A New Lower Bound for the Critical Probability of Site Percolation on the Square Lattice

J. van den Berg and A. Ermakov*
CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

ABSTRACT

The critical probability for site percolation on the square lattice is not known exactly. Several authors have given rigorous upper and lower bounds. Some recent lower bounds are (each displayed here with the first three digits) 0.503 (Tóth [13]), 0.522 (Zuev [15]), and the best lower bound so far, 0.541 (Menshikov and Pelikh [12]). By a modification of the method of Menshikov and Pelikh we get a significant improvement, namely, 0.556. Apart from a few classical results on percolation and coupling, which are explicitly stated in the Introduction, this paper is self-contained. © 1996 John Wiley & Sons, Inc.

1. INTRODUCTION

Site percolation on the square lattice is one of the most studied percolation processes. In contrast to bond percolation on this lattice (for which the critical probability is 1/2 [8]), the critical probability for site percolation is not known exactly. Monte Carlo simulations suggest that it is about 0.593. Several authors have obtained rigorous upper and lower bounds. In this paper we concentrate on lower bounds, and we start with some history: From the general arguments of Hammersley [4], it followed that the above-mentioned critical probability is larger than 1/3. The main result of Harris [6] (combined with a comparison result of Hammersley [5]) yields that it is at least 1/2. About 20 years later, it was

rigorously proved that $1/2$ is also a *strict* lower bound [7]. After this, improvements were made more frequently: 0.503478 (Tóth [13]), 0.522105 (Zuev [15]), and, finally, 0.5416 (Menshikov and Pelikh [12]).

Improving bounds for critical percolation probabilities is not only interesting in itself, but also useful for other fields (see, for instance, van den Berg and Maes [14]).

In the present paper we use the main idea of Menshikov and Pelikh, but modify their method and make more extensive use of stochastic dominance and coupling arguments. This leads to a significant improvement, stated in Theorem 1.1 below. Moreover, while many details in the paper of Menshikov and Pelikh have been omitted, we give a complete account.

Theorem 1.1. *The critical probability for site percolation on the square lattice is larger than 0.556.*

In the remainder of this section we give a short introduction to the key notions in this paper: percolation, coupling, and stochastic dominance. We will state two classical results (Theorems 1.2 and 1.3 below), which will be used later. Apart from these, our paper is self-contained. In Section 2 we discuss the main ideas of the Menshikov–Pelikh method. The proof of Theorem 1.1 involves a "global" comparison argument (Section 3) and is completed in Section 4 by "local" comparison arguments.

- **Percolation.** Let G be a regular (vertex transitive), countably infinite graph, for instance the d-dimensional hypercubic lattice. Suppose each vertex is, independently of all others, open (usually denoted by 1) with probability p and closed (0) with probability $1 - p$. An open cluster is a maximal connected subgraph of which all vertices are open. Let v be any vertex of G and define the percolation probability

$$\theta(p) = P(v \text{ belongs to an infinite open cluster}) . \tag{1}$$

Next we define the critical probability

$$P_c = \inf\{p : \theta(p) > 0\} . \tag{2}$$

To emphasize the dependence of θ and P_c on the graph G, we will sometimes write θ_G and $P_c(G)$.

The model above is called *site percolation*. If not the vertices but the edges are randomly open or closed, we speak of *bond* percolation. For more information on percolation in general, see Grimmett [3] or Kesten [9]. Some papers which deal specifically with bounds on critical probabilities are (besides those mentioned above) Balister, Bollobás, and Stacey [1] and Łuczak and Wierman [11].

- **Site percolation on the square lattice.** The square lattice, which we denote by S in this paper, is the graph whose vertices can be viewed as points with integer coordinates in the plane, and where two vertices v and w share an edge (are adjacent; are neighbors) iff their euclidian distance $\|v - w\|$ equals 1 [see Fig. 1(a)]. The so-called *matching lattice* of S, denoted by S^*, is

Illustrations

a.   b.

c.   d.
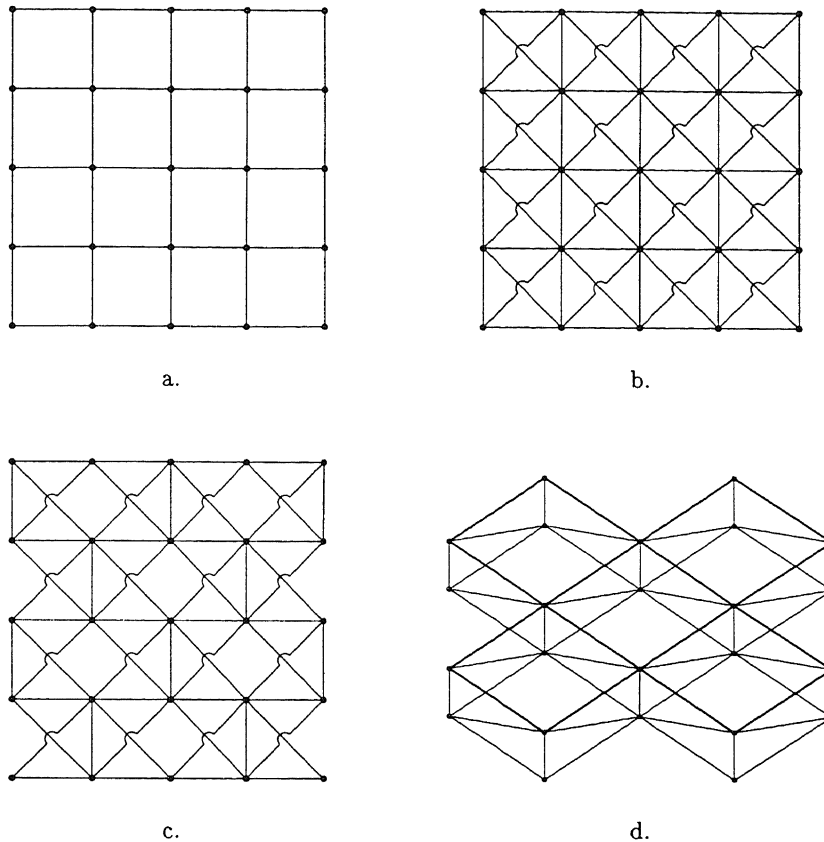
Fig. 1. (a) Lattice S; (b) lattice S^*; (c and d) two spatial representations of the lattice L.

obtained from S by "adding the diagonals" in each unit square [see Fig. 1(b)]. More precisely, S^* has the same vertices as S, but in S^* two vertices v and w share an edge iff $\|v - w\| = 1$ or $\sqrt{2}$. So in S each vertex has four neighbors and in S^* eight neighbors.

Fisher [2], using arguments of Harris [6] proved the following result, which is of essential importance in the Menshikov–Pelikh method.

Theorem 1.2.

$$P_c(S) + P_c(S^*) \geq 1 .$$ (3)

Remarks.

(i) Since the early 1980s, it has been known that the reversed inequality also holds, but that inequality plays no role in this paper.

(ii) Since S^* contains S, it is clear that $P_c(S^*) \leq P_c(S)$. So Theorem 1.2 immediately implies $P_c(S) \geq 1/2$, which is one of the lower bounds listed in the beginning of this section.

● **Stochastic dominance and coupling.** Let $\Omega = \{0, 1\}^K$ with K a finite or countable set. Elements of Ω are typically denoted by $\omega \equiv (\omega_i, i \in K)$. We say that $\omega \le \omega'$ if $\omega_i \le \omega'_i$ for all $i \in K$. A set $A \subset \Omega$ is *increasing* if $\omega \in A$ and $\omega \le \omega'$ implies $\omega' \in A$. Let μ and μ' be two probability measures on Ω (equipped with the natural sigma-algebra generated by finite cylinder events). We say that μ is *stochastically dominated* by μ' (which we denote by $\mu \overset{st}{\le} \mu'$) if $\mu(A) \le \mu'(A)$ for every increasing event A. If the distribution μ depends on a real-valued parameter p, and $p \le p'$ implies $\mu_p \overset{st}{\le} \mu_{p'}$, then we say that μ is stochastically increasing in p. A *coupling of* μ and μ' is a probability measure P on $\Omega \times \Omega$ whose two marginals on Ω are μ and μ' [i.e., which satisfies $P(A \times \Omega) = \mu(A)$ and $P(\Omega \times A) = \mu'(A)$ for all events $A \subset \Omega$]. The following result, which gives a connection between coupling and stochastic dominance, is well known (but not easy to prove) and goes back to more general results by Strassen (see, e.g., Liggett [10], Chap. II, Theorem 2.4).

Theorem 1.3. $\mu \overset{st}{\le} \mu'$ *if and only if there exists a coupling P of μ and μ' with the property* $P(\{(\omega, \omega') \in \Omega \times \Omega : \omega \le \omega'\}) = 1$.

Note that the if-part of this Theorem is obvious.

Since subsets of K correspond to elements of $\{0, 1\}^K$, the above definitions give naturally rise to analogous definitions for random subsets.

2. THE MENSHIKOV–PELIKH METHOD

A key idea in the Menshikov–Pelikh paper [12] is that by deleting every other vertical edge of S^* we obtain the lattice in Figure 1(c), which can also be viewed (as we will do in the remainder of this paper) as the lattice in Figure 1(d), which we denote by L. This lattice L consists of two layers of S with vertical connections, and one diagonal connection in every vertical face. More precisely, the set of vertices of L is $Z^2 \times \{0, 1\}$ and two different vertices $v = (x_1, x_2, x_3)$ and $w = (y_1, y_2, y_3)$ share an edge in L if their euclidian distance is 1, or ($y_1 = x_1$ and $y_2 - x_2 = x_3 - y_3$) or ($y_2 = x_2$ and $y_1 - x_1 = x_3 - y_3$).

From Theorem 1.2 we have

Lemma 2.1. *Suppose p is such that $\theta_L(p) > 0$ or $\theta_L(p) = \theta_S(1 - p) = 0$. Then $P_c(S) \ge 1 - p$.*

Proof. If the assumption of Lemma 2.1 holds, but the conclusion does not, then $\theta_S(1 - p) > 0$ and hence $\theta_L(p) > 0$. Hence $P_c(L) \le p$. Since we also have $P_c(S) < 1 - p$, we get $P_c(L) + P_c(S) < 1$ and hence (since L, as a graph, is contained in S^*), $P_c(S) + P_c(S^*) < 1$, which contradicts Theorem 1.2.   □

If $p = 1/2$, then the assumption of Lemma 2.1 is clearly true (since L contains S), but this only gives the lower bound $1/2$. However, the special structure of L makes it possible to compare percolation on L and S in a suitable way, giving rise

to values of p which are considerably smaller than $1/2$ but still have the property that percolation on L with parameter p is "easier" than percolation on S with the (larger) parameter $1 - p$. By Lemma 2.1 we then get a lower bound for $P_c(S)$ considerably larger than $1/2$. This is, roughly speaking, what is done in the Menshikov–Pelikh paper [12], and will also be done in our paper [but in a different way; see Remark (iii) at the end of Section 4].

In Section 3 we describe a "global comparison" between percolation on S and L involving certain growth processes. In Section 4 we study the local properties of these growth processes and complete the proof of Theorem 1.1.

3. THE GLOBAL COMPARISON

Consider site percolation on L with parameter p. Menshikov and Pelikh [12] observed that, using the interpretation of L as two layers of S (with extra connections between the two layers), this can also be described as "percolation on S with four defect types": Assign to each vertex v of S the random variable $\epsilon(v) \in \{0, 1\}^2$ which describes the states of its "lower corresponding vertex" $(v, 0)$ and "upper corresponding vertex" $(v, 1)$ in L. For instance, if $(v, 0)$ is open (1) and $(v, 1)$ closed (0), then we set $\epsilon(v) = \binom{0}{1}$. So the $\epsilon(v)$, $v \in Z^2$, are i.i.d. random variables taking values $\binom{0}{1}$, $\binom{1}{0}$, $\binom{0}{0}$, and $\binom{1}{1}$ with probability $p(1-p)$, $p(1-p)$, $(1-p)^2$, and p^2, respectively. Further, we say that two adjacent vertices $v = (x_1, x_2)$ and $w = (y_1, y_2)$ of S are ϵ-adjacent if there exist corresponding vertices (x_1, x_2, x_3) and (y_1, y_2, y_3) in L of v and w respectively, which are open and adjacent in L. More precisely, if for two adjacent vertices v and w we denote the one with smallest sum of coordinates by min and the other by max, then v and w are ϵ-adjacent if $(\epsilon(min), \epsilon(max)) = (\binom{1}{*}, \binom{1}{*})$ or $(\binom{*}{1}, \binom{*}{1})$ or $(\binom{1}{*}, \binom{*}{1})$. Here $*$ stands for "0 or 1." We say that two vertices v and w in S are ϵ-connected if there exists a path from v to w such that each pair of consecutive vertices in this path is ϵ-adjacent. This gives, in an obvious way, rise to a definition of ϵ-connected clusters. From the definition of the ϵ-variables it is clear that the occurrence of an infinite open cluster on L implies the occurrence of an infinite ϵ-connected cluster on S and vice versa.

Now consider the following growth process on S. First define a total order on S with minimal element O. Let the process ϵ_v, $v \in Z^2$, be as above. The first step in the growth process consists of "inspecting" the ϵ-value of O. If it is $\binom{1}{*}$ or $\binom{*}{1}$, then we assign $\xi(O) = 1$, otherwise $\xi(O) = 0$. More generally, at the nth step we select the vertex v which is lowest in order, has been inspected before and received ξ-value 1, and has at least one neighbor that has not yet received a ξ-value (such a neighbor will be called a child of v and v its parent). If no such vertex v exists, we say that the process has died out, and each vertex which has not yet received a ξ-value, gets ξ-value 0. Otherwise assign to each child of v the ξ-value 1 or 0, depending on whether or not it is ϵ-adjacent to v. It is clear that each vertex with ξ-value 1 is ϵ-connected to O (although the reverse is not generally true). Therefore, if the above growth process does not die out, the ϵ-cluster of O is infinite, and hence, in the corresponding percolation process on L, $(0, 0, 0)$ or $(0, 0, 1)$ belongs to an infinite open cluster, so that $\theta_L(p) > 0$. Now suppose that p is such that, at each step, the joint conditional distribution of the

ξ-values assigned to the children at that step, given all the ξ-values of the vertices which have been treated at previous steps, stochastically dominates the product distribution with parameter $1 - p$. Then, using Theorem 1.3 for each step (with K the set of children at that step) and standard arguments, we can construct a coupling $\mathscr{P}$ of the set of vertices which eventually have ξ-value 1, and the open cluster containing O in the percolation process on S with parameter $1 - p$, in such a way that, with $\mathscr{P}$-probability 1, the former set contains the latter. In particular, we then have that $\theta_S(1 - p) > 0$ implies $\theta_L(p) > 0$, so that, by Lemma 2.1, $1 - p$ is a lower bound for $P_c(S)$.

In view of the above, it is natural to investigate the conditional "offspring" distributions in the growth process. This will be done in the next section.

4. THE LOCAL COMPARISON AND THE PROOF OF THEOREM 1.1

In this section we study the (conditional) offspring distribution in a step in the growth process described in the previous section. The amount of conditioning information is, in some sense, unbounded: The larger the number of previous steps, the more information we have. By "conditioning out" all information except certain "local information," we get a finite problem as will be pointed out below [see also Remark (iv) at the end of this paper].

First we give some more notation. From now on we let, for $k \in \mathbb{N}$ and $p_1, \ldots, p_k \in [0, 1]$, $\pi_{(p_1, \ldots, p_k)}$ denote the product distribution on $\{0, 1\}^k$ with parameters $p_1, \ldots, p_k$. If $p_1 = p_2 = \cdots = p_k$, and the value of k is clear from the context, we will write just π_{p_1}.

In the previous section we defined the random variables ϵ_v, $v \in Z^2$, and the notion of ϵ-adjacency. Define, for each edge $e = \{v, w\}$ of S,

$$\eta(e) = I(v \text{ and } w \text{ are } \epsilon\text{-adjacent}) , \tag{4}$$

where $I(\cdot)$ denotes the indicator function. Since $\eta(\{v, w\})$ is a function of $\epsilon(v)$ and $\epsilon(w)$ and the ϵ's are independent random variables, we have the following version of a general well-known property:

Lemma 4.1. *Let* $W \subset V$ *be finite sets of vertices of* S, *and* X *a set of edges of* S *with both endpoints in* V. *Let* ∂W *be the set of all* $w \in W$ *for which there exists a* $v \in V \backslash W$ *with* $\{v, w\} \in X$. *Further, let* $X(W)$ *be the edges in the set* X *of which both endpoints are in* W. *Finally, let* Y *be a set of edges with one endpoint in* V^c *and the other in* W. *Then, given* $(\epsilon(w), w \in \partial W)$, *$(\eta(z), z \in Y \cup X(W))$ is independent of* $(\eta(x), x \in X \backslash X(W))$. *Hence, if* Ω *denotes* $\{\binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1}\}^{\partial W}$, *then for any* $\alpha \in \{0, 1\}^Y$ *and* $\beta \in \{0, 1\}^X$,

$$P((\eta(y) = \alpha(y), y \in Y \mid (\eta(x) = \beta(x), x \in X))$$

$$= \sum_{\gamma \in \Omega} P((\epsilon(w) = \gamma(w), w \in \partial W) \mid (\eta(x) = \beta(x), x \in X))$$

$$\times P((\eta(y) = \alpha(y), y \in Y) \mid (\epsilon(w) = \gamma(w), w \in \partial W),$$

$$(\eta(x) = \beta(x), x \in X(W))) . \tag{5}$$

$\square$

This lemma, which is intuitively obvious and easy to prove, will be used frequently in this paper (mostly without explicitly referring to it).

Now suppose we are at a certain step in the growth process described in Section 3. Let b_2 denote the vertex selected at this step (we reserve the notation b_1 and b_3 for possible "brothers" of b_2). Suppose b_2 has three children a_1, a_2, and a_3. In view of the arguments at the end of Section 3, we are interested in whether the conditional distribution of $(\xi(a_1), \xi(a_2), \xi(a_3))$, given all the ξ-values assigned in previous steps, stochastically dominates π_{1-p}. From the information on which we condition, we can reconstruct the past of the growth process in the sense that, for each step in the past, we know the selected vertex at that step and the assignment of ξ-values to its children. In other words, it tells us for each vertex v which already has a ξ-value, who its parent w is and what the value of $\eta(\{v, w\})$ is. In particular, we know from this information who the parent c of b_2 is, who the parent d of c is, and who the "brothers" of b_2 (i.e., other children of c) are (if there are any). To be more explicit, suppose we are in the situation of Figure 2(a), where b_2 has two brothers, named b_1 and b_3. Note that if $\eta(\{b_1, c\}) = 0$, then $\xi(b_1) = 0$, so b_1 has not been the selected vertex at any step, and hence the only neighbor v of b_1 for which we know $\eta(\{b_1, v\})$ is c. The same observation holds for b_3. Now apply Lemma 4.1 (with V the set of all vertices which have already obtained a ξ-value, $W = \{b_1, b_2, b_3, c, d\}$, X the set of edges $\{v, v'\}$ with v, $v' \in V$ and v the parent of v' (or vice versa), and Y the set consisting of the edges $\{a_1, b_2\}$, $\{a_2, b_2\}$, and $\{a_3, b_2\}$. Then it is easy to see that the conditional distribution we are looking at is a mixture of conditional distributions of $\eta(\{a_1, b_2\})$, $\eta(\{a_2, b_2\})$, $\eta(\{a_3, b_2\})$ given $\eta(\{b_2, c\}) = 1$, $\eta(\{c, d\}) = 1$, and the values of $\eta(\{b_1, c\})$, $\eta(\{b_3, c\})$ and $\epsilon(d)$, and the ϵ-values of those brothers of b that are ϵ-adjacent to c. If each such "local conditional" distribution stochastically dominates π_{1-p}, then clearly the mixture itself also dominates π_{1-p}. Therefore, we will study these local conditional distributions. Of course, the configuration in Figure 2(a) is only one of the configurations of a_1, a_2, a_3, b_1, b_2, b_3, c, and d that can occur. However, it is not difficult to see that there is essentially only one other configuration, namely, that in Figure 2(b). This follows from the following arguments: First of all, by the symmetry properties of L, it does not matter which of the four neighbors of b_2 we choose for c. Then there are three choices for d: One where b_2, c and d are on one line, and two where the line through b_2 and c and that through c and d are perpendicular. Of the last two cases, one is, from a graph-theoretical viewpoint, equivalent to the case where b_2, c, and d are on one line. Further, our assumption that b_2 has three children is sufficiently general: If it has fewer children, say only a_2 and a_3, then we can add an "imaginary copy" of a_1 [in fact this has been done in Fig. 2(b), where a_1 is adjacent to d and hence must be a child of d or of a vertex whose ξ-value is "at least as old" as that of d]. This does not disturb our arguments, because, if the joint conditional distribution, which we study, of the three children dominates π_{1-p}, then so do its marginals, in particular the joint conditional distribution of the two "real" children. For similar reasons we have assumed that c has, besides b, two other children. Since, if it has fewer, then we add "imaginary" ones. The conditional distribution which we are studying is, in the real situation, a convex combination of the relevant conditional distributions in the cases with "imaginary" b_i's. If π_{1-p} is stochastically dominated by each of these, then it is also stochastically dominated by the mixture.
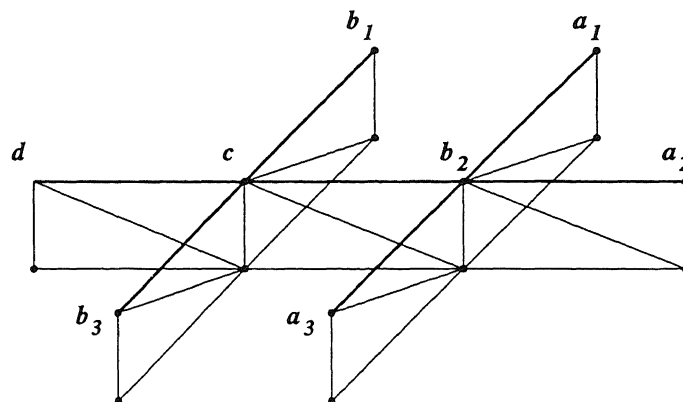
a). Conf. I
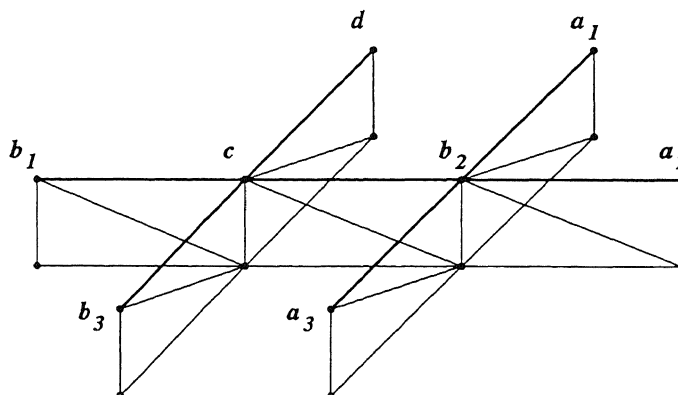
b). Conf. II

Fig. 2. The local configurations I(a) and II(b).

Summarizing the above arguments, we get the proposition below, where we abbreviate $\eta(\{a_i, b_2\})$ by $\eta(a_i)$ and $\eta(\{b_i, c\})$ by $\eta(b_i)$, $i = 1, \ldots, 3$, and $\eta(\{c, d\})$ by $\eta(c)$.

Proposition 4.2. *Let* a_1, a_2, a_3, b_1, b_2, b_3, c *and* d *be as in Figure 2(a) or 2(b). Let* $\epsilon(v)$, *with* v *in the above set, be independent random variables taking values* $\binom{0}{0}$, $\binom{0}{1}$, $\binom{1}{0}$, *and* $\binom{1}{1}$ *with probability* $(1 - p)^2$, $p(1 - p)$, $p(1 - p)$, *and* p^2 *respectively. Let* $\eta(a_i) = I(a_i$ *is* ϵ-*adjacent to* $b_2)$, $i = 1$, 2, 3, $\eta(b_i) = I(b_i$ *is* ϵ-*adjacent to* $c)$, $i = 1$, 2, 3, *and* $\eta(c) = I(c$ *is* ϵ-*adjacent to* $d)$, *with* ϵ-*adjacency as defined in Section 3. Let* $\mathcal{M}$ *be the set of all events of the form* $\{\eta(b_2) = 1$, $\eta(c) = 1\} \cap E_d \cap E_{b_1} \cap E_{b_3}$. *Here* E_d *is one of the events* $\{\epsilon(d) = \binom{1}{0}\}$, $\{\epsilon(d) = $

$\binom{0}{1}$, *or* $\{\epsilon(d) = \binom{1}{1}\}$, *and each* E_{b_i}, $i = 1$, 3 *is one of the events* $\{\eta(b_i) = 0\}$, $\{\eta(b_i) = 1, \epsilon(b_i) = \binom{1}{0}\}$, $\{\eta(b_i) = 1, \epsilon(b_i) = \binom{0}{1}\}$, *or* $\{\eta(b_i) = 1, \epsilon(b_i) = \binom{1}{1}\}$.

If p is such that in both cases [Figs. 2(a) and 2(b)], for each $E \in \mathcal{M}$, the conditional distribution $P((\eta(a_1), \eta(a_2), \eta(a_3)) \in \cdot \mid E)$ stochastically dominates π_{1-p}, then $P_c(S) \geq 1 - p$. $\square$

By this proposition we have reduced our problem to a finite one, but there is still much work to be done. To prove Theorem 1.1, we want to show that the condition of Proposition 4.2 holds for $p = 1 - 0.556 = 0.444$. In principle we could, for each $E \in \mathcal{M}$, check if the stochastic dominance property in the condition of the proposition indeed holds. For each choice of E this would lead (as we will see later) to checking 13 inequalities, for each of the two cases I [Fig. 2(a)] and II [Fig. 2(b)]. Since there are 48 possibilities for E, one would have to check $2 \times 13 \times 48 = 1248$ inequalities. Although this is possible by computer, we proceed in a different way, which leads to checking "only" 13 inequalities. We will construct a suitable distribution $\hat{\mu}$ on $\{0, 1\}^3$ which, for each $E \in \mathcal{M}$, is stochastically dominated by $P((\eta(a_1), \eta(a_2), \eta(a_3)) \in \cdot \mid E)$. The only thing we then have to do is to check if, for $p = 0.444$, $\hat{\mu}$ stochastically dominates π_{1-p}. [One may think that checking the 1248 inequalities may lead to a larger lower bound of $P_c(S)$, but that appears not to be the case. See Remark (ii) at the end of this section.]

We start by giving some extra notation. First, q will denote $1 - p$. Let, for $\alpha \in \{\binom{0}{1}, \binom{1}{0}, \binom{1}{1}\}$, ν_α be the conditional distribution of $(\eta(a_1), \eta(a_2), \eta(a_3))$ given $\epsilon(b_2) = \alpha$. It is easy to see that, in both cases I and II,

$$\nu_{\binom{1}{1}} = \pi_{1-q^2},$$

$$\nu_{\binom{1}{0}} = \pi_{(1-q^2, 1-q^2, p)}, \tag{6}$$

$$\nu_{\binom{0}{1}} = \pi_{(p, p, 1-q^2)}.$$

Moreover, let for $\alpha = \binom{1}{0}$, $\binom{0}{1}$, and $\binom{1}{1}$, μ_α denote the conditional probability of $(\eta(a_1), \eta(a_2), \eta(a_3))$ given $\epsilon(c) = \alpha$ and $\eta(b_2) = 1$. By elementary manipulations [sum over the possibilities for $\epsilon(b_2)$ and use that the conditional distribution of $((\eta(a_1), \eta(a_2), \eta(a_3))$ given $\epsilon(c)$, $\eta(b_2)$ and $\epsilon(b_2)$ depends only on $\epsilon(b_2)$] we can express the μ_α's as mixtures of the ν_α's:

$$\mu_{\binom{1}{*}} = \frac{p}{1+q} \nu_{\binom{1}{1}} + \frac{q}{1+q} \nu_{\binom{1}{0}} + \frac{q}{1+q} \nu_{\binom{0}{1}},$$

$$\mu_{\binom{0}{1}} = p\nu_{\binom{1}{1}} + 0\nu_{\binom{1}{0}} + q\nu_{\binom{0}{1}}, \tag{7}$$

where $\mu_{\binom{1}{*}}$ stands for $\mu_{\binom{1}{1}}$ and $\mu_{\binom{1}{0}}$ which appear to be equal.

Now define

$$\tilde{\mu} := \left(1 - q\frac{2+q}{1+q}\right)\nu_{\binom{1}{1}} + \frac{q}{1+q}\nu_{\binom{1}{0}} + q\nu_{\binom{0}{1}}. \tag{8}$$

From (7) and (8), using that $\nu_{\binom{0}{1}}$ and $\nu_{\binom{1}{0}}$ are both stochastically smaller than $\nu_{\binom{1}{1}}$ and that $q \geq q/(1+q)$, we have

Lemma 4.3.

$$\tilde{\mu} \overset{st}{\leq} \mu_{\binom{1}{*}} \quad and \quad \tilde{\mu} \overset{st}{\leq} \mu_{\binom{0}{1}} \, .$$

Our next step in the construction of a distribution $\hat{\mu}$ as mentioned in the discussion after Proposition 4.2, is the following lemma.

Lemma 4.4. *Let* $r \in [0, 1]$. *If* $min_{E \in \mathcal{M}} P(\epsilon(c) = \binom{1}{*} \mid E) \geq r$, *then, for all* $E \in \mathcal{M}$,

$$P((\eta(a_1), \eta(a_2), \eta(a_3)) \in \cdot \mid E) \overset{st}{\geq} r\mu_{\binom{1}{*}} + (1 - r)\tilde{\mu} \, .$$

Proof. Suppose the assumption of Lemma 4.4 holds. Let $A \subset \{0, 1\}^3$ be an increasing event. We have

$$P((\eta(a_1), \eta(a_2), \eta(a_3)) \in A \mid E)$$

$$= \sum_{\alpha \in \{\binom{0}{1}, \binom{1}{*}\}} P(\epsilon(c) = \alpha \mid E) \, P((\eta(a_1), \eta(a_2), \eta(a_3)) \in A \mid E, \epsilon(c) = \alpha)$$

$$= \sum_{\alpha \in \{\binom{0}{1}, \binom{1}{*}\}} P(\epsilon(c) = \alpha \mid E) \, \mu_\alpha(A)$$

$$\geq P(\epsilon(c) = \binom{1}{*} \mid E) \, \mu_{\binom{1}{*}}(A) + P(\epsilon(c) = \binom{0}{1} \mid E) \, \tilde{\mu}(A)$$

$$\geq r\mu_{\binom{1}{*}}(A) + (1 - r)\tilde{\mu}(A) \, . \tag{9}$$

The first equality is trivial; the second follows from the definition of μ_α and the fact that, given $\epsilon(c)$ and $\eta(b_2)$, $(\eta(a_1), \eta(a_2), \eta(a_3))$ is independent of $\epsilon(d)$, $\eta(c)$, $\eta(b_1)$, $\eta(b_3)$, $\epsilon(b_1)$, and $\epsilon(b_3)$. The first inequality follows from the second statement in Lemma 4.3. The second inequality follows from the first statement in Lemma 4.3 and the assumption of Lemma 4.4. $\square$

Proposition 4.2 and Lemma 4.4 motivate the search for an appropriate r. We find

Lemma 4.5. *Let, for* $p \in (0, 1)$, $r_0(p) = (1 - q^2)/(2 - q^2)$. *We have* $min_{E \in \mathcal{M}} P((\epsilon(c) = \binom{1}{*} \mid E) = r_0(p)$.

Proof. We give separate proofs for case I [Fig. 2(a)] and II [Fig. 2(b)]. We start with case I. First we observe that if $\epsilon(b_1) \equiv \binom{0}{0}$ on E, then E implies $\epsilon(c) = \binom{1}{*}$. Also note that if $\epsilon(b_1) \equiv \binom{*}{1}$ on E, then the information concerning b_1 in E is redundant (for the event $\{\epsilon(c) = \binom{1}{*}\}$). Hence, by the previous observation, $P((\epsilon(c) = \binom{1}{*} \mid E)$ is then a convex combination of 1 and $P((\epsilon(c) = \binom{1}{*} \mid E')$, where $E' \in \mathcal{M}$ satisfies $\eta(b_1) \equiv 0$. Therefore, we will assume $\eta(b_1) \equiv 0$ on E. Also note that if $\epsilon(d) \equiv \binom{0}{1}$, then E implies $\epsilon(c) = \binom{*}{1}$, and the information concerning b_3 is redundant. It is clear that we can then replace E by $E_1 := \{\eta(b_1) = 0, \eta(b_2) = 1, \epsilon(c) = \binom{*}{1}\}$ without changing the conditional probability that $\epsilon(c) = \binom{1}{*}$. This is also the case when $\epsilon(b_3) \equiv \binom{0}{1}$ on E. Further, if $\epsilon(b_3) \equiv \binom{1}{*}$, then the information concerning b_3 is redundant. By these observations it follows that we only have to compute $P(\epsilon(c) = \binom{1}{*} \mid E_i)$, $i = 1, 2$, with E_1 as above and $E_2 = \{\epsilon(d) = \binom{1}{*}, \eta(b_1) = 0, \eta(b_2) = 1, \eta(b_3) = 0, \eta(c) = 1\}$. It is elementary to verify that these are

equal to $r_0(p)$ and $\frac{p(1-q^2)+(1-q^2)}{p(2-q^2)+(1-q^2)}$ [which is clearly larger than $r_0(p)$] respectively. This completes the proof for case I.

As to case II, first note that if $\epsilon(d) \equiv (\begin{smallmatrix}1\\0\end{smallmatrix})$ on E, then E implies $\epsilon(c) = (\begin{smallmatrix}1\\*\end{smallmatrix})$, so we will assume that $\epsilon(d) \equiv (\begin{smallmatrix}*\\1\end{smallmatrix})$. Further, if $\epsilon(b_1) \equiv (\begin{smallmatrix}1\\*\end{smallmatrix})$, then the information concerning b_1 is redundant, and if $\epsilon(b_1) \equiv (\begin{smallmatrix}0\\1\end{smallmatrix})$, then E implies $\epsilon(c) = (\begin{smallmatrix}*\\1\end{smallmatrix})$, and we can replace E by the event $F_1 := \{\epsilon(c) = (\begin{smallmatrix}*\\1\end{smallmatrix}), \eta(b_2) = 1\}$. If $\eta(b_1) \equiv 0$ on E, then $\epsilon(b_1)$ is either $(\begin{smallmatrix}0\\0\end{smallmatrix})$, in which case the information on b_1 is redundant, or $(\begin{smallmatrix}0\\1\end{smallmatrix})$, in which case $\epsilon(c)$ must be $(\begin{smallmatrix}*\\0\end{smallmatrix})$ and hence $(\begin{smallmatrix}1\\0\end{smallmatrix})$. By symmetry, similar observations hold for b_3. By these arguments it suffices to compute $P(\epsilon(c) = (\begin{smallmatrix}1\\*\end{smallmatrix}) \mid F_1)$, which is equal to $(1 - q^2)/(2 - q^2 - p)$, which clearly is larger than $r_0(p)$. $\square$

Proposition 4.6. *Let* $p \in (0, 1)$ *and* $r_0(p) = (1 - q^2)/(2 - q^2)$ *as in Lemma 4.5. Let* $\hat{\mu}$ *be the following mixture of three product distributions on* $\{0, 1\}^3$

$$\hat{\mu} = \left(r_0(p) \frac{p}{1+q} + (1 - r_0(p))\left(1 - q\frac{2+q}{1+q}\right)\right) \pi_{1-q^2}$$

$$+ \frac{q}{1+q} \pi_{(1-q^2, 1-q^2, p)}$$

$$+ \left(r_0(p) \frac{q}{1+q} + (1 - r_0(p))q\right) \pi_{(p, p, 1-q^2)} . \tag{10}$$

If $\hat{\mu} \overset{st}{\geq} \pi_{1-p}$, *then* $P_c(S) \geq 1 - p$.

Proof. Use (6), (7), and (8) to see that

$$\hat{\mu} = r_0(p)\mu_{(\begin{smallmatrix}1\\0\end{smallmatrix})} + (1 - r_0(p))\tilde{\mu} . \tag{11}$$

The proposition is now a straightforward consequence of Proposition 4.2, Lemma 4.4, and Lemma 4.5. $\square$

Proof of Theorem 1.1. Theorem 1.1 will now be proved by checking that if $p = 0.444$, for each increasing event $A \subset \{0, 1\}^3$,

$$\hat{\mu}(A) > \pi_{1-p}(A) . \tag{12}$$

In fact, since $\hat{\mu}$ is symmetric in a_1 and a_2, we do not have to check all increasing events (only up to $a_1 - a_2$ symmetry). It appears that we have to check (12) for 13 events A. These events and their probabilities under $\hat{\mu}$ and π_{1-p} (for $p = 0.444$) are given in the following table [with the notation $A_i = \{(x_1, x_2, x_3) \in \{0, 1\}^3 : x_i = 1\}$]. Although we used the computer to make these calculations, the inequalities can also be checked by hand in a "reasonable" time [note that, for each A, the rhs of (12) is a very simple polynomial and the lhs a relatively simple rational expression in p; also note that the probabilities for cases 6–13 in the table can be very simply expressed in terms of those for cases 1–5]. Therefore, in our opinion, our proof of Theorem 1.1 should be considered as a "classical" proof and not a computer-assisted proof.

Case	A	$\hat{\mu}(A)$ ($p = 0.444$)	$\pi_{0.556}(A)$
1	A_1	0.573647	0.556000
2	A_3	0.602653	0.556000
3	$A_1 \cap A_2$	0.344268	0.309136
4	$A_1 \cap A_3$	0.335370	0.309136
5	$A_1 \cap A_2 \cap A_3$	0.195740	0.171880
6	$A_1 \cup A_2$	0.803026	0.802864
7	$A_1 \cup A_3$	0.840930	0.802864
8	$A_1 \cup (A_2 \cap A_3)$	0.713277	0.693256
9	$(A_1 \cap A_2) \cup A_3$	0.751181	0.693256
10	$(A_1 \cap A_2) \cup (A_1 \cap A_3)$	0.483898	0.446392
11	$(A_1 \cap A_3) \cup (A_2 \cap A_3)$	0.475001	0.446392
12	$A_1 \cup A_2 \cup A_3$	0.930678	0.912472
13	$(A_1 \cap A_2) \cup (A_2 \cap A_3) \cup (A_1 \cap A_3)$	0.623529	0.583649

In each of the 13 cases, the probability under the distribution $\hat{\mu}$ (with $p = 0.444$) is, according to the table, indeed larger than under $\pi_{0.556}$ [note that the difference is minimal for case 6; see also Remark (i) below]. Hence Theorem 1.1 follows. $\square$

Remarks.

(i) The distributions ν_α [see (6)] are clearly stochastically increasing in p. This, together with the obvious facts that $\nu_{(1)}$ dominates both $\nu_{(0)}$ and $\nu_{(0)}^{(0)}$, and that p and $p/(1 + q)$ are increasing in p, implies that also $\mu_{(1)}$ and $\mu_{(1)}^{(0)}$ are stochastically increasing in p. In the same way we see from (8) that $\tilde{\mu}$ is stochastically increasing in p. Finally, since $\mu_{(1)}$ stochastically dominates $\tilde{\mu}$ and both are stochastically increasing in p, and $r_0(p)$ is also increasing in p, we have from (11) that $\hat{\mu}$ is stochastically increasing in p. Hence, for each increasing event A, $\hat{\mu}(A)$ is an increasing function of p. Clearly, $\pi_{1-p}(A)$ is decreasing in p. It is also easy to check that each of these increasing (decreasing) functions is 0 (1) for $p = 0$ and 1 (0) for $p = 1$. So each equation $\hat{\mu}(A) = \pi_{1-p}(A)$ has a unique solution in $[0, 1]$. Our lower bound 0.556 is taken (a little bit smaller than) 1 minus the largest of the 13 solutions (namely, the solution for the event $A_1 \cup A_2$).

(ii) One may think that we may have taken $\tilde{\mu}$ in (8) "too stochastically small" and that checking the 1248 inequalities mentioned after Proposition 4.2 may give a better (i.e., larger) lower bound for $P_c(S)$. This is not true because of the following: As we said in Remark (i), the event $A_1 \cup A_2$ is the worst one, in the sense that if we decrease p, then this event is the first for which (12) will fail. Also, from (6), (7), and (8) it follows immediately that the probability of this event under $\tilde{\mu}$ is the same as under $\mu_{(1)}^{(0)}$. Now let E_0 be an event for which the minimum in Lemma 4.5 is reached. Then it is easy to see that, with $A = A_1 \cup A_2$, $E = E_0$, and $r = r_0(p)$, we have equality in (9), and from (11) that the last expression equals $\hat{\mu}(A)$. Hence, if p decreases, then as soon as (12) fails for some A, then one of the 1248 inequalities mentioned above also fails.

(iii) As we said in the Introduction, many details have been omitted in the Menshikov–Pelikh paper [12]. Therefore, it is not easy to compare our paper with theirs precisely. However, it seems that the improvement we have obtained is mainly due to the fact that we study the *joint* conditional distribution of all children of a parent, while they study the conditional distribution of one child (and therefore have to include information on possible brothers in the event they condition on). The growth process in their paper is slightly different from ours. For instance, in their setup a child can have more than one parent, which is seemingly an advantage but appears, so far, to make thinks only unnecessarily complicated. Generally speaking, the comparison arguments in their paper are formulated in the framework of Markov chain theory rather than coupling theory.

(iv) In the beginning of this section we reduced the problem to a finite one, by "conditioning out all information outside a_1, a_2, a_3, b_1, b_2, b_3, c, and d." This choice looks somewhat arbitrary and one may wander whether including more information leads to a better lower bound for $P_c(S)$. By Remark (ii) above and because of the form of the events for which the minimum in Lemma 4.5 is reached, it follows that, to improve the bound in this way, one has to include not only the parent of d and the brothers of c, but also (at least) the children of b_3. Based on calculations we made, we found the improvement which can be obtained in that way not sufficient to justify the extra amount of work.

ACKNOWLEDGMENT

REFERENCES

[1] P. Balister, B. Bollobás, and A. Stacey, Improved upper bounds for the critical probability of oriented percolation in two dimensions, *Random Struct. Alg.*, 5, 573–589 (1994).

[2] M. E. Fisher, Critical probabilities for cluster size and percolation problems, *J. Math. Phys.*, 2, 620–627 (1961).

[3] G. Grimmett, *Percolation*, Springer-Verlag, New York, 1989.

[4] J. M. Hammersley, Percolation processes. Lower bounds for the critical probability, *Ann. Math. Stat.*, 28, 790–795 (1957).

[5] J. M. Hammersley, Comparison of atom and bond percolation, *J. Math. Phys.*, 2, 728–733 (1961).

[6] T. E. Harris, A lower bound for the critical probability in a certain percolation process, *Proc. Camb. Phil. Soc.*, 56, 13–20 (1960).

[7] Y. Higuchi, Coexistence of the infinite (∗) clusters: a remark on the square lattice site percolation, *Z.W.*, 61, 75–81 (1982).

[8] H. Kesten, The critical probability of bond percolation on the square lattice equals 1/2, *Commun. Math. Phys.*, 74, 41–59 (1980).

[9] H. Kesten, Percolation theory and first-passage percolation, *Ann. Probab.*, **15**, 1231–1271 (1987).

[10] T. M. Liggett, *Interacting Particle Systems*, Springer-Verlag, New York, 1985.

[11] T. Łuczak and J. C. Wierman, Critical probability bounds for two-dimensional site percolation models, *J. Phys. A*, **21**, 3131–3138 (1988).

[12] M. V. Menshikov and K. D. Pelikh, *Mat. Zametki*, **46**, 38–47 [Engl. Transl.: Percolation with several defect types. An estimate of critical probability for a square lattice, *Math. Notes Acad. Sci. USSR*, **46**(4), 778–785 (1990)].

[13] B. Tóth, A lower bound for the critical probability of the square lattice site percolation, *Z.W.*, **69**, 19–22 (1985).

[14] J. van den Berg and C. Maes, Disagreement percolation in the study of Markov fields, *Ann. Probab.*, **22**, 749–763 (1994).

[15] S. A. Zuev, A lower bound for the percolation threshold for the square lattice, *Moscow Univ. Math. Bull.*, **43**(5), 59–61 (66–69 in translation) (1988).