A central limit theorem with applications to percolation, epidemics and Boolean models

MATHEW D. PENROSE
University of Durham

February 2001

Abstract

Suppose $X = (X_x)_{x \in \mathbb{Z}^d}$ is a white noise process, and $H(B)$, defined for finite subsets B of $\mathbb{Z}^d$, is determined in a stationary way by the restriction of X to B. Using a martingale approach, we prove a central limit theorem (CLT) for H as B becomes large, subject to H satisfying a "stabilization" condition (the effect of changing X_x at a single site needs to be local). This CLT is then applied to component counts for percolation and Boolean models, to the size of the big cluster for percolation on a box, and to the final size of a spatial epidemic.

Short title: **CLT with applications**

Keywords. Geometric probability, percolation, Boolean model, central limit theorem, martingale.

American Mathematical Society 1991 Classifications. Primary 60F05, 60D05. Secondary 60K35.

1 Introduction

The Central Limit Theorem for Martingale Difference Arrays (or Martingale CLT for short) says, loosely speaking, that if $(M_0, M_1, M_2, \ldots, M_n)$ is a discrete-time martingale, normalized so that $\mathrm{Var}(M_n) = 1$ and satisfying some mild regularity conditions, then $M_n - M_0$ is approximately standard normal provided the sum of the squares of the successive martingale differences $M_i - M_{i-1}$ is close to its mean with high probability, reducing the problem of normal approximation to proving a law of large numbers. This result is rather classical (i.e. old); the version used here is from McLeish [22], and a related result dates back to Lévy [19]; see also Doob [6].

Much more recent is the observation that the Martingale CLT can be used to prove some previously intractable central limit theorems in spatial probability. Kesten and Lee

[16] used the Martingale CLT to derive a CLT for the length of the minimal spanning tree on uniform random points in the unit cube; subsequently Lee [18] adapted the method to the number of vertices of given degree in this minimal spanning tree. Kesten and Zhang [17] used the Martingale CLT to find a CLT for first passage percolation, and Zhang [34] used it to obtain a CLT for the number of clusters of critical bond percolation on a box, and also for the intersection of the infinite cluster with a box for supercritical bond percolation.

It seems clear that the Martingale CLT is a powerful tool in spatial probability. The first aim of this paper is to formulate a general CLT that we hope is flexible enough to have a wide range of potential applications, but whose proof is nevertheless fairly simple. As well as the Martingale CLT, the proof uses another classical result, the Ergodic Theorem, as a simplifying ingredient. The general CLT and its proof are presented in Section 2. In the three subsequent sections, some applications are described. These include some aspects of lattice percolation not considered in [34] (Section 3), the final size of a spatial epidemic (Section 4), and some quantities concerned with the Boolean model of random sets in $\mathbb{R}^d$ (Section 5).

While the applications considered here are all somewhat related to percolation, further non-percolative applications concerned with the geometry of random points in the unit cube are dealt with in Penrose and Yukich [27], and we anticipate that there will be others. For example, it may be possible to use these methods to obtain CLTs for various estimators on a planar point processes in a large window, under a null hypothesis that it is a homogeneous Poisson process. See section 2.7 of Stoyan *et al.* [32], or section 8.2.2 of Cressie [5], for an introduction to these topics.

Notation. Let $d \geq 1$ be an integer. By a *lattice box* we mean a set $B \subset \mathbb{Z}^d$ of the form $B = \mathbb{Z}^d \cap \prod_{i=1}^d [a_i, b_i]$. Let $\mathcal{B}$ be the collection of all lattice boxes in $\mathbb{Z}^d$.

Except where stated to the contrary, we always define diameter of sets in $\mathbb{Z}^d$ via the l_∞ norm with $\|x\|_\infty$ given by the maximum absolute value of its coordinates. Thus, for finite $A \subset \mathbb{Z}^d$, we set $\mathrm{diam}(A) = \sup_{x,y \in A} \|x - y\|_\infty$. Also let $|A|$ denote the cardinality of A, and for $K > 0$, let $\partial_K A$ denote the set of elements of A at an l_∞ distance at most K from $\mathbb{Z}^d \setminus A$; set $\partial A = \partial_1 A$.

For $t \geq 0$ let $\lfloor t \rfloor$ denote the largest integer not exceeding t, and let $\lceil t \rceil$ denote the smallest integer not smaller than t.

For $\sigma > 0$, let $\mathcal{N}(0, \sigma^2)$ be the normal probability distribution on $\mathbb{R}$ with density $f(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/(2\sigma^2))$. Also, let $\mathcal{N}(0, 0)$ represent the degenerate probability distribution on $\mathbb{R}$ consisting of a unit point mass at zero, which we view as a special case of the normal.

2 A central limit theorem

Let $(E, \mathcal{E}, P_0)$ be an arbitrary probability space. On a suitable probability space $(\Omega, \mathcal{F}, P)$, let $X = (X_x, x \in \mathbb{Z}^d)$ be E-valued white noise with distribution P_0; that is, let $X = (X_x, x \in \mathbb{Z}^d)$ be a family of independent identically distributed random elements of E, each X_x having distribution P_0, indexed by the integer lattice. For existence of such an $(\Omega, \mathcal{F}, P)$ and X, see for example section 8.7 of Williams [33].

Suppose $\mathcal{R}$ is a collection of finite subsets ("regions") of $\mathbb{Z}^d$; for example, $\mathcal{R}$ might be the collection of all lattice boxes. Assume $\mathcal{R}$ is translation-invariant, in the sense that if $B \in \mathcal{R}$ then $\tau_y B \in \mathcal{R}$ for all $y \in \mathbb{Z}^d$, where τ_y denotes translation by y, so $\tau_y B = \{x + y : x \in B\}$. By a *stationary $\mathcal{R}$-indexed functional of X*, we mean a family $(H(X; B), B \in \mathcal{R})$ of real-valued random variables indexed by all regions in the collection $\mathcal{R}$, with the property that $(X_x, x \in B)$ determines the value of $H(X; B)$, and does so in a stationary way, meaning that $H(\tau_y X; \tau_y B) = H(X; B)$ (almost surely) for all $y \in \mathbb{Z}^d$, where $\tau_y X$ is the family of variables $(X_{x-y}, x \in \mathbb{Z}^d)$. We shall give a CLT for $H(X; B)$ as the set B becomes large.

If the variables X_x are real-valued, an example of a stationary $\mathcal{R}$-indexed functional is is the sum $\sum_{x \in B} X_x$; in this case the (very) classical CLT applies. Of more interest to us are cases where the dependence of $H(X; B)$ on $(X_x, x \in B)$ is more complicated. These will be discussed in more detail later on, but to give a flavour, we outline some of the possible interesting choices for X, H.

- *Site percolation:* Set $E = \{0, 1\}$, with $X_x = 1(0)$ representing an open (closed) site. Let $H(X; B)$ be some function of the subgraph of the integer lattice induced by the open sites in B.

- *Bond percolation:* Similar to site percolation, but let X_x have 2^d possible values, encoding the open/closed status of the d edges of the integer lattice having one endpoint at x and the other endpoint lexicographically preceding x.

- *Spatial epidemic with removal*: Let $d = 2$ and let X_x be a quintuple $(T_x, (e_{xy}))$ with y running through the neighbours of x in $\mathbb{Z}^2$. Here T_x represents the time an individual at site x remains infected once it becomes infected, before becoming immune, and e_{xy} represents the time from becoming infected before x has a contact with y.

- *Functionals on Poisson processes*: Let X_x be a homogeneous Poisson process of intensity λ on the unit cube centred at x (strictly speaking, translated to the origin to make $(X_x, x \in \mathbb{Z}^d)$ an i.i.d. family of Poisson processes). Then the union of the point processes $X_x, x \in B$, is a homogeneous Poisson process on the union of unit cubes centred at points in B. Many functionals of this Poisson process are of interest (see [27]), such as the number of components of the union of unit balls centred at the points of this Poisson process.

- *Boolean models:* Similar to the previous example, but with "unit balls" replaced by random shapes such as balls of random radius.

We are interested in normal approximation for $H(X; B_n)$ where $(B_n)_{n\geq 1}$ is an $\mathcal{R}$-valued sequence of regions with $|B_n| \to \infty$ as $n \to \infty$. In many applications we cannot expect to include *all* such sequences; instead we consider some *class* of such sequences to allow for the imposition of regularity on the sets B_n.

If $(B_n)_{n\geq 1}$ is a sequence of subsets of $\mathbb{Z}^d$ we write $\liminf(B_n)$ for the set $\cup_{n\geq 1} \cap_{m\geq n} B_m$. In particular, if $\liminf(B_n) = \mathbb{Z}^d$, we shall say the sequence *tends to* $\mathbb{Z}^d$. Restricting attention to sequences tending to $\mathbb{Z}^d$ turns out to be useful. We now give two other regularity conditions. The second of these is rather technical; some examples are given after the main theorem.

Definition 2.1 *A sequence $(B_n)_{n\geq 1}$ of regions of $\mathbb{Z}^d$ has* vanishing relative boundary *if $|\partial B_n|/|B_n| \to 0$ as $n \to \infty$.*

Observe that if $(B_n)_{n\geq 1}$ has vanishing relative boundary, then for any $K > 0$, we have $|\partial_K B_n|/|B_n| \to 0$ as $n \to \infty$. We shall use this fact below without further comment.

Definition 2.2 *A class $\mathcal{C}$ of sequences of regions of $\mathbb{Z}^d$ is* strongly translation-invariant *if for any sequence $(B_n)_{n\geq 1}$ in $\mathcal{C}$, (i) for any $y \in \mathbb{Z}^d$ the sequence $(\tau_y B_n)_{n\geq 1}$ is in $\mathcal{C}$, and (ii) for any $\varepsilon > 0$, there is a sequence $(B'_n)_{n\geq 1}$ such that (a) $B'_n \subseteq B_n$ for each n; (b) $|B'_n| \geq (1-\varepsilon)|B_n|$ for each n; and (c) for any sequence $(z_n)_{n\geq 1}$ with $z_n \in B'_n$ for all n, the sequence of translates $(\tau_{-z_n}(B_n))_{n\geq 1}$ is itself in $\mathcal{C}$.*

Our condition for a CLT to hold is a general formulation of the notion of *stabilization* introduced in the context of minimal spanning trees by Lee [18]. It says, roughly, that the effect on H of changing the value of X at a single site is local. Let X' be the process X with the value X_0 at the origin replaced by an independent copy X_* of X_0 (that is, an E-valued variable X_* with distribution P_0, independent of X), but with the values at all other sites the same. Let $\Delta_0(B)$ (the "effect of changing X_0") be the increment $H(X; B) - H(X'; B)$. For $x \in \mathbb{Z}^d$, define $\Delta_x(B)$ (the "effect of changing X_x") similarly; that is, let X'' be the process X with X_x replaced by X_*, and let $\Delta_x(X; B) = H(X; B) - H(X''; B)$.

Definition 2.3 *Given a class $\mathcal{C}$ of $\mathcal{R}$-valued sequences, the $\mathcal{R}$-indexed functional $(H(X; B), B \in \mathcal{R})$ is defined to* stabilize *on sequences in $\mathcal{C}$ if there exists a random variable $\Delta_0(\infty)$ such that for any $\mathcal{R}$-valued sequence $(B_n)_{n\geq 1}$ in class $\mathcal{C}$, the variables $\Delta_0(B_n)$ converge in probability to $\Delta_0(\infty)$.*

Definition 2.4 *The $\mathcal{R}$-indexed functional $(H(X; B), B \in \mathcal{R})$ satisfies the* bounded moments condition *if there exists $\gamma > 2$ such that*

$$\sup_{B\in\mathcal{R}} \mathbb{E}[|\Delta_0(B)|^\gamma] < \infty. \tag{2.1}$$

For $x \in \mathbb{Z}^d$, let $\mathcal{F}_x$ be the σ-field generated by $(X_y, y \preceq x)$, where $y \preceq x$ means y precedes or equals x in the lexicographic ordering on $\mathbb{Z}^d$. Write $\xrightarrow{\mathcal{D}}$ for convergence in distribution as $n \to \infty$. Now we can state our main result.

Theorem 2.1 *Let $\mathcal{R}$ be a a translation-invariant collection of finite subsets of $\mathbb{Z}^d$. Let $\mathcal{C}$ be a class of $\mathcal{R}$-valued sequences, such that any sequence in $\mathcal{C}$ tends to $\mathbb{Z}^d$ and has vanishing relative boundary, and such that $\mathcal{C}$ is strongly translation-invariant.*

Suppose $(H(X;B); B \in \mathcal{R})$ is a stationary $\mathcal{R}$-indexed functional of X which stabilizes on sequences in $\mathcal{C}$ and satisfies the bounded moments condition. Suppose that $(B_n)_{n\geq 1}$ is a sequence in class $\mathcal{C}$. Then

$$\lim_{n\to\infty} |B_n|^{-1}\mathrm{Var}(H(X;B_n)) = \sigma^2 \tag{2.2}$$

and

$$|B_n|^{-1/2}(H(X;B_n) - \mathbb{E}\, H(X;B_n)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2), \tag{2.3}$$

with $\sigma^2 = \mathbb{E}\left[(\mathbb{E}\left[\Delta_0(\infty)|\mathcal{F}_0\right])^2\right]$.

Remarks. (i) In practice, for all examples considered here, we take $\gamma = 4$ in the bounded moments condition.

(ii) The conclusion of the theorem allows for the possibility that $\sigma^2 = 0$, in which case (2.2) implies (2.3). The conclusion is perhaps more interesting when $\sigma^2 > 0$. In our examples we check this on a case-by-case basis, either directly using the definition of σ^2 in terms of conditional expectation, or using (2.2).

(iii) Here are some examples of classes that might be of interest. These cover all applications of Theorem 2.1 given in this paper.

If $\mathcal{R}$ is the collection of all finite subsets of $\mathbb{Z}^d$, and $\mathcal{C}$ is the class of all $\mathcal{R}$-valued sequences that tend to $\mathbb{Z}^d$ and have vanishing relative boundary, then $\mathcal{C}$ is strongly translation-invariant. Indeed, if $(B_n) \in \mathcal{C}$, and $\varepsilon > 0$, then if we take K_n to be the largest value of k such that $|\partial_k B_n| \leq \varepsilon |B_n|$, then since (B_n) has vanishing relative boundary, $K_n \to \infty$. If we take $B'_n = B_n \setminus \partial_{K_n} B$, then for any sequence of $x_n \in B'_n$ the translate $\tau_{-x_n}(B_n)$ includes the cube $[-K_n, K_n]^d$, so the sequence $(\tau_{-x_n}(B_n))_{n\geq 1}$ tends to $\mathbb{Z}^d$, giving us strong translation-invariance.

If $\mathcal{R}$ is the collection $\mathcal{B}$ of all *lattice boxes*, and $\mathcal{C}$ is the class of all $\mathcal{B}$-valued sequences tending to $\mathbb{Z}^d$, then each sequence $(B_n)_{n\geq 1}$ in $\mathcal{C}$ has vanishing relative boundary. As in the previous example, $\mathcal{C}$ is strongly translation-invariant.

The next example formalizes the notion of a sequence of boxes which grow at the same rate in all directions, for example concentric nested cubes. For $\delta > 0$, a lattice box $B = \prod_{i=1}^d ([a_i, b_i] \cap \mathbb{Z})$ is to be described as *δ-comparable* if $\min_{i\leq d}(b_i - a_i)/\max_{i\leq d}(b_i - a_i) \geq \delta$. In words, B is δ-comparable if there exists a small cube contained in B, and a large cube containing B, with the ratio between the side length of the small cube and that of the large cube at least δ. Let $\mathcal{B}_\delta$ denote the collection of all δ-comparable lattice boxes.

Suppose $(B_n = \prod_{i=1}^d([-a_{i,n}, b_{i,n}] \cap \mathbb{Z}))_{n\geq 1}$ is a sequence in $\mathcal{B}$, that tends to infinity. Let us say that the sequence $(B_n)_{n\geq 1}$ is *comparable* if we have

$$\liminf_{n\to\infty} \frac{\inf\{a_{1,n}, b_{1,n}, a_{2,n}, b_{2,n}, \ldots, a_{d,n}, b_{d,n}\}}{\sup\{a_{1,n}, b_{1,n}, a_{2,n}, b_{2,n}, \ldots, a_{d,n}, b_{d,n}\}} > 0.$$

This condition implies that there exists $\delta > 0$ such that all the boxes B_n are δ-comparable. It also implies that the origin is not too near the boundaries of the boxes B_n.

Let $\delta > 0$ and let $\mathcal{R} = \mathcal{B}_\delta$. Suppose $\mathcal{C}$ is the class of all comparable $\mathcal{R}$-valued sequences that tend to $\mathbb{Z}^d$. Then any sequence $(B_n)_{n\geq 1}$ in $\mathcal{C}$ has vanishing relative boundary. Also $\mathcal{C}$ is strongly translation-invariant. Indeed, given $(B_n)_{n\geq 1} \in \mathcal{C}$, and $\varepsilon > 0$, we can take $B'_n = B_n \setminus \partial_{K_n} B_n$ with $K_n = \lfloor (\varepsilon\delta/(2d))\mathrm{diam}(B_n)\rfloor$. Then $|B_n \setminus B'_n|$ is a union of $2d$ slabs, each of size at most $(\varepsilon/(2d))|B_n|$, and hence $|B'_n| \geq (1-\varepsilon)|B_n|$. Also, if $x_n \in B'_n$ for each n, then the sequence $(\tau_{-x_n} B_n)_{n\geq 1}$ is comparable because $\tau_{-x_n} B_n$ contains the box $([-K_n, K_n] \cap \mathbb{Z})^d$, and tends to $\mathbb{Z}^d$ for the same reason.

Proof of Theorem 2.1. Let $(B_n)_{n\geq 1}$ be a sequence of of regions in class $\mathcal{C}$. As advertised, the plan is to use the Martingale CLT. To represent $H(X; B_n) - \mathbb{E} H(X; B_n)$ as a sum of martingale differences, let $k_n = |B_n|$ and define the filtration $(\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_{k_n})$ as follows: let $\mathcal{G}_0$ be the trivial σ-field, label the elements of B_n in lexicographic order as $x_1, \ldots, x_{k_n}$, and let $\mathcal{G}_i = \mathcal{F}_{x_i}$ for $1 \leq i \leq k_n$. Then

$$H(X; B_n) - \mathbb{E} H(X; B_n) = \sum_{i=1}^{k_n} D_i$$

where we set $D_i = \mathbb{E}[H(X; B_n)|\mathcal{G}_i] - \mathbb{E}[H(X; B_n)|\mathcal{G}_{i-1}]$. By orthogonality of martingale differences,

$$\mathrm{Var}[H(X; B_n)] = \mathbb{E} \sum_{i=1}^{k_n} D_i^2.$$

By this representation of the variance, along with the Martingale CLT (Theorem (2.3) of [22]) it suffices to prove the conditions

$$\sup_{n\geq 1} \mathbb{E}\left[\max_{1\leq i\leq k_n} \left(k_n^{-1/2}|D_i|\right)^2\right] < \infty, \tag{2.4}$$

$$k_n^{-1/2} \max_{1\leq i\leq k_n} |D_i| \xrightarrow{P} 0, \tag{2.5}$$

and

$$k_n^{-1} \sum_{i=1}^{k_n} D_i^2 \xrightarrow{L^1} \sigma^2. \tag{2.6}$$

We use the following, easily checked, representation of the martingale differences:

$$D_i = \mathbb{E}[\Delta_{x_i}(B_n)|\mathcal{F}_{x_i}]. \tag{2.7}$$

It is not hard to check (2.4) and (2.5). Indeed, by (2.7) and the conditional Jensen's inequality we have

$$k_n^{-1}\mathbb{E}[\max_{i\le k_n} D_i^2] \le k_n^{-1} \sum_{i=1}^{k_n} \mathbb{E}[D_i^2] \le k_n^{-1} \sum_{i=1}^{k_n} \mathbb{E}[\Delta_{x_i}(B_n)^2]$$

which is uniformly bounded by the bounded moments assumption (2.1).

For the second condition (2.5), let $\varepsilon > 0$ and use Boole's and Markov's inequalities to obtain

$$P\left[\max_{1\le i\le k_n} |D_i| \ge k_n^{1/2}\varepsilon\right] \le \sum_{i=1}^{k_n} \frac{\mathbb{E}[|D_i|^\gamma]}{k_n^{\gamma/2}\varepsilon^\gamma},$$

which tends to zero, again by (2.1).

It remains to prove (2.6). Let $x \in \mathbb{Z}^d$. By strong translation-invariance, the sequence $(\tau_{-x}B_n)_{n\ge1}$ of translated regions is also in $\mathcal{C}$. By stationarity of H, the sequence of variables $\Delta_x(B_n)$ is almost surely the same as the sequence $\Delta_0(\tau_{-x}(B_n))$, and therefore by the stabilization assumption, the variables $\Delta_x(B_n)$ converge in probability to a limit, denoted $\Delta_x(\infty)$. For $x \in \mathbb{Z}^d$ and $B \in \mathcal{B}$, let

$$F_x(B) = \mathbb{E}[\Delta_x(B)|\mathcal{F}_x]; \qquad F_x = \mathbb{E}[\Delta_x(\infty)|\mathcal{F}_x].$$

Then $(F_x, x \in \mathbb{Z}^d)$ is a stationary family of random variables, which have finite second moment because of the bounded moments condition. Also, the σ-field of translation-invariant $\sigma(X)$-measurable events is trivial (see Durrett [8], chapter 6, lemma 4.3). We claim that by the Ergodic Theorem ([8], chapter 6, section 2),

$$k_n^{-1} \sum_{x\in B_n} F_x^2 \xrightarrow{L^1} \mathbb{E}[F_0^2] = \mathbb{E}[(\mathbb{E}[\Delta_0(\infty)|\mathcal{F}_0])^2]. \tag{2.8}$$

This is proved as follows. Let $e_1 = (1, 0, \dots, 0) \in \mathbb{Z}^d$. Given $\varepsilon > 0$, by the Ergodic Theorem we can choose $K > 0$ such that for all $n \ge K$, the average of $F_{e_1}^2, F_{2e_1}^2, \dots, F_{ne_1}^2$ is within an L^1 distance at most ε of $\mathbb{E}[F_0^2]$.

Divide B_n into one-dimensional *intervals* by which we mean maximal subsets of B_n of the form $(\mathbb{Z} \cap [a,b]) \times \{z_2\} \times \cdots \times \{z_d\}$, with $a, b, z_1, \ldots, z_d$ in $\mathbb{Z}$. Let B_n^* be the union of constituent intervals of length at least K. Let $k_n = |B_n|$ and $k_n' = |B_n^*|$. Since $(B_n)_{n \geq 1}$ has vanishing relative boundary, $\lim_{n\to\infty}(k_n'/k_n) = 1$. Writing $\|\cdot\|_1$ for the L^1-norm of random variables, we have

$$\left\| (k_n^{-1} \sum_{x \in B_n} F_x^2) - \mathbb{E}[F_0^2] \right\|_1 \leq k_n^{-1} \left\| (\sum_{x \in B_n^*} F_x^2) - k_n' \mathbb{E}[F_0^2] \right\|_1 + k_n^{-1} \left\| \left(\sum_{x \in B_n \setminus B_n^*} F_x^2 \right) - (k_n - k_n') \mathbb{E}[F_0^2] \right\|_1 . \tag{2.9}$$

By the choice of K and tranlsation-invariance, for each interval I of length at least K the average of $F_z^2, z \in I$, is within an L^1-distance ε of $\mathbb{E}[F_0^2]$. Therefore the first term on the right hand side of (2.9) is at most ε, while the second term tends to zero because $(k_n'/k_n) \to 1$. Therefore the left side of (2.9) is less than 2ε for large n, and (2.8) follows.

We need to show that $F_x(B_n)^2$ approximates to F_x^2. For any $B \in \mathcal{R}$, by Cauchy-Schwarz

$$\mathbb{E}[|F_0(B)^2 - F_0^2|] \leq (\mathbb{E}[(F_0(B) + F_0)^2])^{1/2} (\mathbb{E}[(F_0(B) - F_0)^2])^{1/2}.$$

By the definition and the conditional Jensen's inequality,

$$\begin{aligned} \mathbb{E}[(F_0(B) + F_0)^2] &= \mathbb{E}[(\mathbb{E}[\Delta_0(B) + \Delta_0(\infty)|\mathcal{F}_0])^2] \\ \leq \mathbb{E}[\mathbb{E}[(\Delta_0(B) + \Delta_0(\infty))^2|\mathcal{F}_0]] &= \mathbb{E}[(\Delta_0(B) + \Delta_0(\infty))^2] \end{aligned}$$

which is uniformly bounded by the bounded moments condition. Similarly,

$$\mathbb{E}[(F_0(B) - F_0)^2] \leq \mathbb{E}[(\Delta_0(B) - \Delta_0(\infty))^2].$$

By the stabilization and bounded moments conditions this is uniformly bounded, and moreover (see [33] A 13.2(f)), for any sequence $(\tilde{B}_n)_{n \geq 1}$ in class $\mathcal{C}$, $\mathbb{E}[|F_0(\tilde{B}_n)^2 - F_0^2|] \to 0$.

Returning to the given sequence (B_n), we now use the strong translation-invariance of $\mathcal{C}$. Given $\varepsilon > 0$, let B_n' be a sequence of subregions of B_n with relative size at least $1 - \varepsilon$ and with the property that for any sequence of $x_n \in B_n'$, the sequence $(\tau_{-x_n} B_n)_{n \geq 1}$ is in class $\mathcal{C}$. We claim that

$$\lim_{n \to \infty} \sup_{x \in B_n'} \mathbb{E}[|F_x(B_n)^2 - F_x^2|] = 0. \tag{2.10}$$

Indeed, if this were untrue we could take a sequence $(x_n)_{n \geq 1}$ with $x_n \in B_n'$ and $\limsup \mathbb{E}[|F_{x_n}(B_n)^2 - F_{x_n}^2|] > 0$. By translation-invariance, this would imply that

$\limsup \mathbb{E}[|F_0(\tau_{-x_n}(B_n))^2 - F_0^2|] > 0$, which contradicts the conclusion of the previous paragraph.

Using (2.10), the uniform boundedness of $\mathbb{E}[|F_x(B_n)^2 - F_x^2|]$, and the fact that ε can be taken arbitrarily small in the above argument, it is routine to deduce that

$$k_n^{-1} \sum_{x \in B_n} (F_x(B_n)^2 - F_x^2) \xrightarrow{L^1} 0,$$

and therefore (2.8) remains true with F_x replaced by $F_x(B_n)$; that is, (2.6) holds. ■

3 Percolation

In this section we consider percolation on the usual integer lattice $\mathbb{L}^d$, with vertex set $\mathbb{Z}^d$ and edges between all vertex pairs at an l_1-distance of 1, when restricted to a finite box. Theorem 2.1 gives a method for proving CLTs associated for various quantities associated with the percolation process as the dimensions of the box increase to infinity. Such CLTs are important in the statistical estimation of percolation parameters, a subject recently studied by Meester and Steif [24]. The quantities we consider here are the *number of clusters*, the *size of the biggest cluster*, and the *size of the cluster at the origin*. We state our results for site percolation but the method can be adapted to bond percolation, as was mentioned in Section 2. For a general reference on percolation see Grimmett [9].

Throughout this section, we assume $d \geq 2$ and take $X = (X_x, x \in \mathbb{Z}^d)$ to be a family of i.i.d. Bernoulli random variables with parameter $p \in (0,1)$. Sites $x \in \mathbb{Z}^d$ with $X_x = 1(0)$ are denoted *open* (*closed*). We denote by *open clusters* the connected components of the subgraph of $\mathbb{L}^d$ induced by the set of open vertices. Given a region B of $\mathbb{Z}^d$ (for example, a box), we denote by *open clusters in* B the connected components of the subgraph of the integer lattice $\mathbb{L}^d$ induced by the set of open vertices lying in B. Thus, for two vertices to lie in the same open cluster there must be an open path connecting them; for them to lie in the same open cluster in B there must be an open path *within* B connecting them.

Let $C_0(B)$ denote the open cluster in B containing the origin, and let C_0 be the open cluster containing the origin. By the *size* of an open cluster or an open cluster in B we mean the number of vertices it contains; in standard graph-theory terminology this would be called the *order* of the component but this use of the word 'size' is more common (and intuitive) in percolation theory. The *biggest* open cluster in B is the one with the greatest size (not necessarily unique).

Theorem 3.1 *Let $p \in (0,1)$. Let $H(X;B)$ be the number of open clusters in B. Then there exists $\sigma^2 > 0$ such that for any sequence $(B_n)_{n\geq 1}$ of regions of $\mathbb{Z}^d$, with vanishing relative boundary, and with* $\liminf(B_n) = \mathbb{Z}^d$, *we have*

$$\lim_{n\to\infty} |B_n|^{-1}\mathrm{Var}(H(X;B)) = \sigma^2$$

and

$$|B_n|^{-1/2}(H(X;B_n) - \mathbb{E}\,H(X;B_n)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2).$$

This is one of the simplest non-trivial applications of Theorem 2.1, taking $\mathcal{R}$ to consist of all finite subsets of $\mathbb{Z}^d$, and $\mathcal{C}$ to be the class of $\mathcal{R}$-valued sequences with vanishing relative boundary that tend to $\mathbb{Z}^d$. The bounded moments condition holds because changing the value of X_0 cannot alter the number of components by more than $2d$. Also, the functional $H(X;B)$ stabilizes on sequences in $\mathcal{C}$; for example if $X_0 = 0$ and $X_* = 1$, then $\Delta_0(\infty)$ is equal to the number of open clusters of X lying adjacent to the origin, minus 1.

A bond percolation version of this result (for a less general class of sequences $(B_n)_{n\geq 1}$) is in Zhang [34]. In contrast with results of this sort prior to [34], there is no requirement that p be non-critical. Moreover, there is no need for the sets B_n to be cubes or even boxes; all we need is vanishing relative boundary.

Next, consider *the size of the biggest cluster*. Let $p_c = p_c(d)$ be the infimum of all p such that $\theta(p) > 0$, where $\theta(p)$ denotes the probability that there is an infinite component including the origin. It is well-known [9] that $p_c \in (0,1)$. Consider supercritical percolation with $p_c < p < 1$, and suppose $(B_n, n \geq 1)$ is a comparable sequence of lattice boxes tending to $\mathbb{Z}^d$. We look at the biggest cluster in B_n. This cluster (and more especially its bond percolation analogue) is a spatial analogue to the "giant component" much studied in random graph theory (see for example [15]). This analogy has been pursued vigorously by Borgs *et al.* [3].

It is known that for site percolation restricted to B_n, given $\varepsilon > 0$, with high probability the size of the bigest cluster lies in the range $(1 \pm \varepsilon)\theta(p)|B_n|$, and the size of second biggest cluster is at most $\varepsilon|B_n|$. In fact the probability that this fails to happen decays exponentially in $|B_n|^{(d-1)/d}$; see Pisztora [29], or Penrose and Pisztora [28]. Hence, the size of the biggest cluster, divided by the size of the box, is a natural (and consistent) statistical estimator for $\theta(p)$, determined by the behaviour of the process inside B. The size of the biggest cluster is not considered by Zhang [34], who instead looks at the intersection of the infinite cluster with B, which is *not* determined by the process inside B. An intermediate estimator for $\theta(p)$ is suggested in [24], based on the number of sites in a cube which are connected by open paths to the boundary of a cube of twice the side-length; this is determined by the configuration inside the larger of the two cubes. It is possible to apply Theorem 2.1 to get a CLT for this estimator, too, but here we just concentrate on the biggest cluster.

Theorem 3.2 *Let $p \in (p_c, 1)$, and let $H(X;B)$ be the size of the biggest open cluster in B. Then there exists $\sigma^2 > 0$, such that for any comparable sequence (B_n) of lattice boxes tending to $\mathbb{Z}^d$, we have $\lim_{n\to\infty} |B_n|^{-1}\mathrm{Var}(H(X;B_n)) = \sigma^2$ and*

$$|B_n|^{-1/2}(H(X;B_n) - \mathbb{E}\,H(X;B_n)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2). \tag{3.1}$$

We shall prove this using Theorem 2.1; first we need some lemmas. The first one is for the case $d = 2$; in this case for integers $J \geq K > 0$ let $\rho_{J,K}$ be the probability that there is a path of open sites crossing a given $J \times K$ rectangle from left to right (i.e., the long way).

Lemma 3.1 *Suppose $d = 2$ and $p \in (p_c, 1)$. Then $1 - \rho_{3L,L}$ decays exponentially in L, that is,*

$$\limsup_{L \to \infty} L^{-1} \log(1 - \rho_{3L,L}) < 0.$$

Proof. The ingredients of the proof are standard and can be found for example in Durrett [7], Chapter 6; we give a sketch. The proof uses *duality*: for site percolation, the dual of $\mathbb{L}^2$ is the graph $(\mathbb{Z}^2, *)$ with the same vertex set but with edges between any two vertices that are unit distance apart in l_∞ norm. This graph structure induces a notion of $*$-paths and $*$-connected subsets of $\mathbb{Z}^2$.

First, we claim that $\lim_{L\to\infty} \rho_{L,L} = 1$. Suppose this were not true; then by duality, the probability of there being a closed top-to-bottom $*$-crossing of an $L \times L$ square does not tend to zero as $L \to \infty$. By the RSW theorem (see (11.70) of [9]) the probability that there is a closed $*$-circuit in the annulus $[-3L, 3L] \setminus [-L, L]$ does not tend to zero. Hence, there exists an infinite sequence $(A_n)_{n\geq 1}$ of disjoint annuli of this form, such that the probability of there being a closed $*$-circuit in A_n is bounded away from zero. Therefore with probability 1 there are (infinitely many) closed $*$-circuits around the origin, contradicting the assumption that $p > p_c$.

By the preceding claim and a further application of the RSW theorem, we have that $\lim_{L\to\infty} \rho_{2L,L} = 1$. Also, by (8) of [7], page 138, we have that if $0 < \lambda < 1$ and $1 - \rho_{2L,L} \leq \lambda/25$, then $1 - \rho_{4L,2L} \leq \lambda^2/25$. It follows from these two facts that $1 - \rho_{2L,L}$ decays exponentially in L as $L \to \infty$ through powers of 2. We can fill in the gaps between powers of 2 because if $L \leq M < 2L$, then

$$1 - \rho_{2M,M} \leq 1 - \rho_{2M,L} \leq 5(1 - \rho_{2L,L}), \tag{3.2}$$

where the first inequality is trivial and the second comes from (6) of [7], page 137. This gives exponential decay of $1 - \rho_{2L,L}$, and a further application of the second inequality in (3.2) (taking $M = 3L/2$) gives exponential decay of $1 - \rho_{3L,L}$. ■

Lemma 3.2 *Let $d \geq 2$, and let $\delta, \alpha \in (0, 1)$. Then there exists $p \in (0, 1)$ such that for site percolation on $\mathbb{Z}^d$ with parameter p, for any z_1, z_2 in any δ-comparable box in $\mathbb{Z}^d$, the probability that there is a path of open sites in the box from z_1 to z_2 exceeds α.*

Proof. First suppose $d = 2$. The proof is a variant of the well-known Peierls argument (see for example [9]), based on the fact that there is a combinatorial constant κ such that the number of $*$-connected subsets of $\mathbb{Z}^2$ including the origin of size k is at most κ^k.

Given a δ-comparable box B and given $z_1, z_2 \in B$, choose a deterministic lattice path γ in $\mathbb{L}^2$ from z_1 to z_2, of minimal length. This path will remain in B.

If there is no open path in B from z_1 to z_2, then there must be a closed $*$-path in B that separates z_1 from z_2. This $*$-path must pass through some point of γ, and if it passes through a point in γ at a distance at least k (in the l_1 norm) both from z_1 and from z_2, then the the $*$-path must contain at least $k\delta/2$ vertices (here we have used δ-comparability). Therefore, with $q = 1 - p$, the expected number of closed $*$-paths in B separating z_1 from z_2 is bounded by

$$2\sum_{k=1}^{\infty} \sum_{j=\lceil k\delta/2 \rceil}^{\infty} \kappa^j q^j \leq \frac{2(\kappa q)^{\delta/2}}{(1-\kappa q)(1-(\kappa q)^{\delta/2})}$$

which can be made less than $1 - \alpha$ by making p close enough to 1.

This completes the proof for $d = 2$. The case for $d > 2$ can be dealt with by using the case $d = 2$ repeatedly, along with the Harris-FKG inequality (see [7] or [9]). ■

In the case $d \geq 3$, a similar rôle to Lemma 3.1 is played by the following "finite slab lemma". By a "finite slab" we mean a lattice box of the form $\{1, \dots, K\} \times B$, or a rotation or tranlation thereof, where K is fixed and B is arbitrarily large. The finite slab lemma is closely related to a well-known result of Grimmett and Marstrand [10] on percolation in infinite slabs ((7.2) of [9]). A weaker version of the finite slab lemma (with a different proof) is in (7.78) of [9].

Lemma 3.3 *Let $d \geq 3$, and let $p \in (p_c, 1)$. Let $\delta > 0$. Then there exists integer $K = K(p, \delta) > 0$, and $\delta^* \in (0, 1)$, such that for any lattice box B of the form $\{1, 2, \dots, K\} \times B'$, with B' a δ-comparable lattice box in $\mathbb{Z}^{d-1}$, and any z_1, z_2 in B, the probability that there is an open path in B from z_1 to z_2 exceeds δ^*.*

Proof. By Lemma 3.2, there exists $\varepsilon_1 > 0$ such that for site percolation on $\mathbb{Z}^{d-1}$ with parameter $1 - \varepsilon_1$, for any x_1, x_2 in any $(\delta/2)$-comparable box in $\mathbb{Z}^{d-1}$, the probability that there is a path of open sites in the box from x_1 to x_2 exceeds $1/2$.

Let us say that a set $A \subset \mathbb{Z}^d$ is *occupied* if each site in A is open, and set $B(n) = [-n, n]^d$. Consider percolation with parameter $p' \in (p_c, p)$. By (7.9) of [9] one can pick integers $0 < m < n$, such that with probability close to 1, there is an open path from the boundary of $B(m)$ to some point on the boundary of $B(n)$ which is also on the boundary of some occupied translate of $B(m)$ adjacent to $B(n)$.

Set $N = (n + m + 1)$, and let $B = [-2N, 2N] \times B'$ with B' an arbitrary δ-comparable lattice box in $\mathbb{Z}^{d-1}$. If any side of B has length less than $8N$, then all sides have length less than $8N/\delta$, and $|B| \leq (8N/\delta)^d$. From now on we assume that all sides of B have length at least $8N$.

Consider the grid of vertices $\{4Nx : x \in \{0\} \times \mathbb{Z}^{d-1}\}$, and for each such vertex set $B_x = 4Nx + B(N)$. The set of vertices $x \in \{0\} \times \mathbb{Z}^{d-1}$ for which $4Nx + B(2N)$ lies entirely in the slab B is effectively a $(\delta/2)$-comparable box B'' in $\mathbb{Z}^{d-1}$.

Going back to percolation with parameter p rather than p', and arguing as in the proof of (7.2) of [9], define each successive site $x \in B''$ to be occupied if B_x contains an occupied translate of $B(m)$ that is connected by an open path in B'' to $B(m)$, and also to occupied translates of $B(m)$ contained in each of the neighbouring boxes of the form $2Ny + B(N)$, $y \in \mathbb{Z}^d$. By the proof of (7.2) of [9], it is possible to choose n and m, so that the set of occupied sites in the grid sites connected to the origin dominates the cluster at the origin for Bernoulli site percolation with parameter $1 - \varepsilon_1$ in B''.

Suppose z_1, z_2 are arbitrary vertices in B. Take grid points $x_1, x_2 \in B''$ with $z_i \in (4nx_i + B(4N))$ for $i = 1, 2$. By the above, the probability that there is an open path in B from $B(m)$ to a translate of $B(m)$ in B_{x_1} is at least $1/2$, as is the probability that there is an open path in B from $B(m)$ to a tranlate of $B(m)$ in B_{x_2}. Also the probability that $B(m)$ is occupied is $p^{(2m+1)^d}$, while the probability that $4Nx_1 + B(4N)$ is occupied is $p^{(8N+1)d}$, as is the probability that $4Nx_2 + B(4N)$ is occupied. By the Harris-FKG inequality, the intersection of these five events occurs with probability at least the product of their probabilities, and if they all occur then there is an open path in B from z_1 to z_2. This proves the result with $K = 4N + 1$ and

$$\delta^* = \min(p^{(8N/\delta)^d}, (1/4)p^{(2m+1)^d + 2(8N+1)^d}). \qquad \blacksquare$$

We shall use the next lemma to check the stabilization condition in the setting of Theorem 3.2. For $1 \leq i \leq d$ let the *width in the i-direction* of a connected subset of a lattice box B be the cardinality of its projection onto the i-th coordinate. We shall say it *crosses B in the i-direction* if its width in the i-direction is the same as that of B, and *crosses B in all directions* if it crosses B in the i-direction for $i = 1, 2, \dots, d$.

Lemma 3.4 *Let $d \geq 2$. Let $p \in (p_c, 1)$. Let $(B_n)_{n \geq 1}$ be a comparable sequence of boxes tending to $\mathbb{Z}^d$, and let $b_n = \lceil \mathrm{diam}(B_n)^{1/(2d)} \rceil$. Let A_n be the event that there is a unique open cluster in B_n that crosses B_n in all directions, and that all other open clusters in B_n have diameter less than b_n. Then $P[A_n] \to 1$ as $n \to \infty$.*

Proof. First suppose $d = 2$. Set $b_n^- = \lfloor b_n/2 \rfloor$. Let $\mathcal{D}_n$ be the collection of all non-empty rectangles of the form $[rb_n^-, (r+1)b_n^-] \times [sb_n^-, (s+3)b_n^-] \cap B_n$ or $[rb_n^-, (r+3)b_n^-] \times [sb_n^-, (s+1)b_n^-] \cap B_n$, with $r, s \in \mathbb{Z}$. The cardinality of $\mathcal{D}_n$ is $O(b_n^2)$.

By Boole's inequality and Lemma 3.1, with probability approaching 1 as $n \to \infty$, each of the rectangles in $\mathcal{D}_n$ has an open crossing the long way. If this occurs, the open crossing paths in overlapping rectangles must cross each other, so that the union of all such paths is contained in a single open cluster in B_n, that crosses B_n in all directions. Any other open cluster of diameter at least b_n must cross at least one of these rectangles, and so must be connected to the union of paths described above. This proves the result for $d = 2$.

Now suppose $d \geq 3$. A bond percolation version of the lemma is essentially in [29], Theorem 3.1, although the condition there on the B_n is a little stronger than comparability. For convenience we sketch the argument since much of [29] is irrelevant here.

First pick $K = K(p,\delta)$ as in the statement of Lemma 3.3. Then for large n, we can divide B_n into a large number of disjoint slabs of the form $B \cap ([j+1, j+K] \times \mathbb{Z}^{d-1})$. For each slab, the probability thet there is an open cluster in the slab that crosses the slab in the 2-direction is bounded away from zero, so by independence of disjoint slabs the probability that there is at least one open cluster in B_n that crosses B_n in the 2-direction tends to 1.

By a similar argument to the proof of (8.44) of [9], using Lemma 3.3, for any $i \neq j$ in $\{1, 2, \ldots, d\}$ the probability that the open cluster containing x has width at least b_n in the i-direction but does not cross B in the j-direction, decays exponentially in b_n, uniformly over $x \in B_n$. Therefore by an application of Boole's inequality, every cluster with diameter at least b_n crosses B_n in all directions, with probability approaching 1.

Choose distinct x and y in B_n, both on the left edge of B_n, i.e. both with minimal 1-coordinate. The probability that the open clusters including x and including y are distinct but both cross B_n in the 1-direction, decays exponentially in $\mathrm{diam}(B_n)$, uniformly in x, y, again by an argument similar to the proof of (8.44) in [9]. Hence by Boole's inequality, with probabilty tending to 1 there is at most one open cluster crossing B_n in the 1-direction. Combined with the conclusion of the previous paragraph, this completes the proof for $d \geq 3$. ■

We shall use the next lemma to check the bounded moments condition in the setting of Theorem 3.2. Recall that $\mathcal{B}_\delta$ is the set of all δ-comparable lattice boxes.

Lemma 3.5 *Let $d \geq 2$. Let $p \in (p_c, 1)$. Let e_1 and e_2 be distinct elements of $\mathbb{Z}^d$ both lying adjacent to the origin. For each box $B \in \mathcal{B}$ and $r > 0$, let C_{e_i} be the open cluster in B containing e_i, and let $E'(B;r)$ denote the event*

$$\{\mathrm{diam}(C_{e_1}(B)) \geq r\} \cap \{\mathrm{diam}(C_{e_2}(B)) \geq r\} \cap \{e_2 \notin C_{e_1}(B)\}.$$

Then for all $\delta > 0$,

$$\limsup_{r\to\infty}(r^{-1} \sup\{\log P[E'(B;r)] : B \in \mathcal{B}_\delta, \{e_1, e_2\} \subseteq B\}) < 0.$$

Proof. First suppose $d = 2$. For $L > 0$, let Q_L be the square $[-3L, 3L] \times [-3L, 3L]$, and let A_L be the annulus $Q_L \setminus (-L, L) \times (-L, L)$. Then A_L is the union of four overlapping rectangles, each of which has dimensions $6L \times 2L$ or $2L \times 6L$. We denote these four rectangles $A_L^{\mathrm{left}} := [-3L, -L] \times [-3L, 3L]$, $A_L^{\mathrm{up}} := [-3L, 3L] \times [L, 3L]$, and A_L^{right}, A_L^{down}, defined similarly. Let E_L be the event that there is a long-way open crossing of each of A_L^{left}, A_L^{right}, A_L^{up}, and A_L^{down}.

Suppose $r > 0$ and B is a δ-comparable lattice box, such that the event $E'(B;r)$ has non-zero probability. This implies that the box B includes the origin, and also includes a point at an l_∞-distance at least $r/2$ from the origin; otherwise event $E'(B;r)$ is impossible. Then *if $E_{\lfloor \delta r/6 \rfloor}$ occurs, then $E'(B;r)$ does not*, as we shall now show.

Since B includes some point x with $\|x\|_\infty \geq r$, by δ-comparability, it must include at least one of the four corners of $Q_{\lfloor \delta r/6 \rfloor}$. We consider only the case where the top right corner of $Q_{\lfloor \delta r/6 \rfloor}$ is in B; the other three cases are treated similarly.

The first possibility is that $Q_{\lfloor \delta r/6 \rfloor} \subseteq B$. In this case, it is clear that if $E_{\lfloor \delta r/6 \rfloor}$ occurs then there is an open circuit in $A_{\lfloor \delta r/6 \rfloor}$ in which case the event $E'(B; r)$ cannot occur. If this first possibility does not happen, then either the left edge or the lower edge of $Q_{\lfloor \delta r/6 \rfloor}$, or both, lies entirely outside B.

A second possibility is that both the left edge and the lower edge of $Q_{\lfloor \delta r/6 \rfloor}$ lie entirely outside B (see Figure 1). In this case, the lower end of the rectangle $A^{\text{right}}_{\lfloor \delta r/6 \rfloor}$ is "curtailed" by the lower edge of B; but if there is a long-way (i.e. top-bottom) open crossing of the rectangle, then there is also a top-bottom crossing of its intersection with B. Similarly, if there is a right-left crossing of $A^{\text{up}}_{\lfloor \delta r/6 \rfloor}$, then there is also a right-left crossing of its intersection with B. Hence if $E_{\lfloor \delta r/6 \rfloor}$ occurs, the union of the lower and left edges of B, with the open top-bottom crossing of $A^{\text{right}}_{\lfloor \delta r/6 \rfloor} \cap B$ and the left-right crossing of $A^{\text{up}}_{\lfloor \delta r/6 \rfloor} \cap B$ form a circuit around the origin. This prevents $E'(B; r)$ from occurring.

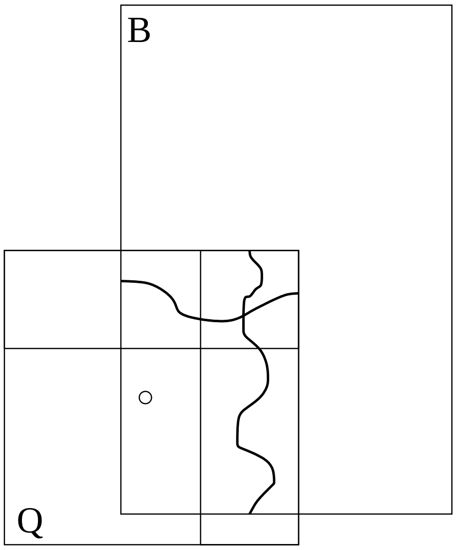

Figure 1.

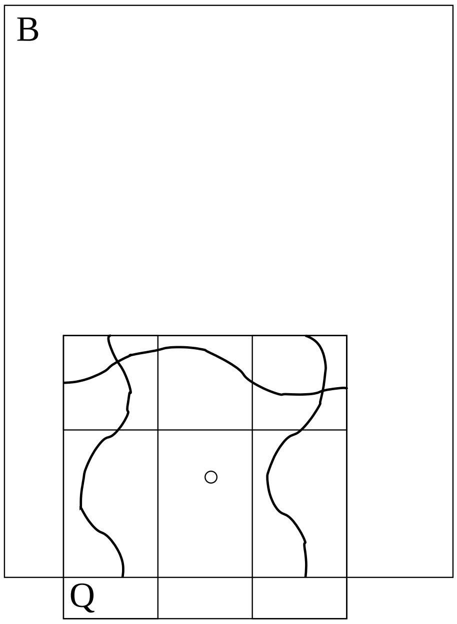

Figure 2.

A third possibility is that the lower edge of $Q_{\lfloor \delta r/6 \rfloor}$ lies entirely outside B, but the left edge does not (see Figure 2). In this case $A^{\text{up}}_{\lfloor \delta r/6 \rfloor}$ lies entirely in B, whilst $A^{\text{left}}_{\lfloor \delta r/6 \rfloor}$ and $A^{\text{right}}_{\lfloor \delta r/6 \rfloor}$ do not. The occurence of $E_{\lfloor \delta r/6 \rfloor}$ implies that there are long-way open crossings of $A^{\text{up}}_{\lfloor \delta r/6 \rfloor}$, and of $A^{\text{left}}_{\lfloor \delta r/6 \rfloor} \cap B$ and $A^{\text{right}}_{\lfloor \delta r/6 \rfloor} \cap B$. Together with the lower edge of B these form a circuit around the origin, preventing the occrurence of $E'(B; r)$.

The argument for the remaining fourth possibility is entirely analagous to that for the third possibility. Thus we have proved the claim that if $E_{\lfloor \delta r/6 \rfloor}$ occurs, then $E'(B;r)$ does not. By Lemma 3.1, $1 - P[E_L]$ decays exponentially in L, so $\sup_B P[E'(B;r)]$ decays exponentially in r, as required. This completes the proof for $d = 2$.

Now suppose $d \geq 3$. The proof uses a different annulus construction, loosely based on Section 5(f) of [10]. By Lemma 3.3, we can (and do) choose $K = K(p, \delta)$ and δ^* such that so that the probability of there being an open path between any two points in $\{1, 2, \ldots, K\} \times B$, with B an arbitrary $(1/2)$-comparable $(d-1)$-dimensional lattice box, is at least δ^*.

For integer $i > 0$, define the cube $Q'_i = [-iK, iK]$ and the annulus $A'_i = Q'_i \setminus Q'_{i-1}$. Each annulus A'_i is a union of $2d$ slabs of thickness K; let us denote the annulus *peculiar* if any of these slabs, when extended to infinity, contains a face of B. At most $2d$ of the annuli are peculiar.

Suppose $r > 0$ and B is a δ-comparable lattice box, such that the event $E'(B;r)$ has non-zero probability. Then the box B includes the origin, and also includes a point at an l_∞-distance at least $r/2$ from the origin. Therefore by δ-comparability, it must include at least one of the 2^d corners of $Q'_{i(r)}$, where we set $i(r) = \lfloor \delta r/(2K) \rfloor$. We consider only the case where $(i(r)K, \ldots, i(r)K) \in B$; the other $2^d - 1$ cases are treated similarly.

Suppose $i \leq i(r)$ and the annulus A'_i is not peculiar. We claim that $A'_i \cap B$ is a connected union of at most $2d$ slabs, each of which is the product of a interval of length K with a $(1/2)$-comparable set in $\mathbb{Z}^{d-1}$. These properties can be seen from a higher-dimensional generalization of Figures 1 and 2. More formally, connectivity follows from the fact that if one starts at an element z of $A'_i \cap B$ and moves in the direction of the j-th coordinate vector $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$, as far as the hyperplane $\{(x_1, \ldots, x_d) : x_j = i(r)K\}$, one remains in $A'_i \cap B$ along the length of this path. Thus by a series of up to d such straight-line paths one can find a path in $A'_i \cap B$ from z to $(i(r)K, \ldots, i(r)K)$. We leave the reader to verify $(1/2)$-comparability.

Suppose $i \leq i(r)$ and A'_i is not peculiar. Then, as a consequence of the preceding claim and the choice of K, given that $C_{e_1}(B)$ and $C_{e_2}(B)$ both extend to the boundary of Q'_{i-1}, the probability that they connect with each other in the intersection of B with A'_i is at least $(\delta^*)^{2d}$. Therefore, if F_i denotes the event that $C_{e_1}(B \cap Q'_i)$ and $C_{e_2}(B \cap Q'_i)$ are disjoint and both extend to the boundary of Q'_i, then $F_{i+1} \subseteq F_i$ and we have

$$P[E'(B;r)] \leq P[F_{i(r)}] \leq \prod_{i=2}^{i(r)} P[F_i | F_{i-1}]$$
$$\leq (1 - (\delta^*)^{2d})^{i(r)-2d-1} \leq \exp(-(\delta^*)^{2d} \delta (4K)^{-1} r)$$

which gives exponential decay for $\sup_B P[E'(B;r)]$, as required. ■

Proof of Theorem 3.2. If $(B_n)_{n \geq 1}$ is a comparable sequence of lattice boxes tending to $\mathbb{Z}^d$, then there exists some $\delta > 0$ such that all the boxes B_n are δ-comparable. Choose such a δ, and set $\mathcal{R} = \mathcal{B}_\delta$, the collection of all δ-comparable boxes. Let $\mathcal{C}$ be the class of all comparable $\mathcal{R}$-valued sequences that tend to $\mathbb{Z}^d$. With $H(X;B)$ denoting the size of the biggest open cluster in B, it is clear that $(H(X;B), B \in \mathcal{R})$ is a stationary $\mathcal{R}$-indexed functional. We need to show that it satisfies the hypotheses of Theorem 2.1. First we prove stabilization on sequences in $\mathcal{C}$. With $X = (X_x, x \in \mathbb{Z}^d)$ the Bernoulli process defined at the start of this section, let $X' = (X'_x, x \in \mathbb{Z}^d)$ be obtained by taking X'_0 to be an independent Bernoulli variable with parameter p, and $X'_x = X_x$ for all $x \neq 0$. Let C'_0 be defined in the same way as C_0 but referring to the process X'.

Define the random variable $\Delta_0(\infty)$ as follows. If $X_0 = X'_0$ then set $\Delta_0(\infty) = 0$. If $X_0 = 1$ and C_0 is finite, or if $X'_0 = 1$ and C'_0 is finite, then set $\Delta_0(\infty) = 0$. If $X_0 = 1, X'_0 = 0$ and C_0 is infinite, then removing 0 from C_0 splits it into at least 1 and at most $2d$ components, and with probability 1, just one of them is infinite, by the "uniqueness of the infinite cluster" property of percolation [9]. In this case set $\Delta_0(\infty)$ to be the total size of the finite components broken off by removing the origin, plus 1. If $X_0 = 0, X'_0 = 1$, and C'_0 is infinite, let $-\Delta_0(\infty)$ be the total size of the finite components broken off by removing the origin from C'_0, plus 1.

Let $(B_n)_{n\geq 1}$ be a comparable sequence of boxes tending to $\mathbb{Z}^d$, and set $b_n = \lceil \mathrm{diam}(B_n)^{1/(2d)} \rceil$. Suppose $X_0 = 1$ and $X'_0 = 0$ (the reverse case being treated similarly). If C_0 is *finite*, then the change $\Delta_0(B_n)$ is equal to zero, for all n large enough, because in this case the biggest cluster will not involve the origin. If C_0 is *infinite*, and also event A_n (defined in Lemma 3.4) occurs, and if n is large enough so the distance from the origin to ∂B_n exceeds b_n^d, then $C_0 \cap B_n$ has possibly several components; one of these includes the origin and is the biggest open cluster in B_n. Removing the origin reduces the size of the biggest cluster by the total size of the finite broken off pieces, plus 1 for the removed site at 0, so $\Delta_0(B_n) = \Delta_0(\infty)$.

By Lemma 3.4 and Borel-Cantelli, for any increasing subsequence of the natural numbers we can take a sub-subsequence such that A_n occurs for all but finitely many n in the sub-subsequence, almost surely. Therefore (see [33], A 13.2(e)) $\Delta_0(B_n) \to \Delta_0(\infty)$ in probability, so that H stabilizes on sequences in $\mathcal{C}$.

Next we check the bounded moments condition. Again suppose $X_0 = 1$ and $X'_0 = 0$. The value of $\Delta_0(B)$ is bounded by $2d$ times the size of the second largest of the pieces created by removing 0 from $C_0(B)$. Given $t > 0$, the probability that this exceeds $2dt$ is bounded above by the probability that there are at least two distinct open X'-clusters in the intersection of B with the cube of side $t^{1/d}$ centred at the origin, each of which have the origin as a neighbour and have diameter at least $(1/2)t^{1/d}$. Hence, by Lemma 3.5 there exists $\alpha > 0$ such that for large enough t, and all δ-comparable boxes B,

$$P[|\Delta_0(B)| > t] \leq \exp(-\alpha t^{1/d}),$$

and the bounded moments condition follows. Hence, by Theorem 2.1, setting $\sigma^2 = \mathbb{E}[(\mathbb{E}[\Delta_0(\infty)|\mathcal{F}_0])^2]$, we have $\lim_{n\to\infty} |B_n|^{-1}\mathrm{Var}(H(X;B_n)) = \sigma^2$, and also (3.1) holds.

It remains to prove $\sigma^2 > 0$. We do this using the definition of σ^2 directly. Let Y be the collection of variables $(X_x, x \prec 0)$. Let μ_{X_0} and μ_Y be the probability distributions of X_0 and Y respectively. Note that Y and X_0 are independent. Then by definition,

$$
\begin{aligned}
\sigma^2 &= \int \mu_{X_0}(dx) \int \mu_Y(dy) (\mathbb{E}[\Delta_0(\infty)|X_0 = x, Y = y])^2 \\
&\geq \int \mu_{X_0}(dx) \left(\int \mu_Y(dy) \mathbb{E}[\Delta_0(\infty)|X_0 = x, Y = y] \right)^2 \\
&= \int \mu_{X_0}(dx) (\mathbb{E}[\Delta_0(\infty)|X_0 = x])^2.
\end{aligned}
$$

If $X_0 = 1$ then $\Delta_0(\infty)$ is non-negative and has a strictly positive probability of being strictly positive; see the description of $\Delta_0(\infty)$ earlier on in this proof. Hence, $\mathbb{E}[\Delta_0(\infty)|X_0 = 1] > 0$, and since $\mu_{X_0}(\{1\}) > 0$ this ensures that $\sigma^2 > 0$. ■

Finally in this section, we derive a CLT for the size $|C_0(B)|$ of the cluster in B including the origin. Again, this quantity is determined entirely by the status of sites inside B. Let $C_b(B)$ denote the biggest open cluster in B; if not unique, select one of the maximal open clusters in B using some pre-specified deterministic rule. Also, let $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-t^2/2) dt$, the distribution function of the standard normal. In effect, the next CLT is for the *conditional* distribution of $|C_0(B)|$ given that $C_0(B)$ is "big"; the actual distribution of $|C_0(B)|$, suitably scaled and centred, converges to a defective distribution with an atom of size $1 - \theta(p)$ at $-\infty$ and with its remaining mass spread over the real line in a Gaussian way.

Theorem 3.3 *Let $p \in (p_c, 1)$. Let $(B_n)_{n \geq 1}$ be a comparable sequence of lattice boxes tending to $\mathbb{Z}^d$. Let $a_n = \mathbb{E}[|C_b(B_n)|]$ and let σ^2 be the same (strictly positive) constant as in Theorem 3.2. Then for $t \in \mathbb{R}$,*

$$
\lim_{n \to \infty} P\left[\frac{|C_0(B_n)| - a_n}{\sigma \sqrt{|B_n|}} > t \right] = \theta(p)(1 - \Phi(t)). \tag{3.3}
$$

Proof. For each n set $b_n = \lceil \mathrm{diam}(B_n)^{1/(2d)} \rceil$. Let $Q_n = ([-b_n, b_n] \cap \mathbb{Z})^d$. We start by showing that

$$
\lim_{n \to \infty} a_n / |B_n| = \theta(p). \tag{3.4}
$$

Since $a_n = \sum_{x \in B_n} P[x \in C_b(B_n)]$, it suffices to show that $P[x \in C_b(B_n)]$ is close to $\theta(p)$ uniformly for $x \in B_n$ at a distance at least b_n from the boundary of B_n. For such x we use the fact that

$$
\begin{aligned}
P[x \in C_b(B_n)] = P[\mathrm{diam} C_x(B_n) \geq b_n] &- P[\{\mathrm{diam} C_x(B_n) \geq b_n\} \cap \{x \in C_b(B_n)\}^c] \\
&+ P[x \in C_b(B_n) \cap \{\mathrm{diam} C_x(B_n) \geq b_n\}^c].
\end{aligned}
$$

In the right hand side, the first probability tends to $\theta(p)$. The second and third probabilities both tend to zero because of Lemma 3.4. This proves (3.4).

Let $t \in \mathbb{R}$ and define the events

$$F = \{C_0(Q_n) \cap \partial Q_n \neq \emptyset\},$$

and

$$G_b = \left\{ \frac{|C_b(B_n)| - a_n}{\sigma\sqrt{|B_n|}} \geq t \right\}; \qquad G_0 = \left\{ \frac{|C_0(B_n)| - a_n}{\sigma\sqrt{|B_n|}} \geq t \right\}.$$

We shall prove that

$$\lim_{n\to\infty} P[G_b|F] = 1 - \Phi(t). \tag{3.5}$$

Before proving this let us see why it is useful. If F^c occurs then $|C_0(B_n)| \leq (2b_n)^d \ll |B_n|$. Therefore by (3.4), for large enough n, we have $(|C_0(B_n)| - a_n)/(\sigma\sqrt{|B_n|}) < t$. Moreover, by definition $G_0 \subseteq G_b$, so for large n, we have

$$P[G_0] = P[G_0 \cap F] = P[G_b \cap F] - P[(G_b \setminus G_0) \cap F].$$

If $(G_b \setminus G_0) \cap F$ occurs then there are two (or more) disjoint clusters in B_n of diameter at least b_n, and the probability of this occurring tends to zero by Lemma 3.4. Hence, since $\lim_{n\to\infty}(P[F]) = \theta(p)$, the desired result (3.3) follows from (3.5).

It remains to prove (3.5). Since events F and G_b are both increasing, by the Harris-FKG inequality we have $P[G_b|F] \geq P[G_b]$, and by Theorem 3.2,

$$\liminf_{n\to\infty} P[G_b|F] \geq \liminf_{n\to\infty} P[G_b] = 1 - \Phi(t).$$

For an inequality the other way, let Q_n^+ be the cube $([-3b_n - 1, 3b_n + 1] \cap \mathbb{Z})^d$. Let U be the event that there is at most a single cluster in the annulus $Q_n^+ \setminus Q_n$ having nonempty intersection both with ∂Q_n^+ and with the exterior boundary of Q_n. Then $P[U] \to 1$ by the same argument as the proof of Lemma 3.5.

Let $B_n^- = B_n \setminus [-b_n, b_n]^d$. Let $|C_b(B_n^-)|$ denote the size of the biggest cluster in B_n^-. We assert that

$$G_b \subseteq \{|C_b(B_n^-)| \geq a_n + t\sigma\sqrt{|B_n|} - (7b_n)^d\} \cup U^c.$$

This is because if U occurs, then removing vertices in Q_n from $C_b(B_n)$ may break it into several components, but only one of these extends to the boundary of Q_n^+, so that the others have total size at most $(7b_n)^d$. Hence, using the independence of F and $|C_b(B_n^-)|$, we have

$$\begin{aligned} P[G_b \cap F] &\leq P[|C_b(B_n^-)| \geq a_n + t\sigma\sqrt{|B_n|} - (7b_n)^d]P[F] + P[U^c] \\ &\leq P[|C_b(B_n)| \geq a_n + t\sigma\sqrt{|B_n|} - (7b_n)^d]P[F] + P[U^c]. \end{aligned}$$

Taking $n \to \infty$, using Theorem 3.2 and the fact that $P[U^c] \to 0$, we have

$$\limsup_{n\to\infty} P[G_b \cap F] \leq (1 - \Phi(t))\theta(p),$$

completing the proof of (3.5). ■

4 The final size of a spatial epidemic

In this section we consider a model for a spatial "Susceptible-Infected-Removed" epidemic. An individual is initially susceptible to disease, and at some stage may become infected by an infected neighbour; after being infected for a period of time, the individual is "removed" from the system, representing either death or recovery with permanent immunity. The model can also be thought of as representing the spread of a forest fire. The spatial aspect of the model comes from the placing of individuals on the lattice $\mathbb{Z}^2$. This model is the subject of Cox and Durrett [4] (also Chapter 10 of Durrett [7]) and our description follows notation from there.

There is one individual at each site of $\mathbb{Z}^2$. At time 0, the individual at the origin becomes infected while all the others are susceptible. Once an individual becomes infected, it remains infected for a nonnegative random period of time with some specified distribution F that is not a point mass at 0, before becoming permanently immune. From time to time, an individual makes "contacts" with its neighbours; whenever an infected individual makes a contact with a susceptible neighbour, the neighbour becomes infected. We think of contacts as being directed; let the distribution of the amount of time, subsequent to x becoming infected, until the first contact from x to y, be denoted F'. In [7] it is assumed that F' is exponential, consistent with contact times between neighbours occuring as homogeneous Poisson processes, but this is not essential to the discussion here.

For a formal setup, suppose on some probability space that there exists a family of independent variables T_x, defined for each $x \in \mathbb{Z}^2$, and $e_{x,y}$, defined for each neighbour pair x, y in $\mathbb{Z}^2$, such that T_x has distribution F and $e_{x,y}$ has distribution F'. The directed edge (x, y) of the lattice with vertex set $\mathbb{Z}^2$ and edges between neighbouring pairs, is deemed *open* if $e_{x,y} < T_x$ and *closed* if not. By $C_0^{\rightarrow}$ we mean the union of $\{0\}$ with the set of sites $x \in \mathbb{Z}^2$ for which there is a directed path of open directed edges starting at 0 and ending at x. Then (see [7]) $C_0^{\rightarrow}$ has the same distribution as the set of individuals ever infected.

Let $\theta = \theta(F, F')$ denote the probability that the epidemic described above affects an infinite number of individuals, that is, the probability that $C_0^{\rightarrow}$ is infinite. We shall give a CLT valid for any pair of distributions (F, F') such that $0 < \theta < 1$; such pairs exist by routine percolation arguments; see [7].

To get a CLT, consider the above model restricted to a lattice box $B \subset \mathbb{Z}^2$, including the origin. Instead of the whole of $\mathbb{Z}^2$, suppose the population consists of individuals lying on the sites of a lattice box B containing the origin. Otherwise, the model is just the same as before. Let $C_0^{\rightarrow}(B)$ denote the set of sites $x \in B$ for which there is a directed path of open edges, starting at 0 and ending at x, and lying entirely within B. Again, $C_0^{\rightarrow}(B)$ has the same distribution as the total number of sites ever infected in the epidemic on B.

By restricting the epidemic to B, we assure ourselves of a finite final size $|C_0^{\rightarrow}(B)|$ of the epidemic. Our result is a CLT for the conditional distribution of this quantity, given that the epidemic is "large", as the box grows. There are a few analogous results in the literature for other epidemic models. For example Martin-Löf [20] has a CLT for the final size of a non-spatial epidemic. Andersson and Djehiche [1] prove a law of large numbers, and conjecture a CLT, for the final size of a different sort of spatial epidemic from that considered here.

Theorem 4.1 *Suppose* $0 < \theta < 1$. *Let* $(B_n)_{n\geq 1}$ *be a comparable sequence of lattice boxes tending to* $\mathbb{Z}^2$. *Then there is a sequence of constants* a_n *and a constant* $\sigma > 0$ *such that*

$$\lim_{n\to\infty} P\left[\frac{|C_0^{\rightarrow}(B_n)| - a_n}{\sigma\sqrt{|B_n|}} > t\right] = (1-\Phi(t))\theta. \tag{4.1}$$

The proof runs along similar lines to that of Theorem 3.3. Note that the above setup gives a form of locally dependent directed bond percolation on the lattice $\mathbb{Z}^2$. We need also to consider the dual lattice, with vertex set given by translating $\mathbb{Z}^2$ by the vector $(1/2, 1/2)$, and edges in both directions between all nearest neighbour pairs. With each directed edge in the original lattice is associated a directed edge in the dual lattice, obtained by rotating the first edge by 90 degrees counterclockwise about its mid-point. We assume that the open/closed status of each directed edge (bond) in the dual lattice is the same as that of the associated bond in the original lattice.

The first step towards a proof is a version of RSW theorem for this model. Given integers $J > 0, L > 0$, let $R_{J,K}$ be the probability that there is a right-to-left crossing of the rectangle $[0, J]\times[0, K]$ by open bonds in the original directed percolation model (unlike some authors, we allow crossings using the boundary). Also let $\tilde{R}_{J,K}$ be the probability that there is a a right-to-left crossing of the rectangle $[1/2, J + 1/2] \times [1/2, K + 1/2]$ by closed bonds in the dual directed percolation model.

Lemma 4.1 *For all even L, we have* $R[3L/2, L] \geq (1-(1-R_{L,L})^{1/2})^3$ *and* $\tilde{R}[3L/2, L] \geq (1-(1-\tilde{R}_{L,L})^{1/2})^3$.

Proof. This result is (2.2) of Cox and Durrett [4]. Here we do not assume that F' is an exponential distribution but this does not affect the proof. Since there is a gap in the proof given in [4] (and also in [7]), we give more details.

Consider the rectangle $[0, (3/2)L]\times[0, L]$ as being made up of two overlapping squares. Let s be a directed path across the left hand square from right to left. Adopting the notation of [4] (also used in [7]), we are concerned with events E_s (that s is the lowest right-to-left crossing of the left hand square), F'_s (that there is a top-to-bottom crossing of the right hand square to the first part of path s), and H (that there is a right-to-left crossing of the right hand square which ends at least half way up). We follow the coupling construction in [4], where a "new" directed percolation process is created by replacing the configuration of bonds starting at vertices in the left square on or below s by an independent copy, while leaving the status of other bonds unchanged. The event F'_s

refers to the "new" process, while events E_s and H refer to the "old" one. As explained in [4], if the first visit of path s to the left boundary of the right square is less than half way up, the occurrence of $E_s \cap F'_s \cap H$ ensures that there is a right-to-left crossing (in the "old" configuration) of the entire rectangle.

We consider the union over all paths s whose first visit to the left boundary of the right square is less than half way up. The event $G = \cup_s(E_s \cap F'_s)$ in [4] is *not* increasing so we cannot be sure that $P[G \cap H] \geq P[G]P[H]$, which is the gap in [4]. Let $\tau = R_{L,L}$, and set $Q(\tau) = 1 - (1-\tau)^{1/2}$. Then

$$\begin{aligned} R_{(3/2)L,L} &\geq P[\cup_s(E_s \cap F'_s \cap H)] \\ &= \sum_s \sum_{\sigma \in A_s} P[E'_\sigma \cap F'_s \cap H] \end{aligned}$$

where A_s is the set of configurations σ of bonds (x, y) with x in the part of the left square lying on and below s, giving rise to event E_s, and E'_σ is the event that the configuration of "old" bonds starting in the left square on and below s is σ. Then

$$P[F'_s \cap H | E'_\sigma] \geq P[F'_s | E'_\sigma] P[H | E'_\sigma] = P[F'_s] P[H | E'_\sigma],$$

where the inequality follows from a version of the Harris-FKG inequality ((2.1) of [4]), and the equality comes from the construction of F'_s, which makes it independent of E'_σ. Hence

$$\begin{aligned} R_{(3/2)L,L} &\geq \sum_s \sum_{\sigma \in A_s} P[E'_\sigma] P[F'_s] P[H | E'_\sigma] \\ &= \sum_s P[F'_s] \sum_{\sigma \in A_s} P[H \cap E'_\sigma] \\ &\geq \sum_s (Q(\tau)) \sum_{\sigma \in A_s} P[H \cap E'_\sigma], \end{aligned}$$

where the last inequality follows from the square root trick (see [4] or [7]). Hence,

$$\begin{aligned} R_{(3/2)L,L} \geq Q(\tau) P[\cup_s(E_s \cap H)] &= Q(\tau) P[(\cup_s E_s) \cap H] \\ &\geq Q(\tau) P[\cup_s E_s] P[H] \geq (Q(\tau))^3 \end{aligned}$$

where the inequalities in the last line come from the Harris-FKG inequality and the square root trick, respectively.

The proof of the corresponding result for dual crossings is similar: see the remarks in [4], page 185. ■

As well as Lemma 4.1 we use the inequalities (see (2.4) and (2.11) of [4])

$$1 - R_{kL,L} \leq 4(1 - R_{(k+1)L/2,L}), \qquad k \geq 1, \tag{4.2}$$

and

$$\tilde{R}_{kL,L} \geq (\tilde{R}_{(k+1)L/2,L})^4, \qquad k \geq 1. \tag{4.3}$$

These are proved by observing Figure 2 of [4] and appealing to Boole's inequality in the first case and to the Harris-FKG inequality in the second.

Lemma 4.2 *Suppose $\theta > 0$. Then $1 - R_{3L,L}$ decays exponentially in L, that is,*

$$\limsup_{L\to\infty} L^{-1} \log(1 - R_{3L,L}) < 0. \tag{4.4}$$

Proof. By duality (see (2.9) of [4]), the events that there is an open right-to-left crossing of $[0, L+1] \times [0, L]$, and that there is a closed top-to-bottom crossing in the dual lattice of $[1/2, L+1/2] \times [-1/2, L+1/2]$, are complementary. Hence by rotation-invariance, $R_{L+1,L} + \tilde{R}_{L+1,L} = 1$, so that

$$R_{L,L} + \tilde{R}_{L,L} \geq 1. \tag{4.5}$$

We claim that $R_{L,L} \to 1$ as $L \to \infty$ through the even integers. Suppose this claim were false. Then by (4.5), $\tilde{R}_{L,L}$ does not tend to zero. By Lemma 4.1, along with (4.3), $\tilde{R}_{3L,L}$ also does not tend to zero. By a further application of the Harris-FKG inequality, there exists an increasing sequence $(L_n)_{n\geq 1}$ such that the probability of there being a closed dual circuit in the annulus $[-3L_n, 3L_n] \setminus (-L_n, L_n)$ is bounded away from zero, and these annuli are disjoint. Therefore with probability 1 there are (infinitely many) closed dual circuits around the origin, contradicting the assumption that $\theta > 0$. This proves the claim.

By a further application of Lemma 4.1, this time for open crossings, along with (4.2), we have $R_{2L,L} \to 1$ as $L \to \infty$ through the even integers. Also, if $0 < \lambda < 1$ and $1 - R_{2L,L} \leq \lambda/49$, then by (2.6) of [4], $1 - R_{4L,2L} \leq \lambda^2/49$. Hence by iteration, $1 - R_{2L,L}$ decays exponentially in L as L increases through powers of 2, and then we can deduce (4.4) using (4.2), as in the proof of Lemma 3.1. ■

Define *clusters* of the directed bond percolation model on $\mathbb{Z}^2$ or on B by setting two vertices x, y to be in the same cluster (the same cluster in B) if there are open directed paths (directed paths in B) both from x to y and from y to x. Let $C_b^{\leftrightarrow}(B)$ denote the biggest cluster in B, choosing via some deterministic rule if the biggest cluster is non-unique.

Lemma 4.3 *Suppose $\theta > 0$. Let $(B_n)_{n\geq 1}$ be a comparable sequence of boxes tending to $\mathbb{Z}^2$, and set $b_n = \lceil \mathrm{diam}(B_n)^{1/4} \rceil$. Let H_n be the event that $C_b^{\leftrightarrow}(B_n)$ has diameter at least $\mathrm{diam}(B_n) - b_n$ and that any open directed path in B_n of diameter at least b_n passes through some point in $C_b^{\leftrightarrow}(B_n)$. Then $P[H_n] \to 1$ as $n \to \infty$.*

Proof. The proof is the same as that of the case $d = 2$ of Lemma 3.4, except that now we consider the event that each of the rectangles in the collection $\mathcal{D}_n$ has long-way open crossings by directed paths *in both directions*. By Lemma 4.2, the probability of this event tends to 1, and its occurrence implies H_n, as in the proof of Lemma 3.4. ■

Proposition 4.1 *Suppose* $0 < \theta < 1$. *There exists* $\sigma > 0$ *such that for any* $\delta > 0$, *if* $(B_n)_{n \geq 1}$ *is a comparable sequence of* δ*-comparable lattice boxes tending to* $\mathbb{Z}^d$, *we have* $\lim_{n\to\infty} |B_n|^{-1}\mathrm{Var}|C_b^{\leftrightarrow}(B_n)| = \sigma^2$ *and*

$$|B_n|^{-1/2}(|C_b^{\leftrightarrow}(B_n)| - \mathbb{E}\,|C_b^{\leftrightarrow}(B_n)|) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2). \tag{4.6}$$

Proof. First note that in the directed percolation model on $\mathbb{Z}^2$, with probability 1 there exists an infinite cluster. This follows from Lemma 4.2; see page 234 and Figure 9.4 of [7], or Figure 4 of [4].

By Lemma 4.2, one can take a sequence of disjoint annuli surrounding the origin, such that there are, almost surely, open circuits surrounding the origin in infinitely many of the annuli. See Figure 9.6 of [7] or Figure 5 of [4] for a picture of such a circuit. Therefore the infinte cluster is almost surely unique; let it be denoted $C_\infty^{\leftrightarrow}$.

The proof of Proposition 4.1 is analogous to that of Theorem 3.2 but we need to take care over directed paths. Take X_x to be open/closed status of the four directed bonds emanating from x, $x \in \mathbb{Z}^2$. As usual X denotes the process $(X_x, x \in \mathbb{Z}^2)$ and X' is the same process with its value at the origin replaced by an independent copy. Let $H(X; B)$ be $|C_b^{\leftrightarrow}(B)|$.

We check stabilization only in the case where all of the four bonds emanating from the origin are closed for X some of them are open for X' (it is not hard to see other cases can be deduced from this one). In this case a finite number of vertices will be added to $C_\infty^{\leftrightarrow}$ by changing from X to X', namely those vertices x for which there are X'-open directed paths both from from x to $C_\infty^{\leftrightarrow}$ and from $C_\infty^{\leftrightarrow}$ to x, but one or both of these paths must pass through the origin. Set $-\Delta_0(\infty)$ to be the number of vertices added to $C_\infty^{\leftrightarrow}$ in this way.

Now consider the sequence of boxes B_n. Let $b_n = \lceil \mathrm{diam}(B_n)^{1/4} \rceil$, and let B_n^o be the set $B_n \setminus \partial_{b_n} B_n$. If n is large enough, then $C_\infty^{\leftrightarrow} \cap B_n^o$ is non-empty. If this occurs and also event H_n defined in Lemma 4.3 occurs, then

$$C_b^{\leftrightarrow}(B_n) \cap B_n^o = C_\infty^{\leftrightarrow} \cap B_n^o$$

and therefore the number of vertices added to $C_b^{\leftrightarrow}(B_n)$ as a result of adding the open bonds of X'_0 is precisely equal to $-\Delta_0(\infty)$. This proves the convergence in probabilty of $\Delta_0(B_n)$ to $\Delta_0(\infty)$.

Next we check the bounded moments condition with $\gamma = 4$. As in the proof of the case $d = 2$ of Lemma 3.5, we consider the square $Q_L = [-3L, 3L]^2$ and the annulus $Q_L \setminus (-L, L)^2$, made up of four overlapping rectangles A_L^{left}, A_L^{right}, A_L^{up}, and A_L^{down}. Modifying the proof of Lemma 3.5, redefine E_L to be the event that each of these four rectangles has long-way open crossings *in both directions*. So Figures 1 and 2 should be modified by replacing each path drawn with two directed paths in opposite directions.

Suppose that $0 \in B$ and one of the corners of Q_L (for example, the top right one) is in B. Suppose that E_L occurs. Suppose x and y are two elements of $B \setminus [-3L, 3L]^2$. If there is a directed open path from x to y passing through the origin, then it must pass through the annulus A_L on the way to and from the origin, and therefore (see for example Figures 1 and 2, modified as described above) there is also a directed open path from x to y that avoids the origin but instead uses the open directed paths in A_L that exist because of event E_L. The same goes for paths from y to x, and therefore x and y lie in the same cluster in B for the process X, if and only if they lie in the same cluster for the process X'.

It follows from the above that if $0 \in B$ and one of the corners of Q_L lies in B, and E_L occurs, then for any cluster C in B for X, the elements of $C \setminus Q_L$ are all in the same cluster in B for X'. Likewise, for any cluster C in B for X', the elements of $C \setminus Q_L$ are all in the same cluster in B for X. Therefore if $0 \in B$ and one of the corners of Q_L is in B, and E_L occurs, then $|\Delta_0(B)| \leq (6L+1)^2$.

If $0 \in B$ but none of the corners of Q_L lies in B, then by δ-comparability, $\mathrm{diam}(B) \leq (6L+1)/\delta$ and $|B| \leq (6L+1)^2/\delta$, providing an upper bound for $\Delta_0(B)$. Therefore, for *any* δ-comparable box B we have

$$|\Delta_0(B)| \leq (6L+1)^2/\delta, \qquad \text{on event} \quad E_L. \tag{4.7}$$

By Lemma 4.2, $1 - P[E_L]$ decays exponentially in L. Together with (4.7) this gives us a bound for $E[(\Delta_0(B))^4]$, holding uniformly over δ-comparable boxes B.

Thus the stabilization and bounded moments conditions hold. We can apply Theorem 2.1 to get the result, and σ^2 is strictly positive by a similar argument to that used in the proof of Lemma 3.2. ■

Proof of Theorem 4.1. Using (4.6), the proof of (4.1) proceeds in much the same way as that of Theorem 3.3, with the rôle of event F in that proof played by the event that there is a directed path from the origin to ∂Q_n. The rôle of event U in that proof is played by the event that there is an open circuit in the annulus $Q_n^+ \setminus Q_n$. ■

5 Boolean models

As can be seen from the volumes by Hall [13], by Meester and Roy [23], by Molchanov [25], and by Stoyan *et al.* [32], Boolean models play a central rôle in stochastic geometry and spatial statistics. We make the following fairly general formulation. Let μ_S be a *shape distribution*, that is, a probability distribution on the space $\mathcal{S}$ of all compact sets in $\mathbb{R}^d$. For measure-theoretic details see Matheron [21] page 27. Assume that μ_S is concentrated on path-connected sets. Assume also that μ_S is concentrated on a uniformly bounded collection of sets, that is, there is a finite constant K such that

$$\mu_S(\{S \in \mathcal{S} : |s| \leq K \quad \forall s \in S\}) = 1. \tag{5.1}$$

Let $\lambda > 0$. On a suitable probability space let $(\xi_i, i \geq 1)$ be an enumeration of the points of a homogeneous Poisson process of rate λ on $\mathbb{R}^d$, and let $(S_i, i \geq 1)$ be a family of random closed sets each with distribution μ_S, independent of each other of the Poisson process. Let $\Xi = \cup_{i \geq 1}(S_i + \xi_i; i \geq 1)$, where the addition operation $+\xi_i$ denotes translation of S_i by the vector ξ_i. Following the literature on this subject, we refer to each Poisson point ξ_i as a "germ", to S_i as a "grain" associated with the germ ξ_i, and to $S_i + \xi_i$ as a *random shape centred at* ξ_i. One is interested in the resulting set Ξ, which we call a Boolean model with intensity λ and shape distribution μ_S. In particular, we consider here the path-connected components of $\Xi \cap B$ ("occupied clusters in B") and of $B \setminus \Xi$ ("vacant clusters in B"), where B is a continuous-space box, or "window". In what follows, we write simply "component" for "path-connected component".

We are concerned with CLTs as the window B becomes large. Such CLTs are important in statistical estimation. For "sparse" limiting regimes with λ becoming small as the window becomes large, Hall ([12] Theorem 3.3, [13] Theorem 4.9) has proved CLTs for the number of occupied clusters of a given order (and also Poisson limit theorems for some sparse regimes). He also observes in [12], Section 3, that the total number of occupied clusters in the window satisfies a CLT in the sparse regime. Moreover, Hall [11] obtained a Poisson limit for the number of vacant clusters in a window for a "dense" limiting regime where λ becomes large as the window grows; see also Molchanov [26].

In contrast with the above, we are interested here in the "moderate-intensity" limiting regime with λ held constant as the window becomes large. Some CLTs are already known in this regime. Hall [13], Theorem 3.5, has a CLT for the total volume of the vacant region in a window (and hence also for the total occupied volume). Hall [12], Theorem 2.2, has a CLT for the number of occupied clusters of order 1 in a window, where the order of a cluster denotes the number of random shapes it comprises. More recently, Heinrich and Molchanov [14] have obtained a general CLT for measures generated by Boolean models (and other models). The method we use here complements that of [14]. We apply it to the number of occupied clusters and to the number of vacant clusters, for which it is not clear that the method of [14] is applicable; our method should also work for various other quantities, such as the number of clusters of given order.

Theorem 5.1 *Let Ξ be a Boolean model as described above. Let $(R_n, n \geq 1)$ be a sequence of sets of the form $\prod_{i=1}^d [a_i - 1/2, b_i + 1/2]$, with all a_i, b_i integer-valued, with $\liminf(R_n) = \mathbb{R}^d$. Let H_n be the number of occupied clusters in R_n for the Boolean model Ξ. Then with $|R_n|$ denoting the Lebesgue measure of R_n,*

$$\lim_{n\to\infty} |R_n|^{-1} \mathrm{Var}(H_n) = \sigma^2 \tag{5.2}$$

and

$$|R_n|^{-1/2}(H_n - \mathbb{E} H_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{5.3}$$

where σ^2 is a strictly positive constant that does not depend on the choice of sequence (R_n).

Proof. Let $X = (X_x)_{x \in \mathbb{Z}^d}$ be a family of independent Poisson processes on $[-\frac{1}{2}, \frac{1}{2}]^d \times \mathcal{S}$, each with mean measure $\lambda \mathrm{Leb} \times \mu_S$. We view (X_x) as a random subset of $[-\frac{1}{2}, \frac{1}{2}]^d \times S$ with a Poisson-distributed number of elements, each of them a pair (ξ, S) with $\xi \in [-\frac{1}{2}, \frac{1}{2}]^d$ and $S \in \mathcal{S}$. Assume the Boolean model Ξ is generated by setting

$$\Xi = \cup_{x \in \mathbb{Z}^d}(\Xi_x) \quad \text{with} \quad \Xi_x = \cup_{(\xi, S) \in X_x}(S + (x + \xi)).$$

Assume without loss of generality that K, given in (5.1), is a strictly positive integer. Given a lattice box $B \in \mathcal{B}$, let $\tilde{R}(B)$ be the continuous-space box given by taking the union of the unit cubes centred at points in B, and let $R(B)$ be the set of points in $\tilde{R}(B)$ at an l_∞ distance at least K from the complement of $\tilde{R}(B)$. Define $H(X; B)$ to be the number of occupied clusters in $R(B)$ of Ξ. Clearly $(H(X; B), B \in \mathcal{B})$ is a stationary $\mathcal{B}$-indexed functional of X. Also, if B_n is chosen so that $R(B_n) = R_n$, then $|R_n|/|B_n| \to 1$. Our aim is to apply Theorem 2.1 to the functional H defined above, with $\mathcal{R}$ being the collection $\mathcal{B}$ of all lattice boxes, and $\mathcal{C}$ being the class of all $\mathcal{B}$-valued sequences tending to $\mathbb{R}^d$.

First we show stabilization for this functional. Let Ξ^- be the set $\Xi \setminus \Xi_0$. Let Ξ_0' be an independent copy of Ξ_0. Then $\Delta_0(B)$ is the number of occupied components in B for $\Xi^- \cup \Xi_0$, minus the number of occupied components in B for $\Xi^- \cup \Xi_0'$.

Let $\mathcal{K}$ be the (almost surely finite) collection of all components of Ξ^- which intersect the set $[-3K, 3K]^d$. The configuration of Ξ_0 induces an adjacency relation on $\mathcal{K}$ whereby two components are adjacent if there is a path connecting them lying within the set Ξ_0. Similarly, Ξ_0' induces an adjacency relation on $\mathcal{K}$ in the same way. These two adjacency relations induce two different graph structures on vertex set $\mathcal{K}$; let the number of components be denoted κ for the adjacency induced by Ξ_0, κ' for the adjacency induced by Ξ_0'. Let ζ be the number of components of Ξ contained entirely within Ξ_0, and let ζ' be the number of components of $\Xi^- \cup \Xi_0'$ contained entirely within Ξ_0'.

There are only finitely many pairs of random shapes in Ξ^- that intersect $[-3K, 3K]^d$. Take R_1 big enough so that for every such pair that is connected by a path in Ξ^-, there exists such a path that stays in the set $[-R_1, R_1]^d$. Define R_2 the same way using the set Ξ rather than Ξ^-, and let R_3 be defined the same way using the set $\Xi^- \cup \Xi_0$ rather than Ξ^- or Ξ. Let $R_4 = \max(R_1, R_2, R_3)$. Then for any box B that is large enough to contain $[-R_4, R_4]^d$, we have $\Delta_0(B) = \kappa + \zeta - \kappa' - \zeta'$; thus H stabilizes.

Next we check the bounded moments condition (with $\gamma = 4$ as usual). The value of $\Delta_0(B)$ is bounded by the number of germs of Ξ in the set $[-3K, 3K]^d$, plus the number of germs of Ξ_0'. This bound has a Poisson distribution with finite fourth moments. Thus Theorem 2.1 applies, and (5.2) and (5.3) hold for some $\sigma^2 \geq 0$.

Finally, we must show that the limiting variance σ^2 is strictly positive. We shall do this using (5.2), along with an adaptation of the method of Avram and Bertsimas [2] for finding lower bounds for variance.

The value of σ^2 is independent of the choice of sequence (R_n), and therefore to show $\sigma^2 > 0$ using (5.2) we are at liberty to choose any sequence (R_n). Our choice is to take $R_n = [-n(2K+1)+\frac{1}{2}, n(2K+1)+\frac{1}{2}]^d$. Divide R_n into $(2n)^d$ non-overlapping "big cubes" of side $(2K+1)$. For each of these "big cubes", define an annulus by removing the "small cube" of side 1 centred at the centre of the big cube.

Let N_n be the number of big cubes for which there are no germs lying in the corresponding annulus. Then $\mathbb{E}[N_n]/|R_n|$ is bounded away from zero. Let $\mathcal{F}$ be the σ-field generated by the value of N_n, along with the positions of the germs not lying in the big cubes contributing to N_n, and the values of their associated grains. Then

$$\mathrm{Var}(H(R_n)) = \mathrm{Var}(\mathbb{E}[H(R_n)|\mathcal{F}]) + \mathbb{E}[\mathrm{Var}(H(R_n|\mathcal{F}))] \geq \mathbb{E}[\mathrm{Var}(H(R_n|\mathcal{F}))]. \tag{5.4}$$

Suppose we are given the value of N_n and the configuration of random shapes centred outside the big cubes contributing to N_n. The only remaining variability is from the random shapes centred inside the inner small cubes corresponding to those big cubes contributing to N_n; because each of these inner cubes is surrounded by a big empty annulus, the shapes with germs in these inner cubes have no overlap with any of the other shapes. Hence $\mathrm{Var}(H(R_n)|\mathcal{F}) = N_n \mathrm{Var}(U)$, where U is the number of components of Ξ_0. Since $P[U = 0] > 0$ and $P[U \geq 1] > 0$, we have $\mathrm{Var}(U) > 0$. It follows that $\mathrm{Var}(H(R_n)) \geq \mathbb{E}[N_n]\mathrm{Var}(U)$, and this divided by $|R_n|$ is bounded away from zero. Hence $\sigma^2 > 0$ by (5.2). ■

Vacant clusters. To obtain an analogous result to Theorem 5.1 for vacant clusters, we need some control over the number of vacant regions within a finite continuous-space box. One way to do this is to insist that all the shapes be Euclidean balls of random radius; then we can use the following result of Meester and Roy [23].

Lemma 5.1 ([23], lemma 4.5) *If k d-dimensional Euclidean balls intersect the unit cube $[0,1]^d$ then the vacant region inside the unit cube has at most $c_d k^d$ components, where c_d is a constant which depends only on the dimension.*

If the shape distribution μ_S is concentrated on Euclidean balls, then without loss of generality one can assume it is concentrated on balls centred at the origin (see [12], page 426).

Another way to control the number of vacant clusters in a bounded region is to assume that $d = 2$ and all the shapes of the Boolean model are line segments with random length and random direction. This is the *Poisson sticks* model studied previously by Roy [30] (see also [23]). In this case the analogue to Lemma 5.1 is the following:

Lemma 5.2 *Suppose $L_1, \dots, L_k$ are line segments in $\mathbb{R}^2$ of arbitrary orientations. Then $(0,1)^2 \setminus \cup_{i=1}^k L_i$ has at most 2^k components.*

Proof. Each line segment L_i, if extended to infinity, divides $\mathbb{R}^2$ into two half-spaces, denoted F_i and G_i. Enumerate as $A_1, \dots, A_\nu$ all the nonempty sets of the form $(0,1)^2 \cap \cap_{i=1}^k H_i$ with each H_i being either F_i or G_i. Then $\nu \leq 2^k$. Each component of $(0,1)^2 \setminus \cup_{i=1}^k L_i$, being open, must have nonempty intersection with at least one of the sets A_j, and since each A_j is convex so is connected, it follows that the number of components is at most ν. ■

The preceding lemmas enable us to prove a CLT for the number of vacant clusters in the interesting special cases, either of Euclidean balls or of Poisson sticks. We conjecture that Lemma 5.2 can be extended from sticks to general convex shapes in $\mathbb{R}^2$, or possibly even in $\mathbb{R}^d$; if true, the CLT could be extended to these. Controlling the number of components of the complement, in general, seems however to be harder than one might at first think.

Theorem 5.2 *Let Ξ be a Boolean model with shape distribution μ_S concentrated on Euclidean balls centred at the origin of uniformly bounded radius. Let $(R_n)_{n\geq 1}$ be a sequence of sets of the form $\prod_{i=1}^d [a_i - 1/2, b_i + 1/2]$, with all a_i, b_i integer-valued, satisfying $\liminf(R_n) = \mathbb{R}^d$. Let H_n be the number of vacant clusters in R_n of the Boolean model Ξ. Then*

$$\lim_{n\to\infty} |R_n|^{-1} \mathrm{Var}(H_n) = \sigma^2 \tag{5.5}$$

and

$$|R_n|^{-1/2}(H_n - \mathbb{E}\, H_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{5.6}$$

where σ^2 is a strictly positive constant that does not depend on the choice of sequence (R_n).

Proof. We assume that Ξ is generated as defined in the proof of Theorem 5.1, and use notation from that proof, except that now we set $H(X; B)$ to be the number of vacant clusters in $R(B)$ of Ξ. Again, this is a stationary $\mathcal{B}$-indexed functional of X, and we take $\mathcal{C}$ to consist of all $\mathcal{B}$-valued sequences tending to $\mathbb{Z}^d$.

The argument for stabillization is similar to that used in proving Theorem 5.1. Let Ξ^- be the set $\cup_{x\in\mathbb{Z}^d\setminus\{0\}}(\Xi_x)$, and let $\mathcal{K}$ be the set of all components of $\mathbb{R}^d \setminus \Xi^-$ which intersect the set $[3K, 3K]^d$. Then $\mathcal{K}$ is a finite set by Lemma 5.1. Let $\mathcal{K}_1$ (respectively $\mathcal{K}_1'$), a subset of $\mathcal{K}$, consist of those components which are obliterated by Ξ_0, that is, contained in Ξ_0 (respectively Ξ_0'). Some of the remaining components comprising $\mathcal{K}$ may be split into two or more pieces by adding Ξ_0, incrementing the number of vacant components; let ν (respectively ν') be total of such increments caused by Ξ_0 (respectively Ξ_0'); this is also finite by Lemma 5.1. Then for any box B large enough to contain all the bounded components from the collection $\mathcal{K}$, we have $\Delta_0(X;B) = \nu - \nu' - \text{Card}(\mathcal{K}_1) + \text{Card}(\mathcal{K}'_1)$. This shows stabilization.

Next we check the bounded moments condition. Let N_- (respectively N) be the number of Poisson points in $[-3K,3K]^d \setminus [-1/2,1/2]^d$ (respectively, in $[-3K,3K]^d$). For all B, the number of components obliterated by Ξ_0 is bounded by $c_d N_-^d$ because of Lemma 5.1, so that its fourth moment is bounded by $c_d^4 \mathbb{E}[N_-^{4d}]$, and likewise for components obliterated by Ξ'_0. Moreover, the increment in the number of vacant components caused by adding Ξ_0 is bounded by $c_d N^d$, so its fourth moment is bounded by $c_d^4 E[N^{4d}]$, and likewise for Ξ'_0. The bounded moments condition now follows from the fact that the Poisson distribution has a finite $(4d)$-th moment.

Finally we wish to show the limiting variance σ^2 is strictly positive. This time we consider the particular sequence of cubes $R_n = [-6K + \frac{1}{2}, 6K + \frac{1}{2}]^d$, and divide R_n into $(2n)^d$ non-overlapping "big cubes" of side $6K$. Also, for each big cube, take a concentric "small cube" of side $3K$, whose removal from the big cube leaves an annulus. This time, let N_n be the number of these annuli such that the configuration of shapes centred in the annulus covers the entire annulus. Let $\mathcal{F}$ be the σ-field generated by N_n along with the positions and associated grain values of all germs except those in small cubes lying inside the annuli contributing to N_n. Again we use (5.4) (with the new definition of $\mathcal{F}$). We have $\text{Var}(H(R_n)|\mathcal{F}) = \sum_{i=1}^{N_n} \text{Var}(U_i)$, where U_i is the change in number of vacant components in the i-th small cube when the shapes centred in that cube are added. Then the variances of the U_i are bounded below, since there is some chance that there are no germs in the small cube and some chance the the grains centred in that cube cover it entirely. Then we can use (2.2) to deduce that $\sigma^2 > 0$. ■

Theorem 5.3 *Let $d = 2$. Let Ξ be a Boolean model with shape distribution μ_S concentrated on line segments of uniformly bounded length, centred at the origin with random orientation having a nondegenerate distribution. Let $(R_n, n \geq 1)$ be a nondecreasing sequence of sets of the form $\prod_{i=1}^d [a_i - 1/2, b_i + 1/2]$, with all a_i, b_i integer-valued, having union $\mathbb{R}^d$. Let H_n be the number of vacant clusters in R_n of the Boolean model Ξ. Then $\lim_{n\to\infty} |R_n|^{-1}\text{Var}(H_n) = \sigma^2$ and*

$$|R_n|^{-1/2}(H_n - \mathbb{E}\, H_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where σ^2 is a strictly positive constant that does not depend on the choice of sequence (R_n).

The proof is much the same as for the previous theorem, so we just give a sketch. This time we use Lemma 5.2 instead of Lemma 5.1, and we use the fact that the Poisson distribution has finite moment generating function. For the proof that $\sigma > 0$, choose $\delta > 0$ such that the distribution μ_S, restricted to sticks of length greater than 3δ, has a nondegenerate distribution. Take N_n to be the number of small cubes for which the configuration outside the small cube induces a circuit whose boundary lies entirely within a distance δ of the boundary of the small cube. Then there is a positive chance that the sticks centred inside the small cube will cut the interior of this circuit into several pieces, providing the required lower bound on the conditional variance. ■

6 Concluding remarks

We have seen here, and see further in [27], that Theorem 2.1 has a variety of applications, albeit tending to involve a certain amount of effort in checking the stabilization and bounded moment conditions.

Various extensions to Theorem 2.1 would be of interest. One would involve relaxing the condition that the stationary functional $H(X;B)$ be determined entirely by the restriction of X to B, requiring instead some kind of bound to hold on the effect of $(X_x)_{x \notin B}$ on $H(X;B)$. There may be some scope for adapting the proof of Theorem 2.1 to this setting. Such an extension would be useful, for example, for studying the intersection of the infinite cluster with B for percolation, or for studying Boolean models where there is no uniform bound on the shape radius.

It would also be of interest to relax the condition that the underlying variables X_x be independent. An extension to cases where the process $(X_x)_{x \in \mathbb{Z}^d}$ is stationary and has a finite range dependence, or has a spatial Markov property, might be useful. Such extensions could be an interesting challenge; one approach might be to look for coupled process X, X' conditioned to take different values at 0 but nevertheless differing from each other only on a finite collection of sites. This is easy for our i.i.d. case; we just replace the value of X at 0.

Since the first version of this paper, Sheridan [31] has used related methods to obtain CLTs for percolation and random-cluster models on more general graphs than $\mathbb{Z}^d$.

Acknowledgements. This research began during a visit in April–June 1999 to the Fields Institute in Toronto, whose hospitality and support is much appreciated. Additionally, I thank the following people: Joe Yukich for encouraging a "generalist" approach to CLTs; Iain MacPhee for suggesting the idea for Lemma 5.2; Frank Ball for some references on epidemics, and Yu Zhang for providing a preprint of [34]. Last but not least, I thank an anonymous referee for pointing out an error in a previous version of the paper.

References

[1] Andersson, H. and Djehiche, B. (1997) Limit theorems for the total size of a spatial epidemic. *J. Appl. Prob.* **34**, 698-710.

[2] Avram, F. and Bertsimas, D. (1993) On central limit theorems in geometrical probability. *Ann. Appl. Prob.* **3**, 1033-1046.

[3] Borgs, C., Chayes, J.T., Kesten, H., and Spencer, J. (1999) The birth of the infinite cluster: finite-size scaling in percolation. Preprint.

[4] Cox, J.T. and Durrett, R. (1988) Limit theorems for the spread of epidemics and forest fires. *Stoch. Proc. Appl.* **30**, 171-191.

[5] Cressie, N.A.C. (1991) *Statistics for Spatial Data.* Wiley, New York.

[6] Doob, J.L. (1953) *Stochastic Proceses.* John Wiley, New York.

[7] Durrett, R. (1988) *Lecture Notes on Particle Systems and Percolation.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

[8] Durrett, R. (1991) *Probability: Theory and Examples* Wadsworth & Brooks/Cole, Pacific Grove, CA.

[9] Grimmett, G. (1999) *Percolation (Second Edition).* Springer, Berlin

[10] Grimmett, G. and Marstrand, J.M. (1990) The supercritical phase of percolation is well behaved. *Proc. Royal Soc. London A* **430**, 439-457.

[11] Hall, P. (1985) Distribution of size, structure and number of vacant regions in a high-intensity mosaic. *Z. Warsch. verw. Gebiete* **70**, 237-261.

[12] Hall, P. (1986) Clump counts in a mosaic. *Ann. Probab.* **13**, 1250-1266.

[13] Hall, P. (1988) *Introduction to Coverage Processes.* John Wiley, New York.

[14] Heinrich, L., and Molchanov, I.S. (1999) Central Limit Theorem for a class of random measures associated with germ-grain models. *Adv. Appl. Probab.* **31**, 283-314.

[15] Janson, S., Knuth, D.E., Łuczak, T. and Pittel, B. (1993) The Birth of the Giant Component. *Rand. Struct. and Alg.* **4**, 233-358.

[16] Kesten, H. and Lee, S. (1996) The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6**, 495-527.

[17] Kesten, H. and Zhang, Y. (1997) A central limit theorem for critical first passage percolation in two dimensions. *Probab. Th. Rel. Fields* **107**, 137-160.

[18] Lee, S. (1997) The central limit theorem for Euclidean minimal spanning trees I. *Ann. Appl. Probab.* **7**, 996-1020

[19] Lévy, P. (1937) *Théorie de l'Addition des Variables Aléatoires.* Gauthier-Villars, Paris.

[20] Martin-Löf, A. (1986) Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.* **23**, 265-282.

[21] Matheron, G. (1975) *Random Sets and Integral Geometry.* Wiley, New York.

[22] McLeish, D.L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 620-628.

[23] Meester, R. and Roy, R. (1995) *Continuum Percolation.* Cambridge University Press.

[24] Meester, R. and Steif, J. (1998) Consistent estimation of percolation quantities. *Statistica Neerlandica* **52**, 226-238.

[25] Molchanov, I. (1997) *Statistics of the Boolean Model for Practicioners and Mathematicians.* John Wiley, Chichester.

[26] Molchanov, I. (1998) A limit theorem for scaled vacancies of the Boolean model. *Stochastics and Stochastics Reports* **58**, 45-65.

[27] Penrose, M.D. and Yukich, J.E. (200?) Central limit theorems for some graphs in computational geometry. *Ann. Appl Probab.*, to appear.

[28] Penrose, M.D. and Pisztora, A. (1996) Large deviations for discrete and continuous percolation. *Adv. Appl. Probab.* **28**, 29-52.

[29] Pisztora, A. (1996) Surface order large deviations for Ising, Potts and Percolation models. *Probab. Th. Rel. Fields* **104**, 427-466.

[30] Roy, R. (1991) Percolation of Poisson sticks on the plane. *Probab. Th. Rel. fields* **89**, 503-517.

[31] Sheridan, D. (2000) Central limit theorems for graphical models. Preprint, Statistical Laboratory, University of Cambridge.

[32] Stoyan, D., Kendall, W.S., and Mecke, J. (1995) *Stochastic Geometry and its Applications (Second Edition).* Wiley, Chichester.

[33] Williams, D. (1991) *Probability with Martingales.* Cambridge University Press, Cambridge.

[34] Zhang, Y. (2001) A martingale approach in the study of percolation clusters on the Z^d lattice. *J. Theoret. Probab.* **14**, 165-187.

Department of Mathematical Sciences
University of Durham
South Road
Durham DH1 3LE
United Kingdom
Mathew.Penrose@durham.ac.uk