ALEA, Lat. Am. J. Probab. Math. Stat. **10** (2), 667–679 (2013)

Confidence intervals for the critical value in the divide and color model

András Bálint, Vincent Beffara and Vincent Tassion

Chalmers University of Technology
S-41296 Göteborg, Sweden

UMPA – ENS Lyon
46 allée d'Italie
F-69364 Lyon cedex 07, France
E-mail address: abalint@few.vu.nl, vbeffara@ens-lyon.fr, vincent.tassion@ens-lyon.fr

Abstract. We obtain confidence intervals for the location of the percolation phase transition in Häggström's divide and color model on the square lattice $\mathbb{Z}^2$ and the hexagonal lattice $\mathbb{H}$. The resulting probabilistic bounds are much tighter than the best deterministic bounds up to date; they give a clear picture of the behavior of the DaC models on $\mathbb{Z}^2$ and $\mathbb{H}$ and enable a comparison with the triangular lattice $\mathbb{T}$. In particular, our numerical results suggest similarities between DaC model on these three lattices that are in line with universality considerations, but with a remarkable difference: while the critical value function $r_c(p)$ is known to be constant in the parameter p for $p < p_c$ on $\mathbb{T}$ and appears to be linear on $\mathbb{Z}^2$, it is almost certainly non-linear on $\mathbb{H}$.

1. Introduction

Our object of study in this paper is the critical value function in Häggström's divide and color (DaC) model, see Häggström (2001). This is a stochastic model that was originally motivated by physical considerations (see Häggström (2001); Chayes et al. (2007)), but it has since then been used for biological modeling in Gravner et al. (2007) as well and inspired several generalizations (see, *e.g.*, Hsu and Han (2008); Bálint et al. (2009); Graham and Grimmett (2011)). Our results concerning the location of the phase transition give a clear picture of the behavior of the DaC model on two important lattices and lead to intriguing open questions.

Received by the editors October 14, 2012; accepted July 4, 2013.
2010 *Mathematics Subject Classification.* 60K35, 82B43, 82B20.
Key words and phrases. Percolation, divide and color model, critical value, locality, stochastic domination.
AB was supported by grants of the Netherlands Organization for Scientific Research (NWO) and the Swedish Research Council (VR), and VB and VT by ANR grant 2010-BLAN-0123-01.

Our analysis will be based on the same principles as Riordan and Walters (2007), where confidence intervals were obtained for the critical value of Bernoulli bond and site percolation on the 11 Archimedes lattices by a modification of the approach of Balister et al. (2005). The main idea in Balister et al. (2005); Riordan and Walters (2007) is truly multidisciplinary and attractive, namely to reduce a problem which has its roots in theoretical physics by deep mathematical theorems to a situation in which a form of statistical testing by numerical methods becomes possible. Our other main goal with this paper is to demonstrate the strength of this strategy by applying it to a system which is essentially different from those in its previous applications. In particular, in the DaC model, as opposed to the short-range dependencies in Balister et al. (2005) and the i.i.d. situation in Riordan and Walters (2007), one has to deal with correlations between sites at arbitrary distances from each other. We believe that the method of Balister et al. (2005); Riordan and Walters (2007) has a high potential to be used in a number of further models (see *e.g.* Deijfen et al. (2011) where a very similar approach is followed) and deserves higher publicity than it enjoys at the moment.

Given a graph G with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$ and parameters $p, r \in [0,1]$, the DaC model on G is defined in two steps: first, Bernoulli bond percolation with density p is performed on G, and then the resulting open clusters are independently colored black (with probability r) or white (a more detailed definition will follow in the next paragraph). Note that this definition resembles the so-called random-cluster (or FK) representation of the ferromagnetic Ising model, with two important differences: a product measure is used in the DaC model in the first step instead of a random-cluster measure with cluster weight 2 and the second step is more general here in that all $r \in [0,1]$ are considered instead of only $1/2$.

Now we set the terminology that is used throughout, starting with an alternative (equivalent) definition of the DaC model which goes as follows. First, an *edge configuration* $\eta \in \{0,1\}^{\mathcal{E}}$ is drawn according to the product measure $\nu_p^{\mathcal{E}}$ where ν_p is the probability measure on $\{0,1\}$ with $\nu_p(\{1\}) = 1 - \nu_p(\{0\}) = p$. In the second step, a *site configuration* $\xi \in \{0,1\}^{\mathcal{V}}$ is chosen by independently assigning state 1 with probability r or otherwise 0 to each vertex, conditioning on the event that there exists no edge $e = \langle v, w\rangle \in \mathcal{E}$ such that $\eta(e) = 1$ and $\xi(v) \neq \xi(w)$. We denote the probability measure on $\{0,1\}^{\mathcal{V}} \times \{0,1\}^{\mathcal{E}}$ associated to this procedure by $\mathbb{P}^G_{p,r}$. An edge e (a vertex v) is said to be *open* or *closed* (*black* or *white*) if and only if it is in state 1 or 0, respectively. We will call the maximal subsets of $\mathcal{V}$ connected by open edges *bond clusters*, and the maximal monochromatic connected (via the edge set of $\mathcal{E}$, not only the open edges!) subsets of $\mathcal{V}$ *black* or *white clusters*. We write $C_v(\eta)$ for the bond cluster of a vertex v in the edge configuration η and use Ω_S to denote $\{0,1\}^S$ for arbitrary sets S.

Note that the measure $\mathbb{P}^G_{p,r}$ is concentrated on the set of pairs (η, ξ) such that for all edges $e = \langle v, w\rangle \in \mathcal{E}$, $\xi(v) = \xi(w)$ whenever $\eta(e) = 1$. When this compatibility condition is satisfied, we write $\eta \sim \xi$.

For infinite graphs G, there are two types of phase transitions present in the DaC model in terms of the appearance of infinite 1-clusters; first, there exists $p_c = p_c^G \in [0,1]$ such that $\mathbb{P}^G_{p,r}$(there exists an infinite bond cluster) is 0 for $p < p_c$ and 1 for $p > p_c$. Second, for each fixed p, there exists $r_c = r_c^G(p)$ such that $\mathbb{P}^G_{p,r}$(there exists an infinite black cluster) is 0 for $r < r_c$ and *positive* for $r > r_c$. For more on the different character of these two types of phase transitions, see Bálint

et al. (2013). A key feature of the DaC model (as noted in Häggström (2001)) is that while it is close in spirit to the Ising model, its simulation is straightforward from the definition and does not require sophisticated MCMC algorithms. In this paper, we will exploit this feature in order to learn about the values and various features of the *critical value function* $r_c^G(p)$.

Monotonicity and continuity properties of the function $r_c^G(p)$ for general graphs have been studied in Bálint et al. (2013). Here we will focus on two specific graphs, namely the *square lattice* $\mathbb{Z}^2$ and the *hexagonal lattice* $\mathbb{H}$ (see Figure 1.1), for which $p_c^{\mathbb{Z}^2} = 1/2$ and $p_c^{\mathbb{H}} = 1 - 2\sin(\pi/18) \approx 0.6527$ (see Kesten (1982)). Our reason for this restriction is twofold: first, these two are the most commonly considered planar lattices (apart from the triangular lattice $\mathbb{T}$, for which the critical value function $r_c^{\mathbb{T}}$ has been completely characterized in Bálint et al. (2009)), whence results about these cases are of the greatest interest. On the other hand, the DaC model on these lattices enjoys a form of duality (described in Section 2.2) which is a key ingredient for the analysis we perform in this paper.

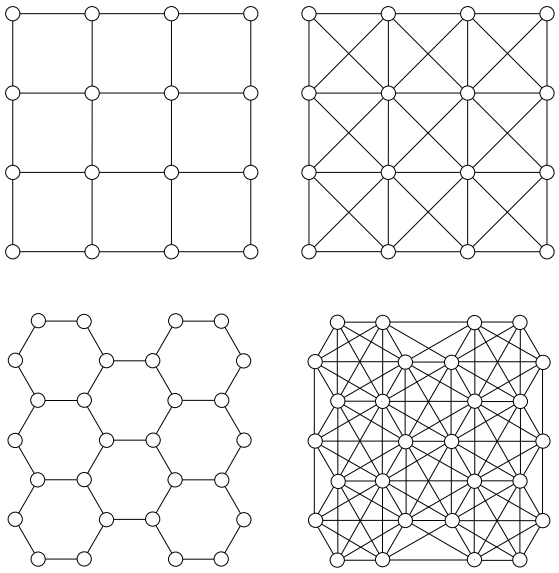

FIGURE 1.1. A finite sublattice of the square lattice $\mathbb{Z}^2$ (above left) and the hexagonal lattice $\mathbb{H}$ (below left) and their respective matching lattices (right).

Fixing $\mathcal{L} \in \{\mathbb{Z}^2, \mathbb{H}\}$, it is trivial that $r_c^{\mathcal{L}}(p) = 0$ for all $p > p_c^{\mathcal{L}}$, and it easily follows from classical results on Bernoulli bond percolation that $r_c^{\mathcal{L}}(p_c^{\mathcal{L}}) = 1$ (see Bálint et al. (2009) for the case $\mathcal{L} = \mathbb{Z}^2$). However, there are only very loose theoretical bounds for the critical value when $p < p_c^{\mathcal{L}}$: the duality relation (2.2) in Section 2.2 below and renormalization arguments as in the proof of Theorem 2.6 in Häggström (2001) give that $1/2 \le r_c^{\mathcal{L}}(p) < 1$ for all such p, and Proposition 1 in Bálint et al. (2013) gives just a slight improvement of these bounds for very small values of p. Therefore, our ultimate goal here is to get good estimates for $r_c^{\mathcal{L}}(p)$ with $p < p_c^{\mathcal{L}}$.

We end this section with an outline of the paper. Section 2 contains a crucial reduction of the infinite-volume models to a *finite* situation by a criterion that is

stated in terms of a finite sublattice but nonetheless implies the existence of an infinite cluster. This method, often called static renormalization in percolation, is a particular instance of coarse graining. We then describe in Section 3 how the occurrence of this finite size criterion can be tested in an efficient way and obtain confidence intervals for $r_c^{\mathcal{L}}(p)$ as functions of uniform random variables (Proposition 3.1). Finally, we implement this method using a (pseudo)random number generator, and present and discuss the numerical results in Section 4.

2. Finite size criteria

2.1. *An upper bound for $r_c(p)$.* In this section, we will show how to obtain an upper bound for $r_c^{\mathcal{L}}(p)$ by deducing a finite size criterion for percolation in the DaC model (Proposition 2.3). This criterion, which is a quantitative form of Lemma 2.10 in Bálint et al. (2009), will play a key role in Sections 3–4. To enhance readability, we will henceforth focus on the case $\mathcal{L} = \mathbb{Z}^2$ and mention $\mathcal{L} = \mathbb{H}$ only when the analogy is not straightforward. Accordingly, we will write $\mathbb{P}_{p,r}$ and $r_c(p)$ for $\mathbb{P}_{p,r}^{\mathbb{Z}^2}$ and $r_c^{\mathbb{Z}^2}(p)$ respectively, and denote the edge set of $\mathbb{Z}^2$ by $\mathcal{E}^2$. Let us first recall a classical result (Lemma 2.2 below) concerning 1-dependent percolation.

Definition 2.1. Given a graph $G = (\mathcal{V}, \mathcal{E})$, a probability measure ν on $\{0,1\}^{\mathcal{E}}$ is called *1-dependent* if, whenever $S \subset \mathcal{E}$ and $T \subset \mathcal{E}$ are vertex-disjoint edge sets, the state of edges in S is independent of that of edges in T under ν.

It follows from standard arguments or from a general theorem of Liggett, Schonmann and Stacey (Liggett et al. (1997)) that if each edge is open with a sufficiently high probability in a 1-dependent bond percolation on $\mathbb{Z}^2$, then the origin is with positive probability in an infinite bond cluster. Currently the best bound is given by Balister, Bollobás and Walters:

Lemma 2.2. (Balister et al. (2005)) *Let ν be any 1-dependent bond percolation measure on $\mathbb{Z}^2$ in which each edge is open with probability at least* 0.8639*. Then the probability under ν that the origin lies in an infinite bond cluster is positive.*

Now, suppose that the lattice $\mathbb{Z}^2$ is embedded in the plane the natural way (so that $v = (i, j) \in \mathbb{Z}^2$ has coordinates i and j). We consider the following partition of $\mathbb{R}^2$ (see Figure 2.2): given parameters $s \in \mathbb{N} = \{1, 2, \ldots\}$ and $\ell \in \mathbb{N}$, we take $k = s + 2\ell$ and define, for all $i, j \in \mathbb{Z}$, the $s \times s$ squares

$$S_{i,j} = [ik + \ell, ik + \ell + s] \times [jk + \ell, jk + \ell + s],$$

the $s \times 2\ell$ rectangles

$$H_{i,j} = [ik + \ell, ik + \ell + s] \times [jk - \ell, jk + \ell],$$

the $2\ell \times s$ rectangles

$$V_{i,j} = [ik - \ell, ik + \ell] \times [jk + \ell, jk + \ell + s],$$

and what remains are the $2\ell \times 2\ell$ squares $[ik - \ell, ik + \ell] \times [jk - \ell, jk + \ell]$.

We will couple $\mathbb{P}_{p,r}$ to a 1-dependent bond percolation measure. Define $f : \Omega_{\mathcal{E}^2} \times \Omega_{\mathbb{Z}^2} \to \Omega_{\mathcal{E}^2}$, as follows. To each horizontal edge $e = \langle (i, j), (i + 1, j) \rangle \in \mathcal{E}^2$, we associate a $(2\ell + 2s) \times s$ rectangle $R_e = S_{i,j} \cup V_{i+1,j} \cup S_{i+1,j}$ and the event E_e that there exists a left-right black crossing in R_e (*i.e.*, a connected path of vertices all of which are black which links the left side of R_e to its right side) and

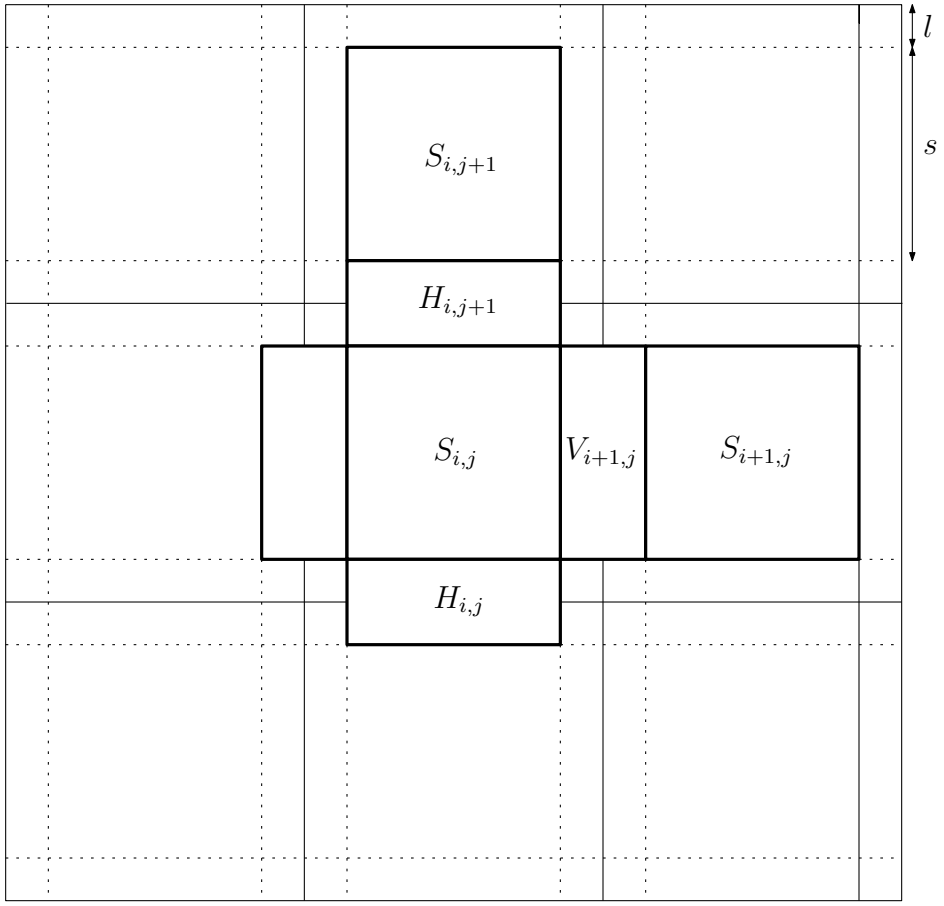

FIGURE 2.2. A partition of $\mathbb{R}^2$.

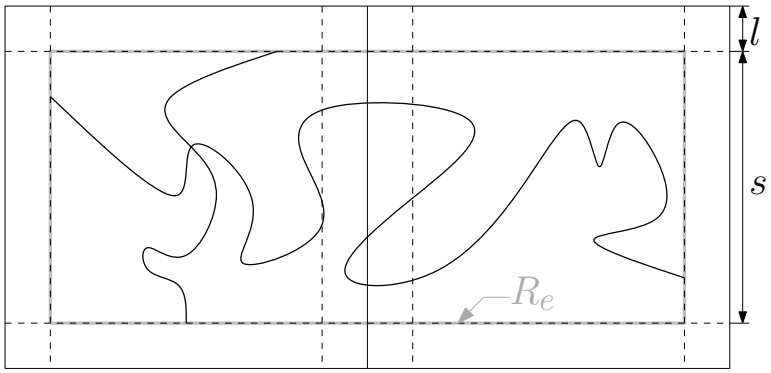

FIGURE 2.3. A black component in R_e witnesses the occurrence of E_e.

an up-down black crossing in $S_{i,j}$ (see Figure 2.3). Here and below, a vertex in the

corner of a rectangle is understood to link the corresponding sides in itself. For each vertical edge $e = \langle (i,j),(i,j+1)\rangle \in \mathcal{E}^2$, we define the $s \times (2\ell + 2s)$ rectangle $R_e = S_{i,j} \cup H_{i,j+1} \cup S_{i,j+1}$ and the event $E_e = \{$up-down black crossing in R_e and left-right black crossing in $S_{i,j}\} \subset \Omega_{\mathcal{E}^2} \times \Omega_{\mathbb{Z}^2}$. For each edge $e \in \mathcal{E}^2$, we also consider the event $F_e = \{$there exists a bond cluster which contains a vertex in R_e and a vertex at graph distance at least ℓ from $R_e\} \subset \Omega_{\mathcal{E}^2} \times \Omega_{\mathbb{Z}^2}$, and define $\tilde{E}_e = E_e \cap F_e^c$. Now for each configuration $\omega = (\eta, \xi) \in \Omega_{\mathcal{E}^2} \times \Omega_{\mathbb{Z}^2}$, we determine a corresponding bond configuration $f(\omega) = \gamma \in \Omega_{\mathcal{E}^2}$ as follows: for all $e \in \mathcal{E}^2$, we declare e open if and only if $\tilde{E}_e$ holds (*i.e.*, we define $\gamma(e) = 1$ if and only if $\omega \in \tilde{E}_e$). Finally, we define the probability measure $\nu = f_* \mathbb{P}_{p,r}$ on $\Omega_{\mathcal{E}^2}$.

It is not difficult to check that ν is a 1-dependent bond percolation measure. Indeed, if e and e' are two vertex-disjoint edges in $\mathcal{E}^2$, then the corresponding rectangles R_e and $R_{e'}$ are at graph distance at least 2ℓ from one another, hence F_e^c and $F_{e'}^c$ are independent. Given that F_e and $F_{e'}$ do not hold, the bond clusters in R_e and $R_{e'}$ are colored independently of each other. Keeping this in mind, a short computation proves the independence of $\tilde{E}_e$ and $\tilde{E}_{e'}$ under $\mathbb{P}_{p,r}$, which implies the 1-dependence of ν.

Note also that the function f was chosen in such a way that if $\gamma = f(\omega) \in \Omega_{\mathcal{E}^2}$ contains an infinite open bond cluster, then ω contains an infinite black cluster. Such configurations have zero $\mathbb{P}_{p,r}$-measure for $r < r_c(p)$. Finally, note that $\mathbb{P}_{p,r}(\tilde{E}_e)$ is the same for all edges $e \in \mathcal{E}^2$. These observations combined with Lemma 2.2 imply that, denoting $\langle (0,0),(0,1)\rangle \in \mathcal{E}^2$ by e_1, we have the following result.

Proposition 2.3. *Given any values of the parameters $s, \ell \in \mathbb{N}$, if p and r are such that*

$$\mathbb{P}_{p,r}(\tilde{E}_{e_1}) \geq 0.8639, \tag{2.1}$$

then $r_c(p) \leq r$.

Note that Proposition 2.3 is indeed a finite size criterion since the event $\tilde{E}_{e_1}$ depends on the state of a finite number of edges and the color of a finite number of vertices. A similar criterion, which will imply a lower bound for $r_c(p)$, will be given in Section 2.3.

2.2. *Duality.* A concept that is essential in understanding site percolation models on $\mathcal{L} \in \{\mathbb{Z}^2, \mathbb{H}\}$ is that of the *matching lattice* $\mathcal{L}^*$ which is a graph with the same vertex set, $\mathcal{V}$, as $\mathcal{L}$ but more edges: the edge set $\mathcal{E}^*$ of $\mathcal{L}^*$ consists of all the edges in $\mathcal{E}$ plus the diagonals of all the faces of $\mathcal{L}$ (see Figure 1.1). The finiteness of a monochromatic cluster in $\mathcal{L}$ can be rephrased in terms of circuits of the opposite color in $\mathcal{L}^*$ and vice versa; see Kesten (1982) for further details. We say that $B \subset \mathcal{V}$ is a *black $*$-component* in a color configuration $\xi \in \Omega_{\mathcal{V}}$ if it is a black component in terms of the lattice $\mathcal{L}^*$ (*i.e.*, $\xi(v) = 1$ for all $v \in B$ and B is connected via $\mathcal{E}^*$).

Accordingly, there is yet another phase transition in the DaC model on $\mathcal{L}$ at the point where an infinite black $*$-component appears; formally, for each fixed $p \in [0,1]$, one can define $r_c^*(p,\mathcal{L})$ as the value such that $\mathbb{P}_{p,r}^{\mathcal{L}}$(there exists an infinite black $*$-component) is 0 for $r < r_c^*(p,\mathcal{L})$ and positive for $r > r_c^*(p,\mathcal{L})$. It was proved in Bálint et al. (2009) that there is an intimate connection between all the critical values in the DaC model that we mentioned so far; namely, for all $p < p_c^{\mathcal{L}}$,

$$r_c^{\mathcal{L}}(p) + r_c^*(p,\mathcal{L}) = 1. \tag{2.2}$$

Actually, this relation was proved only for $\mathcal{L} = \mathbb{Z}^2$, but essentially the same proof gives the result for $\mathcal{L} = \mathbb{H}$ as well. The importance of this result here is that due to the duality relation (2.2), a lower bound for $r_c^{\mathcal{L}}(p)$ may be obtained by giving an upper bound for $r_c^*(p, \mathcal{L})$.

2.3. *A lower bound for* $r_c(p)$. As in Section 2.1, we will focus on $\mathcal{L} = \mathbb{Z}^2$ since the case $\mathcal{L} = \mathbb{H}$ is analogous; we denote $r_c^*(p, \mathbb{Z}^2)$ here and in the next section by $r_c^*(p)$. Obviously $r_c(p)$ itself is an upper bound for $r_c^*(p)$. However, a better bound may be obtained by a slight modification of the approach given in Section 2.1. For each $e \in \mathcal{E}^2$, let R_e and F_e be as in Section 2.1, define E_e^* by substituting black $*$-component for black component in the definition of E_e, and take $\tilde{E}_e^* = E_e^* \cap F_e^c$. Then, by similar arguments as those before Proposition 2.3 and using (2.2), we get the following:

Proposition 2.4. *Given any values of the parameters* $s, \ell \in \mathbb{N}$*, if* p *and* r *are such that*

$$\mathbb{P}_{p,r}(\tilde{E}_{e_1}^*) \geq 0.8639, \tag{2.3}$$

then $r_c^*(p) \leq r$*, and hence* $r_c(p) \geq 1 - r$.

3. The confidence interval

The main idea in Balister et al. (2005); Riordan and Walters (2007) is to reduce a stochastic model to a new model in finite volume by criteria similar in spirit to those in Section 2 and do repeated (computer) simulations of the new model to test whether the corresponding criteria hold. The point is that after a sufficiently large number of simulations, one can see with an arbitrarily high level of confidence whether or not the probability of an event exceeds a certain threshold. By the special nature of the events in question, statistical inferences regarding the original, infinite-volume model may be made from the simulation results.

To be able to follow this strategy, we will have to refine Propositions 2.3–2.4 as those are concerned with the state of finitely many objects, but still in the infinite-volume model. The adjusted criteria that truly are of finite size are given below, see (3.1) and (3.2). Finding an efficient way of performing the simulation step involves further obstacles. The main problem is that it would be unfeasible to run a large number of separate simulations for different values of r to find, for a fixed p, the lowest value of r such that both (3.1) and (3.2) seem sufficiently likely to hold. We will tackle this difficulty with a stochastic coupling, which is the simultaneous construction of several stochastic models on the same probability space. Such a construction will enable us to deal with *all* values of $r \in [0, 1]$ at the same time and is very related to the model of invasion percolation.

After the description of the coupling, a "theoretical" confidence interval (meaning a confidence interval as a function of i.i.d. random variables) for $r_c(p)$ is given in Proposition 3.1. The numerical confidence intervals obtained by this method using computer simulations will be presented in Section 4. Note also that the inequalities (3.1) and (3.2) implicitly involve the parameters s and ℓ whose choices may influence the width of the confidence intervals obtained; this issue is addressed before the proof of Proposition 3.1. Our methods in this section work for a general $p \in [0, p_c^{\mathcal{L}})$; we note that substantial simplifications are possible in the case $p = 0$ (*i.e.*, in the absence of correlations), see Riordan and Walters (2007).

Fix $p \in [0,1/2)$ and $s, \ell \in \mathbb{N}$, and define the rectangle $\tilde{R}_{e_1} = [0, 2s+4\ell] \times [0, s+2\ell]$. Note that for a configuration $\omega \in \Omega_{\mathcal{E}^2} \times \Omega_{\mathbb{Z}^2}$, one can decide whether $\omega \in \tilde{E}_{e_1}$ (respectively $\omega \in \tilde{E}^*_{e_1}$) holds by checking the restriction of ω to $\tilde{R}_{e_1}$. In fact, defining $\tilde{G} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ as the minimal subgraph of $\mathbb{Z}^2$ which contains $\tilde{R}_{e_1}$ and considering the DaC model on $\tilde{G}$, it is easy to see that for any $r \in [0,1]$, $\mathbb{P}^{\tilde{G}}_{p,r}(\tilde{E}_{e_1}) = \mathbb{P}_{p,r}(\tilde{E}_{e_1})$ and $\mathbb{P}^{\tilde{G}}_{p,r}(\tilde{E}^*_{e_1}) = \mathbb{P}_{p,r}(\tilde{E}^*_{e_1})$. (These equalities hold despite the fact that $\mathbb{P}^{\tilde{G}}_{p,r}$ is *not* the same distribution as the projection of $\mathbb{P}_{p,r}$ on $\tilde{G}$.) Therefore, by Propositions 2.3 and 2.4,

$$\mathbb{P}^{\tilde{G}}_{p,r}(\tilde{E}_{e_1}) \geq 0.8639 \tag{3.1}$$

would imply that $r_c(p) \leq r$, and

$$\mathbb{P}^{\tilde{G}}_{p,r}(\tilde{E}^*_{e_1}) \geq 0.8639 \tag{3.2}$$

would imply that $r_c(p) \geq 1 - r$. Below we shall describe a method which tests whether (3.1) or (3.2) holds, simultaneously for all values of $r \in [0,1]$.

We construct the DaC model on $\tilde{G}$ with parameters p and an arbitrary $r \in [0,1]$ as follows. Fix an arbitrary deterministic enumeration $v_1, v_2, \ldots, v_{|\tilde{\mathcal{V}}|}$ of the vertex set $\tilde{\mathcal{V}}$, and for $V \subset \tilde{\mathcal{V}}$, let $\min(V)$ denote the vertex in V of the smallest index. For all $r \in [0,1]$, we define the function

$$\begin{array}{rccc} \Psi_r : & \Omega_{\tilde{\mathcal{E}}} \times [0,1]^{\tilde{\mathcal{V}}} & \to & \Omega_{\tilde{\mathcal{E}}} \times \Omega_{\tilde{\mathcal{V}}}, \\ & (\eta, U) & \mapsto & (\eta, \xi_r), \end{array}$$

where

$$\xi_r(v) = \begin{cases} 1 & \text{if } U(\min(C_v(\eta))) < r, \\ 0 & \text{if } U(\min(C_v(\eta))) \geq r. \end{cases}$$

Now, if $\mathbb{U}$ denotes uniform distribution on the interval $[0,1]$ and $(\eta, U) \in \Omega_{\tilde{\mathcal{E}}} \times [0,1]^{\tilde{\mathcal{V}}}$ is a random configuration with distribution $\nu_p^{\tilde{\mathcal{E}}} \otimes \mathbb{U}^{\tilde{\mathcal{V}}}$, then it is not difficult to see that $(\eta, \xi_r) = \Psi_r((\eta, U))$ is a random configuration with distribution $\mathbb{P}^{\tilde{G}}_{p,r}$.

We are interested in the following question: for what values of r does $(\eta, \xi_r) \in \tilde{E}_{e_1}$ (respectively, $(\eta, \xi_r) \in \tilde{E}^*_{e_1}$) hold? The first step is to look at the edges in η in $\tilde{R}_{e_1} \setminus R_{e_1}$ to see if there is a bond cluster which connects R_{e_1} and the boundary of $\tilde{R}_{e_1}$. If no such connection is found, it is easy to see that there exists a *threshold value* $r_1 = r_1(\eta, U) \in [0,1]$ such that for all $r \in [0, r_1)$, $(\eta, \xi_r) \notin \tilde{E}_{e_1}$, and for all $r \in (r_1, 1]$, we have that $(\eta, \xi_r) \in \tilde{E}_{e_1}$. Indeed, the color configurations are coupled in such a way that if $r' \geq r$ and $(\eta, \xi_r) \in \tilde{E}_{e_1}$ then $(\eta, \xi_{r'}) \in \tilde{E}_{e_1}$, since all vertices that are black in ξ_r are black in $\xi_{r'}$ as well. A similar argument shows that in case of $\eta \notin F_{e_1}$, there exists $r_1^* = r_1^*(\eta, U) \in [0,1]$ such that $(\eta, \xi_r) \notin \tilde{E}^*_{e_1}$ for all $r \in [0, r_1^*)$, whereas $(\eta, \xi_r) \in \tilde{E}^*_{e_1}$ for all $r \in (r_1^*, 1]$. Otherwise, *i.e.*, if there is a connection in η between R_{e_1} and the boundary of $\tilde{R}_{e_1}$, we know that neither of $\tilde{E}_{e_1}$ or $\tilde{E}^*_{e_1}$ has occurred. Hence, in that case, we define $r_1 = r_1^* = 1$, which preserves the above "threshold value" properties as $(r_1, 1] = (r_1^*, 1] = \varnothing$.

Now, if we want a confidence interval with confidence level $1 - \varepsilon$ where $\varepsilon > 0$ is fixed, we choose positive integers m and n in such a way that the probability of having at least m successes among n Bernoulli experiments with success probability 0.8639 each is smaller than (but close to) $\varepsilon/2$. For instance, for a 99.9999% confidence interval, we can choose $n = 400$ and $m = 373$. By repeating the above experiment n times, each time with random variables that are independent of all

the previously used ones, we obtain threshold values $r_1, r_2, \ldots, r_n$ and $r_1^*, r_2^*, \ldots, r_n^*$. Then we sort them so that $\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_n$, and $\tilde{r}_1^* \leq \tilde{r}_2^* \leq \ldots \leq \tilde{r}_n^*$.

Proposition 3.1. *Each of the inequalities $r_c(p) \leq \tilde{r}_m$ and $1 - \tilde{r}_m^* \leq r_c(p)$ occurs with probability at least $1 - \varepsilon/2$, hence $[1 - \tilde{r}_m^*, \tilde{r}_m]$ is a confidence interval for $r_c(p)$ of confidence level $1 - \varepsilon$.*

Before turning to the proof, we remark that the above confidence interval does not necessarily provide meaningful information. In fact, with very small ($< \varepsilon$) probability, $\tilde{r}_m < 1 - \tilde{r}_m^*$ can occur. Otherwise, for unreasonable choices of s and ℓ, taking a too small ℓ in particular, it could happen that there is a connection in the bond configuration between R_{e_1} and the boundary of $\tilde{R}_{e_1}$ in at least $n - m + 1$ experiments out of the n, in which case $[1 - \tilde{r}_m^*, \tilde{r}_m] = [0, 1]$ indeed contains $r_c(p)$ but gives no new information.

However, the real difficulty is that although a confidence interval with an arbitrarily high confidence level may be obtained with the above algorithm, we do not know in advance how *wide* the confidence interval is. The width of the interval depends on s and ℓ, and it is a difficult problem to find good parameter values. A way to make the confidence interval narrower is to decrease the value of m, but that comes at the price of having a lower confidence level.

The choices we made for the parameters s and ℓ in our simulations, together with some intuitive reasoning advocating these choices, are given in the Appendix.

Proof of Proposition 3.1. Let $\mathbb{S}$ be the probability measure on the sample space $[0,1]^{2n}$ which corresponds to the above experiment, where a realization $(\tilde{r}_1, \tilde{r}_1^*, \tilde{r}_2, \tilde{r}_2^*, \ldots, \tilde{r}_n, \tilde{r}_n^*)$ contains the (already ordered) threshold values. Let $\mathbb{B}_{0.8639}$ denote the binomial distribution with parameters n and 0.8639, and $\mathbb{B}_{a(r)}$ the binomial distribution with parameters n and $a(r) = \mathbb{P}_{p,r}^{\tilde{G}}(\tilde{E}_{e_1})$.

For $r \in [0,1]$, let N_r denote the number of trials among the n such that $\tilde{E}_{e_1}$ occurs at level r. Note that N_r has distribution $\mathbb{B}_{a(r)}$. Since $a(r) \geq 0.8639$ implies $r \geq r_c(p)$ (see inequality (3.1)), we have that $r < r_c(p)$ implies $a(r) < 0.8639$. Therefore, for all $r < r_c(p)$, $\mathbb{B}_{a(r)}$ is stochastically dominated by $\mathbb{B}_{0.8639}$. This implies that for all $r < r_c(p)$, we have that

$$\begin{aligned} \mathbb{S}(\tilde{r}_m < r) &\leq \mathbb{S}(N_r \geq m) \\ &= \mathbb{B}_{a(r)}(\{m, m+1, \ldots, n\}) \\ &\leq \mathbb{B}_{0.8639}(\{m, m+1, \ldots, n\}) \\ &\leq \varepsilon/2, \end{aligned}$$

by the definition of m and n.

Hence, for all $\delta > 0$, we have that $\mathbb{S}(\tilde{r}_m < r_c(p) - \delta) \leq \varepsilon/2$, which easily implies that $\mathbb{S}(\tilde{r}_m < r_c(p)) \leq \varepsilon/2$. We also have $\mathbb{S}(\tilde{r}_m^* < r_c^*(p)) \leq \varepsilon/2$ by a completely analogous computation, which implies by equation (2.2) that $\mathbb{S}(1 - \tilde{r}_m^* > r_c(p)) \leq \varepsilon/2$. Therefore,

$$\mathbb{S}(1 - \tilde{r}_m^* \leq r_c(p) \leq \tilde{r}_m) \geq 1 - \varepsilon,$$

which is exactly what we wanted to prove. □

4. Results of the simulations

We implemented the method described in the previous section in a computer program, and the results for parameter values $\varepsilon = 10^{-6}$, $n = 400$, $m = 373$ are given below.[1] We stress again that although the method in Section 3 that determines a confidence interval for $r_c^{\mathcal{L}}(p)$ is mathematically rigorous, the results below are obtained by using the random number generator (Mersenne Twister, available at `http://www.math.sci.hiroshima-u.ac.jp/`), therefore their correctness depends on "how random" the generated numbers are. The simulations ran on the computers of the ENS-Lyon, and yielded the confidence intervals represented in Figure 4.4.

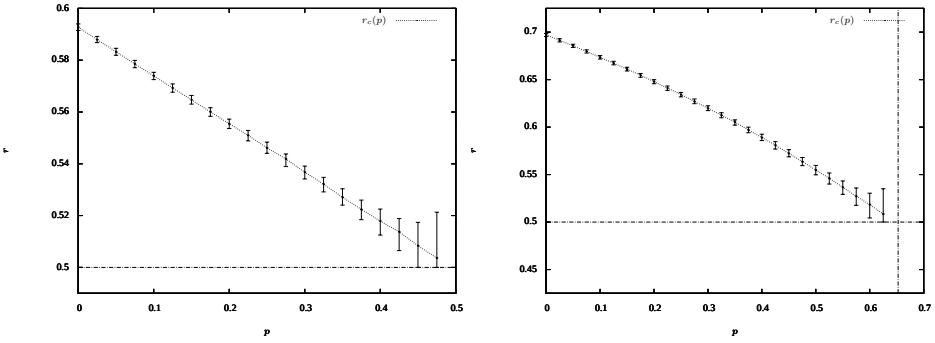

FIGURE 4.4. Simulation results for different values of $p < p_c^{\mathcal{L}}$ (left: on the square lattice; right: on the hexagonal lattice). The dashed line was obtained via a non-rigorous correction method.

Having looked at Figure 4.4, we conjecture the following concerning the behavior of $r_c^{\mathcal{L}}(p)$ as a function of p:

Conjecture 4.1. *For $\mathcal{L} \in \{\mathbb{Z}^2, \mathbb{H}\}$, in the interval $p \in [0, p_c^{\mathcal{L}})$, $r_c^{\mathcal{L}}(p)$ is a strictly decreasing function of p and*

$$\lim_{p \to p_c^{\mathcal{L}}-} r_c^{\mathcal{L}}(p) = \frac{1}{2}.$$

Since it is rigorously known that $r_c^{\mathcal{L}}(0) > 1/2$ and $r_c^{\mathcal{L}}(p) \geq 1/2$ for all $p \in [0, p_c^{\mathcal{L}})$, Conjecture 4.1 would imply that $r_c^{\mathcal{L}}(p) > 1/2$ for all $p < p_c^{\mathcal{L}}$. This suggests that the DaC model on $\mathbb{Z}^2$ or $\mathbb{H}$ is qualitatively different from the DaC model on the triangular lattice, where the critical value of r is $1/2$ for all subcritical p (see Theorem 1.6 in Bálint et al. (2009)). However, $\lim r_c^{\mathcal{L}}(p) = 1/2$ would mean that the difference disappears as p converges to $p_c^{\mathcal{L}}$.

The fact that the difference should disappear was conjectured by one of the authors (VB) and Federico Camia, based on the following heuristic reasoning. Near $p = p_c^{\mathcal{L}}$, the structure of the random graph determined by the bond configuration (whose vertices correspond to the bond clusters, and there is an edge between two vertices if the corresponding bond clusters are adjacent in $\mathcal{L}$) is given by the geometry of "near-critical percolation clusters," which is expected to be universal for 2-dimensional planar graphs. This suggests that the critical r for p close to its

[1] These results — without the description of the method — have been included in Bálint et al. (2013) as well.

critical value should not depend much on the original underlying lattice, and we expect the convergence of $r_c^{\mathcal{L}}(p)$ to $1/2$ to be universal and hold in the case of any 2-dimensional lattice.

There is an additional, strange feature appearing in the case of the square lattice: $r_c(p)$ seems to be close to being an affine function of p on the interval $[0, 1/2)$. This is not at all the same on the hexagonal lattice, and we have not found any interpretation of this observation, or of the special role $\mathbb{Z}^2$ seems to play here.

Open question 4.2. *Is $r_c^{\mathbb{Z}^2}(p)$ an affine function of p for $p < 1/2$?*

Acknowledgments. We thank Federico Camia and Ronald Meester for interesting and useful discussions and valuable comments on an earlier version of this paper. V.T. thanks for the hospitality of the VU University Amsterdam where much of the work reported here was derived during an internship provided by ENS Lyon.

Appendix A.

The algorithm in Section 3 is described for general values of s and ℓ, and the concrete values of these parameters will not affect the correctness of the simulation results. However, a reasonable choice is important for the tightness of the bounds obtained and the efficiency of the algorithm, *i.e.*, the running time of the program. The heuristic arguments given here are somewhat arbitrary, and it is quite possible that there exist other choices that would give at least as good results as ours.

Applying the method described in Section 3 requires to simulate a realization of the DaC model on the graph $\tilde{G}$, which is a $2L \times L$ rectangular subset of the square lattice where

$$L = s + 2\ell. \tag{A.1}$$

We will keep this value fixed while we let ℓ and s depend on p. Since we want to estimate the critical value for a phase transition, it is natural to take the largest L possible. After having performed various trials of our program, we chose $L = 8000$, which was estimated to be the largest value giving a reasonable time of computation.

Having fixed the size of the graph, we want to choose the parameters so that the probability of $\tilde{E}_{e_1}$ is as high as possible. We need to find a balanced value for ℓ as small values favor E_{e_1}, but a large ℓ might be required to prevent F_{e_1} from happening. The exponential decay theorem in Aizenman and Barsky (1987); Menshikov (1986) for subcritical Bernoulli bond percolation ensures the existence of an appropriate ℓ of moderate size. In our context, we decided that a good $\ell = \ell(p)$ would be one that ensures

$$\mathbb{P}_{p,r}^{\tilde{G}}(F_{e_1}) \approx 0.001. \tag{A.2}$$

We did simulations in order to find an ℓ such that (A.2) holds, then chose s according to equation (A.1). The values we used in our simulations are summed up in Table A.1.

	$\mathbb{Z}^2$		$\mathbb{H}$	
p	s	ℓ	s	ℓ
0	7998	1	7998	1
0.025	7986	7	7986	7
0.05	7986	8	7984	8
0.075	7982	9	7982	9
0.1	7980	10	7980	10
0.125	7978	11	7978	11
0.15	7976	12	7976	12
0.175	7974	13	7974	13
0.2	7970	15	7970	15
0.225	7964	18	7968	16
0.25	7962	19	7964	18
0.275	7956	22	7962	19
0.3	7948	26	7954	23
0.325	7938	31	7952	24
0.35	7926	37	7946	27
0.375	7904	48	7940	30
0.4	7876	62	7932	34
0.425	7822	89	7924	38
0.45	7704	148	7908	46
0.475	7260	370	7896	52
0.5			7876	62
0.525			7844	78
0.55			7790	105
0.575			7710	145
0.6			7538	231
0.625			7002	499

TABLE A.1. Parameters chosen

References

M. Aizenman and D. J. Barsky. Sharpness of the phase transition in percolation models. *Comm. Math. Phys.* **108** (3), 489–526 (1987). MR874906.

A. Bálint, V. Beffara and V. Tassion. On the critical value function in the divide and color model. *ALEA* **10**, 653–666 (2013).

A. Bálint, F. Camia and R. Meester. Sharp phase transition and critical behaviour in 2D divide and colour models. *Stochastic Process. Appl.* **119** (3), 937–965 (2009). MR2499865.

P. Balister, B. Bollobás and M. Walters. Continuum percolation with steps in the square or the disc. *Random Structures Algorithms* **26** (4), 392–403 (2005). MR2139873.

L. Chayes, J. L. Lebowitz and V. Marinov. Percolation phenomena in low and high density systems. *J. Stat. Phys.* **129** (3), 567–585 (2007). MR2351418.

M. Deijfen, A. E. Holroyd and Y. Peres. Stable Poisson graphs in one dimension. *Electron. J. Probab.* **16**, no. 44, 1238–1253 (2011). MR2827457.

B. Graham and G. Grimmett. Sharp thresholds for the random-cluster and Ising models. *Ann. Appl. Probab.* **21** (1), 240–265 (2011). MR2759201.

J. Gravner, D. Pitman and S. Gavrilets. Percolation on fitness landscapes: effects of correlation, phenotype, and incompatibilities. *J. Theoret. Biol.* **248** (4), 627–645 (2007). MR2899086.

O. Häggström. Coloring percolation clusters at random. *Stochastic Process. Appl.* **96** (2), 213–242 (2001). MR1865356.

C. Hsu and D. Han. Asymptotic behavior of percolation clusters with uncorrelated weights. *Theoret. and Math. Phys.* **157** (2), 309–320 (2008). MR2493783.

H. Kesten. *Percolation theory for mathematicians*, volume 2 of *Progress in Probability and Statistics*. Birkhäuser Boston, Mass. (1982). ISBN 3-7643-3107-0. MR692943.

T. M. Liggett, R. H. Schonmann and A. M. Stacey. Domination by product measures. *Ann. Probab.* **25** (1), 71–95 (1997). MR1428500.

M. V. Menshikov. Coincidence of critical points in percolation problems. *Dokl. Akad. Nauk SSSR* **288** (6), 1308–1311 (1986). MR852458.

O. Riordan and M. Walters. Rigorous confidence intervals for critical probabilities. *Phys. Rev. E* **76**, 011110 (2007). DOI: 10.1103/PhysRevE.76.011110.