons

$$Q_N := (1+2\theta_N)\,Q(N/P_\infty, p), \qquad Q'_N := (1+\theta_N)\,Q(N/P_\infty, p)$$

as in Theorem 4.1. Let $n := [c_{42}(\log N)^2]$ (with c_{42} to be specified later) and let

$$B_N := \{x \in Q'_N \cap \mathbb{Z}^2 : |C(x)| > n\}.$$

Then

$$E|B_N| \geqq |Q'_N \cap \mathbb{Z}^2|\, P_\infty \geqq (1+\theta_N)^2 N - c_{43} N^{1/2} \geqq N + \theta_N N$$

so that, using Lemma 5.1 and the lower bound on θ_N,

$$\begin{aligned} P_p[|B_N| \leqq N] &\leqq P_p[|B_N| - E|B_N| \leqq -\theta_N N] \\ &\leqq \exp(-c_{44}\,\theta_N^2\, N/n^2) \to 0 \end{aligned} \tag{5.6}$$

as $N \to \infty$. Further, if not all sites in B_N are connected together within Q_N then there must be an occupied dual path within Q_N either connecting the interior of Q'_N to the boundary of Q_N or surrounding a large cluster within Q_N. Either way (part of) this path has endpoints separated by g-distance of order $c_{42}^{1/2} \log N$ or more, since

$$d_H(\gamma_{Q'_N}, \gamma_{Q_N}) \geqq c_{45}\,\theta_N\, N^{1/2} \gg \log N.$$

As there are only of order N^2 possible pairs of endpoints for such a path, its existence is unlikely if c_{42} is large, by (1.1). More precisely,

$$P_p[B_N \text{ intersects two distinct clusters}] \to 0 \quad \text{as} \quad N \to \infty.$$

Thus from (5.6),

$$P_p[\text{there exists a cluster of size } N \text{ or more in the configuration inside } Q_N] \to 1. \tag{5.7}$$

The event in (5.7) is independent of the event that there exists an occupied dual circuit enclosing Q_N, and is positively correlated (by the Harris-FKG inequality, [10]) with the event $[0 \in B_N]$ which has probability at least P_∞. All three events together ensure $N \leqq |C(0)| < \infty$. Therefore we obtain the desired lower bound, using Theorem 4.1 and the upper bound on θ_N:

$$\begin{aligned} P_p[N \leqq |C(0)| < \infty] &\geqq P_\infty \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2}(1+\theta_N) - c_{46} N^{1/6}(\log N)^{2/3}) \\ &\geqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} - c_{47} N^{1/6}(\log N)^{2/3}), \end{aligned} \tag{5.8}$$

and (1.11) follows. □

Proof of (1.10). Suppose $N \leqq |C(0)| < \infty$ and let γ denote innermost occupied dual circuit enclosing 0. We must show that the total probability of those configurations in which $\rho(\gamma_W, (P_\infty/N)^{1/2}\gamma)$ exceeds (const.) $N^{-1/6}(\log N)^{1/3}$ is small relative to the lower bound in (5.8).

First, from (1.11), (5.4), and (5.8) we see that, for κ_N defined in the proof of (1.11).

$$P_p[N \leqq |C(0)| < \infty, |R(\gamma)| \leqq (1-\kappa_N) N/P_\infty] = o(P_p[N \leqq |C(0)| < \infty]).$$

Let $\delta_N := c_{48} N^{-1/6} (\log N)^{1/3}$, with c_{48} to be specified later. Second, from Theorem 3.1, (1.11), and (5.8) we obtain

$$\begin{aligned}(5.9)\ & P_p[N \leqq |C(0)| < \infty, |R(\gamma)| > (1+\delta_N^2) N/P_\infty] \\ & \leqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} - \sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} \delta_N^2/3 + c_{49} N^{1/6} (\log N)^{2/3}) \\ & = o(P_p[N \leqq |C(0)| < \infty]),\end{aligned}$$

provided c_{48} is chosen sufficiently large. And third, from Theorem 3.2, (1.11), and (5.8) we get

$$\begin{aligned}(5.10)\quad & P_p[N \leqq |C(0)| < \infty, |R(\gamma)| > (1-\kappa_N) N/P_\infty, \rho(\gamma_W, \gamma/|R(\gamma)|^{1/2}) \geqq \delta_N] \\ & \leqq \exp(-\sigma \mathscr{W} P_\infty^{-1/2} N^{1/2} + c_{50} \kappa_N N^{1/2} - c_{51} N^{1/2} \delta_N^2) \\ & = o(P_p[N \leqq |C(0)| < \infty]).\end{aligned}$$

again provided c_{48} is chosen sufficiently large.

On the other hand, if both

$$\rho(\gamma_W, \gamma/|R(\gamma)|^{1/2}) < \delta_N \quad \text{and} \quad (1-\kappa_N) N/P_\infty \leqq |R(\gamma)| \leqq (1+\delta_N^2) N/P_\infty$$

then

$$\begin{aligned}\rho(\gamma_W, (P_\infty/N)^{1/2} \gamma) & \leqq \rho(\gamma_W, \gamma/|R(\gamma)|^{1/2}) + \rho(\gamma/|R(\gamma)|^{1/2}, (P_\infty/N)^{1/2} \gamma) \\ & \leqq \delta_N + \big||R(\gamma)|^{-1/2} - (P_\infty/N)^{1/2}\big| \operatorname{diam}(\gamma) \\ & \leqq \delta_N + 2|(N/P_\infty |R(\gamma)|)^{1/2} - 1| \operatorname{diam}(\gamma_W) \\ & \leqq \delta_N + 3(\kappa_N + \delta_N^2) \operatorname{diam}(\gamma_W) \\ & = O(N^{-1/6} (\log N)^{1/3})\end{aligned}$$

and (1.10) follows. □

VI Error terms for large deviation probabilities

In this section we will prove the error estimates (1.12) and (1.13) for the probabilities in Theorem 1.2. Our approach is similar to that of the previous section: we first prove the estimate (1.13) for the error term in the probability (1.9) of the large deviation event $F_L(\lambda)$, then show that the probability that the shape approximation (1.8) fails, for $\zeta(L)$ as in (1.12), is much smaller than the probability in (1.9). We will tacitly assume in this section that L is large, where "large" may depend on λ and p. For this section we allow the constants c_i to depend on λ in addition to p.

If γ is the outermost occupied dual circuit surrounding some point of $\mathbb{Z}^2$, let us call $R(\gamma)$ a *bubble*. Fix L (the scale of the square Λ_L); we will call a bubble *large* if $|R(\gamma)| \geqq c_{52} (\log L)^2$, and *small* otherwise. Here c_{52} is a constant to be specified later. Notice that by Theorems 3.1 and 4.1 $(\log L)^2$ is the order of magnitude of the area of the largest bubble which "typically" appears in

Λ_L. The idea of Theorem 1.2 is that when $|\Lambda_L \cap C_\infty|$ is smaller than its typical value by a fraction λ, it is usually because there is a single large bubble shaped like W occupying a fraction λ of Λ_L.

Let us first prove an extension of Theorem 3.1. Let

$$u(A) := \sigma \mathscr{W} A^{1/2} - c_1 A^{1/6} (\log A)^{2/3}$$

denote the negative of the exponent in that theorem.

Theorem 6.1 *For each $p \in (p_c, 1)$ there exists a constant $c_{53} = c_{53}(p) > 0$ such that for every L, A, and a with $L \geqq 2$ and $A \geqq a \geqq c_{53}(\log L)^2$,*

(6.1) P_p [there exists a set Γ of occupied dual circuits γ with disjoint regions $R(\gamma)$, with $|R(\gamma) \cap \Lambda_L| \geqq a$ for each $\gamma \in \Gamma$, and with $\sum_{\gamma \in \Gamma} |R(\gamma) \cap \Lambda_L| \geqq A$]

$$\leqq |\Lambda_L|^4 \exp(-u(A)).$$

Proof. We decompose the event in (6.1) according to the number $k = |\Gamma|$ of circuits, their total area N in Λ_L, their individual areas $a_1, \ldots, a_k$, and a site x_i inside each circuit. Using the van den Berg-Kesten inequality ([5]) and Theorem 3.1, the left side of (6.1) is thereby bounded above by

$$(6.2) \qquad \sum_{k \leqq \min(|\Lambda_L|, A/a)} \; \sum_{A \leqq N \leqq |\Lambda_L|} \; \sum_{a_1 + \ldots + a_k = N, a_i \geqq a} \; \sum_{x_1, \ldots, x_k \in \Lambda_L} \prod_{i=1}^{k} \exp(-u(a_i))$$

$$\leqq \sum_{k \leqq \min(|\Lambda_L|, A/a)} \; \sum_{A \leqq N \leqq |\Lambda_L|} N^k |\Lambda_L|^k \exp\left(- \min_{a_1 + \ldots + a_k = N, a_i \geqq a} \sum_{i=1}^{k} u(a_i) \right).$$

Now for $x \geqq$ some c_{54} we have $u(x)$ concave and $x u'(x) \leqq 3u(x)/4$. As in the proof of Theorem 6.1 of [4], this concavity implies that if we fix k and $N = a_1 + \ldots + a_k$, and fix all but two of the a_i's, the minimum of $\sum_{i=1}^{k} u(a_i)$ is obtained by taking one of the two remaining a_i's as large as possible and the other as small as possible (i.e. equal to a). It follows that the minimum on the right side of (6.2) is equal to $(k-1) u(a) + u(N - (k-1) a)$. This can be interpreted as meaning that the most likely configuration of k bubbles, of minimum area a and total area N, consists of $k-1$ bubbles of area a and one large bubble with all of the remaining area. The right side of (6.2) is now bounded by

$$(6.3) \qquad \sum_{k \leqq \min(|\Lambda_L|, A/a)} |\Lambda_L|^{2k+1} \exp(-(k-1) u(a) - u(A - (k-1) a)).$$

Using monotonicity of u' and the above bound on $x u'(x)$ it is easily seen that (the logarithm of) the summand in (6.3) is a decreasing function of $k \geqq 1$, provided $a \geqq c_{53} (\log L)^2$. Replacing each term with the $k = 1$ term gives (6.1). □

We need to establish some notation and terminology for the proof of (1.13). Let B_L denote the union of the (necessarily large) bubbles $R(\gamma)$ for which

$$|R(\gamma) \cap \Lambda_L| \geqq c_{52} (\log L)^2;$$

let

$$S_L := \{x \in \Lambda_L : d(x, \partial \Lambda_L) \leqq 2 c_{52} (\log L)^2\}$$

denote the boundary strip, and define the *foam region*

$$G_L(B_L):=\Lambda_L \backslash (B_L \cup S_L);$$

$G_L(B_L)$ consists of small bubbles and (a neighborhood of) part of C_∞.

There are two modes of occurrence for the large-deviation event $F_L(\lambda)$. One is for large bubbles to enclose a fraction at least approximately λ of Λ_L, leaving a fraction at most approximately $1-\lambda$ for the foam region. Then typically a fraction P_∞ of the sites in the foam region, and of course none of the sites in the large bubbles, will be in C_∞, for an overall fraction of at most $(1-\lambda) P_\infty$. The other mode is for the large bubbles (if any) to occupy a fraction less than λ of Λ_L, while the foam in the foam region is fluffier that typical; a fraction significantly greater than $1-P_\infty$ of the sites in G_L must be in (necessarily small) bubbles, hence also in small clusters. This second mode, as we shall see, is far less likely than the first, while the first is dominated, analogously to Theorem 1.1, by configurations in which there is only one large bubble shaped like the Wulff set W.

Proof of (1.13). Let B denote a possible value of the set B_L of large bubbles. Let $H=H(B)$ denote the set of all sites which are in small bubbles contained entirely in B^c. If $F_L(\lambda)$ occurs with $B_L=B$, then for the corresponding value $G_L(B)$ of the foam region, we have

$$|G_L(B) \cap H(B)| + |S_L \cup (B \cap \Lambda_L)| \geqq |\Lambda_L|(1-P_\infty+\lambda P_\infty),$$

as these are upper and lower bounds for the number of sites in Λ_L in finite clusters.

Define $\theta_L=\theta_L(B)$ by

$$(1-P_\infty+\theta_L)|G_L(B)| + |S_L \cup (B \cap \Lambda_L)| = |\Lambda_L|(1-P_\infty+\lambda P_\infty).$$

θ_L is defined so that if $F_L(\lambda)$ occurs with $B_L=B$, then

$$|G_L(B) \cap H(B)| \geqq (1-P_\infty+\theta_L)|G_L(B)|,$$

i.e. the fraction of $G_L(B)$ in small bubbles is too big by at least θ_L.

The event that $B_L=B$ tells us two things:

(i) every dual bond in ∂B is occupied;

(ii) every site in the corresponding foam region $G_L(B)$ either is connected to ∞ in B^c or is in a small bubble contained entirely in B^c.

In (ii) we use $G_L(B)$ and not $\Lambda_L \backslash B$ because sites in the boundary strip S but outside B could be in large bubbles. Notice that events (i) and (ii) are independent; this can be used, just as in the proof of Theorem 6.1 in [4], in showing that

$$P_p(F_L(\lambda)) \leqq \sum_B P_p(B_L=B)\, P_p[|G_L(B) \cap H(B)| \geqq (1-P_\infty+\theta_L(B))|G_L(B)|] \tag{6.4}$$

where the sum is over all possible values B of B_L. (In the earlier paper values of B are specified by a subscript j; (6.4) here corresponds to (6.14) there.)

Let $n := [c_{52}(\log L)^2]$; then the last event in (6.4) implies that the fraction $f_{\leqq n}(G_L(B))$ of the sites in $G_L(B)$ which are in clusters of size n or less satisfies $f_{\leqq n}(G_L(B)) - E\, f_{\leqq n}(G_L(B)) \geqq \theta_L(B)$. By Lemma 5.1 the probability therefore satisfies

$$(6.5) \qquad \begin{aligned} &P_p[|G_L(B) \cap H(B)| \geqq (1 - P_\infty + \theta_L(B))|G_L(B)|] \\ &\quad \leqq 18 \exp(-c_{55}\, \theta_L(B)^2 |G_L(B)|/(\log L)^4), \end{aligned}$$

provided $\theta_L(B) > 0$.

Analogously to κ_N of the proof of (1.11), let τ_L satisfy

$$L^{-5/6}(\log L)^{7/3} \ll \tau_L \ll L^{-2/3}(\log L)^{2/3}.$$

Consider first the class $\mathscr{B}'(L)$ consisting of those B with

$$|B \cap \Lambda_L| \geqq (1 - \tau_L)\, \lambda |\Lambda_L|.$$

These correspond to the first mode of occurrence mentioned above for $F_L(\lambda)$. From Theorem 6.1 and the upper bound on τ_L,

$$(6.6) \qquad \begin{aligned} P_p[B_L \in \mathscr{B}'(L)] &\leqq |\Lambda_L|^4 \exp(-u((1 - \tau_L)\, \lambda |\Lambda_L|)) \\ &\leqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L + c_{56} L^{1/3}(\log L)^{2/3}). \end{aligned}$$

Next for $j \geqq 0$ consider the class $\mathscr{B}_j(L)$ of those B for which

$$(1 - 2^{j+1}\tau_L)\, \lambda |\Lambda_L| < |B \cap \Lambda_L| \leqq (1 - 2^j \tau_L)\, \lambda |\Lambda_L|.$$

These correspond to the second mode of occurrence for $F_L(\lambda)$. Using Theorem 6.1 again,

$$(6.7) \qquad P_p[B_L \in \mathscr{B}_j(L)] \leqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L + c_{57}\, 2^j \tau_L L + c_{58} L^{1/3}(\log L)^{2/3}).$$

Now (6.7) and (6.5) can be used to bound the right side of (6.4), if we can obtain lower bounds for $|G_L(B)|$ and $\theta_L(B)$. Let $B \in \mathscr{B}_j(L)$. Note that necessarily $2^j \tau_L < 1$. The first bound is easy:

$$(6.8) \qquad \begin{aligned} |G_L(B)| &\geqq |\Lambda_L| - |B \cap \Lambda_L| - |S_L| \geqq (1 - \lambda)|\Lambda_L| + \tau_L \lambda |\Lambda_L| - |S_L| \\ &\geqq (1 - \lambda)|\Lambda_L|. \end{aligned}$$

For the second bound, if $P_\infty < \theta_L$ then the probability in (6.5) is 0 so we may assume $P_\infty > \theta_L$. Then

$$\begin{aligned} (1 - P_\infty + \lambda P_\infty)|\Lambda_L| &= (1 - P_\infty + \theta_L)|G_L(B)| + |S_L \cup (B \cap \Lambda_L)| \\ &= (1 - P_\infty + \theta_L)|\Lambda_L| + (P_\infty - \theta_L)|S_L \cup (B \cap \Lambda_L)| \\ &\leqq (1 - P_\infty + \theta_L)|\Lambda_L| + (P_\infty - \theta_L)(|S_L| + (1 - 2^j \tau_L)\, \lambda |\Lambda_L|) \\ &\leqq (1 - P_\infty + \theta_L)|\Lambda_L| + (P_\infty - \theta_L)(1 - 2^{j-1}\tau_L)\, \lambda |\Lambda_L|. \end{aligned}$$

Solving for θ_L gives

$$(6.9) \qquad \theta_L(B) \geqq \lambda P_\infty\, 2^{j-1} \tau_L \qquad \text{for all } B \in \mathscr{B}_j(L).$$

Combining (6.4), (6.5), (6.7), (6.8), and (6.9) we obtain, analogously to (5.3):

$$(6.10)\qquad \begin{aligned} &P_p(F_L(\lambda), B_L \notin \mathscr{B}'(L)) \\ &\quad \leqq \sum_{j \geqq 0} 18 \exp(-\sigma \mathscr{W} \lambda^{1/2} L + c_{57} 2^j \tau_L L + c_{58} L^{1/3} (\log L)^{2/3} \\ &\qquad - c_{59} 2^{2j} \tau_L^2 L^2/(\log L)^4) \\ &\quad \leqq \sum_{j \geqq 0} \exp(-\sigma \mathscr{W} \lambda^{1/2} L - c_{60} 2^{2j} \tau_L^2 L^2/(\log L)^4) \\ &\quad \leqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L - c_{61} \tau_L^2 L^2/(\log L)^4). \end{aligned}$$

Now we turn to lower bounds on $P_p(F_L(\lambda))$. The condition $\lambda < (\operatorname{diam}(\gamma_W))^{-2}$ ensures that the curve $\lambda^{1/2} L \gamma_W$ is entirely inside Λ_L; hence (for large L) so is γ_Q, where $Q = Q(1+\tau_L)\lambda|\Lambda_L|, p)$ is the polygon from Theorem 4.1 which approximates $\lambda^{1/2} L \gamma_W$. If there exists an occupied dual circuit enclosing Q, let γ_I denote the innermost such circuit. Given any dual circuit α, let C_∞^α denote the set of those sites connected to ∞ by a path of occupied bonds entirely outside $R(\alpha)$. If for some dual circuit α, both

(i) $$\gamma_I = \alpha,$$

and

(ii) $$|C_\infty^\alpha \cap (\Lambda_L \backslash R(\alpha))| \leqq (1-\lambda) P_\infty |\Lambda_L|$$

then $F_L(\lambda)$ occurs. Now (i) and (ii) are independent, and we will show (ii) is a very likely event. In fact it is easily checked that (ii) will occur provided

$$(6.11)\qquad f_{\leqq n}(\Lambda_L \backslash Q) - E f_{\leqq n}(\Lambda_L \backslash Q) > -c_{62} \tau_L$$

where $n = [c_{63} (\log L)^2]$; here c_{63} is chosen (using Theorem 1.1) so that $P[n < |C(0)| < \infty]$ is much smaller than τ_L. From Lemma 5.1 the probability of (6.11) approaches one as $L \to \infty$. Thus from Theorem 4.1,

$$(6.12)\qquad \begin{aligned} P_p(F_L(\lambda)) &\geqq \sum_\alpha P_p[\gamma_I = \alpha]\, P_p[|C_\infty^\alpha \cap (\Lambda_L \backslash R(\alpha))| \leqq (1-\lambda) P_\infty |\Lambda_L|] \\ &\geqq P_p[\text{there exists an occupied dual circuit enclosing } Q]/2 \\ &\geqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L - c_{64} L^{1/3}/(\log L)^{2/3}). \end{aligned}$$

Together (6.6), (6.10), and (6.12) prove (1.13). □

Proof of (1.12). Suppose $F_L(\lambda)$ occurs and let $\gamma_1, \ldots, \gamma_k$ denote those occupied dual circuits, if any, which bound a (necessarily large) bubble for which

$$|R(\gamma_i) \cap \Lambda_L| \geqq c_{52} (\log L)^2;$$

of course K is random. We may assume γ_1 maximizes $|R(\gamma_i) \cap \Lambda_L|$ over $i = 1, \ldots, K$. As in the proof of (1.10) we must show that the total probability of those configurations in which $\rho(\gamma_W, \gamma_1/\lambda^{1/2} L)$ exceeds (const.) $L^{-1/3} (\log L)^{1/3}$ is small relative to the lower bound

$$(6.13)\qquad \exp(-\sigma \mathscr{W} \lambda^{1/2} L - c_{64} L^{1/3}/(\log L)^{2/3})$$

in (6.12).

Let

$$\delta'_L := c_{65}\, L^{-1/3} (\log L)^{1/3}$$

with c_{65} to be specified. Let τ_L be as in the proof of (1.13). If the desired event

$$\rho(\gamma_W, \gamma_1/\lambda^{1/2} L) \leqq \delta'_L$$

(which says that the one largest bubble is of the right size and shape to account for $F_L(\lambda)$ and (1.8)) does not occur, there are four possibilities:

(i) $$|B_L| = \sum_{i=1}^{K} |R(\gamma_i) \cap \Lambda_L| < (1-\tau_L)\,\lambda |\Lambda_L|,$$

i.e. there is not enough volume in all the large bubbles together;

(ii) $$|R(\gamma_1) \cap \Lambda_L| < (1-3\tau_L)\,\lambda |\Lambda_L|,$$

but not (i), i.e. there is more than one bubble of significant size;

(iii) $$|R(\gamma_1) \cap \Lambda_L| > (1+(\delta'_L)^2)\,\lambda |\Lambda_L|,$$

i.e. there is too much volume in the largest bubble;

(iv) both $|R(\gamma_1) \cap \Lambda_L| \geqq (1-3\tau_L)\,\lambda |\Lambda_L|$ and $\rho(\gamma_W, \gamma_1/\lambda^{1/2} L) > \delta'_L$,

i.e. the largest bubble is big enough but of the wrong shape.

From (6.10) we see that the probability of (i) is much less than (6.13). From the proof of (1.10) (compare (5.9) and (5.10)) we see that the probabilities of (iii) and (iv) are much less than (6.13), provided c_{65} is sufficiently large.

Under (ii) let M be the least index such that

$$\sum_{i=1}^{M} |R(\gamma_i) \cap \Lambda_L| \geqq \tau_L\, \lambda |\Lambda_L|.$$

Then for some $1 \leqq j < J := [\tau_L^{-1}] - 2$, both

$$j\tau_L\, \lambda |\Lambda_L| \leqq \sum_{i=1}^{M} |R(\gamma_i) \cap \Lambda_L| < (j+1)\,\tau_L\, \lambda |\Lambda_L| \tag{6.14}$$

and

$$(J-j)\,\tau_L\, \lambda |\Lambda_L| \leqq \sum_{i=M+1}^{K} |R(\gamma_i) \cap \Lambda_L| \tag{6.15}$$

with the occupied dual circuits in these two events occurring disjointly. By Theorem 6.1 the probabilities of (6.14) and (6.15) are bounded above by

$$|\Lambda_L|^4 \exp(-u(j\tau_L\, \lambda |\Lambda_L|)) \quad \text{and} \quad |\Lambda_L|^4 \exp(-u((J-j)\,\tau_L\, \lambda |\Lambda_L|))$$

respectively. Using the van den Berg-Kesten inequality ([5]) it follows that the probability of (ii) is bounded above by

$$\sum_{j=1}^{J-1} |A_L|^8 \exp(-[u(j\tau_L \lambda|A_L|)+u((J-j)\tau_L \lambda|A_L|)]). \tag{6.16}$$

From the property of $u(.)$ mentioned after (6.2), the maximum term in this sum occurs for $j=1$ (or $j=J-1$), so that (6.16) is bounded above by

$$\begin{aligned} &J|A_L|^8 \exp(-[u((1-4\tau_L)\lambda|A_L|)+u(\tau_L \lambda|A_L|)]) \\ &\leqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L + c_{66}\tau_L L + c_{67} L^{1/3}(\log L)^{2/3} - c_{68}\tau_L^{1/2} L) \\ &\leqq \exp(-\sigma \mathscr{W} \lambda^{1/2} L - c_{69}\tau_L^{1/2} L) \end{aligned}$$

which is much less than (6.13). □

References

1. Aizenman, M., Chayes, J.T., Chayes, L., Newman, C.M.: Discontinuity of the magnetization in the $1/|x-y|^2$ Ising and Potts models. J. Stat. Phys. **50**, 1–40 (1988)
2. Aizenman, M., Kesten, H., Newman, C.M.: Uniqueness of the infinite cluster and continuity of connectivity functions for short- and long-range percolation. Commun. Math. Phys. **111**, 505–532 (1987)
3. Alexander, K.S.: Lower bounds on the connectivity function in all direction for Bernoulli percolation in two and three dimensions. Ann. Probab. **18**, 1547–1562 (1989)
4. Alexander, K.S., Chayes, J.T., Chayes, L.: The Wulff construction and asymptotics of the finite cluster distribution for two-dimensional Bernoulli percolation. Commun. Math. Phys. **131**, 1–50 (1990)
5. van den Berg, J., Kesten, H.: Inequalities with applications to percolation and reliability. J. Appl. Probab. **13**, 293–313 (1985)
6. Bonnesen, T.: Les problemes des isoperimetres et des isepiphanes. Paris: Gauthier-Villars 1929
7. Dobrushin, R.L., Kotecky, R., Shlosman, S.B.: Equilibrium crystal shapes – a microscopic proof of the Wulff construction. Research announcement; (preprint, 1988)
8. Fortuin, C., Kasteleyn, P.: On the random cluster model. I. Physica **57**, 536–564 (1972)
9. Hammersley, J.M.: Percolation processes. Lower bounds for the critical probability. Ann. Math. Stat. **28**, 790–795 (1957)
10. Harris, T.E.: A lower bound for the critical probability in certain percolation processes. Proc. Cambridge Philos. Soc. **56**, 13–20 (1960)
11. Kesten, H.: The critical probability of bond percolation on the square lattice equals 1/2. Commun. Math. Phys. **74**, 41–79 (1980)
12. Shlosman, S.B.: Wulff construction justified. In: Simon, B., Truman, A., Davies, I.M. (eds.) IXth International Congress on Mathematical Physics. Proceedings of the Congress Held at University College of Swansea, Swansea, July 17–27, 1988. Adam Hilger Ltd., Bristol 1989
13. Taylor, J.E.: Existence and structure of solutions to a class of nonelliptic variational problems. Symp. Math. **14**, 499–508 (1974)
14. Taylor, J.E.: Unique structure of solutions to a class of nonelliptic variational problems. Proc. Symp. Pure Math. **27**, 419–427 (1975)
15. Wallen, L.J.: All the way with Wirtinger: a short proof of Bonnesen's inequality. Am. Math. Mont. **94**, 440–442 (1987)