Commun. Math. Phys. 131, 1–50 (1990)

Communications in Mathematical Physics



# The Wulff Construction and Asymptotics of the Finite Cluster Distribution for Two-Dimensional Bernoulli Percolation

K. Alexander[1]*, J. T. Chayes[2]** and L. Chayes[2]**

[1] Department of Mathematics, University of Southern California, Los Angeles, California 90089, USA
[2] Department of Mathematics, University of California, Los Angeles, California 90024, USA

**Abstract.** We consider two-dimensional Bernoulli percolation at density $p > p_c$ and establish the following results:

1. The probability, $P_N(p)$, that the origin is in a *finite* cluster of size $N$ obeys

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \log P_N(p) = -\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}},$$

where $P_\infty(p)$ is the infinite cluster density, $\sigma(p)$ is the (zero-angle) surface tension, and $\omega(p)$ is a quantity which remains positive and finite as $p \downarrow p_c$. Roughly speaking, $\omega(p)\sigma(p)$ is the minimum surface energy of a "percolation droplet" of unit area.

2. For all supercritical densities $p > p_c$, the system obeys a microscopic Wulff construction: Namely, if the origin is conditioned to be in a finite cluster of size $N$, then with probability tending rapidly to 1 with $N$, the shape of this cluster—measured on the scale $\sqrt{N}$—will be that predicted by the classical Wulff construction. Alternatively, if a system of finite volume, $N$, is restricted to a "microcanonical ensemble" in which the infinite cluster density is below its usual value, then with probability tending rapidly to 1 with $N$, the excess sites in finite clusters will form a single large droplet, which—again on the scale $\sqrt{N}$—will assume the classical Wulff shape.

## 1. Introduction

We consider Bernoulli bond percolation on the square lattice in which bonds are independently occupied with density $p$ and vacant with density $1-p$. This model is known to have a phase transition at density $p_c = 1/2$, below which the occupied clusters are finite with probability one (w.p. 1) and above which there is a unique

* Work supported in part by NSF Grant No. DMS-87-02906
** Work supported in part by NSF Grant No. DMS-88-06552

infinite cluster w.p. 1. In this paper, we study the finite occupied clusters throughout the high-density or percolating phase, i.e. whenever $p > p_c$. Specifically, we obtain detailed estimates on the distribution of sizes and shapes of asymptotically large finite clusters.

In order to motivate our questions and our results, it is worth noting at the outset that the study of large finite clusters in the high-density phase of percolation has an analogue in other statistical mechanics models. The high-density phase of percolation corresponds to the ordered, and hence low-temperature phase of models such as the Ising magnet; the infinite cluster density is the analogue of the spontaneous magnetization [FK] (see also [ACCN]). Thus, in a distributional sense, the infinite cluster in a percolation configuration corresponds to the collection of excess plus spins in a low-temperature plus-state Ising configuration. Similarly, an anomalously large finite cluster in a percolation configuration corresponds to an anomalously large droplet of minus spins in a plus-state Ising configuration; i.e. the asymptotically large finite clusters may be viewed as "droplets of the wrong phase." In a more general context, the study of the shapes of these clusters is related to the question of crystal formation in other systems: What are the equilibrium shapes of crystals of one phase immersed in another?

*1.A. Previous Results.* Let us first discuss the size distribution of large finite clusters in percolation. This is typically described by the so-called finite cluster distribution:

$$P_N(p) = \mathbf{P}_p(|C(0)| = N), \tag{1.1}$$

where $\mathbf{P}_p(-)$ denotes Bernoulli measure at density $p$, and $|C(0)|$ denotes the size of the occupied cluster of the origin. There has been a good deal of previous work on the large-$N$ behavior of $P_N(p)$. It has been known for some time that below threshold $P_N(p)$ obeys the bounds

$$e^{-c_1(p)N} \leqq P_N(p) \leqq e^{-c_2(p)N} \quad (p < p_c) \tag{1.2}$$

in all dimensions, with $c_1(p)$ and $c_2(p)$ positive, finite, dimension-dependent constants. The lower bound in (1.2) is trivial; the upper bound was originally derived in [H], and then rederived in [K1] and [AN]. The behavior above threshold is of a very different form; for $d$ dimensions, $P_N(p)$ is expected to satisfy

$$e^{-c_3(p)N^{(d-1)/d}} \leqq P_N(p) \leqq e^{-c_4(p)N^{(d-1)/d}} \quad (p > p_c) \tag{1.3}$$

with $c_3(p)$ and $c_4(p)$ positive, finite, dimension-dependent constants. Both bounds in (1.3) were originally derived only for $p$ near 1 [KS]. The lower bound was later shown to hold for all $p > p_c$ in [ADS]. That the upper bound in (1.3) holds for all $p > p_c$ was demonstrated for two dimensions in [K3] (see also [CC2]). Still later, in [CCN], an upper bound of the form (1.3) with logarithmic modifications (i.e. with $c_4(p)$ replaced by $c_4(p)/\log N$) was shown to hold in dimensions $d \geqq 3$ whenever $p$ is above a value[1] which was conjectured to coincide with the percolation

[1] Very recently, there have been two independent proofs that this value coincides with the half-space percolation threshold ([BGN], [Z])

threshold. Recently, in [KZ], an upper bound of the form (1.3) was derived without a logarithmic modification, but $p$ is still restricted to lie above the value used in the [CCN) proof. In any case, we note that (1.3) has been established completely for two-dimensional percolation.

To our knowledge there has not been any previous work on the distribution of shapes of finite clusters in percolation, although there has been work on the analogous problem in other systems. As explained above, this problem is related to the classic question of determining the equilibrium shape of a crystal of one phase immersed in another. Under the assumption that the shape is determined only by a single intrinsic property, namely the surface tension (thus neglecting extrinsic effects such as gravity), one arrives at the following variational problem: For a fixed crystal volume, determine the shape which minimizes the surface energy. The solution of the continuum version of this problem was given at the turn of the century in the so-called Wulff construction [Wu]: Let $\sigma(\mathbf{n})$ denote the surface tension of a flat interface orthogonal to the outward normal $\mathbf{n}$. Then the equilibrium shape $W$ of a crystal of fixed volume is given by the convex set

$$W = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \cdot \mathbf{n} \leqq \sigma(\mathbf{n}) \text{ for all } \mathbf{n}\}. \tag{1.4}$$

For a crystal of volume V, the linear dimension of the Wulff shape (1.4) is simply scaled by the multiplicative factor $(V/|W|)^{1/d}$, where $|W|$ denotes the volume of $W$. That (1.4) is the unique minimizer of the variational problem has been proved by Taylor ([T1], [T2]).

The Wulff construction described above, and variants of this construction which account for the effects of gravity [ATZ] or the effects of substrates in the system [W], [ZAT], provide a good explanation of the observed thermodynamic properties of equilibrium crystal shapes. Changes in specific features of the Wulff shape have been related to various phase transitions: the roughening transition is believed to coincide with the disappearance of facets in $W$; if $W'$ is the Wulff shape for crystals in the presence of a substrate (see e.g. [ZAT]), the transition to complete wetting can be formulated as the vanishing of $|W'|$. See [RW], [BN], [A] for reviews on the study of equilibrium crystal shapes in various models; work on the Wulff construction for constrained $(1+1$ dimensional) models can be found in [DD] and [DDR].

On the other hand, from the viewpoint of statistical mechanics, the Wulff construction alone does not provide a theory of equilibrium crystal shapes. In many microscopic models, it is of course possible to extract an angle-dependent surface tension by studying the behavior of asymptotically large surfaces. However, it is not entirely obvious that the shapes of finite crystals will be distributed about the Wulff shape determined by this surface tension; furthermore, even if this is the case, one would like a probabilistic description of the deviation of the actual crystal shapes from the Wulff shape.

There has been progress on a microscopic theory of the Wulff construction for the two-dimensional Ising magnet at low temperatures. Minlos and Sinai [MS] studied a finite-volume microcanonical system with plus boundary conditions in which the magnetization was fixed at a value below the plus-state spontaneous magnetization—thereby forcing excess minus spins into the system. Roughly

speaking, they proved that in typical configurations at very low temperature, most of the excess minus spins form a single large droplet of essentially square shape. In this regard, it should be noted that as the temperature tends to zero, the Wulff shape $W$ of the two-dimensional Ising magnet tends to a square. More precisely, let $m(T)$ be the spontaneous magnetization of the two-dimensional Ising magnet at temperature $T$. Minlos and Sinai studied a system of volume $L^2$ with plus boundary conditions in the microcanonical ensemble with magnetization $m = (1 - 2\alpha)m(T)$, which therefore had an excess volume fraction $\alpha$ of minus spins. For $T$ very small, they showed that in the $L \uparrow \infty$ limit, a typical configuration contains a large (dual) contour $\gamma$ separating plus and minus spins such that

$$|\mathscr{A}(\gamma) - \alpha L^2| < c_5(T)L^{3/2}, \tag{1.5}$$

$$|\mathscr{L}(\gamma) - 4\sqrt{\alpha}L| < c_6(T)L, \tag{1.6}$$

where $\mathscr{A}(\gamma)$ denotes the area enclosed by $\gamma$, and $\mathscr{L}(\gamma)$ denotes the length of $\gamma$. The constants $c_5(T)$ and $c_6(T)$ tend to zero as $T \downarrow 0$. They also had estimates similar to (1.5) which showed that the magnetization inside $\gamma$ tends to $-m(T)$, while that outside $\gamma$ tends to $m(T)$ as $T \downarrow 0$, thus establishing that "most" of the excess minus spins are indeed enclosed by $\gamma$. That $\gamma$ tends to the boundary of a square is clear from the factor $4\sqrt{\alpha}L$ in the length bound (1.6). The Minlos–Sinai droplet theorem is thus a microscopic verification of the Wulff construction for the two-dimensional Ising magnet in the limit $T \downarrow 0$.

Simultaneously with the work presented here, Dobrushin, Kotecky and Shlosman [DKS] have announced a substantial refinement of the Minlos–Sinai result which deals directly with the Wulff construction for the two-dimensional Ising magnet and which closely parallels our work on two-dimensional percolation. Dobrushin, Kotecky and Shlosman again consider a system of volume $L^2$ (with periodic boundary conditions) in the microcanonical ensemble at magnetization $m = (1 - 2\alpha)m(T)$, $0 < \alpha < 1/4$. Again, they show that in the $L \uparrow \infty$ limit, a typical configuration contains a large contour $\gamma$. However, rather than comparing $\gamma$ to the boundary of the zero-temperature Wulff shape (i.e. the square), they compare it to the boundary, $\partial W \equiv \gamma_w$, of the actual Wulff shape $W$ at temperature $T$. (The natural means of comparison, namely the Hausdorff distance, is also used in our work and will be explained below.) For very low temperature, they show that the Hausdorff distance between $\gamma$ and a translate of $\gamma_w$ is bounded above by a sublinear power of $L$; since the length of $\gamma$ itself is only of order $L$, the deviation of $\gamma$ from $\gamma_w$—i.e. the ratio of the Hausdorff distance between $\gamma$ and $\gamma_w$ to the length $\mathscr{L}(\gamma)$—tends to zero like a power of $L$. Thus their work is a strong microscopic proof of the Wulff construction for the two-dimensional Ising magnet at small but positive temperature. As we will see below, the work presented here for two-dimensional percolation is complementary to the Dobrushin, Kotecky and Shlosman work on the two-dimensional Ising magnet; we do not obtain as sharp an estimate on the deviation from the Wulff shape, but we are able to prove a microscopic form of the Wulff construction for all supercritical values of the parameter $p$.

In order to formulate a microscopic Wulff construction for percolation, we

must first identify the "crystals of one phase immersed in another" and then define a surface tension for these crystals. As mentioned earlier, our candidate crystals are just the large finite clusters when $p > p_c$—these are the analogues of the droplets of minus spins in a sea of plus spins in the Ising magnet at $T < T_c$. In the latter system, surface tension is the exponential decay rate of the probability of an asymptotically large dual surface separating the plus and minus spins. Similarly in bond percolation one can define a dual model in which a dual $(d-1)$-cell is occupied whenever the corresponding bond is vacant; then the surface tension is just the exponential decay rate of the probability of large surfaces composed of these dual $(d-1)$-cells. More precisely, the zero-angle surface tension is the decay rate for dual surfaces spanning large loops in a lattice hyperplane, while an angle-dependent surface tension is the decay rate for dual surfaces spanning loops oriented at some non-zero angle to a coordinate axis. In both the Ising magnet and percolation, one expects that the Wulff shape (1.4) derived from this surface tension is not spherical (except at the critical point), reflecting the anisotropy of the system.

*1.B. Statement of Results and Discussion.* The upper and lower bounds in (1.3) suggest that for $p > p_c$, $\log P_N(p)/N^{(d-1)/d}$ should approach some well-defined value as $N \uparrow \infty$. Our first result is that this is indeed the case in two dimensions; moreover, we can identify the limiting value in terms of other quantities in percolation. Here we will briefly define these quantities; precise definitions are given in Sects. 2 and 3. First, let $P_\infty(p)$ denote the infinite cluster density at bond probability $p$. Next, let $\sigma(\mathbf{n}, p)$ be the angle-dependent surface tension at bond density $p$, obtained by considering the probability of dual surfaces oriented orthogonally to the outward normal $\mathbf{n}$. We will denote the standard zero-angle surface tension at density $p$ by $\sigma(p) \equiv \sigma(\hat{e}_y, p)$; i.e. $\sigma(p)$ is the surface tension for a surface along the $\hat{e}_x$-axis, and thus orthogonal to the outward normal $\hat{e}_y$. Note that, in two dimensions, $\sigma(p)$ is just the inverse of the on-axis dual correlation length. We also define a new quantity, $\omega(p)$, which we call the Wulff constant, as follows. Consider the Wulff variational problem, as described earlier, for the surface tension $\sigma(\mathbf{n}, p)$: namely, given $\sigma(\mathbf{n}, p)$, what is the minimum surface energy of a "droplet" of unit area? (Recall that the surface energy of a droplet is the integral of the surface tension $\sigma(\mathbf{n}, p)$ over the boundary of the droplet.) We define $\omega(p)$ to be the ratio of this minimum surface energy to the zero-angle surface tension $\sigma(p)$. At first, it may seem rather unnatural to divide out the surface tension $\sigma(p)$; however, this is the appropriate scaling from the viewpoint of critical phenomena. Indeed, we can show that for all $p$

$$\sqrt{2\pi} \leqq \omega(p) \leqq 4, \tag{1.7}$$

so that $\omega(p)$ remains finite as $p \downarrow p_c$. Our first main result is:

**Theorem 1.** *In the two-dimensional Bernoulli bond percolation model on the square lattice, for every* $p > p_c$

$$\lim_{N \to \infty} \frac{1}{\sqrt{N}} \log P_N(p) = -\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}}.$$

Of course, this theorem completely determines the leading critical behavior of the decay rate of the finite cluster distribution as $p \downarrow p_c$.

It is worth noting that we also prove upper bounds on $P_N(p)$ which are somewhat stronger than Theorem 1 implies: namely

$$P_{\geqq N}(p) \leqq \exp\{-[\omega(p)\sigma(p)/\sqrt{P_\infty(p)}]\sqrt{N}[1-N^{-1/4}(\log N)^4]\}, \tag{1.8}$$

where $P_{\geqq N}(p) \geqq P_N(p)$ is the probability of the event $N \leqq |C(0)| < \infty$. (See Theorem 1.A in Sect. 4.) However, we cannot yet supplement this with a lower bound which is stronger than that implied by Theorem 1.

Our next set of results constitute a microscopic proof of the two-dimensional Wulff construction for all $p > p_c$. Roughly speaking, we show that in the (unlikely) configurations in which the origin is in a finite cluster of size exceeding $N$, this cluster assumes a Wulff shape of linear scale $\sqrt{N/P_\infty}$. Of course, this shape will not be achieved exactly on the scale of the lattice; the best that can be expected is that, with high probability, the cluster achieves this shape on scales which are large relative to the lattice spacing, but small relative to the size of the cluster.

To be precise, consider the angle-dependent surface tension, $\sigma(\mathbf{n}, p)$, for two-dimensional percolation at density $p$. Let $W = W(p)$ be the *unit area* continuum Wulff shape for surface tension $\sigma(\mathbf{n}, p)$, defined via Eq. (1.4), and let $\gamma_w = \gamma_w(p)$ be the boundary of this shape.[2] We use the Hausdorff distance to compare the boundary of our cluster to a translate of $\gamma_w$. The reader should recall that the Hausdorff distance between two fixed curves, $\gamma_1$ and $\gamma_2$, is simply

$$D_H(\gamma_1, \gamma_2) = \max\left\{\max_{\mathbf{x}\in\gamma_1}\min_{\mathbf{y}\in\gamma_2}|\mathbf{x}-\mathbf{y}|, \max_{\mathbf{x}\in\gamma_2}\min_{\mathbf{y}\in\gamma_1}|\mathbf{x}-\mathbf{y}|\right\}. \tag{1.9}$$

We use the metric

$$\rho(\gamma_1, \gamma_2) = \min_{\mathbf{x}\in\mathbb{R}^2} D_H(\gamma_1, \gamma_2 + \mathbf{x}), \tag{1.10}$$

i.e., we translate the curves until their Hausdorff distance is minimized.

Our second principal result is:

**Theorem 2.** *Consider the two-dimensional Bernoulli bond percolation model on the square lattice with $p > p_c$, and condition on the event $N \leqq |C(0)| < \infty$. Then there exists a function $\eta(N) = \eta(N; p)$, with $\eta(N) \downarrow 0$ monotonically as $N \uparrow \infty$, such that, with conditional probability tending rapidly to one with $N$, there is an occupied circuit of dual bonds, $\gamma$, encircling the origin satisfying*

$$\sqrt{\frac{1}{N}}\ \rho(\sqrt{N/P_\infty}\gamma_w, \gamma) \leqq \eta(N).$$

In the statement of Theorem 2, and in later theorems, we use the term " tending rapidly to one with $N$" to mean tending to one faster than any inverse power of $N$.

For future reference, it is worth noting that we also derived a variant of

[2] We will often make no distinction between a curve $\gamma$ and its image

Theorem 2 in which the final inequality is replaced by

$$\rho\left(\gamma_w, \frac{\gamma}{\sqrt{\mathscr{A}(\gamma)}}\right) \leqq \tilde{\eta}(N), \tag{1.11}$$

where $\mathscr{A}(\gamma)$ denotes the area enclosed by $\gamma$, and $\tilde{\eta}(N)$ is another function which tends monotonically to zero with $N$.

In order to prove Theorem 2, we first had to establish a stability result for the Wulff variational problem, which may be of independent interest. Roughly speaking, the stability result says that if a curve, $\gamma$, enclosing unit area, differs from the minimizer by an amount $\eta > 0$ (i.e. if $\rho(\gamma_w, \gamma) \geqq \eta > 0$), then the value of the surface energy functional for $\gamma$ exceeds the minimum by some strictly positive function $f(\eta)$. See Theorem 5.2 for more details. It is worth noting that such a stability result will fail for dimension $d > 2$ due to the existence of arbitrarily thin filaments. Thus an extension of the microscopic Wulff construction to higher dimensions may require a new formulation of the problem.

In Theorem 2, we achieved a cluster of large size by directly conditioning on its existence—i.e. by conditioning on the event $N \leqq |C(0)| < \infty$. While this is perfectly reasonable from the viewpoint of percolation theory, it is rather unnatural from the viewpoint of statistical mechanics. In the latter case, one would restrict to a microcanonical ensemble, as in the Minlos–Sinai droplet theorem [MS] and in the Wulff construction theorem of Dobrushin, Kotecky and Shlosman [DKS]. In percolation, the analogue of restricting to the microcanonical ensemble at the "wrong magnetization" is to condition on the event that the volume fraction of the infinite cluster which intersects a large finite box has the "wrong density," e.g. this volume fraction is strictly less than $P_\infty(p)$. In order to compensate for this, the configurations must have more sites in finite clusters within the box than would be the case in the unconditioned measure. Roughly speaking, our result states that under this conditioning, typical configurations have "most" of these excess sites in one connected component: a single droplet. Moreover, this droplet behaves like the cluster in Theorem 2: with probability tending to one in the size of the system, the droplet—when appropriately scaled—approaches the Wulff shape.

More precisely, let $C_\infty \equiv C_\infty(\omega)$ denote the sites of the (w.p. 1 unique) infinite cluster in configuration $\omega$. Let $\Lambda_L \subset \mathbb{Z}^2$ denote the square of side $L$ centered at the origin. We will condition on the event

$$F_L(\lambda) = \left\{ \omega \;\middle|\; \frac{|C_\infty \cap \Lambda_L|}{|\Lambda_L|} \leqq (1-\lambda) P_\infty(p) \right\} \tag{1.12}$$

with $0 < \lambda < [\mathrm{diam}\,(\gamma_w)]^{-2}$, where $\mathrm{diam}\,(\gamma_w)$ is the maximum distance between any pair of points in $\gamma_w$. Here our $\lambda$ is analogous to the $\alpha$ in the Ising problem discussed above. The occupied bonds of $\omega$ partition the sites of $\mathbb{Z}^2 \setminus C_\infty(\omega)$ into an infinite number of connected components: the finite clusters. Those finite clusters which are sufficiently large (presumably of linear dimension exceeding the correlation length) qualify as "droplets of the wrong phase." Our single-droplet result is:

**Theorem 3.** *Consider the two-dimensional Bernoulli bond percolation model on the square lattice with $p > p_c$, and condition on the event $F_L(\lambda)$ with $0 < \lambda <$*

$[\mathrm{diam}(\gamma_w)]^{-2}$. *Then there exist functions* $\phi_L(\lambda) = \phi_L(\lambda; p)$, $\zeta_L(\lambda) = \zeta_L(\lambda; p)$ *and* $\mu_L(\lambda) = \mu_L(\lambda; p)$ *tending monotonically to zero as* $L \uparrow \infty$, *such that, with conditional probability tending rapidly to one with* $L$, *there is an occupied circuit of dual bonds,* $\gamma$, *satisfying*

(*a*) $\mathscr{A}(\gamma) \geqq [1 - \phi_L(\lambda)]\lambda |\Lambda_L|$,

(*b*) $\rho\left(\gamma_w, \dfrac{\gamma}{\sqrt{\mathscr{A}(\gamma)}}\right) \leqq \zeta_L(\lambda)$,

(*c*) $\mathrm{Int}(\gamma)$ *contains a connected cluster of size exceeding* $P_\infty(p)[1 - \mu_L(\lambda)]\lambda |\Lambda_L|$.

It is worth remarking that in order to prove Theorem 3, we first required an estimate on the probability of $F_L(\lambda)$—which is, of course, just a large deviations estimate for the random variable $|C_\infty \cap \Lambda_L|/|\Lambda_L|$. For $0 < \lambda < [\mathrm{diam}(\gamma_w)]^{-2}$, our estimate is:

$$\lim_{L \to \infty} \frac{1}{L} \log \mathbf{P}_p[F_L(\lambda)] = -\sqrt{\lambda}\sigma(p)\omega(p). \tag{1.13}$$

See Theorem 6.1 for more details.

This paper is organized as follows. In Sect. 2, we set our notation and review a few basic results in percolation theory. Section 3 is devoted to geometrical preliminaries: There we define the angle-dependent surface tension for percolation, and use it to construct a norm on $\mathbb{R}^2$. This norm is then used to formulate the Wulff variational problem for percolation. Theorems 1, 2 and 3 are the contents of Sects. 4, 5 and 6, respectively. Some of the more tedious aspects of our stability result for the two-dimensional Wulff problem are relegated to appendices.

## 2. Notation, Definitions and Preliminaries

In this section, we will set our notation and review some of the basic results in percolation.

We will consider the square site lattice $\mathbb{Z}^2$, the dual square site lattice $(\mathbb{Z}^*)^2 \equiv \mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$, and the plane $\mathbb{R}^2$. Points of $(\mathbb{Z}^*)^2$ will sometimes be denoted with a $*$ superscript, e.g. for $x = (x_1, x_2) \in \mathbb{Z}^2$, $x^* \equiv (x_1 + \frac{1}{2}, x_2 + \frac{1}{2}) \in (\mathbb{Z}^*)^2$; however, for notational convenience, we will often omit the $*$. On $\mathbb{Z}^2$ and $(\mathbb{Z}^*)^2$, we will use the lattice $L^1$, $L^2$ and $L^\infty$ norms: i.e. for $x = (x_1, x_2) \in \mathbb{Z}^2$,

$$|x|_1 = |x_1| + |x_2|, \tag{2.1a}$$

$$|x|_2 = \sqrt{x_1^2 + x_2^2}, \tag{2.1b}$$

$$|x|_\infty = \max\{|x_1|, |x_2|\}. \tag{2.1c}$$

On $\mathbb{R}^2$, we will generally use the Euclidean (i.e. $L^2$) norm, in addition to another norm to be introduced in Sect. 3. For $S \subset \mathbb{Z}^2$, $|S|$ will denote the cardinality (i.e. number of points) of $S$, while for $R \subset \mathbb{R}^2$, $|R|$ will denote the Euclidean area of $R$.

For $\gamma: [0, T] \to \mathbb{R}^2$ a continuous curve in $\mathbb{R}^2$, let $\mathscr{L}(\gamma)$ denote the (Euclidean) *arclength* of $\gamma$. Recall that the curve $\gamma$ is said to be *rectifiable* if $\mathscr{L}(\gamma) < \infty$. Let $\mathscr{J}$ denote the set of rectifiable Jordan curves in $\mathbb{R}^2$. For $\gamma \in \mathscr{J}$, let $\mathrm{Int}(\gamma) \subset \mathbb{R}^2$ denote

the interior of $\gamma$, and let

$$\mathscr{A}(\gamma) \equiv |\mathrm{Int}(\gamma)| \tag{2.2}$$

denote the (Euclidean) are enclosed by $\gamma$.

The set of all *bonds* between nearest-neighbor sites of $\mathbb{Z}^2$, i.e. pairs $x, y \in \mathbb{Z}^2$ with $|x-y|_1 = 1$, will be denoted by $\mathbb{B}_2$. A *path* in $\mathbb{B}_2$ is a sequence (finite or infinite) of bonds $b_1, b_2, \ldots$, with no repetitions, such that $b_n$ and $b_{n+1}$ have a common endpoint. A *contour* in $\mathbb{B}_2$ is a finite closed path: $b_1, b_2, \ldots, b_N$ such that the initial endpoint of $b_1$ is the final endpoint of $b_N$. Two paths are said to be *disjoint* if they have no bonds in common. For $x, y \in \mathbb{Z}^2$, a set $S \subset \mathbb{Z}^2$ is said to *separate* $x$ from $y$ if every path from $x$ to $y$ includes at least one bond with an endpoint in $S$. Similarly, the set of all bonds between nearest-neighbor sites of $(\mathbb{Z}^*)^2$ will be denoted by $\mathbb{B}_2^*$; we can, of course, define paths, contours, etc. in $\mathbb{B}_2^*$.

The nearest-neighbor Bernoulli bond percolation model on the square lattice at density $p$ is defined by independently choosing each bond of $\mathbb{B}_2$ to be *occupied* with probability $p$ or *vacant* with probability $1-p$. We denote by $\mathbf{P}_p$ the product measure on $\Omega$ at density $p$, and by $\mathbf{E}_p$ the expectation with respect to $\mathbf{P}_p$. We will often suppress the subscript $p$ in $\mathbf{P}_p$ and $\mathbf{E}_p$. For $S_1, S_2 \subset \mathbb{Z}^2$, we say that $S_1$ is *connected* to $S_2$ in the configuration $\omega$ if there is a path of occupied bonds in $\omega$ from a site in $S_1$ to a site in $S_2$. If such a path occurs within a set of bonds $B \subset \mathbb{B}_2$, we say that $S_1$ is connected to $S_2$ in $B$. The maximal connected subsets of $\omega$ are called the (occupied) *clusters* of $\omega$. Note that, as defined, these clusters are sets of sites in $\mathbb{Z}^2$, not bonds in $\mathbb{B}_2$. For $x \in \mathbb{Z}^2$, we denote by $C(x) = C(x; \omega)$ the cluster containing $x$ in $\omega$. If $x$ is not connected to any other site by occupied bonds, then $C(x) = \{x\}$. Consistent with the notation defined above, $|C(x)|$ denotes the cardinality of $C(x)$.

In $d$ dimensions, bond percolation at density $p$ is dual to $(d-1)$-cell percolation at density $1-p$. In two dimensions, the model is self-dual: if a given $b \in \mathbb{B}_2$ is vacant (occupied), then we can view the unique $b^* \in \mathbb{B}_2^*$ which intersects $b$ as occupied (vacant). It is often convenient to view a given configuration $\omega$ not as a collection of occupied and vacant bonds, but rather as a collection of occupied bonds and occupied dual bonds. Then it is clear that each finite cluster of $\omega$ is surrounded by an innermost contour of occupied dual bonds in $\omega$. Henceforth, unless otherwise specified, when we speak of clusters of $\omega$, we will mean clusters connected by occupied bonds, and when we speak of contours in $\omega$, we will mean occupied dual bond contours.

Next we review a few basic results concerning bond percolation on the square lattice. It is known ([BH], [Har], [K2]) that the model has a phase transition at density $p_c = 1/2$, below which the occupied clusters are finite w.p. 1 and above which there is a unique infinite occupied cluster. Let us denote the infinite cluster by $C_\infty = C_\infty(\omega) \subset \mathbb{Z}^2$. The order parameter for the transition is the *infinite cluster density*:

$$P_\infty(p) = \mathbf{P}_p(0 \in C_\infty). \tag{2.3}$$

It is known that (in $d = 2$), $P_\infty(p) \downarrow 0$ continuously as $p \downarrow p_c$ [R].

The analogue of the two-point correlation in a spin system is the connectivity

function: For $x, y \in \mathbb{Z}^2$, the connectivity event is

$$t_{x,y} = \{\omega | x \text{ is connected to } y \text{ by occupied bonds}\} \tag{2.4a}$$

and the *connectivity function* is

$$\tau_{x,y}(p) = \mathbf{P}_p(t_{x,y}). \tag{2.4b}$$

Similarly, for $x^*, y^* \in (\mathbb{Z}^*)^2$, the dual connectivity event and dual connectivity function are given by

$$t^*_{x^*,y^*} = \{\omega | x^* \text{ is connected to } y^* \text{ by occupied dual bonds}\}, \tag{2.5a}$$

$$\tau^*_{x^*,y^*}(p) = \mathbf{P}_p(t^*_{x^*,y^*}). \tag{2.5b}$$

By duality (in $d = 2$), $\tau^*_{x^*,y^*}(p) = \tau_{x,y}(1-p)$. For $p < p_c$, the *correlation length*, $\xi(p)$, is defined by the behavior of the on-axis connectivity function:

$$\lim_{n\to\infty} \frac{1}{n} \log \tau_{0,(n,0)}(p) = -\frac{1}{\xi(p)}; \tag{2.6}$$

furthermore, even for the off-axis connectivity function, $\xi(p)$ provides the following a priori bound:

$$\tau_{x,y}(p) \leqq e^{-[1/\xi(p)]|x-y|_\infty}. \tag{2.7}$$

It is known that $\xi(p) < \infty$ for $p < p_c$ [K2], and that $\xi(p) \uparrow \infty$ continuously as $p \uparrow p_c$. (For a general review of t e properties of $\tau$, see e.g. [CC1] or [G]). A final useful fact concerning the connectivity function is that it obeys the *Hammersley–Simon inequality* ([H], [S]): For $x, y \in \mathbb{Z}^2$, let $S \subset \mathbb{Z}^2$ be a surface which separates $x$ from $y$. Then

$$\tau_{x,y}(p) \leqq \sum_{z\in S} \tau_{x,z}(p)\tau_{z,y}(p). \tag{2.8}$$

The behavior of the finite cluster distribution for percolation has already been discussed in some detail in the introduction. Here, let us just set some additional notation. We define

$$P_N(p) = \mathbf{P}_p(|C(0)| = N), \tag{2.9a}$$

$$P_{\leqq N}(p) = \sum_{n\leqq N} P_n(p), \tag{2.9b}$$

$$P_{\geqq N}(p) = \sum_{N\leqq n<\infty} P_n(p). \tag{2.9c}$$

Although Theorem 1 of the Introduction is stated in terms of $P_N(p)$, the quantity with which we will be working most often is the tail of the finite cluster distribution, $P_{\geqq N}(p)$.

Finally, we review a few general notions and inequalities. We denote the indicator function of an event $A \subset \Omega$ by $\mathbb{1}_A$:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}. \tag{2.10}$$

*Definition 2.1.* Let $\omega_1, \omega_2 \in \Omega$. There is a natural partial order on $\Omega$ defined by the relation $\omega_1 \prec \omega_2$ if all occupied bonds in $\omega_1$ are also occupied in $\omega_2$. An event $A \subset \Omega$ is said to be *positive* or *increasing* (respectively, *negative* or *decreasing*) if $\mathbb{1}_A$ is nondecreasing (respectively, nonincreasing) with respect to this partial order.

The *Harris-FKG inequality* ([Har], [FKG]) says that if $A_1, A_2 \subset \Omega$ are both positive (or both negative) events, then

$$\mathbf{P}_p(A_1 \cap A_2) \geqq \mathbf{P}_p(A_1)\mathbf{P}_p(A_2). \tag{2.11}$$

*Definition 2.2.* Let $\omega \in A \subset \Omega$ and $B \subset \mathbb{B}_2$. The event $A$ is said to *occur on the set* $B$ in configuration $\omega$ if $A$ occurs in $\omega$ restricted to $B$, regardless of the configuration in $\mathbb{B}_2 \setminus B$; more precisely, we define

$$A|_B = \{\omega \in A \mid \tilde{\omega} \in A \text{ for all } \tilde{\omega} \text{ such that } \tilde{\omega} = \omega \text{ on all bonds in } B\}. \tag{2.12}$$

Two events $A_1, A_2 \subset \Omega$ are said to *occur disjointly*, denoted by $A_1 \circ A_2$, if there are (bond) disjoint sets on which they occur:

$$A_1 \circ A_2 = \{\omega \in A_1 \cap A_2 \mid \exists B_1, B_2 \subset \mathbb{B}_2, B_1 \cap B_2 = \varnothing, \omega \in A_1|_{B_1} \cap A_2|_{B_2}\}. \tag{2.13}$$

Similarly, three or more events are said to occur disjointly if they are pairwise mutually disjoint, e.g.

$$A_1 \circ A_2 \circ A_3 = (A_1 \circ A_2) \cap (A_2 \circ A_3) \cap (A_1 \circ A_3). \tag{2.14}$$

The *van den Berg–Kesten inequality* [BK] says that if $A_1, A_2 \subset \Omega$ are both positive (or both negative) events, then

$$\mathbf{P}_p(A_1 \circ A_2) \leqq \mathbf{P}_p(A_1)\mathbf{P}_p(A_2). \tag{2.15}$$

The inequality (2.15) was extended to the case of the $A_i$ being intersections of positive and negative events by van den Berg and Fiebig [BF]. By induction, an analogue of (2.15) clearly holds for the disjoint union of three or more (say) negative events.

## 3. Geometrical Properties

In this section, we define the surface tension and establish some geometrical properties of its angular dependence. We then formulate the Wulff variational problem for percolation.

We begin by defining the zero-angle surface tension, $\sigma(p)$. Since we are in two dimensions, we need only consider the (on-axis) dual connectivity function $\tau^*_{0^*,(n,0)^*}(p)$, as defined in Eq. (2.5). The following proposition is an immediate consequence of the duality relation $\tau^*_{x^*,y^*}(p) = \tau_{x,y}(1-p)$ and well-known properties of the ordinary connectivity function (cf. Eq. (2.6) and (2.7)).

**Proposition 3.1.** *For the two-dimensional Bernoulli bond percolation model, the limit*

$$\sigma(p) = -\lim_{n \to \infty} \frac{1}{n} \log \tau^*_{0^*,(n,0)^*}(p)$$

*exists with* $\sigma(p) > 0$ *for* $p > p_c$ *and* $\sigma(p) \downarrow 0$ *as* $p \downarrow p_c$. *Furthermore, for finite* $n$

$$\tau^*_{0^*,(n,0)^*}(p) \leqq e^{-\sigma(p)n}.$$

The limiting constant in Proposition 3.1 is called the surface tension by analogy to similar quantities in spin systems, e.g. the Ising ferromagnet. There one considers spin configurations in a cube of scale $L$ in which the top and bottom regions correspond to distinct pure phases separated by an interface. The Gibbsian weight of such configurations—relative to the total weight of all allowed configurations—has the purported scaling $e^{-\sigma L^{d-1}}$. The constant $\sigma$ has the interpretation of an excess (surface) free energy or surface tension. In percolation, rather than studying the Gibbsian weight of configurations with an interface, one considers the probability that the top surface of a cube of scale $L$ is disconnected from the bottom. This has the dual representation of an interface separating the top and bottom of the cube. In $d=2$, it is readily established that this probability has the desired scaling (modulo power law corrections in front of the exponential factor), with $\sigma$ given as in Proposition 3.1. By duality, it is clear that $\sigma(p)=[\xi(1-p)]^{-1}$, where $\xi(p)$ is the correlation length (cf. Eq. (2.6)). What is not so obvious, but has nevertheless been established [CCGKS], is that $\sigma(p)=\frac{1}{2}[\xi'(p)]^{-1}$, where $\xi'(p)$ is the *above threshold* correlation length, i.e. the decay rate of a truncated connectivity function. The higher-dimensional problems are not on quite this sound a footing; for further discussion of the zero-angle surface tension in $d \geqq 3$, see [ACCFR], [CC2].

An angle-dependent surface tension can be defined by considering the behavior of the off-axis dual correlation function $\tau^*_{0^*,x^*}(p)$. Although such a surface tension (or the correlation length) has been discussed previously, both in the context of percolation [CC1] and spin systems [CCS], here we will treat the problem from a somewhat more geometrical perspective.

**Proposition 3.2.** *Consider the two-dimensional Bernoulli bond percolation model at density p.*

A) *Let* $x \in \mathbb{Q}^2$ (*where* $\mathbb{Q}$ *denotes the rationals*) *and let k be any integer for which* $kx \in \mathbb{Z}^2$. *Then the limit*

$$g(x;p) \equiv -\lim_{n\to\infty} \frac{1}{nk\sigma(p)} \log \tau^*_{0^*,nkx^*}(p)$$

*exists and is independent of k. Furthermore,* $\forall x \in \mathbb{Z}^2$, $g(x;p)$ *provides the a priori bound*:

$$\tau^*_{0^*,x^*}(p) \leqq e^{-\sigma(p)g(x;p)}.$$

B) *For each p, the function* $g(x) \equiv g(x;p)$ *has the following properties*:

(i) *Scaling* (*or homogeneity*): *For each* $\lambda \in \mathbb{Q}$,

$$g(\lambda x) = |\lambda| g(x).$$

(ii) *Symmetry*: $g(x)$ *is invariant under interchange of the components of x or sign reversal of either component of x; that is, if* $x=(x_1,x_2) \in \mathbb{Q}^2$,

$$g(x) = g(-x_1,x_2) = g(x_1,-x_2) = g(-x_1,-x_2) = g(x_2,x_1).$$

(iii) *Convexity*: *For each* $x, y \in \mathbb{Q}^2$ *and each* $\lambda \in \mathbb{Q}$, *with* $0 \leqq \lambda \leqq 1$,

$$g(\lambda x + (1-\lambda)y) \leqq \lambda g(x) + (1-\lambda) g(y).$$

*Proof.* Let $x \in \mathbb{Q}^2$, and denote by $k$ (say) the smallest integer for which $kx \in \mathbb{Z}^2$.

Then existence of the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \tau^*_{0^*,nkx^*}(p) \tag{3.1}$$

follows from the (log) subadditive inequality

$$\tau^*_{0^*,(n_1+n_2)kx^*}(p) \geqq \tau^*_{0^*,n_1kx^*}(p)\ \ \tau^*_{0^*,n_2kx^*}(p). \tag{3.2}$$

Denoting the limit in (3.1) by $k\sigma(p)g(x;p)$, (3.2) directly implies the *a priori* bound. Furthermore, the scaling in (3.2) shows that $g(x;p)$ is independent of $k$, and establishes property (i). Property (ii) is a $\mathbb{Z}^2$ lattice symmetry which $g(x)$ clearly inherits. Convexity is established by observing that for $x, y \in \mathbb{Q}^2$, $\lambda \in \mathbb{Q}$, with $0 \leqq \lambda \leqq 1$, and $m$ any integer for which $\lambda mx \in \mathbb{Z}^2$ and $(1-\lambda)my \in \mathbb{Z}^2$, we have

$$\begin{aligned} \tau^*_{0^*,m(\lambda x+(1-\lambda)y)^*} &\geqq \tau^*_{0^*,m\lambda x^*}\tau^*_{m\lambda x^*,m(\lambda x+(1-\lambda)y)^*} \\ &= \tau^*_{0^*,m\lambda x^*}\tau^*_{0^*,m(1-\lambda)y^*}, \end{aligned} \tag{3.3}$$

which, after logs and limits, is the desired result. ■

Collecting the above properties, we have:

**Corollary.** *The function $g(x)$ may be extended to a convex, continuous function on $\mathbb{R}^2$, where it defines a norm equivalent to the Euclidean norm.*

*Proof.* Convexity implies that $g(x)$ is continuous on $\mathbb{Q}^2$; hence it may be extended to a continuous function on all of $\mathbb{R}^2$, where it enjoys properties (i), (ii) and (iii). Obviously, $g(0) = 0$. Furthermore $g(1,0) = g(0,1)$ and convexity imply that if $x \neq 0$, then $g(x) \neq 0$. Finally, convexity at $\lambda = \frac{1}{2}$ and the scaling property imply the triangle inequality. Hence, $g$ defines a norm on $\mathbb{R}^2$. Finally observe that if $x \in \mathbb{R}^2$, then

$$\frac{1}{\sqrt{2}}|x|_2 \leqq |x|_\infty \leqq g(x) \leqq |x|_1 \leqq \sqrt{2}|x|_2, \tag{3.4}$$

which demonstrates the equivalence of $g$ to the Euclidean norm. ■

*Remark.* The norm $g$ is, of course, closely related to the angle-dependent surface tension. Assume we have an interface oriented at an angle $\theta$ to the $\hat{e}_x$-axis. It is customary to track the direction dependence of the surface tension in terms of the outward normal $\mathbf{n} = \hat{e}_\theta$ to the interface: $\sigma(\mathbf{n};p) = \sigma(\hat{e}_\theta;p)$. In two dimensions, since there is only a single tangent vector, $\mathbf{t} = \hat{e}_r$, to the interface, it is just as convenient to track the direction dependence of the surface tension as a function of $\hat{e}_r$. The relationship between our $g$ and the conventional direction-dependent surface tension is

$$\sigma(p)g(\hat{e}_r;p) = \sigma(\hat{e}_\theta;p), \tag{3.5}$$

where, as usual, $\sigma(p)$ is the surface tension for an interface oriented at angle $\theta = 0$ to the $\hat{e}_x$-axis: that is, $\sigma(p) = \sigma(\hat{e}_\theta = \hat{e}_y;p)$, so that $g(\hat{e}_r = \hat{e}_x;p) = 1$. It is also conventional to define $\sigma(\hat{e}_\theta;p)$ only as a function of the normal vector $\mathbf{n} = \hat{e}_\theta$, whereas here we define $g$ for all $x \in \mathbb{R}^2$. If $x = r\hat{e}_r = r(\cos\theta, \sin\theta)$, then

$$g(x;p) = rg(\hat{e}_r;p) = |x|_2 g(\hat{e}_r;p). \tag{3.6}$$

Thus the ratio of $g$ to the Euclidean norm is a measure of the angle-dependence of the surface tension.

Now let us formulate the Wulff variational problem for percolation. Given a continuous curve $\gamma:[0,T]\to\mathbb{R}^2$, it is possible to define the $g$-length of $\gamma$ by analogy to the (Euclidean) arclength $\mathscr{L}(\gamma)$: Let $\mathscr{P}=(t_0,\ldots,t_N)$, $0=t_0\leqq t_1\leqq\cdots\leqq t_N=T$, denote a partition of $[0,T]$, and let $\gamma_{\mathscr{P}}=(\gamma(t_0),\ldots,\gamma(t_N))$ be the corresponding polygonal approximation to $\gamma$. The $g$-length of $\gamma$ at density $p$ is given by

$$g_p(\gamma)=\sup_{\mathscr{P}} g_p(\gamma_{\mathscr{P}})\equiv\sup_{\mathscr{P}}\sum_{t_n\in\mathscr{P}} g(\gamma(t_{n+1})-\gamma(t_n);p). \tag{3.7}$$

We will sometimes omit the subscript $p$ in our notation for the $g$-length. By equivalence of the norms, it is clear that $g_p(\gamma)<\infty$ if and only if $\mathscr{L}(\gamma)<\infty$ (i.e. $\gamma$ rectifiable). We remind the reader that $\mathscr{J}$ is our notation for the set of rectifiable Jordan curves in $\mathbb{R}^2$, and $\mathscr{A}(\gamma)$ denotes the Euclidean area enclosed by $\gamma$.

Guided by the underlying lattice model, one is led to study the (continuum) variational problem of minimizing the surface energy subject to the constraint of enclosing unit area:

$$\omega(p)=\inf_{\gamma\in\mathscr{J}}\{g_p(\gamma)|\mathscr{A}(\gamma)=1\}. \tag{3.8}$$

We call $\omega(p)$ the Wulff constant at density $p$. It is easy to see that the constraint $\mathscr{A}(\gamma)=1$ may be replaced by the inequality $\mathscr{A}(\gamma)\geqq 1$ without changing the value of the functional, i.e.

$$\omega(p)=\inf_{\gamma\in\mathscr{J}}\{g_p(\gamma)|\mathscr{A}(\gamma)\geqq 1\}. \tag{3.8'}$$

We note that (3.8) or (3.8′) is a problem of the isoperimetric type, but that it is somewhat more difficult since the length and area of $\gamma$ are measured in different—though equivalent—norms. However, from this equivalence and the solution to the (Euclidean) isoperimetric problem, it follows immediately that

$$\sqrt{2\pi}\leqq\omega(p)\leqq 2\sqrt{2\pi}. \tag{3.9}$$

Indeed, using the sharper upper bound in Eq. (3.4), one easily obtains the improved estimate

$$\omega(p)\leqq 4. \tag{3.9'}$$

What is not so obvious is that there is a minimizer of (3.8) and that this minimizer is unique in $\mathscr{J}$. This a consequence of the Wulff construction [Wu], and general existence [T1] and uniqueness theorems [T2] due to Taylor, which we state here only for the case at hand.

**Theorem 3.3.** *Let $\mathscr{J}$ be the set of all rectifiable Jordan curves in $\mathbb{R}^2$, let $g_p(\gamma)$ denote the density-$p$ $g$-length of the curve $\gamma$ as defined by Eq. (3.7) and Proposition 3.2, let $\mathscr{A}(\gamma)$ denote the Euclidean area enclosed by $\gamma$, and let $\omega(p)$ be the Wulff constant as defined in Eq. (3.8). Then there is a unique curve $\gamma_w=\gamma_w(p)\in\mathscr{J}$ with $\mathscr{A}(\gamma_w)=1$ such that*

$$\omega(p)=g_p(\gamma_w).$$

*Furthermore, this curve is given by the Wulff construction,* i.e. $\gamma_w$ *is the boundary of the Wulff shape* (1.4) *for the surface tension* $\sigma(\hat{e}_\theta; p)$ *given in Eq.* (3.5).

It will turn out that in order to prove a microscopic Wulff construction for percolation, we will need somewhat more than Theorem 3.3: In addition, we will need a *stability result* which says that if $\gamma$ is a unit area curve at a Hausdorff distance $\eta > 0$ from $\gamma_w$, then $|g_p(\gamma) - g_p(\gamma_w)| \geqq f(\eta)$, where $f$ is a strictly positive function. In order to establish this, we will first prove that $\gamma_w$ is the unique minimizer of (3.8) over a somewhat larger class than just Jordan curves. See Sect. 5 for more details.

## 4. Bounds on the Finite Cluster Distribution

In this section, we establish our principal analytic result: (exponentially) optimal bounds on the finite cluster distribution for all $p > p_c$, as contained in Theorem 1 of the introduction. Although Theorem 1 is stated in terms of $P_N(p)$, here we will be working with the somewhat more natural tail of the finite cluster distribution, $P_{\geqq N}(p)$. In particular, we will first derive Theorem 1 for $P_{\geqq N}(p)$; then, at the end of this section, we will use a variant of the subadditivity argument in [KS] to extend the result to $P_N(p)$. The proof of Theorem 1 divides naturally into two parts: upper bounds and lower bounds. As in previous work on the finite cluster distribution, the proof of the upper bound is more difficult, although here we obtain sharper upper bounds than lower bounds.

*4.A. The Upper Bound.* The results of this subsection is:

**Theorem 1.A.** *In the two-dimensional Bernoulli bond percolation model on the square lattice, for every* $p > p_c$ *and all* $N \in \mathbb{Z}^+$ *sufficiently large*

$$P_{\geqq N}(p) \leqq \exp\{-[\omega(p)\sigma(p)/\sqrt{P_\infty(p)}]\sqrt{N}[1 - N^{-1/4}(\log N)^4]\},$$

*where* $P_\infty(p)$, $\sigma(p)$ *and* $\omega(p)$ *are the infinite cluster density, surface tension and Wulff constant, as defined in Eq.* (2.3), *Proposition 3.1 and Eq.* (3.8).

Our strategy for proving Theorem 1.A is as follows: If the origin is in a cluster of size at least $N$, then there must be some dual ring $\gamma$ surrounding the cluster. This cluster can be thought of as a "broken off" portion of the infinite cluster; it therefore should have density roughly $P_\infty(p)$ [ADS]. Now there are two possibilities: either $\gamma$ is large enough to enclose a cluster of the "correct density": $\mathscr{A}(\gamma) \gtrsim N/P_\infty$; or $\gamma$ encloses an "overdense" cluster: $\mathscr{A}(\gamma) \lesssim N/P_\infty$. In the former case, we get the desired bound simply by estimating the probability of the ring $\gamma$, omitting any estimate on the probability of the cluster itself. Our estimate on the ring probability is given in Lemma 4.1. In the latter case, we use a large deviations estimate (Lemma 4.2) to show that an "overdense" cluster is far too costly. We then combine these lemmas to give a proof of Theorem 1.A.

Before obtaining our ring estimate, let us introduce the notion of an *m-skeleton* of a (dual) lattice contour $\gamma$. To this end, let $U \subset \mathbb{R}^2$ denote the unit ball in the $g$-norm (cf. Proposition 3.2):

$$U = \{x \in \mathbb{R}^2 \mid g(x) \leqq 1\}, \tag{4.1}$$

and, for $m \in \mathbb{R}$, let $mU$ denote the $m$-ball:

$$mU = \{mx | x \in U\}. \tag{4.2}$$

Observe that, by the symmetry and convexity properties of $g$, these sets are convex and four-fold symmetric. Next, consider the lattice $m$-ball

$$\mathbb{U}_m = mU \cap \mathbb{Z}^2. \tag{4.3}$$

It follows from the convexity and symmetry of $U$ that $\mathbb{U}_m$ is a $\mathbb{Z}^2$-connected set. We denote the external boundary of this set by $\partial \mathbb{U}_m$:

$$\partial \mathbb{U}_m = \{x \in \mathbb{Z}^2 \setminus \mathbb{U}_m | \exists x' \in \mathbb{U}_m \text{ with } |x - x'|_1 = 1\}, \tag{4.4}$$

and the union of the set and its boundary by $\overline{\mathbb{U}_m}$:

$$\overline{\mathbb{U}_m} = \mathbb{U}_m \cup \partial \mathbb{U}_m. \tag{4.5}$$

Finally, for $x^* \in (\mathbb{Z}^*)^2$, we denote the translation of these sets by the lattice vector $x^*$ by $\mathbb{U}_m(x^*)$ and $\partial \mathbb{U}_m(x^*)$:

$$\mathbb{U}_m(x^*) = T^{x^*} \mathbb{U}_m, \tag{4.6a}$$

$$\partial \mathbb{U}_m(x^*) = T^{x^*} \partial \mathbb{U}_m. \tag{4.6b}$$

It is worth observing that if $x^* \in \partial \mathbb{U}_m(0^*)$, then by Proposition 3.2A,

$$\tau^*_{0^*, x^*}(p) \leqq e^{-\sigma(p)m}. \tag{4.7}$$

Now suppose that $\gamma : [0, T] \to \mathbb{R}^2$ is a self-avoiding (dual) lattice loop encircling the origin, i.e. that the image of $\gamma$ is a contour in $\mathbb{B}^*_2$ enclosing $0^*$. Let $t_0 = 0$ and, for definiteness, let us take $\gamma(0) = s_0 \in (\mathbb{Z}^*)^2$ to be the lowest site of $\gamma$ on the positive $y^*$-axis. For $n \geqq 1$, we define

$$t_{n+1} = \inf\{t > t_n | \gamma(t) \in \partial \mathbb{U}_m(s_n)\}, \tag{4.8a}$$

$$s_{n+1} = \gamma(t_{n+1}) \tag{4.8b}$$

provided that $t_{n+1}$ exists. Thus $s_{n+1}$ is the earliest site of $(\mathbb{Z}^*)^2$ on $\gamma$ after $s_n$ which is a $g$-distance at least $m$ units from $s_n$. We let $J = J(m; \gamma)$ be the largest $n$ such that $t_n$, and hence $s_n$, exists. In other words, the sequence $(s_0, \ldots, s_J)$ exhausts the curve $\gamma$ and $g(s_J - s_0) < m$. We call the sequence

$$S_m(\gamma) = (s_0, s_1, \ldots, s_J, s_0) \tag{4.9}$$

the (lattice) *m-skeleton* of $\gamma$. It is clear that each contour $\gamma$ has a unique $m$-skeleton. On the other hand, a given sequence of (dual) points $(s_k)$ represents the $m$-skeleton of many distinct contours. We let

$$\Gamma^m_{(s_k)} = \{\gamma | S_m(\gamma) = (s_k)\} \tag{4.10}$$

denote the collection of all contours with $m$-skeleton $(s_k)$.

Our ring estimate is as follows:

**Lemma 4.1.** *Let $A_0 \in \mathbb{Z}^+$ and consider the event*

$$\mathscr{R}(A_0) = \{\omega | \exists \text{ an occupied dual ring } \gamma \text{ surrounding the origin with } A(\gamma) \geqq A_0\}.$$

*Then for all $p > p_c$ and $A_0$ sufficiently large, $\exists c = c(p) > 0$ such that*

$$\mathbf{P}_p[\mathscr{R}(A_0)] \leqq \exp\left[ -\omega(p)\sigma(p)\sqrt{A_0}\left[ 1 - \frac{\operatorname{clog} A_0}{A_0^{1/4}} \right] \right].$$

*Proof*. We begin with a rather straightforward coarse-grained Peierls estimate using the notion of an $m$-skeleton defined above. The subtlety occurs in relating the result of the Peierls estimate to the area $A_0$.

First, let $\tilde{\Gamma}^m_{(s_k)}$ denote the event that there is a dual contour in $\Gamma^m_{(s_k)}$ (cf. Eq. (4.10)):

$$\tilde{\Gamma}^m_{(s_k)} = \{\omega | \exists \text{ an occupied dual ring } \gamma \in \Gamma^m_{(s_k)}\}. \tag{4.11}$$

Obviously, the event $\tilde{\Gamma}^m_{(s_k)}$ is contained in the disjoint union of the successive events $t^*_{s_k, s_{k+1}}$:

$$\tilde{\Gamma}^m_{(s_k)} \subset t^*_{s_1,s_2} \circ t^*_{s_2,s_3} \circ \cdots \circ t^*_{s_J,s_1}. \tag{4.12}$$

Hence by the van den Berg–Kesten inequality and (4.7), we have

$$\mathbf{P}_p[\tilde{\Gamma}^m_{(s_k)}] \leqq e^{-\sigma(p)(J+1)m}. \tag{4.13}$$

Now given that $\mathscr{R}(A_0)$ has occurred, there must be dual contour $\gamma$ surrounding the origin with an $m$-skeleton $(s_k)$ of $J \geqq J_{\min}(A_0)$ points. For fixed $J$, let $\mathscr{S}(J;m)$ denote the collection of all sequences $(s_k)$ of $J$ points which form the $m$-skeleton of some curve surrounding the origin. Let us determine the size $|\mathscr{S}(J;m)|$ of $\mathscr{S}(J;m)$. First, it follows from the definition of an $m$-skeleton that given a particular point $s_k$, the number of possible "target points" is $|\partial \mathbb{U}_m|$, which is easily bounded above by $\kappa_1 m$ with e.g. $\kappa_1 < 8\pi$. Hence, for fixed initial point $s_0$, the number of $m$-skeletons is less than $(\kappa_1 m)^J$. Furthermore, it is clear that $s_0$ cannot have $y^*$-coordinate exceeding $\frac{1}{2}\sqrt{2}mJ < mJ$. Thus

$$|\mathscr{S}(J;m)| \leqq mJ(\kappa_1 m)^J. \tag{4.14}$$

We have

$$\mathscr{R}(A_0) = \bigcup_{J \geqq J_{\min}(A_0)} \bigcup_{(s_k) \in \mathscr{S}(J;m)} \tilde{\Gamma}^m_{(s_k)}. \tag{4.15}$$

Thus, by subadditivity of the measure and the bounds (4.13) and (4.14):

$$\begin{aligned} \mathbf{P}_p[\mathscr{R}(A_0)] &\leqq \sum_{J \geqq J_{\min}(A_0)} \sum_{(s_k) \in \mathscr{S}(J;m)} \mathbf{P}_p[\tilde{\Gamma}^m_{(s_k)}] \\ &\leqq \sum_{J \geqq J_{\min}(A_0)} mJ(\kappa_1 m)^J e^{-\sigma m J} \\ &\leqq \kappa_2 m^{J_{\min}} e^{-\sigma m J_{\min}} \end{aligned} \tag{4.16}$$

for $m$ sufficiently large. Here $\kappa_2 = \kappa_2(p) < \infty$.

Now we must determine $J_{\min}(A_0)$. To this end, let $(s_k)$ be the $m$-skeleton of some dual contour which, if occupied, would contribute to the event $\mathscr{R}(A_0)$. Let $\gamma_{(s_k)}$ denote the unique (continuum) polygonal curve with the sequence of vertices $(s_k)$, in order. In general, $\gamma_{(s_k)}$ will not be self-avoiding, although it will be the union of a finite number of self-avoiding polygons, and possibly also degenerate polygons. Let us denote by $\mathscr{A}(\gamma_{(s_k)})$ the area of the union of these polygons. It follows from

the variational principle (cf. Eq. (3.8)) and concavity of the square root that

$$g(\gamma_{(s_k)}) \geqq \omega(p)\sqrt{\mathscr{A}(\gamma_{(s_k)})}. \tag{4.17}$$

Roughly speaking, we would like to bound $g(\gamma_{(s_k)})$ from above by $Jm$, and bound $\mathscr{A}(\gamma_{(s_k)})$ from below by $A_0$. However, technically, neither of these bounds is quite correct.

First, $Jm$ is actually a lower bound on $g(\gamma_{(s_k)})$ since the lattice points of $\partial\mathbb{U}_m$ are a $g$-distance greater than $m$ from the origin. However, this is easily rectified by noting that $\partial\mathbb{U}_m \subset \mathbb{U}_{m+1}$. Then, taking into account that $s_J$ may be separated by as much as $m+1$ from $s_0$, we have

$$g(\gamma_{(s_k)}) \leqq (m+1)(J+1). \tag{4.18}$$

Next, we note that the area of any curve $\gamma \in \Gamma^m_{(s_k)}$ can differ from $\mathscr{A}(\gamma_{(s_k)})$ by at most the area of $(J+1)$ $m$-balls:

$$\mathscr{A}(\gamma_{(s_k)}) + (J+1)|\mathbb{U}_m| \geqq \mathscr{A}(\gamma). \tag{4.19}$$

Since the $g$-norm is equivalent to the Euclidean norm, we may write $|\mathbb{U}_m| \leqq \kappa_3(p)m^2$, where $0 < \kappa_3(p) < \infty$. Then, if $\gamma$ "contributes" to the event $\mathscr{R}(A_0)$, we have $\mathscr{A}(\gamma) \geqq A_0$, so that (4.19) implies

$$\mathscr{A}(\gamma_{(s_k)}) \geqq A_0 - \kappa_3(J+1)m^2. \tag{4.20}$$

Thus, by Eqs. (4.17), (4.18) and (4.20), $J_{\min}$ satisfies

$$(m+1)(J_{\min}+1) \geqq \omega(p)\sqrt{A_0 - \kappa_3(J_{\min}+1)m^2}. \tag{4.21}$$

Using $\sqrt{1-x} \geqq 1-x$ for $0 < x < 1$, (4.21) implies

$$J_{\min} \geqq \omega(p)\frac{\sqrt{A_0}}{m+1} - \kappa_4 \tag{4.22}$$

for some $0 < \kappa_4(p) < \infty$. This, together with (4.16), gives us a bound on $\mathbf{P}_p[\mathscr{R}(A_0)]$ in terms of $m$. Now, however, we can choose $m$ to optimize (4.16) subject to the constraint (4.22). A nearly optimal solution occurs when we take $m = A_0^{1/4}$, from which the statement of the lemma follows. ■

The previous lemma provides our basic estimate for the case in which the ring encloses area $\mathscr{A}(\gamma) \gtrsim N/P_\infty$. Next, we attend to the case in which the ring encloses area $\mathscr{A}(\gamma) \lesssim N/P_\infty$. To this end, we denote by $\mathscr{C}_{\leqq n}(\omega)$ the set of sites in the configuration $\omega$ which belong to clusters of size no larger than $n$:

$$\mathscr{C}_{\leqq n}(\omega) = \{x \in \mathbb{Z}^2 \mid C(x;\omega) \leqq n\}, \tag{4.23}$$

and, for any finite set $A \subset \mathbb{Z}^2$, we denote by $f_{\leqq n}(A)$ the fraction of sites in $A$ belonging to clusters no larger than $n$:

$$f_{\leqq n}(A) = \frac{|A \cap \mathscr{C}_{\leqq n}|}{|A|}. \tag{4.24}$$

Note that, by translation invariance, $P_{\leqq n} = \mathbf{E}(f_{\leqq n})$. The following lemma shows the deviation of $f_{\leqq n}(A)$ from $P_{\leqq n}$.

**Lemma 4.2.** *Let* $n \in \mathbb{Z}^+$, $n \geqq 2$, *and* $A \subset \mathbb{Z}^2$, $|A| < \infty$. *Then for every* $\varepsilon \in (0,1)$,

$$\mathbf{P}(|f_{\leqq n}(A) - P_{\leqq n}| \geqq \varepsilon) \leqq c_1 \exp\left(-\frac{c_2 \varepsilon^2 |A|}{n^2}\right),$$

*where* $c_1 = 18$ *and* $c_2 = 1/324$.

*Proof.* Let $S_0$ denote the $3n \times 3n$ square $S_0 = \{(x_1, x_2) \in \mathbb{Z}^2 | 0 \leqq x_1, x_2 \leqq 3n\}$ and consider the set of translations of $S_0$: $S_y = T^{3ny}(S_0)$ which disjointly tile the lattice—i.e. consider the translations $S_y$ with $y = (m_1, m_2)$, $m_1$ and $m_2$ integers. (See Fig. 1.)

We further divide each $S_y$ into nine smaller $(n \times n)$ squares, $S_y^{[1]}, S_y^{[2]}, \ldots, S_y^{[9]}$ thereby forming nine disjoint sublattices. Observe that if $u \in S_y^{[k]}$ and $v \in S_{y'}^{[k]}$ $(y \neq y')$ then the events $C(u) \leqq n$ and $C(v) \leqq n$ are independent.

Let $A \subset \mathbb{Z}^2$, $|A| < \infty$. For $S_y^{[k]} \cap A \neq \varnothing$, consider the random variables

$$\mathbf{V}_y^{[k]} = |\mathscr{C}_{\leqq n} \cap S_y^{[k]} \cap A|, \tag{4.25}$$

and define

$$b_k = \max_y |S_y^{[k]} \cap A|. \tag{4.26}$$

Observe that

$$\mathbf{E}[\mathbf{V}_y^{[k]}] = P_{\leqq n} |S_y^{[k]} \cap A|, \tag{4.27}$$

while

$$\mathbf{V}_y^{[k]} \leqq b_k \leqq (n+1)^2 \leqq 4n^2 \tag{4.28}$$

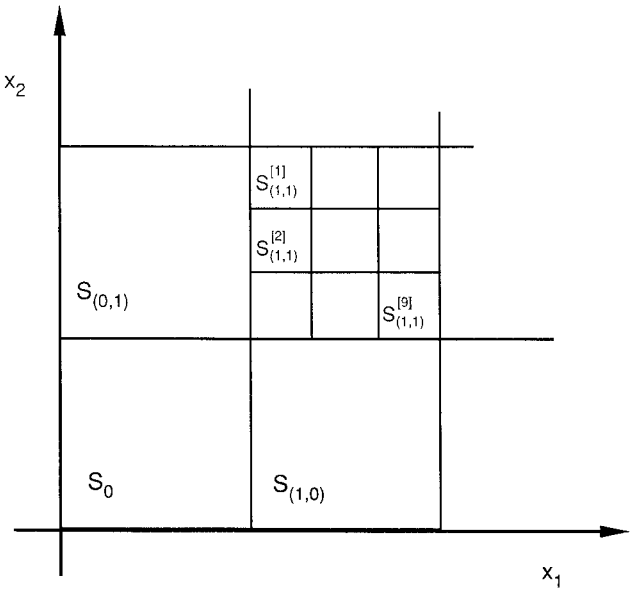


**Fig. 1.** A disjoint tiling of the lattice

(provided that $n \geqq 2$). We have

$$\begin{aligned}\mathbf{E}[(\mathbf{V}_y^{[k]})^2] &= \mathbf{E}\left[\sum_{u,v \in S_y^{[k]} \cap A} \mathbb{1}_{|C(u)| \leqq n} \mathbb{1}_{|C(v)| \leqq n}\right] \\ &\leqq |S_y^{[k]} \cap A| \mathbf{E}\left[\sum_{u \in S_y^{[k]} \cap A} \mathbb{1}_{|C(u)| \leqq n}\right] \\ &= |S_y^{[k]} \cap A| \mathbf{E}[\mathbf{V}_y^{[k]}] = P_{\leqq n} |S_y^{[k]} \cap A|^2. \end{aligned} \tag{4.29}$$

Thus

$$\begin{aligned}\mathrm{Var}(\mathbf{V}_y^{[k]}) &\leqq (P_{\leqq n} - P_{\leqq n}^2) |S_y^{[k]} \cap A|^2 \\ &\leqq (\tfrac{1}{4})(4n^2) |S_y^{[k]} \cap A| = n^2 |S_y^{[k]} \cap A|, \end{aligned} \tag{4.30}$$

or

$$\sum_y \mathrm{Var}(\mathbf{V}_y^{[k]}) \leqq n^2 |A|. \tag{4.31}$$

Take $\varepsilon \in (0,1)$. Now it is clear that whenever $|f_{\leqq n}(A) - P_{\leqq n}| \geqq \varepsilon$ (or equivalently when $|\sum_{y,k} (\mathbf{V}_y^{[k]} - P_{\leqq n} |S_y^{[k]} \cap A|)| > \varepsilon |A|$), then on at least one of the nine sublattices, the "excess" is larger than $\frac{1}{9}\varepsilon |A|$:

$$|f_{\leqq n}(A) - P_{\leqq n}| \geqq \varepsilon \Rightarrow \exists k \text{ such that } \left|\sum_y (\mathbf{V}_y^{[k]} - \mathbf{E}[\mathbf{V}_y^{[k]}])\right| \geqq \tfrac{1}{9}\varepsilon |A|. \tag{4.32}$$

Now, according to a lemma of Bernstein (see e.g. [B]), if $Z_1, Z_2, \ldots, Z_R$ are independent random variables satisfying

$$\max |Z_j| \leqq b \quad \text{and} \quad \sum_j \mathrm{Var}(Z_j) \leqq R s^2, \tag{4.33}$$

then

$$\begin{aligned}\mathbf{P}\left[\left|\sum_{j=1}^R (Z_j - \mathbf{E}[Z_j])\right| \geqq \lambda \sqrt{R}\right] &\leqq 2 \exp\left(-\frac{\lambda^2}{2s^2\left(1 + \dfrac{\lambda b}{3\sqrt{R} s^2}\right)}\right) \\ &\leqq 2 \max\left[\exp\left(-\frac{\lambda^2}{4s^2}\right), \exp\left(-\frac{\lambda \sqrt{R}}{2b}\right)\right]. \end{aligned} \tag{4.34}$$

Substituting the estimates (4.28) and (4.31) into Bernstein's inequality (and using $\lambda \sqrt{R} = \frac{1}{9}\varepsilon |A|$), we have

$$\mathbf{P}\left[\left|\sum_y (\mathbf{V}_y^{[k]} - \mathbf{E}[\mathbf{V}_y^{[k]}])\right| \geqq \tfrac{1}{9}\varepsilon |A|\right] \leqq 2 \exp\left(-\frac{\varepsilon^2 |A|}{324 n^2}\right), \tag{4.35}$$

where we have used $\varepsilon < 1$ to facilitate the calculations. Summing over all $k$, we obtain the desired result. ■

**Corollary.** *Let $A \subset \mathbb{Z}^2$, $|A| < \infty$. Suppose $p > p_c$ and consider the Bernoulli configurations at density $p$ restricted to the set $A$. Let $N$ be an integer with $P_\infty(p)|A| < N < |A|$ and define $P_{\geq N|A}(p)$ to be the probability that on the set $A$ there*

*is a connected cluster of size at least as large as N. Define* $\kappa = \kappa(p, N, A) > 0$ *by*

$$P_\infty(p)|A| = (1-\kappa)N,$$

*and, for* $n \in \mathbb{Z}^+$, *define* $\Delta_n$ *by*

$$P_{\leqq n}(p) = 1 - (1+\Delta_n)P_\infty(p).$$

*Suppose* $\Delta_n < \kappa/(1-\kappa)$ *and* $n < N$. *Then*

$$P_{\geqq N|A}(p) \leqq c_1 \exp\left[ -c_2 \frac{(1-\kappa)}{n^2}\left(\frac{\kappa}{1-\kappa} - \Delta_n\right)^2 NP_\infty \right],$$

*where* $c_1$ *and* $c_2$ *are defined in the statement of Lemma 4.2.*

*Proof.* Let $A \subset \mathbb{Z}^2$, $|A| < \infty$. If (in some configuration) the set $A$ contains a connected cluster of size at least as large as $N$, then the volume fraction of remaining sites in $A$ is no more than

$$1 - \frac{N}{|A|} = 1 - \frac{P_\infty}{1-\kappa}. \tag{4.36}$$

Thus for $n < N$, a cluster of size $N$ could only emerge if

$$f_{\leqq n}(A) \leqq 1 - \frac{P_\infty}{1-\kappa}. \tag{4.37}$$

Evidently

$$\begin{aligned} P_{\geqq N|A}(p) &\leqq \mathbf{P}\left[ f_{\leqq n}(A) \leqq 1 - \frac{P_\infty}{1-\kappa} \right] \\ &\leqq \mathbf{P}\left[ P_{\leqq n} - f_{\leqq n}(A) \geqq P_\infty\left(\frac{\kappa}{1-\kappa} - \Delta_n\right) \right]. \end{aligned} \tag{4.38}$$

The statement in the corollary follows directly from (4.38) and from Lemma 4.2. ■

We will now establish our upper bound.

*Proof of Theorem 1.A.* Take $p > p_c$, and suppose that the origin belongs to a finite cluster. Then the origin is surrounded by an occupied ring of dual bonds. Furthermore, exploiting the exponential decay of the dual connectivity function, w.p.1 there is a finite *outermost* occupied dual ring encircling the origin.[3] Let us enumerate all dual lattice rings about the origin: $r_\alpha$, $\alpha = 1, 2, \ldots$, and define the event

$$\bar{r}_\alpha = \{\omega | r_\alpha \text{ is the outermost occupied dual ring surrounding the origin}\}. \tag{4.39}$$

It is important to observe that given the event $\bar{r}_\alpha$, the statistical behavior of the configurations in $\mathrm{Int}(r_\alpha)$ is identical to that of the unconditional measure restricted

[3] In $d > 2$, the analogous statement may fail due to a possible condensation of closed filaments of dual $(d-1)$-cells. See [CC1, Sect. 3] for further discussion

to $\mathrm{Int}(r_\alpha)$. Finally, we will denote by $R_S$ the event

$$R_S = \bigcup_{\alpha:\mathscr{A}(r_\alpha)=S} \bar{r}_\alpha. \tag{4.40}$$

Let us define $S(\kappa) = (N/P_\infty)(1-\kappa)$ and

$$\kappa^* = \kappa^*(N) = N^{-1/4}(\log N)^4. \tag{4.41}$$

Obviously, the event $\{N \leqq |C(0)| < \infty\}$ may be (disjointly) decomposed according to whether the outermost occupied dual ring surrounding the origin encloses area greater than or less than $S(\kappa^*)$. In the former case, we omit any estimate on the probability of the cluster itself, and simply use Lemma 4.1 to obtain the upper bound:

$$\mathbf{P}\left[\bigcup_{S \geqq S(\kappa^*)} R_S\right] \leqq \mathbf{P}\left[\mathscr{R}\left(\frac{N}{P_\infty}(1-\kappa)\right)\right]$$

$$\leqq \tfrac{1}{2}\exp\{-[\omega(p)\sigma(p)/\sqrt{P_\infty(p)}]\sqrt{N}[1-N^{-1/4}(\log N)^4]\} \tag{4.42}$$

for $N$ sufficiently large. In deriving (4.42), we have used the fact that, for $N$ large enough, the leading correction in Lemma 4.1 is dominated by the difference between $\sqrt{1-\kappa^*}$ and $1-\kappa^*$.

We now focus our attention on the "smaller" rings—i.e. rings enclosing area less that $S(\kappa^*)$. Let $\kappa < \kappa^*$ be any number such that $S(\kappa)$ is an integer. Given the event $R_{S(\kappa)}$, the configuration inside must now struggle to produce a connected cluster at least as large as $N$. An estimate on the probability of this is the exact topic of the corollary to Lemma 4.2; namely

$$\mathbf{P}[N \leqq |C(0)| < \infty | R_{S(\kappa)}] \leqq c_1 \exp\left[-c_2 \frac{(1-\kappa)}{n^2}\left(\frac{\kappa}{1-\kappa} - \Delta_n\right)^2 NP_\infty\right], \tag{4.43}$$

where $n$ is any integer smaller than $N$. Let us choose $n$ to satisfy

$$\Delta_n \leqq \tfrac{1}{2}\kappa. \tag{4.44}$$

Using the known upper bound on $P_{\geqq n}$ (Eq. (1.3)), this can be accomplished without violating $n \leqq N$ (or $\kappa \leqq \kappa^*$) by choosing

$$n = H(\log \tfrac{1}{2}\kappa)^2 \tag{4.45}$$

for some huge constant $H$, provided $N$ is large enough to ensure $N \gg (\log N)^2$. Thus we have, for $N$ large enough,

$$\mathbf{P}[N \leqq |C(0)| < \infty | R_{S(\kappa)}] \leqq c_1 e^{-\phi(\kappa)N} \tag{4.46a}$$

with

$$\phi(\kappa) = (\text{const.})\left(\frac{\kappa^2}{(\log \kappa)^2}\frac{1}{(1-\kappa)}\right). \tag{4.46b}$$

Obviously the worst case is $\kappa = \kappa^*$, for which we get

$$\mathbf{P}[N \leqq C(0)| < \infty | R_{S(\kappa)}] \leqq c_1 \exp[-(\text{const})\sqrt{N}(\log N)^4]. \tag{4.47}$$

Since there are only of the order of N possible values of $\kappa$ such that $S(\kappa)$ is an integer, we can multiply the right-hand side (4.47) by N to bound the probabilities of the "small ring" cases. For $N$ large enough, the result is smaller than the right-hand side of Eq. (4.42), which establishes the desired result. ■

*4.B. The Lower Bound.* The result of this subsection is:

**Theorem 1.B.** *In the two-dimensional Bernoulli bond percolation model on the square lattice, for every* $p > p_c$

$$\liminf_{N\to\infty}\left(\frac{\log P_{\geqq N}(p)}{\sqrt{N}}\right) \geqq -\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}},$$

*where* $P_\infty(p)$, $\sigma(p)$ and $\omega(p)$ *are the infinite cluster density, surface tension and Wulff constant, as defined in Eq. (2.3), Proposition 3.1 and Eq. (3.8).*

Our strategy for proving the lower bound on $P_{\geqq N}$ is straightforward: We will first explicitly construct an approximate (lattice) Wulff curve of occupied dual bonds at essentially the right probability. We will then demonstrate that with probability of order unity (more precisely $P_\infty(p)$) the origin belongs to a cluster at least as large as $N$. To simplify the final proof, and for later reference, we will start by establishing the following auxiliary result concerning the event $\bar{r}_\alpha$ (cf. Eq. (4.39)).

**Lemma 4.3.** *Suppose* $p > p_c$. *Let* $\varepsilon \in \mathbb{R}^+$ *and* $M \in \mathbb{Z}^+$. *For* $M$ *large enough* $\exists \delta(\varepsilon) > 0$ *such that, with probability exceeding* $\exp[-(1+\varepsilon)\sigma(p)\omega(p)\sqrt{M}]$, *the event* $\bar{r}_\alpha$ *occurs for some* $r_\alpha$ *which lies entirely outside a (convex) shape enclosing area exceeding* $[1+\delta]M$.

*Proof.* We divide $\mathbb{R}^2$ into square unit cells centered at the sites of $(\mathbb{Z}^*)^2$; to avoid (zero-probability) possible ambiguities, we suppose that each cell includes its upper and right-hand boundaries, as well as its lower right-hand corner, but no other portion of its boundary. Then each point in $\mathbb{R}^2$ belongs to a unique cell (or site of $(\mathbb{Z}^*)^2$), and, for $u, v \in \mathbb{R}^2$, we may define

$$\theta(u,v) = \mathbf{P}(\text{the site of } u \text{ is connected to the site of } v \text{ by a path of occupied dual bonds}). \tag{4.48}$$

Note that we have suppressed the $p$-dependence in our notation for $\theta(u,v)$. Although $\theta(u,v)$ is not strictly translation invariant, it is approximately so. For example, one may easily obtain

$$(1-p)^2\theta(0,u-v) \leqq \theta(u,v) \leqq \frac{1}{(1-p)^2}\theta(0,u-v). \tag{4.49}$$

Using (4.49), it is straightforward to show that if $r \in \mathbb{R}$ and $v \in \mathbb{R}^2$, then

$$\lim_{r\to\infty}\left(-\frac{\log\theta(0,rv)}{r}\right) = \sigma g(v). \tag{4.50}$$

Of somewhat more relevance is the "half-space" version of $\theta(u,v)$. Let $u, v \in \mathbb{R}^2$. The line passing through $u$ and $v$ divides $\mathbb{R}^2$ into two half-spaces, the "upper"

one of which contains $x_2 = +\infty$ (or, in case of a vertical line, $x_1 = +\infty$). We define $\bar{\theta}(u,v)$ to be the probability that $u$ and $v$ are connected by a path of occupied dual bonds which travels exclusively via sites whose cells have non-zero intersection with the upper half-space relative to $u$ and $v$. We define $\underline{\theta}(u,v)$ analogously in terms of dual connections in the lower half-space. It is not terribly difficult to show that

$$\lim_{r\to\infty}\left(-\frac{\log\bar{\theta}(0,rv)}{r}\right)=\lim_{r\to\infty}\left(-\frac{\log\underline{\theta}(0,rv)}{r}\right)=\sigma g(v). \tag{4.51}$$

Indeed, e.g. for $v$ with rational coordinates, the existence of some limit no smaller than $\sigma g(v)$ follows from subadditivity. That this limit is $\sigma g(v)$ may be established by considering the restriction of the connectivity events to one-dimensional regions of finite width; this provides a decreasing sequence of rates which converge to $\sigma g(v)$ as the width of the regions increases. Any one of these rates may be used as an upper bound on the rate for the half-space connectivity. The extension to irrational coordinates follows immediately from convexity. (See e.g. [CC1] for arguments of this sort.)

Let $\gamma_w$ denote the Wulff curve centered at the origin. (See Theorem 3.3 for a definition of $\gamma_w$.) We will parameterize $\gamma_w$ by $t\in[0,T]$. Let $\mathscr{P}=(t_0,\ldots,t_N)$, $0=t_0\leqq t_1\leqq\cdots\leqq t_N=T$ denote a partition of $[0,T]$, and denote by $\gamma_w^{[\mathscr{P}]}$ the (convex) polygonal curve obtained by joining the points $\gamma(t_j)$ and $\gamma(t_{j+1})$ by straight line segments. As the partition becomes more refined, we get

$$\mathscr{A}(\gamma_w^{[\mathscr{P}]})\uparrow 1 \tag{4.52a}$$

and

$$g(\gamma_w^{[\mathscr{P}]})\uparrow w. \tag{4.52b}$$

Consider the curve $\gamma_w^{*[\mathscr{P}]}$ which is $\gamma_w^{[\mathscr{P}]}$ (linearly) rescaled by $\sqrt{2/(\mathscr{A}(\gamma_w^{[\mathscr{P}]})[1+\mathscr{A}(\gamma_w^{[\mathscr{P}]})])}$. This curve encloses area

$$\mathscr{A}(\gamma_w^{*[\mathscr{P}]})=\frac{2}{1+\mathscr{A}(\gamma_w^{[\mathscr{P}]})}\equiv 1+\delta(\mathscr{P}); \tag{4.53}$$

its $g$-length is

$$\begin{aligned} g(\gamma_w^{*[\mathscr{P}]}) &= g(\gamma_w^{[\mathscr{P}]})\sqrt{\frac{2}{\mathscr{A}(\gamma_w^{[\mathscr{P}]})[1+\mathscr{A}(\gamma_w^{[\mathscr{P}]})]}}\leqq w\sqrt{\frac{2}{\mathscr{A}(\gamma_w^{[\mathscr{P}]})[1+\mathscr{A}(\gamma_w^{[\mathscr{P}]})]}} \\ &= w\frac{1+\delta(\mathscr{P})}{\sqrt{1-\delta(\mathscr{P})}}\leqq w(1+\tfrac{1}{2}\varepsilon) \end{aligned} \tag{4.54}$$

if the partition is sufficiently refined. Now let us again focus on the lattice. By the Harris-FKG inequality, a curve of occupied dual bonds which encircles the origin and (save for the vertices) stays out of the polygon $\gamma_w^{*[\mathscr{P}]}$ scaled up linearly by $\sqrt{M}$ can be produced with a probability exceeding

$$\prod_{j=0}^{N}\Theta[\sqrt{M}\gamma_w^{*[\mathscr{P}]}(t_j),\sqrt{M}\gamma_w^{*[\mathscr{P}]}(t_{j+1})], \tag{4.55}$$

where $\Theta(-)$ denotes $\bar{\theta}(-)$ or $\underline{\theta}(-)$ as appropriate. At the expense of an additional finite ($M$-independent) factor, one can actually ensure that the curve stays *completely* outside the polygon. Then, using the existence of the limit in (4.51) and the bound (4.54), it follows that for $M$ sufficiently large, the $M$-independent factor times the estimate in Eq. (4.55) exceeds $\exp[-(1+\varepsilon)\sigma(p)\omega(p)\sqrt{M}]$, which is the desired result. ■

We will now prove our lower bound on $P_{\geqq N}$.

*Proof of Theorem 1.B.* Let $p > p_c$, $\varepsilon \in (0,1)$ and $N \in \mathbb{Z}^+$. Then by Lemma 4.3, for $N$ sufficiently large, with probability exceeding $\exp[-(1+\varepsilon)\sigma\omega\sqrt{N/P_\infty}]$, the event $\bar{r}_\alpha$ occurs for some $r_\alpha$ lying outside a convex shape which contains more than $(1+\delta)(N/P_\infty)$ sites. By Lemma 4.1, it is clear that we will not significantly degrade our estimate by assuming also that $\mathscr{A}(r_\alpha) < (\text{const})N$, for some sufficiently large constant. Recalling that the measure for those bonds with both endpoints in $\text{Int}(r_\alpha)$ is unconditioned, let us now consider the behavior of Bernoulli configurations restricted to $\text{Int}(r_\alpha)$.

We will first partition the sites of $\text{Int}(r_\alpha)$ into two (deterministic) sets. To this end, let $n \gg 1$ be an integer with $n$ small compared to the linear dimensions of $r_\alpha$ (e.g. we may regard $n$ as a small power of $N$), and let $D > 1$ be a constant of order unity. We write $\text{Int}(r_\alpha) = A_\alpha(n) \cup A_\alpha^c(n)$, where $A_\alpha(n)$ consists of those sites in $\text{Int}(r_\alpha)$ which are a distance greater than $2Dn$ from $r_\alpha$. For any Bernoulli configuration in $\bar{r}_\alpha$, the sites in $A_\alpha(n)$ may be further partitioned into two disjoint categories depending on the size of the cluster to which they belong: each $x \in A_\alpha(n)$ has either

(1) $|C(x)| \leqq n$; or

(2) $|C(x)| > n$.

Let us first show that category (1) does not exhaust too many of the sites of $A_\alpha(n)$. Indeed, by Lemma 4.2

$$\mathbf{P}(f_{\leqq n}(A_\alpha(n)) - P_{\leqq n} \geqq \tfrac{1}{3}\delta P_\infty) \leqq c_1 \exp\left(-\frac{c_2 \delta^2 P_\infty^2 |A_\alpha(n)|}{9n^2}\right). \tag{4.56}$$

In particular, if $n$ is a small power of $N$, and $N$ is sufficiently large, then $|A_\alpha(n)|$ exceeds $(1+\frac{1}{2}\delta)(N/P_\infty)$. Using the fact that $P_{\leqq n} \leqq 1 - P_\infty$, and taking $\delta < \frac{1}{4}$, it follows that, with (conditional) probability tending rapidly to one, more than $(1+\frac{1}{8}\delta)N$ sites belong to category (2).

Next, we claim that, with probability tending rapidly to one with $N$, all of the sites of category (2) belong to a single cluster. Indeed, let us suppose that two sites in category (2) belong to distinct clusters. Since, by definition, both of these sites are further than $2Dn$ from $r_\alpha$, and both belong to clusters of size exceeding $n$, then (for appropriate choice of $D$) there must be a dual interface in $\text{Int}(r_\alpha)$ of linear extent exceeding $Dn$. However, the probability of such an interface is less than $(\text{const})N^2 e^{-Dn\sigma}$, which tends rapidly to zero with $N$.

Now we note that the absence of a dual interface is positively correlated (in the sense of FKG) with the event $\{f_{\leqq n}(A_\alpha(n)) - P_{\leqq n} < \frac{1}{3}\delta P_\infty\}$. Thus with (conditional) probability rapidly approaching one, more than $(1+\frac{1}{8}\delta)N$ belong to a single cluster. Furthermore, given the pair of events discussed above, it is not

difficult to see that the origin belongs to this cluster (as does any site in $A_\alpha(n)$) with probability not smaller than $P_\infty(p)$.

The simultaneous occurrence of an $\bar{r}_\alpha$ of the appropriate type, and the three events discussed above, produces the event $\infty > |C(0)| \geqq N$; all of this occurs with probability exceeding $(\text{const}) \exp(-(1+2\varepsilon)\sigma\omega\sqrt{N/P_\infty}]$, with the constant uniform in $N$. This establishes Theorem 1.B. ■

*4.C. The Combined Bound.* Putting together the results of this section, we have:

**Theorem 1′.** *In the two-dimensional Bernoulli bond percolation model on the square lattice, for every* $p > p_c$

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \log P_{\geqq N}(p) = -\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}}.$$

*Remark.* We have no intuition about the nature of the convergence of $-\log P_{\geqq N}(p)/\sqrt{N}$ to $\omega(p)\sigma(p)/\sqrt{P_\infty(p)}$—we do not even know the sign of the correction. Of course, the derivation in the proof of Theorem 1.A provides a lower bound on the difference. Similarly, if we had good (though not necessarily optimal) lower bounds on $\tau^*_{0^*,x^*}$ which were *uniform in direction*, these could be used to obtain an upper bound on the difference. We consider the nature of these corrections to be an important open problem.

For the time being, we will have to make due with the tautology:

**Corollary.** $\exists \varepsilon(N) \geqq |[\omega(p)\sigma(p)]/\sqrt{P_\infty(p)} + \log P_{\geqq N}(p)/\sqrt{N}| > 0$ *such that* $\varepsilon(N) \downarrow 0$. *Explicitly,* $\forall N$,

$$\exp\left(-[1+\varepsilon(N)]\left[\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}}\right]\sqrt{N}\right)$$
$$\leqq P_{\geqq N}(p) \leqq \exp\left(-[1-\varepsilon(N)]\left[\frac{\omega(p)\sigma(p)}{\sqrt{P_\infty(p)}}\right]\sqrt{N}\right). \tag{4.57}$$

The reader may recall that Theorem 1 of the Introduction was expressed in terms of the actual finite cluster distribution $P_N(p)$. Obviously, since $P_N(p) \leqq P_{\geqq N}(p)$, (4.57) automatically implies an upper bound on $P_N(p)$. In [KS], a subadditivity argument was presented which shows that upper and lower bounds of the form $e^{-b(p)\sqrt{N}}$ on $P_{\geqq N}(p)$ also give lower bounds of this form on $P_N(p)$. Here, however, we need a slight refinement of the Kunz and Souillard estimate to ensure that we do not degrade the constant $\omega(p)\sigma(p)/\sqrt{P_\infty(p)}$. This is provided in the following:

**Proposition 4.4.** *Suppose that* $P_N(p)$ *is known to obey the bound*

$$P_N(p) \geqq e^{-b(p)\sqrt{N}}$$

*with* $b(p)$ *positive. Suppose also that there exists a positive, finite constant* $a(p)$, *and positive function* $\varepsilon(N)$ *with* $\varepsilon(N) \downarrow 0$ *such that* $P_{\geqq N}(p)$ *satisfies*

$$e^{-[1+\varepsilon(N)]a(p)\sqrt{N}} \leqq P_{\geqq N}(p) \leqq e^{-[1-\varepsilon(N)]a(p)\sqrt{N}}.$$

*Then either* $\varepsilon(N) \leqq O(N^{+3/2})$ *or there are positive, finite constants* $k_1(p)$ *and* $k_2(p)$ *such that*

$$P_N(p) \geqq \frac{k_1(p)}{\varepsilon(N)N^{3/2}} \exp\{-[1 + k_2(p)\sqrt{\varepsilon(N)}]a(p)\sqrt{N}\}.$$

*Proof.* As in the [KS] proof, the key is the subadditivity relation:

$$\frac{P_{N+M}}{N+M} \geqq \frac{P_N}{N}\frac{P_M}{M}. \tag{4.58}$$

Let us choose $\delta(N) = D\varepsilon(N)$, with $D$ a constant of order unity to be determined later. By the upper and lower bounds on $P_{\geqq N}(p)$, we have

$$P_{\geqq N(1-\delta(N))} - P_{\geqq N} \geqq e^{-a(p)[1+\varepsilon^*(N)][1-\delta(N)]^{1/2}\sqrt{N}} - e^{-a(p)[1-\varepsilon(N)]\sqrt{N}}, \tag{4.59}$$

where

$$\varepsilon^*(N) \equiv \varepsilon(N(1-\delta(N))). \tag{4.60}$$

By monotonicity of $\varepsilon(N)$, we may replace $\varepsilon(N)$ in the second term on the right-hand side of (4.59) with $\varepsilon^*(N)$. Then, choosing the constant $D$ so that

$$[1+\varepsilon^*(N)][1-\delta(N)]^{1/2} < 1 - \varepsilon(N) - \tfrac{1}{4}\delta(N), \tag{4.61}$$

Eq. (4.59) implies

$$P_{\geqq N(1-\delta(N))} - P_{\geqq N} \geqq (\text{const})\sqrt{N}\delta(N)e^{-a(p)\sqrt{N}}. \tag{4.62}$$

Noting that $P_{\geqq N(1-\delta(N))} - P_{\geqq N}$ represents the sum of $\delta(N)N$ terms, let us define $M^*$ by

$$P_{N-M^*} = \max\{P_{N-M} | 0 \leqq M \leqq \delta(N)N\}, \tag{4.63}$$

so that (4.62) gives

$$P_{N-M^*} \geqq \frac{(\text{const})}{\sqrt{N}} e^{-a(p)\sqrt{N}}. \tag{4.64}$$

By (4.64) and the subadditivity relation (4.58), we have

$$P_N \geqq \frac{(\text{const})}{\delta(N)N^{3/2}} e^{-a(p)\sqrt{N}} P_{M^*}. \tag{4.65}$$

Thus, using the known lower bound

$$P_{M^*} \geqq e^{-b(p)\sqrt{M^*}} \geqq e^{-b(p)\sqrt{\delta(N)N}}, \tag{4.66}$$

we finally obtain

$$P_N \geqq \frac{(\text{const})}{\delta(N)N^{3/2}} e^{-a(p)\sqrt{N}[1+(b(p)/a(p))\sqrt{\delta(N)}]}. \tag{4.67}$$

Recalling that $\delta(N) = O(\varepsilon(N))$, this is the desired result. Note that, although the leading exponential decay rate has been preserved, this lower bound on $P_N$ has a larger correction to the leading rate than does the lower bound on $P_{\geqq N}$. ■

Noting that we may always take $\varepsilon(N) \geqq N^{-3/2}$ in the Corollary to Theorem 1', the lower bound of Theorem 1 now follows from Theorem 1', Proposition 4.4 and the nonoptimal lower bound (1.3) on $P_N$.

## 5. A Microscopic Wulff Construction

In this section, we give our first formulation of the Wulff construction for two-dimensional percolation, as contained in Theorem 2 of the Introduction: Namely, we condition on the (unlikely) event $N \leqq |C(0)| < \infty$, and show that with probability tending to one as $N$ tends to infinity, an interface surrounding the origin is arbitrarily close, in distance measured in units of $\sqrt{N}$, to some translate of the Wulff shape scaled by $\sqrt{N/P_\infty}$. An alternative formulation, in terms of a "microcanonical ensemble," is given in Sect. 6.

Not surprisingly, in order to prove the microscopic Wulff construction, we require a stability result for the Wulff minimum: Namely, if $\gamma$ encloses unit area and if $\rho(\gamma_w, \gamma) \geqq \eta > 0$, then the value of the surface energy functional for $\gamma$ differs from the minimum by a strictly positive function $f(\eta)$. Were this not the case, then the cluster could assume a shape which differs substantially from the Wulff shape at essentially no cost. Two points are worth noting: (1) Although our stability result is sufficient for our purposes, it is far from optimal—see the Remark following the proof of Theorem 5.2. (2) Given the uniqueness of the minimizer, this stability may seem obvious; however, it actually fails for $d > 2$. This suggests that one must formulate another, less stringent notion of the "difference between two contours" in order to prove a higher-dimensional microscopic Wulff construction.

This section is organized as follows: The proof of stability is given in Subsect. 5.A; the Wulff construction (i.e. the proof of Theorem 2) is given in Subsect. 5.B.

*5.A. Stability of the Wulff Minimum.* As explained above, the principal result of this subsection (Theorem 5.2) is that there exists a strictly positive function, $f(\eta)$, such that if $\gamma$ is any acceptable unit area contour satisfying $\rho(\gamma_w, \gamma) \geqq \eta > 0$, then $g(\gamma) \geqq \omega(p) + f(\eta)$. The strategy of our proof is to consider a variant of the standard Wulff variational problem (cf. Eq. (3.8)) with the additional constraint $\rho(\gamma_w, \gamma) \geqq \eta$, and to show: (1) an actual minimizer exists for this modified problem; and (2) this minimizer is not $\gamma_w$. By uniqueness (Theorem 3.3), one would then expect that this minimizer, not being $\gamma_w$, must have a surface energy strictly larger than $\omega(p)$. Unfortunately, it is conceivable—particularly for large $\eta$—that the new minimizer will be found among a larger class than the rectifiable Jordan curves $\mathscr{J}$, so that Theorem 3.3 could not be applied. Thus we must first extend Theorem 3.3, i.e. extend the analysis of the standard (unconstrained) Wulff variational problem, to a larger class of curves.

Let us define the appropriate "larger class." We remind the reader that Jordan curves are closed and non-self-intersecting. Now let $\mathscr{K} \supset \mathscr{J}$ denote the set of all rectifiable closed curves in $\mathbb{R}^2$, i.e. $\mathscr{K} \setminus \mathscr{J}$ are the rectifiable closed curves which do self-intersect. In order to formulate a Wulff variational problem over $\mathscr{K}$, we must

generalize the notion of the "area enclosed" by a contour. For $\gamma \in \mathscr{K}$, we define

$$\mathscr{A}(\gamma) = \inf_{\Gamma \in \mathscr{J}} \{\mathscr{A}(\Gamma) | \gamma \subset \overline{\mathrm{Int}(\Gamma)}\} = \inf_{\Gamma \in \mathscr{J}} \{\mathscr{A}(\Gamma) | \gamma \subset \mathrm{Int}(\Gamma)\}, \tag{5.1}$$

where, as usual, for $\Gamma \in \mathscr{J}$, $\mathscr{A}(\Gamma)$ is simply the Euclidean area enclosed by $\Gamma$. For polygonal paths, this definition agrees with that given above Eq. (4.17). The first step in our proof of stability is to show that, given a chance to vary over $\mathscr{K}$, the standard Wulff variational problem still has the unique minimizer $\gamma_w$.

**Proposition 5.1.** *Let $\mathscr{J}$ be the set of all rectifiable Jordan curves in $\mathbb{R}^2$, let $\mathscr{K}$ be the set of all rectifiable closed curves in $\mathbb{R}^2$, let $g_p(\gamma)$ denote the density-p g-length of the curve $\gamma$ as defined by Eq. (3.7) and Proposition 3.2, let $\mathscr{A}(\gamma)$ denote the area enclosed by $\gamma$ as defined in Eq. (5.1), and let $\omega(p)$ be the Wulff constant as defined in Eq. (3.8). Consider the variational problem*

$$\omega^*(p) = \inf_{\gamma \in \mathscr{K}} \{g_p(\gamma) | \mathscr{A}(\gamma) \geqq 1\}.$$

*Then $\omega^*(p) = \omega(p)$, and, in particular, the unique minimizer of this variational problem is the Wulff curve $\gamma_w = \gamma_w(p)$, as defined in Theorem 3.3.*

*Proof.* It suffices to show that if $\gamma \in \mathscr{K} \setminus \mathscr{J}$ and $\mathscr{A}(\gamma) \geqq 1$, then there is a $\gamma' \in \mathscr{J}$ with $\mathscr{A}(\gamma') \geqq 1$ such that $g(\gamma') < g(\gamma)$. Actually, we will prove the somewhat stronger statement that the minimizer is a *convex* contour in $\mathscr{J}$.

We begin by extending the notion of convexity to curves in $\mathscr{K} \setminus \mathscr{J}$. Suppose that $\gamma \in \mathscr{K}$ is parameterized by $t \in [0, T]$. Then we can define the convex hull of $\gamma$, $H(\gamma)$, in the usual fashion:

$$H(\gamma) = \{x \in \mathbb{R}^2 | x = \lambda\gamma(t_1) + (1-\lambda)\gamma(t_2); 0 \leqq \lambda \leqq 1, 0 \leqq t_1 \leqq t_2 \leqq T\}. \tag{5.2}$$

We say that the curve $\gamma \in \mathscr{K} \setminus \mathscr{J}$ is *convex* if each for $\Gamma \in \mathscr{J}$ such that $\gamma \subset \overline{\mathrm{Int}(\Gamma)}$, we also have $H(\gamma) \subset \overline{\mathrm{Int}(\Gamma)}$. Obviously when $\gamma \in \mathscr{J}$, this is the usual notion of convexity.

Our strategy is to divide $\mathscr{K} \setminus \mathscr{J}$ into two sets: convex and non-convex curves. For curves in the latter (and easier) class, we will show that there is a convex curve in $\mathscr{J}$ which has the same $g$-length as the original curve, but encloses more area. For curves in the former class, we will show that there is a (convex) curve in $\mathscr{J}$ which encloses the same area as the original curve, but has a shorter $g$-length.

For $\gamma \in \mathscr{K} \setminus \mathscr{J}$ (or $\gamma \in \mathscr{J}$ not convex), let us define the curve $\gamma_H$ via

$$\gamma_H = \partial H(\gamma). \tag{5.3}$$

It is obvious that $\gamma_H$ is rectifiable, since its length is bounded above by the smallest circle which circumscribes the convex set $H(\gamma)$. Thus $\gamma_H$ is a convex contour in $\mathscr{J}$. We will show that the contour $\gamma_H$, perhaps modified by a scale factor, will provide a better variational candidate than the original $\gamma$.

We first show that

$$g(\gamma_H) \leqq g(\gamma). \tag{5.4}$$

Indeed, let us parameterize the curve $\gamma_H$ by $s \in [0, S]$. Let $s_1, s_2, \ldots, s_k$ be a partition

of $[0, S]$ for which $\Sigma_i g(\gamma_H(s_i) - \gamma_H(s_{i+1}))$ is an approximation to the $g$-length of $\gamma_H$. We claim that, without loss of generality, we may choose the times $s_i$ in such a way that the points $\{\gamma_H(s_i)\}$ are extreme points of $\gamma_H$. For example, if $s_i$ does not correspond to an extreme point of $\gamma_H$, denote by $s_{i'}$ and $s_{i''}$ the earliest time after and latest time before $s_i$—using cyclic boundary conditions if necessary—such that $\gamma_H(s_{i'})$ and $\gamma_H(s_{i''})$ are extreme points of $\gamma_H$. Obviously this is a refinement of the original partition; however, it is seen that this approximation to the $g$-length is identical to that resulting from the partition in which all times between $s_{i'}$ and $s_{i''}$, including $s_i$, are removed.

Now observe that, by the definition of $\gamma_H$, all extreme points of $\gamma_H$ are on the contour $\gamma$ itself. Thus, by the reasoning of the previous paragraph, we may take the points $\gamma_H(s_i)$ to correspond to times $t_{j(i)}$, not necessarily unique, which form a partition of $[0, T]$ for the original contour $\gamma$. Of course, these times will not, in general, fall in order on the contour $\gamma$. Thus, a favorable comparison of the approximations to $g(\gamma_H)$ and $g(\gamma)$ from the partitions $(s_i)$ and $(t_j)$, i.e. a proof of the inequality

$$\sum_i g(\gamma_H(s_i) - \gamma_H(s_{i+1})) \leqq \sum_j g(\gamma(t_j) - \gamma(t_{j+1})), \tag{5.5}$$

amounts to showing that the shortest $g$-length contour which touches all the vertices of a convex polygon passes through those vertices in order. The proof of this (which is done graphically) is elementary and identical to that for the Euclidean case; for completeness, it has been included in the appendix (Proposition A.1). Clearly, Eq. (5.5) establishes the inequality (5.4).

Now suppose that $\gamma \in \mathscr{K}$ is not convex. Then it is easy to see that

$$\mathscr{A}(\gamma) < \mathscr{A}(\gamma_H). \tag{5.6}$$

Indeed, if $\gamma$ is not convex, we may find a non-extreme point $x \in \gamma_H$ and a disk of radius $a > 0$ about $x$ which is disjoint from $\gamma$. Half this disk belongs to the interior of $\gamma_H$, and hence

$$\mathscr{A}(\gamma_H) \geqq \mathscr{A}(\gamma) + \tfrac{1}{2}\pi a^2. \tag{5.7}$$

We may now rescale the curve $\gamma_H$ by $\sqrt{\mathscr{A}(\gamma)/\mathscr{A}(\gamma_H)}$ and call the new curve $\gamma'$. Then

$$\mathscr{A}(\gamma') = \mathscr{A}(\gamma), \tag{5.8}$$

while by Eq. (5.4)

$$g(\gamma') = g(\gamma_H)\sqrt{\frac{\mathscr{A}(\gamma)}{\mathscr{A}(\gamma_H)}} < g(\gamma_H) \leqq g(\gamma). \tag{5.9}$$

This establishes that the minimizer is convex.

Now we need only consider those contours $\gamma \in \mathscr{K} \setminus \mathscr{J}$ which are convex, i.e. $\gamma_H \subset \gamma$. We will further divide this into two cases: either $\gamma_H$ and $\gamma$ are identical, or $\gamma_H$ is a strict subset of $\gamma$. In either case, given our definition of the area, it is clear that

$$\mathscr{A}(\gamma_H) = \mathscr{A}(\gamma). \tag{5.10}$$

If the sets $\gamma_H$ and $\gamma$ are identical, that is $\gamma_H([0, S]) = \gamma([0, T])$, then the

assumption $\gamma \in \mathscr{K} \setminus \mathscr{J}$ necessarily implies a continuum of double points. In this case, it is quite easy to establish that $g(\gamma) > g(\gamma_H)$. If the entire curve is double covered, then $g(\gamma) \geqq 2g(\gamma_H)$. Otherwise, assume without loss of generality that $\gamma(0)$ is not double covered, let $t_1$ denote the earliest time for which $\gamma(t_1)$ is double covered, and let $t_2 > t_1$ be the earliest time for which $\gamma(t_1) = \gamma(t_2)$. For $t \in (t_1, t_2)$, it is seen that

$$g(\gamma) \geqq g(\gamma_H) + 2g(\gamma(t_1) - \gamma(t)), \tag{5.11}$$

which, together with (5.10), establishes the desired result.

Finally, consider the case in which $\exists x \in \gamma$ such that $x \notin \gamma_H$. Let us denote by $a > 0$ the $g$-distance between $x$ and $\gamma_H$:

$$a = \min_{y \in \gamma_H} g(x - y). \tag{5.12}$$

Although the inequality

$$g(\gamma) \geqq g(\gamma_H) + 2a \tag{5.13}$$

is intuitively clear, we have been unable to find a straightforward proof. A proof of (5.13) which relies on a $g$-based Hausdorff measure of the sets $\gamma$ and $\gamma_H$ has been relegated to the appendix (Proposition A.2). Obviously, (5.10) and (5.13) imply the desired result, and complete the proof. ■

We can now prove the necessary stability of the variational minimum.

**Theorem 5.2.** *Let $\mathscr{K}$ be the set of all rectifiable closed curves in $\mathbb{R}^2$, let $g_p(\gamma)$ denote the density-p g-length of the curve $\gamma$ as defined by Eq. (3.7) and Proposition 3.2, let $\mathscr{A}(\gamma)$ denote the area enclosed by $\gamma$ as defined in Eq. (5.1), let $\omega(p)$ be the Wulff constant as defined in Eq. (3.8), and let $\gamma_w = \gamma_w(p)$ be the Wulff curve as given in Theorem 3.3. Consider the variational problem*

$$\omega(p) + f_p(\eta) = \inf_{\gamma \in \mathscr{K}} \{g_p(\gamma) | \mathscr{A}(\gamma) \geqq 1; \rho(\gamma, \gamma_w) \geqq \eta\}.$$

*Then for all $p > p_c$, $f_p$ is a strictly positive function.*

*Proof.* Let $(\gamma_n | n = 1, 2, \ldots)$ denote a minimizing sequence of contours in $\mathscr{K}$. Without loss of generality, we may assume that each $\gamma_n$ is translated so as to minimize its Hausdorff distance from some fixed $\gamma_w$, i.e. $\rho(\gamma_n, \gamma_w) = D_H(\gamma_n, \gamma_w)$, where the Hausdorff distance $D_H$ is defined in Eq. (1.9). Also without loss of generality, we may assume that the lengths of these curves are bounded: $|\gamma_n| < M < \infty$; hence the $\gamma_n$ may be parameterized by $t \in [0, 1]$, where $t$ is proportional to the arclength, i.e. $\gamma_n : [0, 1] \to \mathbb{R}^2$.

It follows from the above properties that $(\gamma_n)$ is a family of uniformly bounded equicontinuous functions on $[0, 1]$. Hence, by the Ascoli theorem, there is a subsequence—here again denoted by $(\gamma_n)$—which converges uniformly to some $\gamma^* \in \mathscr{K}$. According to Proposition 5.1, it suffices to show that $\gamma^*$ is an actual minimizer of this modified Wulff problem and that $\gamma^* \neq \gamma_w$.

First, observe that $\rho(\gamma^*, \gamma_w) \geqq \eta$ (and hence $\gamma^* \neq \gamma_w$), since otherwise the uniform convergence would imply that the constraint was violated at some finite $n$.

Next, let $t_1, \ldots, t_k$ be a partition of the unit interval, which provides an approximation to $g(\gamma^*)$. Observing that, for each $j$,

$$\gamma_n(t_j) \to \gamma^*(t_j), \tag{5.14}$$

it is readily established that

$$g(\gamma^*) \leqq \lim_{n \to \infty} g(\gamma_n) = \omega + f(\eta). \tag{5.15}$$

Finally, let $\Gamma \in \mathscr{J}$ be a Jordan curve for which

$$\gamma^* \subset \mathrm{Int}(\Gamma), \tag{5.16}$$

and which thus provides an approximation to $\mathscr{A}(\gamma^*)$. Since all the contours $\gamma_n$ live on some finite ball $B$, it is clear that only finitely many of them have points in the (compact) set $\bar{B} \backslash \mathrm{Int}(\Gamma)$. Using this, it is easy to verify that

$$\mathscr{A}(\gamma^*) \geqq \lim_{n \to \infty} \mathscr{A}(\gamma_n) \geqq 1. \tag{5.17}$$

Equations (5.15) and (5.17) show that $\gamma^*$ has all the required properties; thus, it is indeed a minimizer. ■

*Remark.* Given the variational stability proved above, one is tempted to suspect that a stronger statement is true. In particular, a perturbative ("second variational") calculation—based on the fact that the Wulff curve is an extremum—suggests that

$$f(\eta) = O(\eta^2). \tag{5.18}$$

With additional hypotheses on the function $g$, such results should be straightforward to establish; we suspect that (5.18) holds in the general case. A strong stability statement of this sort would represent the first (and easiest) step in obtaining concrete estimates on various convergence rates which appear in this work only as existential quantities.

In any case, for future reference, we note that for small $\eta$,

$$f(\eta) \leqq \alpha\eta, \tag{5.19}$$

with, say, $\alpha < 10$. Equation (5.19) is easily verified by using a trial Wulff shape with a "spike" of length $\alpha\eta$.

*5.B. The Wulff Construction.* We now establish the Wulff construction for the Bernoulli system conditioned on the event $N \leqq |C(0)| < \infty$. Theorem 5.3, below, gives the result mentioned in Eq. (1.11); Theorem 2 of the Introduction then follows quite easily from Theorem 5.3.

**Theorem 5.3.** *Consider the two-dimensional Bernoulli bond percolation model on the square lattice with $p > p_c$, and condition on the event $N \leqq |C(0)| < \infty$. Then there exists a function $\eta(N) = \eta(N; p)$, with $\eta(N) \downarrow 0$ monotonically as $N \uparrow \infty$, such that, with conditional probability tending rapidly to one with $N$, there is an occupied circuit of dual bonds, $\gamma$, encircling the origin satisfying*

$$\rho\left(\gamma_w, \frac{\gamma}{\sqrt{\mathscr{A}(\gamma)}}\right) \leqq \eta(N).$$

*Proof*. Take $p > p_c$ and suppose that the origin belongs to a finite cluster of size at least $N$. Then (cf. Eq. (4.39)) the event $\bar{r}_\alpha$ occurs for some ring $r_\alpha$. Roughly speaking, there are only three possibilities for the ring $r_\alpha$: (i) it may enclose inadequate area to properly support a cluster of size $N$; or (ii) it may enclose an area sufficient to support $|C(0)| \geqq N$, but be of an unfavorable shape; or (iii) (most probably) under rescaling, it may actually be close to the Wulff shape. To deal with the second—and most troublesome—possibility, we will employ the variational stability derived in Theorem 5.2. In order that this can be best exploited, the correct scale for the comparison shape is not necessarily $\sqrt{N/P_\infty}$, but rather the scale of the ring itself. Ultimately, the system will select a ring of the proper size.

In order to quantify the above discussion, let us assume that $N$ is large, and recall the definitions of $\kappa^*(N)$ and $\varepsilon(N)$ from Eq. (4.41) and the corollary at the end of Sect. 4. Let $f(\eta) = f_p(\eta)$ be the "stability function" given in Theorem 5.2, and define $\eta = \eta(N) = \eta(N; p)$ to be the smallest possible number for which

$$f(\tfrac{1}{2}\eta) \geqq 4[\varepsilon(N) + \kappa^*(N)]. \tag{5.20}$$

The three possibilities for the event $\bar{r}_\alpha$ are:

i) $r_\alpha$ encloses area smaller than $[1 - \kappa^*(N)]N/P_\infty$,

ii) $r_\alpha$ encloses area exceeding $[1 - \kappa^*(N)]N/P_\infty$, but $\rho(r_\alpha/\sqrt{\mathscr{A}(r_\alpha)}, \gamma_w) > \eta$,

iii) $r_\alpha$ encloses area exceeding $[1 - \kappa^*(N)]N/P_\infty$ and $\rho(r_\alpha/\sqrt{\mathscr{A}(r_\alpha)}, \gamma_w) \leqq \eta$.

Case (i) has been discussed in the proof of Theorem 1.A. Under these circumstances, the event $|C(0)| \geqq N$ occurs with probability not larger than $\exp[-O(\sqrt{N}(\log N)^4)]$ (cf. Eq. (4.47)).

We will handle case (ii) by a variant of the argument used in the proof of Lemma 4.1: We will bound from above the probability of observing an $r_\alpha$ satisfying the conditions of case (ii) by summing over all possible $m$-skeletons of such curves. First, we observe that if $\gamma$ and $\gamma'$ are *fixed* contours, then

$$\rho(\gamma', \gamma_w) \leqq D_H(\gamma, \gamma') + \rho(\gamma, \gamma_w), \tag{5.21}$$

where $D_H$ is the (untranslated) Hausdorff distance. Thus, if $m$ is not large compared with $\eta$ times the typical length scale of the rings $r_\alpha$, then the $m$-skeletons of these rings will also have a reasonable separation from the Wulff minimizers.

Let us follow the reasoning of Eqs. (4.11)–(4.22). If a given $m$-skeleton takes $J$ steps, the cost is at least $e^{-\sigma J m}$. This quantity (multiplied by the insignificant combinatorial factor $mJ(8\pi m)^J$) must be summed over the allowed range of $J$. Although this necessarily means that the smallest term dominates, let us start by summing away $J > J^*$, where $J^*$ is the least integer satisfying

$$J^* m \geqq 2\omega\sqrt{N/P_\infty}; \tag{5.22}$$

not surprisingly, this tail is negligible relative to $P_{\geqq N}$. Thus, we must perform the summation

$$\sum_{J_{\min} \leqq J \leqq J^*} e^{-\sigma m J}(mJ)(8\pi m)^J, \tag{5.23}$$

where $J_{\min}$ is determined with the help of the variational principle. The sum in

(5.23)—which again is essentially the first term—will provide our bound on case (ii). (Cf. Eq. (4.16)).

Next we derive a lower bound on $J_{\min}$. Let us suppose that $h_m = h_m(r_\alpha)$ is the $m$-skeleton of some lattice curve $r_\alpha$ in category (ii). Now, according to Eq. (4.18), the $g$-length of an $m$-skeleton of $J$ steps satisfies

$$(J+1)(m+1) \geqq g(h_m). \tag{5.24}$$

Thus it suffices to obtain a lower bound on $g(h_m)$. We consider two cases:

(a) $\mathscr{A}(h_m) > \mathscr{A}(r_\alpha)$

(b) $\mathscr{A}(h_m) \leqq \mathscr{A}(r_\alpha)$.

By (4.20), in the second (and more difficult) case, we can replace condition (b) by

(b′) $\mathscr{A}(r_\alpha) - (\text{const})Jm^2 \leqq \mathscr{A}(h_m) \leqq \mathscr{A}(r_\alpha)$.

In both cases, we will use the condition

$$\rho\left(\frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}, \gamma_w\right) \geqq \eta. \tag{5.25}$$

First let us attend to case (b′), which will require more stringent conditions on $m$. By (5.25) and the observation (5.21), we have

$$\rho\left(\frac{h_m}{\sqrt{\mathscr{A}(h_m)}}, \gamma_w\right) \geqq \eta - D_H\left(\frac{h_m}{\sqrt{\mathscr{A}(h_m)}}, \frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}\right). \tag{5.26}$$

Although the Hausdorf distance between $r_\alpha$ and its $m$-skeleton cannot exceed $m$, this distance is not quite the quantity appearing on the right-hand side of Eq. (5.26). However,

$$\begin{aligned} D_H\left(\frac{h_m}{\sqrt{\mathscr{A}(h_m)}}, \frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}\right) &= \frac{1}{\sqrt{\mathscr{A}(r_\alpha)}} D_H\left(\frac{\sqrt{\mathscr{A}(r_\alpha)}}{\sqrt{\mathscr{A}(h_m)}} h_m, r_\alpha\right) \\ &\leqq \frac{1}{\sqrt{\mathscr{A}(r_\alpha)}}\left[D_H(h_m, r_\alpha) + D_H\left(\frac{\sqrt{\mathscr{A}(r_\alpha)}}{\sqrt{\mathscr{A}(h_m)}} h_m, h_m\right)\right] \\ &\leqq \frac{1}{\sqrt{\mathscr{A}(r_\alpha)}}\left[2m + Jm\left|\frac{\sqrt{\mathscr{A}(r_\alpha)}}{\sqrt{\mathscr{A}(h_m)}} - 1\right|\right]. \end{aligned} \tag{5.27}$$

Now using the lower bound on $\mathscr{A}(h_m)$ from condition (b′), together with the facts $\mathscr{A}(r_\alpha) \geqq [1 - \kappa^*(N)]N/P_\infty$ and $Jm \leqq 2\omega\sqrt{N/P_\infty}$, we can bound the right-hand side of Eq. (5.27) by $\Delta m/\sqrt{\mathscr{A}(r_\alpha)}$, where $\Delta > 2$ is a constant of order unity. Thus we obtain

$$\rho\left(\frac{h_m}{\sqrt{\mathscr{A}(h_m)}}, \gamma_w\right) \geqq \eta - \frac{\Delta m}{\sqrt{\mathscr{A}(r_\alpha)}}. \tag{5.28}$$

Recalling that $\sqrt{\mathscr{A}(r_\alpha)}$ scales like $\sqrt{N}$, let us choose $m$ to be the largest integer which satisfies

$$\frac{\Delta m}{\sqrt{\mathscr{A}(r_\alpha)}} \leqq \nu f(\tfrac{1}{2}\eta), \tag{5.29}$$

with $v$ a (small) constant of order unity to be determined later. In any case, by (5.19), we may easily choose $v$ small enough to ensure

$$vf(\tfrac{1}{2}\eta) \leqq \tfrac{1}{2}\eta \tag{5.30}$$

so that (5.28)–(5.30) imply:

$$\rho\left(\frac{h_m}{\sqrt{\mathscr{A}(h_m)}}, \gamma_w\right) \geqq \tfrac{1}{2}\eta. \tag{5.31}$$

Then by the variational principle and the stability Theorem 5.2, we have

$$g(h_m) \geqq \omega(1 + f(\tfrac{1}{2}\eta))\sqrt{\mathscr{A}(h_m)} \geqq \omega(1 + f(\tfrac{1}{2}\eta))\sqrt{\mathscr{A}(r_\alpha) - (\text{const})Jm^2}. \tag{5.32}$$

Now suppose instead that $\mathscr{A}(h_m) > \mathscr{A}(r_\alpha)$ (i.e. case (a)). Then (5.25) and the observation (5.21) give

$$\rho\left(\frac{h_m}{\sqrt{\mathscr{A}(r_\alpha)}}, \gamma_w\right) \geqq \eta - D_H\left(\frac{h_m}{\sqrt{\mathscr{A}(r_\alpha)}}, \frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}\right). \tag{5.33}$$

Here we can simply use the fact that the Hausdorff distance between a curve and its $m$-skeleton is bounded by $2m$ to obtain

$$\rho\left(\frac{h_m}{\sqrt{\mathscr{A}(r_\alpha)}}, \gamma_w\right) \geqq \eta - \frac{\Delta m}{\sqrt{\mathscr{A}(r_\alpha)}} \geqq \tfrac{1}{2}\eta, \tag{5.34}$$

where the final inequality follows from (5.29) and (5.30) and the fact that $\Delta > 2$. Since $\mathscr{A}(h_m) > \mathscr{A}(r_\alpha)$, the curve in the first argument of $\rho$ in (5.34) has more than unit area. Thus here the variational inequality and the stability Theorem 5.2 directly imply that

$$g(h_m) \geqq \omega(1 + f(\tfrac{1}{2}\eta))\sqrt{\mathscr{A}(r_\alpha)}. \tag{5.35}$$

Evidently case (a) gives a stronger bound on $g(h_m)$ than does case (b'); thus we can use (5.32) in both cases.

By (5.24), (5.32) and the category (ii) condition: $\mathscr{A}(r_\alpha) \geqq [1 - \kappa^*(N)]N/P_\infty$, the minimum $J$ satisfies

$$(J_{\min} + 1)(m + 1) \geqq \omega(1 + f(\tfrac{1}{2}\eta))\sqrt{\frac{N}{P_\infty}[1 - \kappa^*(N)] - (\text{const})J_{\min}m^2}. \tag{5.36}$$

Next, one can use the bounds (5.22), (5.30) and (5.20) to translate (5.36) into an inequality concerning $mJ_{\min}$ alone. (See the analogous manipulations in Eqs. (4.21)–(4.22)). It is then straightforward to demonstrate that if a sufficiently small $v$ is selected in Eq. (5.30), then, for $N$ large enough, the sum in (5.23) is bounded above by

$$\tfrac{1}{2}\exp[-\sigma\omega\sqrt{N/P_\infty}(1 + f(\tfrac{1}{2}\eta))(1 - \kappa^*(N))], \tag{5.37}$$

so that (for $N$ sufficiently large) the probability of observing case (ii) is no more than twice this amount.

Using the worst case scenario (4.57) for lower bounds on $P_{\geqq N}(p)$, it is seen that the conditional probability of case (ii) does not exceed $\exp[-(\text{const})(\kappa^*(N)\sqrt{N})]$, which tends rapidly to zero. As discussed earlier, the conditional probability of case (i) is far smaller than this. Thus the only reasonable prospect is case (iii), which is a subset of the event described in the statement of this theorem. ■

As a corollary to the above theorem, we obtain Theorem 2 of the Introduction:

*Proof of Theorem 2.* We must produce a function $\eta'(N)$, tending monotonically to zero with $N$, such that

$$\sqrt{\frac{1}{N}}\rho(\sqrt{N/P_\infty}\gamma_w, \gamma) \leqq \eta'(N). \tag{5.38}$$

To this end, let $N''$ be chosen so that

$$\sqrt{N''/P_\infty}\left(1 - \frac{2c\log N''}{[N'']^{1/4}}\right) \geqq \sqrt{N/P_\infty}(1 + 2\varepsilon(N)), \tag{5.39}$$

where the constant $c$ is given in Lemma 4.1, and let us define $\eta''$ via

$$N'' = N(1 + \eta''). \tag{5.40}$$

(For example, (5.39) and (5.40) are satisfied if we choose $\eta'' = (\text{const})(N^{-1/4}\log N + \varepsilon(N))$ with a sufficiently large constant.) It follows from Eq. (5.39), Lemma 4.1 and Eq. (4.57) that

$$\mathbf{P}[\bar{r}_\alpha, \mathscr{A}(r_\alpha) > N''/P_\infty | N \leqq C(0) < \infty] \tag{5.41}$$

is negligibly small for $N$ sufficiently large. From this, and the proof of Theorem 5.3, we see that essentially the only $\bar{r}_\alpha$'s which contribute to the event $\{N \leqq C(0) < \infty\}$ are those for which the ring $r_\alpha$ satisfies:

$$\rho\left(\frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}, \gamma_w\right) < \eta \tag{5.42a}$$

and

$$1 - \kappa^*(N) \leqq \sqrt{\frac{P_\infty \mathscr{A}(r_\alpha)}{N}} < 1 + \eta''. \tag{5.42b}$$

However, for such rings, obvious scaling properties and the observation (5.21) imply

$$\begin{aligned}\rho\left(\frac{r_\alpha}{\sqrt{N/P_\infty}}, \gamma_w\right) &= \sqrt{\frac{P_\infty \mathscr{A}(r_\alpha)}{N}}\,\rho\left(\frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}, \sqrt{\frac{N}{P_\infty \mathscr{A}(r_\alpha)}}\gamma_w\right) \\ &\leqq [1 + \eta'']\left[\rho\left(\frac{r_\alpha}{\sqrt{\mathscr{A}(r_\alpha)}}, \gamma_w\right) + D_H\left(\gamma_w, \sqrt{\frac{N}{P_\infty \mathscr{A}(r_\alpha)}}\gamma_w\right)\right] \\ &\leqq [1 + \eta''][\eta + (\kappa^*(N) + \eta'')\,\mathrm{diam}(\gamma_w)].\end{aligned} \tag{5.43}$$

The result follows by defining $\eta'$ via the right-handside of Eq. (5.43). ■

## 6. A Single Droplet Theorem

In this section, we give a second formulation of the Wulff construction, as contained in Theorem 3 of the Introduction. Recall that in Sect. 5 we conditioned on the event $N \leqq |C(0)| < \infty$. Here we focus on a somewhat different set of (unusual) circumstances which result in the formation of a Wulff shape: We consider a large square, $\Lambda_L$, centered at the origin:

$$\Lambda_L = \{(x_1, x_2) \in \mathbb{Z}^2 | -L/2 \leqq x_1, x_2 \leqq +L/2\} \tag{6.1}$$

and examine the configurations in which $\Lambda_L$ has an atypically low infinite cluster density. In particular, we will examine the configurations $F_L(\lambda)$ in which the infinite cluster density in $\Lambda_L$ is depleted by a volume fraction $\lambda$:

$$F_L(\lambda) = \left\{ \omega \left| \frac{|C_\infty \cap \Lambda_L|}{|\Lambda_L|} < (1-\lambda) P_\infty(p) \right. \right\}. \tag{6.2}$$

Our principal result (Theorem 3) is that, under such circumstances, with probability tending to one as $L$ tends to infinity, the system develops a large dual contour which (1) encloses area roughly $\lambda|\Lambda_L|$; (2) is approximately of the Wulff shape, measured in units of the linear dimension of the contour; and (3) contains a single large droplet of roughly $\lambda P_\infty(p)|\Lambda_L|$ sites.

In order to prove Theorem 3, we must first estimate the probability of the event $F_L(\lambda)$ on which we are conditioning. This is a large deviations estimate, which is given in Theorem 6.1. Since the physically relevant case is $0 < \lambda < \lambda_c \equiv [\mathrm{diam}(\gamma_w)]^{-2}$, we restrict our large deviations estimate to these values of $\lambda$. However, it is also possible to obtain estimates for $\lambda \in [\lambda_c, 1]$; see the Remark following the statement of Theorem 6.1. Once the large deviations estimate has been established, the proof of Theorem 3 closely parallels the proofs of various theorems in Sects. 4 and 5; therefore, we omit many of the repetitive details.

The large deviations estimate is:

**Theorem 6.1.** *Consider the two-dimensional Bernoulli bond percolation model on the square lattice with $p > p_c$. Let $F_L(\lambda)$ be the event defined in Eq. (6.2) and take $0 < \lambda < \lambda_c \equiv [\mathrm{diam}(\gamma_w)]^{-2}$. Then*

$$\lim_{L \to \infty} \frac{1}{L} \log \mathbf{P}_p[F_L(\lambda)] = -\sqrt{\lambda}\sigma(p)\omega(p),$$

*where $\sigma(p)$ and $\omega(p)$ are the surface tension and Wulff constant, as defined in Proposition 3.1 and Eq. (3.8).*

*Remark.* It is easy to see that

$$\mathrm{diam}(\gamma_w) \leqq \sqrt{2} \tag{6.3}$$

so that the restriction in this theorem is no worse than $\lambda \in (0, \frac{1}{2})$. Since our $\lambda$ is analogous to the $2\alpha$ in the Ising system studied in [MS] and [DKS] (see discussion in the Introduction), our restriction is equivalent to the Ising restriction $\alpha \in (0, \frac{1}{4})$.

The reason for the restriction $\lambda < \lambda_c \equiv [\mathrm{diam}(\gamma_w)]^{-2}$ is clear: As we will show, if $\lambda < \lambda_c$, then $\Lambda_L$ absorbs the excess sites in finite clusters by forming a single

droplet bounded by the curve $\sqrt{\lambda}L\gamma_w(p)$; the surface energy of this curve is $\sqrt{\lambda}L\sigma(p)\omega(p)$. On the other hand, if $\lambda > \lambda_c$, then the curve $\sqrt{\lambda}L\gamma_w(p)$ will leak out of the box; at this point, it is more efficient for the system to absorb the excess sites in a (single) droplet bounded by a curve which is not simply a scaled $\gamma_w$, but which does fit entirely within $\Lambda_L$. Thus, for any $\lambda \in (0,1)$, we can define $\sqrt{\lambda}\omega_\lambda(p)$ to be the minimum $g$-length of a curve within the unit square enclosing area $\lambda$. It is then possible to show that Theorem 6.1 holds with $\omega(p)$ replaced by $\omega_\lambda(p)$. However, since the case $\lambda > \lambda_c$ is not relevant to the Wulff construction, we do not include the more general result here.

*Proof of Theorem 6.1.* We must produce upper and lower bounds on $\mathbf{P}[F_L(\lambda)]$. We begin with the lower bounds. Our strategy here is (1) to show that with a lower bound of the desired type, there is a single large contour in $\Lambda_L$ containing roughly $\lambda|\Lambda_L|$ sites; and (2) to show that outside this contour, the fraction of sites in the infinite cluster does not deviate significantly from $P_\infty$.

Let us first estimate the probability of a contour containing approximately $\lambda|\Lambda_L|$ sites. As in the proof of Theorem 1.A, we will enumerate all dual rings surrounding the origin: $r_\alpha$, $\alpha = 1, 2, \ldots$. Now, however, rather than considering outermost occupied rings, we define:

$$\underline{r_\alpha} = \{\omega | r_\alpha \text{ is the } \textit{innermost} \text{ occupied dual ring surrounding the origin}\}. \tag{6.4}$$

By a variant of the argument used in the proof of Lemma 4.3, it is not difficult to show that for $\varepsilon \in \mathbb{R}^+$, $\exists \delta'(\varepsilon) \geqq \frac{1}{2}\delta(\frac{1}{2}\varepsilon)$ such that for $N$ large, the probability of observing the event $\underline{r_\alpha}$ with $\mathscr{A}(r_\alpha) \geqq [1 + \delta'(\varepsilon)]N$ is larger than $\exp[-(1+\varepsilon)\sigma(p)\cdot\omega(p)\sqrt{N}]$. Indeed, first using Lemma 4.3, one produces the event $\bar{r}_\beta$ (cf. Eq. (4.39)) for $r_\beta$ outside some convex polygonal approximation, $\zeta$, to $\gamma_w$ of area exceeding $(1 + \delta(\frac{1}{2}\varepsilon))N$. As in the proof of Theorem 1.B, we may take $\mathscr{A}(r_\beta) < (\text{const})N$, for some sufficiently large constant, without significantly altering the probabilistic estimate. Then, for some $n$ which is itself large, but only on the order of $\log N$, one ensures that with probability tending to one (as $e^{-O(n)}$), no dual site which is inside the polygon $\zeta$ and a distance further than $n$ from it belongs to a dual path of linear extent as large as $n$. This easily gives the event $\underline{r_\alpha}$ for some $r_\alpha$ with $\mathscr{A}(r_\alpha) \geqq [(1+\delta)N - (\text{const})\sqrt{N}\log N]$ at a cost no larger than $(\text{const})\exp[-(1+\frac{1}{2}\varepsilon)\cdot\sigma(p)\sqrt{N}]$. The desired statement is now seen to hold for all $N$ large enough.

Applying the above result with $N = \lambda L^2$, we see that if $\lambda < \lambda_c$, and $L$ is large enough, then with probability exceeding

$$\exp[-(1+\varepsilon)\sqrt{\lambda}\sigma(p)\omega(p)L] \tag{6.5}$$

the event $\underline{r_\alpha}$ occurs for some $r_\alpha$ with $|\text{Int}(r_\alpha) \cap \Lambda_L| \geqq (1 + \delta')\lambda|\Lambda_L|$.

Next, we must show that in $\Lambda_L \backslash \text{Int}(r_\alpha)$, the infinite cluster density does not exceed $P_\infty$ by more than enough to compensate for the depletion of the infinite cluster density within $r_\alpha$. To this end, we observe that if $r_\alpha$ is the *innermost* circuit surrounding the polygon $\zeta$, here it is the region *outside* $r_\alpha$ that is unconditioned. Thus we can apply the reasoning of previous theorems to show that, with probability tending rapidly to one, the infinite cluster in $\Lambda_L \backslash \text{Int}(r_\alpha)$ will not be dense enough to prevent $F_L(\lambda)$ from occurring. Explicitly, we pick $n$ of the order

of $\log L$, and note that, for any $c$, the event $\{f_{\leqq n}(\Lambda_L \backslash \mathrm{Int}(r_\alpha)) - P_{\leqq n} > c\}$ is positively correlated (in the sense of Harris-FKG) with the event that $r_\alpha$ is the innermost such circuit. Then, using Lemma 4.2, it can be shown that with probability tending to one at least as fast as $\exp[-(\mathrm{const})(\delta'^2|\Lambda_L|/n^2)]$, enough of the remaining sites belong to clusters smaller than $n^2$ to produce the event $F_L(\lambda)$. This completes the proof of the lower bound.

Now let us establish the upper bound on $\mathbf{P}[F_L(\lambda)]$. We begin with the observation, mentioned earlier, that in any configuration $\omega$, any site which does not belong to the infinite cluster is, w.p. 1, surrounded by a finite contour of dual bonds. Let us denote by $\Gamma(\omega)$ the collection of outermost contours in $\omega$ whose interiors have relatively large intersection with the box $\Lambda_L$:

$$\Gamma(\omega) = \{\gamma_1(\omega), \ldots, \gamma_{n(\omega)}(\omega) \,|\, \forall j, 1 \leqq j \leqq n(\omega), |\mathrm{Int}(\gamma_j) \cap \Lambda_L| \geqq (D \log L)^2, \text{ and } \gamma_j \text{ is the outermost dual contour surrounding some point in } \Lambda_L\}. \tag{6.6}$$

In the above $D$ is a large constant to be specified later. Let $\kappa^*(L)$ be the function defined in Eq. (4.41), and define the event

$$Q = Q(\lambda; L) = \left\{\omega \,|\, |\Lambda_L \backslash \bigcup_j \mathrm{Int}(\gamma_j)| \leqq (1-\lambda)|\Lambda_L|[1 + 2\kappa^*(L^2)]\right\}. \tag{6.7}$$

From the estimates of Lemma 4.2, it should be plausible that unless $Q$ occurs, not enough volume has been isolated in large clusters to permit $F_L(\lambda)$ to occur with any reasonable probability.

In order to explicitly prove the above statement, let us first pause to consider the following percolation-type problem. Let $B \subset \mathbb{Z}^2$ be any collection of sites. We will focus on those configurations $\Omega(B^c)$ of bonds with both endpoints in $B^c$. The sites $x \in B^c$ of any such configuration fall into three disjoint categories:

(1) $|C(x)| = \infty$;
(2) $|C(x)| < \infty$ and $C(x) \cap \partial B \neq \varnothing$;
(3) $|C(x)| < \infty$ and $C(x) \cap \partial B = \varnothing$.

We say that a site $x$ in category (3) is in a "truly finite cluster," in the sense that $C(x)$ is unchanged by altering the status of any bond emanating from $B$ (i.e. any bond with one endpoint in $B$). Now take $A \subset B^c$, $n$ an integer, and $\alpha < 1$, and consider the event:

$$\{f^{③}_{\leqq n}(A) \geqq \alpha\} = \{\omega \in \Omega(B^c) \,|\, \text{the fraction of sites in } A \text{ belonging to truly finite clusters of size no larger that } n \text{ exceeds } \alpha\}. \tag{6.8}$$

Although the problem of directly estimating $\mathbf{P}_{p[\Omega(B^c)]}[f^{③}_{\leqq n}(A) \geqq \alpha]$ may seem formidable, it is obvious that this is bounded by the probability of $\{f_{\leqq n}(A) \geqq \alpha\}$ in the usual percolation problem:

$$\mathbf{P}_{p[\Omega(B^c)]}[f^{③}_{\leqq n}(A) \geqq \alpha] \leqq \mathbf{P}_p[f_{\leqq n}(A) \geqq \alpha] \tag{6.9}$$

To see Eq. (6.9), one need only observe that a Bernoulli configuration on the full lattice can be constructed in a two-step process: first draw a Bernoulli configuration in $\Omega(B^c)$, and then independently draw a Bernoulli configuration on the remaining

lattice. If the event $\{f^{③}_{\leqq n}(A) \geqq \alpha\}$ occurs in the first step of the process, then (by definition) the additional bonds cannot decrease the fraction of sites in clusters smaller than $n$.

Let us now consider the event $Q^c$. Denote by $\Gamma_j$ any collection of dual contours such that the event $\Gamma_j^* = \{\omega | \Gamma(\omega) = \Gamma_j\}$ implies the event $Q^c$. Letting $\mathbb{L}(\Gamma_j)$ denote the set of sites inside the contours of $\Gamma_j$, it is seen that we may describe the configurations outside $\mathbb{L}(\Gamma_j)$ as a $\Omega(\mathbb{L}(\Gamma_j)^c)$-percolation process subject to the two constraints:

($\alpha$) There are no large contours in $\Lambda_L$ outside the contours of $\Gamma_j$ (where "large" is specified by the condition in Eq. (6.6)).

($\beta$) All sites in $\partial\mathbb{L}(\Gamma_j)$ are connected to infinity (outside $\mathbb{L}(\Gamma_j)$).

The second constraint follows from the fact that $\Gamma(\omega)$ is a set of outermost contours. We denote the events in conditions ($\alpha$) and ($\beta$) by $\alpha_j$ and $\beta_j$, respectively. Thus if we define $\Xi_j$ to be the event that the contours of $\Gamma_j$ are actually formed, we may write:

$$\Gamma_j^* = \Xi_j \cap \alpha_j \cap \beta_j. \tag{6.10}$$

It is worth observing that the events $\Xi_j$ and $\alpha_j \cap \beta_j$ are independent, and that both $\alpha_j$ and $\beta_j$ are FKG positive events.

Now observe that, given the event $\Gamma_j^*$, the only possible mechanism for $F_L(\lambda)$ to occur is that the fraction of sites belonging to small clusters (i.e. clusters whose intersection with $\Lambda_L$ is less than $(D \log L)^2$) far exceeds its typical value. Let us denote by $A_j$ the sites of $\Lambda_L \setminus \mathbb{L}(\Gamma_j)$ a distance further than $(D \log L)^2$ from $\partial\Lambda_L$. We claim that if $\omega \in \Gamma_j^*$, the event $F_L(\lambda)$ will not occur in $\omega$ unless, in that part of $\omega$ belonging to $\Omega(\mathbb{L}(\Gamma_j)^c)$, the event $\{f_{\leqq (D \log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*\}$ occurs.

The above statement can be verified as follows: Each site in $A_j$ either belongs to the infinite cluster or is in a cluster whose intersection with $\Lambda_L$ is smaller than $(D \log L)^2$. Note that, by the definition of $A_j$, any site in the second category is in a cluster whose total size is actually smaller than $(D \log L)^2$. According to the condition $F_L(\lambda)$, there cannot be more than $(1 - \lambda) P_\infty |\Lambda_L|$ sites in the first category. Thus we find

$$f_{\leqq (D \log L)^2}(A_j) \geqq 1 - (1 - \lambda) P_\infty |\Lambda_L| / |A_j|. \tag{6.11}$$

On the other hand, any site in $\Lambda_L \setminus A_j$ is either a distance $(D \log L)^2$ from the boundary—which accounts for fewer than $(4L)(D \log L)^2$ sites—or is sealed in a large contour—which accounts for fewer than $|\Lambda_L|[\lambda - 2(1 - \lambda)\kappa^*(L^2)]$ sites. Evidently, if $\omega \in F_L(\lambda) \cap \Gamma_j^*$,

$$\begin{aligned} |A_j| &\geqq |\Lambda_L|[1 - [\lambda - 2(1 - \lambda)\kappa^*(L^2)] - 4(D \log L)^2 / L] \\ &= (1 - \lambda)|\Lambda_L|[1 + 2\kappa^*(L^2) - 4(D \log L)^2 / [(1 - \lambda)L]]. \end{aligned} \tag{6.12}$$

For $L$ sufficiently large, Eqs. (6.11) and (6.12) imply that the desired event occurs.

Now recall categories (1), (2) and (3), defined earlier in the context of a general percolation problem on a "depleted" lattice. It should be observed that because the event $\Gamma_j^*$ includes the event $\beta_j$ (which connects all sites in $\partial\mathbb{L}(\Gamma_j)$ to infinity),

there is no category (2) in this system—i.e. all finite clusters are truly finite. It is thus clear that the events

$$\Gamma_j^* \cap \{f_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*\} \tag{6.13a}$$

and

$$\begin{aligned} &\Gamma_j^* \cap \{f^{③}_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*\} \\ &\quad \equiv \Xi_j \cap \alpha_j \cap \beta_j \cap \{f^{③}_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*\} \end{aligned} \tag{6.13b}$$

are equivalent. We note that the event $\Xi_j$ is independent from the other three events on the right-hand side of (6.13b), and that $\{f^{③}_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*\}$ is FKG negative, while, as previously mentioned, $\alpha_j \cap \beta_j$ is FKG positive. Thus, by the Harris-FKG inequality, we have

$$\begin{aligned} &\mathbf{P}_p[F_L(\lambda) \cap \Gamma_j^*] \\ &\quad \leqq \mathbf{P}_p[\Xi_j] \mathbf{P}_{p[\Omega(\mathbb{L}(\Gamma_j)^c)]}[\alpha_j \cap \beta_j] \mathbf{P}_{p[\Omega(\mathbb{L}(\Gamma_j)^c)]}[f^{③}_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*] \\ &\quad = \mathbf{P}_p[\Gamma_j^*] \mathbf{P}_{p[\Omega(\mathbb{L}(\Gamma_j)^c)]}[f^{③}_{\leqq (D\log L)^2}(A_j) \geqq 1 - P_\infty + P_\infty \kappa^*]. \end{aligned} \tag{6.14}$$

Summing over all $j$ (and observing that the events $\Gamma_j^*$ partition $Q^c$), we can use (6.9) and Lemma 4.2 to conclude that the probability of observing $F_L(\lambda)$ in $Q^c$ is negligibly small relative to the anticipated upper bound.

We can therefore obtain an upper bound on $\mathbf{P}_p[F_L(\lambda)]$ simply by estimating the probability of the event $Q$. We start by recalling the upper bound on $\mathbf{P}[\mathscr{R}(A_0)]$ of Lemma 4.1. It is easy to see that (if $A_0$ is not terribly small) this upper bound is log convex and monotone, i.e. for $x$ large enough,

$$u(x) \equiv \sigma(p)\omega(p)\sqrt{x}\left(1 - \frac{c\log x}{x^{1/4}}\right) \tag{6.15}$$

is concave and monotone. Indeed, since $u(x)$ is asymptotically just a square root, it is clear that for $x$ large enough

$$u(x) - xu'(x) \geqq \tfrac{1}{4}u(x). \tag{6.16}$$

As we will see later, it is the concavity of the function $u$ which forces individual clusters of moderate size to coalesce into a single large cluster.

We again divide things up according to which large outer contours are present. Explicitly, we denote by $\Gamma_j$ any collection of dual contours such that the event $\hat{\Gamma_j} = \{\omega | \Gamma(\omega) = \Gamma_j\}$ implies the event $Q$:

$$\mathbf{P}[Q] = \sum_j \mathbf{P}[\hat{\Gamma_j}]. \tag{6.17}$$

However, this time we need only estimate the probability of observing the rings alone:

$$\mathbf{P}[\hat{\Gamma_j}] \leqq \mathbf{P}[\Xi_j]. \tag{6.18}$$

To estimate this, we consider the event:

$$B_k(N) = \{\omega | \exists k+1 \text{ disjoint occupied dual rings } r_0, \ldots, r_k, \text{ none of which are contained in one another, with } |\mathrm{Int}(r_j) \cap \Lambda_L| \geqq (D \log L)^2 \text{ and } \sum_j |\mathrm{Int}(r_j) \cap \Lambda_L| = N\}. \tag{6.19}$$

It is not terribly difficult to see that

$$\sum_{\substack{0 \leqq k \leqq k_{\max}(N) \\ N_0 \leqq N \leqq |\Lambda_L|}} \mathbf{P}[B_k(N)] = \sum_j \mathbf{P}[\Xi_j], \tag{6.20}$$

where

$$k_{\max}(N) = [N/(D \log L)^2] - 1, \tag{6.21}$$

and

$$N_0 = |\Lambda_L|[\lambda - 2(1-\lambda)\kappa^*(L^2)] \equiv \lambda |\Lambda_L|(1 - \kappa_L^*(\lambda)) \tag{6.22}$$

is the minimum enclosed area consistent with the event $Q$.

Let us define $a_j \equiv |\mathrm{Int}(r_j) \cap \Lambda_L|$ and recall the definition (6.15) of the function $u$. By the van den Berg–Kesten inequality and Lemma 4.1, we have

$$\mathbf{P}(B_k(N)) \leqq \sum_{\substack{\{a_j\}: \Sigma a_j = N \\ a_j \geqq (D \log L)^2}} |\Lambda_L|^{k+1} \exp\left\{ - \sum_j u(a_j) \right\}, \tag{6.23}$$

where $|\Lambda_L|^{k+1}$ accounts for all possible placements of the rings. Now, by monotonicity and concavity of $u$, it is clear that the sum in the argument of the exponent in (6.23) is maximized by putting as much mass as possible in a single ring. Thus, using $N^{k+1}$ as an (over)estimate of the number of ways to partition $N$, we have

$$\mathbf{P}[B_k(N)] \leqq N^{k+1} |\Lambda_L|^{k+1} \exp\{ - [u(N - k(D \log L)^2) + ku((D \log L)^2)]\}, \tag{6.24}$$

where we have tacitly assumed $N \geqq (k+1)(D \log L)^2$. We degrade the estimate further by saying $N \leqq |\Lambda_L|$, and thus $(N|\Lambda_L|)^{k+1} \leqq |\Lambda_L|^2 e^{4k \log L}$. Permitting $k$ to assume continuous values in the allowed range, let us attempt to minimize (the negative of) the function in the exponent in (6.24). The derivative of this function with respect to $k$ is

$$-(D \log L)^2 u'(N - k(D \log L)^2) + u((D \log L)^2) - 4 \log L. \tag{6.25}$$

However, $u'$ is decreasing (by convexity) and $N - k(D \log L)^2 \geqq (D \log L)^2$. Thus, using (6.16), this derivative is bounded below by

$$u((D \log L)^2) - (D \log L)^2 u'((D \log L)^2) - 4 \log L \geqq \tfrac{1}{4} u((D \log L)^2) - 4 \log L > 0, \tag{6.26}$$

provided that $D$ has been chosen large enough. Thus the worst case occurs for $k = 0$, for which we have

$$\mathbf{P}[B_k(N)] \leqq (\text{const}) L^4 \exp\{ - u(N)\}. \tag{6.27}$$

Since for each $N$, there are only of order $L$ allowed values of $k$, we may freely sum (6.20) first over $k$, then over $N \geqq N_0$ to obtain an upper bound of the stated form. ■

We can now establish the single droplet result:

*Proof of Theorem 3.* Take $p > p_c$ and condition on the event $F_L(\lambda)$. Then according to the statement of Theorem 3, we must find functions $\phi_L(\lambda)$, $\zeta_L(\lambda)$ and $\mu_L(\lambda)$ tending monotonically to zero as $L\uparrow\infty$, such that, with conditional probability tending rapidly to one with $L$, there is an occupied circuit of dual bonds, $\gamma$, in $\Lambda_L$ satisfying

$(a)$ $\mathscr{A}(\gamma) \geqq [1-\phi_L(\lambda)]\lambda|\Lambda_L|$,

$(b)$ $\rho\left(\gamma_w, \dfrac{\gamma}{\sqrt{\mathscr{A}(\gamma)}}\right) \leqq \zeta_L(\lambda)$,

$(c)$ $\mathrm{Int}(\gamma)$ contains a connected cluster of size exceeding $P_\infty(p)[1-\mu_L(\lambda)]\lambda|\Lambda_L|$.

We first establish property $(a)$. To this end, we note that by Theorem 6.1, there exists a sequence $\psi_L(\lambda) = \psi_L(\lambda; p)$ with $\psi_L \downarrow 0$ as $L\uparrow\infty$ satisfying

$$\exp(-[1+\psi_L(\lambda)][\sqrt{\lambda}\sigma\omega L]) \leqq \mathbf{P}[F_L(\lambda)] \leqq \exp(-[1-\psi_L(\lambda)][\sqrt{\lambda}\sigma\omega L]). \tag{6.28}$$

Let us now define another positive monotone sequence, $\phi_L(\lambda) = \phi_L(\lambda; p)$, with $\phi_L \downarrow 0$ as $L\uparrow\infty$ satisfying

$$\phi_L(\lambda) \geqq 2\max\{\psi_L(\lambda), \kappa_L^*(\lambda)\}, \tag{6.29}$$

where $\kappa_L^*(\lambda)$ was defined in Eq. (6.22). We claim that any such function $\phi_L(\lambda)$ satisfies $(a)$.

To prove the above claim, we first note that, according to the arguments in the proof of Theorem 6.1, the condition $F_L(\lambda)$ means that we may restrict attention to configurations in $B_k(N)$ for $N_0 \leqq N \leqq |\Lambda_L|$ (cf. Eqs. (6.19) and (6.22)). Let us estimate the probability that $B_k(N)$ occurs, but that we do *not* have a ring enclosing sufficient area to imply the result $(a)$, i.e. consider the event:

$$\begin{aligned}(a)^c_{N,k} = \{\omega\in B_k(N) | &\text{ all occupied dual rings } r \text{ satisfy} \\ &|\mathrm{Int}(r)\cap\Lambda_L| \leqq (1-\phi_L)\lambda|\Lambda_L|\}.\end{aligned} \tag{6.30}$$

As in the bound on $\mathbf{P}(B_k(N))$ in Theorem 6.1, $\mathbf{P}[(a)^c_{N,k}]$ can be estimated using the van den Berg–Kesten inequality, Lemma 4.1 and concavity of the function $u$. Now, however, since the event $(a)^c_{N,k}$ imposes a maximum ring size, concavity implies that the optimal configurations will have the maximum number $m = m(N)\in\mathbb{Z}^+$ of large rings—as large as the constraint permits—to absorb most of the area $N$. There will then be a single ring of intermediate scale, containing as much additional area as possible. The remaining $k-m$ rings will have interiors as small as permitted. This translates into the estimate

$$\begin{aligned}\mathbf{P}[(a)^c_{N,k}] \leqq N^{k+1}|\Lambda_L|^{k+1}\exp\{-[&mu((1-\phi_L)\lambda|\Lambda_L|) + (k-m)u((D\log L)^2) \\ &+ u(N - m(1-\phi_L)\lambda|\Lambda_L| - (k-m)(D\log L)^2)]\}.\end{aligned} \tag{6.31}$$

Summing over $k$ then $N$, it is found that any excess of $N$ or $k$ over the minimum allowed values is unnecessarily costly; essentially the entire sum is contained in

the terms $N = N_0$ (and thus $m = 1$) and $k = 1$ (cf. Eqs. (6.24)–(6.27)):

$$\mathbf{P}[(a)^c] \leqq \sum_{\substack{N_0 \leqq N \leqq |\Lambda_L| \\ m(N) \leqq k \leqq k_{\max}(N)}} \mathbf{P}[(a)^c_{N,k}]$$

$$\leqq (\text{const.})|\Lambda_L|^4 \exp\{-[u((1-\phi_L)\lambda|\Lambda_L|) + u((\phi_L - \kappa_L^*)\lambda|\Lambda_L|)]\}. \quad (6.32)$$

Examining the definitions (6.15), (6.22) and (6.29) of $u(x)$, $\kappa_L^*$ and $\phi_L$, it is seen that $\mathbf{P}[(a)^c]$ is very small relative to $\mathbf{P}[F_L(\lambda)]$. This establishes the first claim.

Results ($b$) and ($c$) can now be taken over from previous derivations. Given the existence of the circuit $\gamma$, as $\mathscr{A}(\gamma) \to \infty$, the estimates of Theorem 5.3 which demonstrate that "case (ii)" is highly unlikely relative to "case (iii)" can be applied directly. Indeed, we need only construct our "$\eta$" (here denoted by $\zeta_L(\lambda)$) and $\psi_L(\lambda)$, and then use an identical argument. Now, given the existence of *this* ring, there must be an ample unconditioned region of a fairly regular shape (i.e. convex), providing us with an analogue of Lemma 4.3. Translating *mutatis mutandis* the proof of Theorem 1.B, one can show that, within this region, there is a large cluster of the stated specifications. ■

## Appendix

Here we provide proofs of a few "obvious" geometrical facts which were used in our proof of stability of the Wulff minimum (Theorem 5.2).

### *A.1. Uncrossing of Polygons*

**Proposition A.1.** *Let $v_1, \ldots, v_k \in \mathbb{R}^2$ denote the vertices (extreme points) of a convex polygon. We assume that the vertices are labeled in (cyclic) order,* i.e. *the curve $\gamma_{(v_j)}$ composed of the segments joining successive vertices is self-avoiding. Let $(v'_j)$ denote any reordering of the vertices, and $\gamma_{(v'_j)}$ the curve passing through the $\{v_j\}$ in the new order. Then*

$$g(\gamma_{(v_j)}) \leqq g(\gamma_{(v'_j)}), \quad (A.1)$$

*where $g$ is the norm constructed in Proposition 3.2.*

*Proof.* If the curve $\gamma_{(v'_j)}$ is composed only of line segments joining neighbors in the original ordering (i.e. if $\gamma_{(v'_j)}$ is a reparameterization of $\gamma_{(v_j)}$), then there is nothing to prove. Otherwise, the curve $\gamma_{(v'_j)}$ contains crossing lines: Indeed, suppose that $v_j$ is connected to $v_{j+s}$ with $1 < s < k-1$. Since each vertex belongs to two line segments, there must be an $r < s$ and an $r' > s$ such that $v_{j+r}$ is connected to $v_{j+r'}$. By convexity, the two segments will cross in the interior of the polygon. (See Fig. 2.)

Let $X$ denote the number of crosses in the curve $\gamma_{(v'_j)}$. We claim that there is yet another ordering, $(v''_j)$, with no more than $X-1$ crosses such that

$$g(\gamma_{(v''_j)}) \leqq g(\gamma_{(v'_j)}). \quad (A.2)$$

Indeed, suppose that $v_j$ is connected to $v_{j+s}$, while $v_{j+r}$ is connected to $v_{j+r'}$ with $1 \leqq r < s < r' \leqq k-1$, as illustrated in Fig. 2. Clearly, we may "uncross" the diagram in one of two ways: either attach $v_j$ to $v_{j+r}$ and $v_{j+s}$ to $v_{j+r'}$, or attach $v_j$ to $v_{j+r'}$

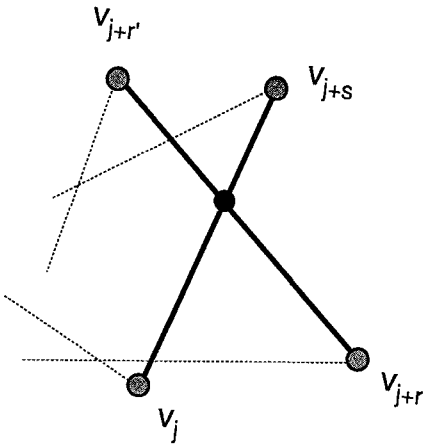


**Fig. 2.** Crossing segments

and $v_{j+s}$ to $v_{j+r}$. Having done so, either way, the result is a collection of line segments which, by the triangle inequality, have total $g$-length no longer than $g(\gamma_{(v_j')})$. However, this does not quite establish the conclusion in Eq. (A.2). Indeed, we must demonstrate that:

(i) the resulting line segments may be assembled into a single closed curve; and
(ii) the uncrossing procedure strictly decreases the number of crosses.

Concerning the first issue, it is seen that before and after the uncrossing, each vertex is connected to two lines; thus we can still construct closed curves from the resulting segments. However, this does not preclude the possibility that, by the uncrossing, we have broken the original curve into two disconnected pieces, i.e. two sets of vertices with no interconnecting segments. To see that this problem may be circumvented, let us exhibit the only possible mechanism for disconnecting the original curve. Suppose that $\{v_j\}$ can be divided into two sets, $A$ and $B$, with $A \cap B = \varnothing$ and (say) $v_j, v_{j+r} \in A$, while $v_{j+s}, v_{j+r'} \in B$. Now if only the segments between $v_j$ and $v_{j+s}$ and between $v_{j+r}$ and $v_{j+r'}$ connect the sets $A$ and $B$ in the original curve $\gamma_{(v_j')}$, then the curve will split if we use the first choice of how to uncross. However, before we contemplate such a move, it is worth observing that since $\gamma_{(v_j')}$ is a closed curve, it must be the case that $v_{j+s}$ and $v_{j+r'}$ are the endpoints of a curve threading through all the other vertices of $B$. A similar statement holds for the vertices $v_j$ and $v_{j+r}$. Thus the alternative choice for uncrossing necessarily implies that the resulting curve will have two lines connecting $A$ to $B$.

The second issue is dispensed with by means of an elementary convexity argument. Suppose that we choose to uncross via the first option, so that $v_j$ ends up connected to $v_{j+r}$. Obviously, this procedure removes the dark cross shown in Figure A.1. Let us show that it does not introduce any new crosses, except if it also removes at least a compensating number of crosses from the original curve. Consider the two "half line segments" starting at $v_j$ and $v_{j+r}$ (before we uncross), as well as the final segment connecting $v_j$ to $v_{j+r}$. Denote by $C$ the set of vertices $v_{j+1}, \ldots, v_{j+r-1}$, and by $D$ the set $v_{j+r+1}, \ldots, v_{j-1}$. It is clear, from the convex

arrangement of the vertices, that any line joining two points in the $C$ group crosses none of the three line segments under question. On the other hand, it is conceivable that a segment joining a pair of vertices in the $D$ group touches one or both of the half segments; it certainly does not touch the line joining $v_j$ and $v_{j+r}$. In this case, the number of crosses can only decrease. Finally, if a $C$ vertex is connected to a $D$ vertex, there is an inevitable single intersection with one of the two half segments before uncrossing, as well as an inevitable single intersection with the segment connecting $v_j$ to $v_{j+r}$ after uncrossing. Thus there is no net change in the number of crosses in this group.

Having verified the statement (A.2), it seen that if the uncrossing procedure is repeated (no more than) $X$ times, the desired inequality is established. ■

*A.2. A g-based Hausdorff Measure.* There is a classic result which states that if $\gamma$ is a rectifiable plane curve of (Euclidean) length $\mathscr{L}(\gamma)$ and $\mu(-)$ is the standard one-dimensional Hausdorff measure, then

$$\mu(\gamma) \leqq \mathscr{L}(\gamma). \tag{A.3}$$

Furthermore, the inequality in (A.3) is an equality if $\gamma$ is a self-avoiding curve. By analogy, if $\mu_g(-)$ is the one-dimensional Hausdorff measure constructed from the metric $g$, one would expect

$$\mu_g(\gamma) \leqq g(\gamma) \tag{A.4}$$

with equality if $\gamma$ is self-avoiding. The derivation of this result involves only minor modifications of the standard derivation of (A.3). The result is also a consequence of Theorem 2.10.13 of [F]. Nevertheless, for the sake of completeness, we will present it as a formal proposition.

*Definition A.* Let $g(x)$ be defined as in Proposition 3.2 and $U_\varepsilon(x) = \{y \in \mathbb{R}^2 | g(x-y) < \varepsilon\}$. For $A \subset \mathbb{R}^2$, we define

$$\mu_{g,\varepsilon}(A) = \inf\left\{2\sum_{j=1}^{N} \varepsilon_j \,|\, A \subset \bigcup_{j=1}^{N} U_{\varepsilon_j};\ \varepsilon_j < \varepsilon\right\}, \tag{A.5a}$$

where the infimum extends over all countable coverings of $A$ by $g$-balls of radius less then $\varepsilon$. The $\mu_{g,\varepsilon}(A)$ are clearly monotone in $\varepsilon$, so that

$$\mu_g(A) = \lim_{\varepsilon \to 0} \mu_{g,\varepsilon}(A) \tag{A.5b}$$

exists. The function $\mu_g(-)$ is called the *one-dimensional g-Hausdorff measure.*

**Lemma A.2.** *Let $\gamma: [0, T] \to \mathbb{R}^2$ be a rectifiable curve. Then*

$$\mu_g(\gamma) \leqq g(\gamma)$$

*(where, as usual, we also use $\gamma$ to denote the range of $\gamma$). Furthermore, if $\gamma$ is self-avoiding,*

$$\mu_g(\gamma) = g(\gamma).$$

*Proof.* We first show that, in general,

$$g(\gamma) \geqq \mu_g(\gamma). \tag{A.6}$$

To this end, take $\gamma$ and denote by $(x_1, \ldots, x_{n(\varepsilon)})$ the sequence of points along $\gamma$, with $x_1 = \gamma(0)$ and $x_{n(\varepsilon)} = \gamma(T)$, such that the $g$-length of the portion of $\gamma$ between $x_i$ and $x_{i+1}$ is equal to $2\varepsilon$, except for the final pair, $x_{n-1}$ and $x_n$, which will in general be separated by a $g$-length along $\gamma$ of less than $2\varepsilon$. Then it is clear that $\gamma$ is covered by the union of balls of $g$-radius $= \varepsilon$ centered at the points $(x_i)$. Up to an additive factor of $3\varepsilon$, the sum of the diameters of the balls in this cover provides an upper bound on $\mu_{g,\varepsilon}(\gamma)$. We have

$$g(\gamma) + 3\varepsilon \geqq \mu_{g,\varepsilon}(\gamma), \tag{A.7}$$

which establishes (A.6).

It remains to be shown that if $g$ is self-avoiding, then $g(\gamma) = \mu_g(\gamma)$. To this end, we first establish the intermediate step

$$\mu_g(\gamma) \geqq g(\gamma(0) - \gamma(T)) \tag{A.8}$$

which holds regardless of any stipulations concerning self-avoidance. It is trivial to show that (A.8) holds if $\gamma$ is a straight line; indeed, in this case, for each $\varepsilon$, $\mu_{g,\varepsilon}(\gamma) = g(\gamma(0) - \gamma(T))$ $(\equiv g(\gamma)$ for a *non-retracing* straight line). In general, let $\gamma$ denote any rectifiable curve assumed, with no loss of generality, to have $\gamma(0) = 0$, and let $U_{\varepsilon_j}(x_j)$, $j = 1, 2, \ldots, N$ be a collection of $g$-balls with $\varepsilon_j < \varepsilon$ and

$$\gamma \subset \bigcup_{j=1}^{N} U_{\varepsilon_j}(x_j). \tag{A.9}$$

For each $x \in \mathbb{R}^2$, consider the level curves

$$\mathbf{C}_x = \{y \in \mathbb{R}^2 \mid g(y) = g(x)\}, \tag{A.10}$$

and denote by $s_j$ the intersection of $\mathbf{C}_{x_j}$ with the line joining 0 $(\equiv \gamma(0))$ and $\gamma(T)$. (See Fig. 3.) Finally, denote by $L$ the straight line segment which runs between 0

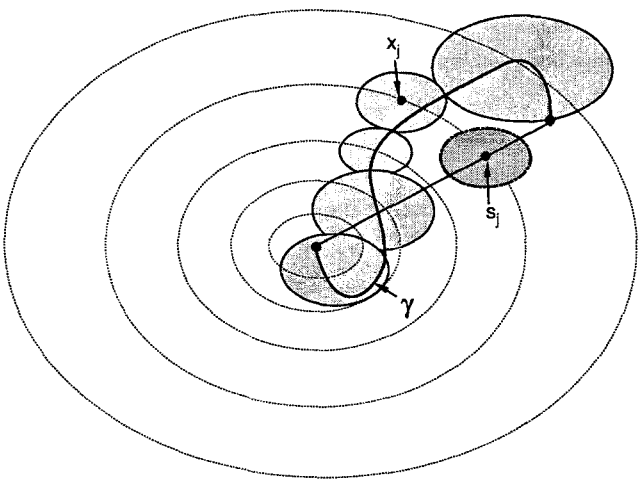


**Fig. 3.** A $g$-covering of the curve $\gamma$

and $\gamma(T)$. We claim that

$$L \subset \bigcup_{j=1}^{N} U_{\varepsilon_j}(s_j), \tag{A.11}$$

from which (A.8) follows immediately. Indeed, for any $y \in L$, $\exists\, y^* \in \mathbf{C}_y$ such that $y^* \in U_{\varepsilon_j}(x_j)$ for some element $j$ of the covering—in particular, any element of $\mathbf{C}_y \cap \gamma$ is such a point $y^*$. But then

$$\begin{aligned} \varepsilon_j &\geqq g(y^* - x_j) \\ &\geqq |g(y^*) - g(x_j)| \\ &= |g(y) - g(s_j)| \\ &= g(y - s_j). \end{aligned} \tag{A.12}$$

Here the third line follows from the fact that these points lie along level $g$-curves, and the fourth line is a consequence of the fact that $y$ and $s_j$ are connected by the straight line $L$. This establishes (A.11).

Given (A.8) and (A.6), the equality $g(\gamma) = \mu_g(\gamma)$ for self-avoiding curves follows easily by considering the Hausdorff measure of polygonal approximations to $\gamma$. ■

**Corollary.** *Let $\gamma \in \mathscr{K} \setminus \mathscr{J}$ denote a "convex" curve (as defined in the proof of Theorem 5.2) and let $\gamma_H$ be the boundary of the convex hull of $\gamma$: $\gamma_H = \partial H(\gamma)$. If $\gamma_H \neq \gamma$ in the sense that $\exists x \in \gamma$ such that $\min_{y \in \gamma_H} g(x - y) \equiv a > 0$, then*

$$g(\gamma) \geqq g(\gamma_H) + 2a.$$

*Proof.* This corollary basically follows from the fact that $\gamma_H$ is self-avoiding and that the Hausdorff measure is an outer measure. Explicitly, let us assume that $\gamma(0) \in \gamma_H$ and denote by $t_x$ the first time when $\gamma(t_x) = x$. We define

$$t_x^- = \inf\{s \mid s < t_x, \gamma(s, t_x) \cap \gamma_H = \varnothing\} \tag{A.13a}$$

and

$$t_x^+ = \sup\{s \mid s > t_x, \gamma(s, t_x) \cap \gamma_H = \varnothing\}. \tag{A.13b}$$

Then

$$\begin{aligned} g(\gamma) &= g(\gamma(0, t_x^-)) + g(\gamma(t_x^-, t_x)) + g(\gamma(t_x, t_x^+)) + g(\gamma(t_x^+, T)) \\ &\geqq g(\gamma(0, t_x^-)) + 2a + g(\gamma(t_x^+, T)) \\ &\geqq 2a + \mu_g(\gamma(0, t_x^-) \cup \gamma(t_x^+, T)) \\ &\geqq 2a + \mu_g(\gamma_H) \\ &= 2a + g(\gamma_H). \quad \blacksquare \end{aligned} \tag{A.14}$$

*Acknowledgements.* Two of us (J. T. C. and L. C.) would like to thank J. E. Avron, R. L. Dobrushin, R. Kotecky and S. B. Shlosman and J. E. Taylor for enjoyable discussions on various facets of the Wulff construction.

## References

[A] Abraham, D. B.: Surface structures and phase transitions—Exact results. In: Phase Transitions and Critical Phenomena, Vol. 10. Domb, C., Lebowitz, J. L. (eds.), London: Academic Press 1987

[ACCFR] Aizenman, M., Chayes, J. T., Chayes, L., Fröhlich, J., Russo, L.: On a sharp transition from area law to perimeter law in a system of random surfaces. Commun. Math. Phys. **92**, 19 (1983)

[ACCN] 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 (1988)

[ADS] Aizenman, M., Delyon, F., Souillard, B.: Lower bounds on the cluster size distribution. J. Stat. Phys. **23**, 267 (1980)

[AN] Aizenman, M., Newman, C. M.: Tree graph inequalities and critical behavior in percolation models. J. Stat. Phys. **36**, 107 (1984)

[ATZ] Avron, J. E., Taylor, J. E., Zia, R. K. P.: Equilibrium shapes of crystals in a gravitational field: Crystals on a table. J. Stat. Phys. **33**, 493 (1983)

[B] Bennett, T.: Probabilistic inequalities for the sum of independent random variables. J. Am. Stat. Assoc. **57**, 33 (1962)

[BF] van den Berg, J., Fiebig, U.: On a combinatorial conjecture concerning disjoint occurrence of events. Ann. Probab. **15**, 354 (1987)

[BGN] Barsky, D. J., Grimmett,G., Newman, C. M.: in preparation

[BH] Broadbent, S. R., Hammersley, J. M.: Percolation processes I: Crystals and mazes. Proc. Camb. Phil. Soc. **53**, 629 (1957)

[BK] van den Berg, J., Kesten, H.: Inequalities with applications to percolation and reliability theory. J. Appl. Prob. **22**, 556 (1985)

[BN] van Beijeren, H., Nolden, I.: The roughening transition in: Current Topics in Physics, Vol. **43**, Schommers, W., von P., Blanckenhagen, (eds.), p. 259, Berlin, Heidelberg, New York: Springer 1987

[CC1] Chayes, J. T., Chayes, L.: Percolation and random media. In: Les Houches Session XLIII: Critical Phenomena, Random Systems and Gauge Theories. Osterwalder, K., Stora, R. (eds.), pp 1001–1142, Amsterdam: Elsevier 1986

[CC2] Chayes, J. T., Chayes, L.: On the upper Critical dimension of Bernoulli percolation. Commun. Math. Phys. **113**, 27 (1987)

[CCGKS] Chayes, J. T., Chayes, L., Grimmett, G. R., Kesten, H., Schonmann, R. H.: The correlation length for the high density phase of Bernoulli percolation. Ann. Probab **17**, 1277 (1989)

[CCN] Chayes, J. T., Chayes, L., Newman, C. M.: Bernoulli percolation above threshold: An invasion percolation analysis. Ann. Probab. **15**, 1272 (1987)

[CCS] Chayes, J. T., Chayes, L., Schonmann, R. H.: Exponential decay of connectivities in the two-dimensional Ising model. J. Stat. Phys. **49**, 433 (1987)

[DD] De Coninck, J., Dunlop, F.: Partial to complete wetting: A microscopic derivation of the Young relation, Ecole Polytechnique preprint

[DDR] De Coninck, J., Dunlop, F., Rivasseau, V.: On the microscopic validity of the Wulff construction and of the generalized Young equation, Ecole Polytechnique preprint

[DKS] Dobrushin, R. L., Kotecky, R., Shlosman, S. B.: In preparation; announcements to appear in the Proceedings of the Karpacz (Poland) Winter School, 1988: Equilibrium crystal shapes—A microscopic proof of the Wulff construction; and in the contributions of Kotecky, R., of Shlosman, S. B. in the Proceedings of the 9th International Congress on Mathematical Physics, Swansea (Wales), July 1988

[F] Federer, H.: Geometric measure theory. Berlin, Heidelberg, New York: Springer 1969

[FK] Fortuin, C., Kasteleyn, P.: On the random cluster model I. Physica **57**, 536 (1972)

[FKG] Fortuin, C., Kasteleyn, P., Ginibre, J.: Correlation inequalities on some partially ordered sets, Commun. Math. Phys. **22**, 89 (1971)

[G] Grimmett, G.: Percolation, Berlin, Heidelberg, New York, Springer (1989)

[H] Hammersley, J. M.: Percolation processes. Lower bounds for the critical probability, Ann. Math. Stat. **28**, 790 (1957)

[Har] Harris, T. E.: A lower bound for the critical probability in certain percolation processes. Proc. Camb. Phil. Soc. **56**, 13 (1960)

[K1] Kesten, H.: Analyticity properties and power law estimates of functions in percolation theory. J. Stat. Phys. **25**, 717 (1981)

[K2] Kesten, H.: The critical probability of bond percolation on the square lattice equals 1/2. Commun. Math. Phys. **74**, 41 (1980)
[K3] Kesten, H.: Percolation Theory for Mathematicians Boston: Birkhäuser, 1982
[KS] Kunz, H., Souillard, B.: Essential singularity in percolation problems and asymptotic behavior of the cluster size distribution. J. Stat. Phys. **19**, 77 (1978)
[KZ] Kesten, H., Zhang, Y.: The probability of a large finite cluster in supercritical Bernoulli percolation, Cornell preprint
[MS] Minlos, R. A., Sinai, Ya. G.: The phenomenon of phase separation at low temperatures in some lattice models of a gas II. Tr. Mosk. Mat. Obshch. **19**, 113 (1968) (English translation: Trans. Moscow Math. Soc. **19**, 121 (1968))
[R] Russo, L.: On the critical percolation probabilities. Z. Warsch. Verw. Geb. **56**, 229 (1981)
[RW] Rottman, C., Wortis, M.: Statistical mechanics of equilibrium crystal shapes: Interfacial phase diagrams and phase transitions. Phys. Rep. **103**, 59 (1984)
[S] Simon, B.: Correlation inequalities and decay of correlations in ferromagnets. Commun. Math. Phys. **77**, 111 (1980)
[T1] Taylor, J. E.: Existence and structure of solutions to a class of nonelliptic variational problems. Symposia Math. **XIV**, 499 (1974)
[T2] Taylor, J. E.: Unique structure of solutions to a class of nonelliptic variational problems. Proc. Symposia Pure Math. **27**, 419 (1975)
[W] Winterbottom, W. L.: Equilibrium shape of a small particle in contact with a foreign substrate. Acta Metal. **15**, 303 (1967)
[Wu] Wulff, G.: Zur Frage der Geschwindigkeit des Wachsthums und der Auflösung der Krystallflachen Z. Krist. **34**, 449 (1901)
[Z] Zhang, Y.: unpublished
[ZAT] Zia, R. K. P., Avron, J. E., Taylor, J. E.: The summertop construction: Crystals in a corner. J. Stat. Phys. **50**, 727 (1988).

Communicated by M. Aizenman

Received March 8, 1989; in revised form November 21, 1989