

Line percolation

Paul Balister, Béla Bollobás, Jonathan D. Lee, and Bhargav P. Narayanan

ABSTRACT. We study a geometric bootstrap percolation model, *line percolation*, on the d -dimensional grid $[n]^d$. In line percolation with infection parameter r , infection spreads from a subset $A \subset [n]^d$ of initially infected lattice points as follows: if there is an axis parallel line \mathcal{L} with r or more infected lattice points on it, then every lattice point of $[n]^d$ on \mathcal{L} gets infected and we repeat this until the infection can no longer spread. The elements of the set A are usually chosen independently, with some density p , and the main question is to determine $p_c(n, r, d)$, the density at which percolation (infection of the entire grid) becomes likely. In this paper, we determine $p_c(n, r, 2)$ up to a factor of $1 + o(1)$ and $p_c(n, r, 3)$ up to multiplicative constants as $n \rightarrow \infty$ for every fixed $r \in \mathbb{N}$. We also determine the size of the minimal percolating sets in all dimensions and for all values of the infection parameter.

1. INTRODUCTION

Bootstrap percolation models and arguments have been used to study a range of phenomena in various areas, ranging from crack formation, clustering phenomena, the dynamics of glasses and sandpiles to neural nets and economics; see [19, 4, 12] for a small sample of such applications. In this paper, we shall study a new geometric bootstrap percolation model defined on the d -dimensional grid $[n]^d$ with infection parameter $r \in \mathbb{N}$ which we call *r -neighbour line percolation*. Given $v \in [n]^d$, write $L(v)$ for the set of d axis parallel lines through v and let

$$L([n]^d) = \bigcup_{v \in [n]^d} L(v)$$

be the set of all axis parallel lines that pass through the lattice points of $[n]^d$. In line percolation, infection spreads from a subset $A \subset [n]^d$ of initially infected lattice points as follows: if there is a line $\mathcal{L} \in L([n]^d)$ with r or more infected lattice points on it, then every lattice point of $[n]^d$ on \mathcal{L} gets infected. In other words, we have a sequence $A = A^{(0)} \subset A^{(1)} \subset \dots A^{(t)} \subset \dots$ of subsets of $[n]^d$ such that

$$A^{(t+1)} = A^{(t)} \cup \{v \in [n]^d : \exists \mathcal{L} \in L(v) \text{ such that } |\mathcal{L} \cap A^{(t)}| \geq r\}.$$

Date: 12 December 2013.

2010 *Mathematics Subject Classification.* Primary 60K35; Secondary 60C05.

The *closure* of A is the set $[A] = \bigcup_{t \geq 0} A^{(t)}$ of eventually infected points. We say that the process *terminates* when no more newly infected points are added, i.e., when $A^{(t)} = [A]$. If all the points of $[n]^d$ are infected when the process terminates, i.e., if $[A] = [n]^d$, then we say that A *percolates*.

The classical model of *r-neighbour bootstrap percolation on a graph* was introduced by Chalupa, Leath and Reich [10] in the context of disordered magnetic systems and has since been extensively studied not only by mathematicians but by physicists and sociologists as well; for a small sample of papers, see, for instance, [1, 13, 14, 20]. In this model, a vertex of the graph gets infected if it has at least r previously infected neighbours in the graph. The model is usually studied in the random setting, where the main question is to determine the critical threshold at which percolation occurs. If the elements of the initially infected set are chosen independently at random, each with probability p , then one aims to determine the value p_c at which percolation becomes likely. In this regard, the *r-neighbour bootstrap percolation model* on $[n]^d$, with edges induced by the integer lattice \mathbb{Z}^d , has been the subject of large body of work; see [17, 7, 6], and the references therein.

On account of its inherent geometric structure, it is possible to construct other interesting bootstrap percolation models on the d -dimensional grid. In the past, this has involved endowing the grid with a graph structure other than the one induced by the integer lattice (which, in other words, is a Cartesian product of paths).

In this direction, Holroyd, Liggett and Romik [18] considered *r-neighbour bootstrap percolation* on $[n]^2$ where the neighbourhood of a lattice point v is taken to be a ‘cross’ centred at v , consisting of $r - 1$ points in each of the four axis directions. Sharp thresholds for a model with an anisotropic variant of these cross neighbourhoods were obtained recently by Duminil-Copin and van Enter [11]. Gravner, Hoffman, Pfeiffer and Sivakoff [15] studied the *r-neighbour bootstrap percolation model* on $[n]^d$ with the edges induced by the Hamming torus where $u, v \in [n]^d$ are adjacent if and only if $u - v$ has exactly one nonzero coordinate; the Hamming torus, in other words, is the Cartesian product of complete graphs, which is perhaps the second most natural graph structure on $[n]^d$ after the grid. They obtained bounds on the critical exponents (i.e., $\log_n(p_c)$) which are tight in the case $d = 2$ and for small values of the infection parameter when $d = 3$.

The line percolation model we consider is a natural variant of the bootstrap percolation model on the Hamming torus studied by Gravner, Hoffman, Pfeiffer and Sivakoff. However, we should note that while all the other models mentioned above are *r-neighbour bootstrap percolation models on some underlying graph*, the line percolation model *is not*.

Morally, line percolation is better thought of as coming from the very general *neighbourhood family percolation* model introduced by Bollobás, Smith and Uzzell [9].

In the neighbourhood family percolation model, one starts by specifying a homogeneous, finite collection of subsets of the grid for each point of the grid; a point of the grid becomes infected if all the points of some set in the collection associated with the point are previously infected. In their paper, Bollobás, Smith and Uzzell prove a classification theorem for two-dimensional neighbourhood family models and show that every such model is of one of three types: *supercritical*, *critical* or *subcritical*. In particular, they show that a model is supercritical if and only if there exist finite sets from which the infection can spread to the whole lattice. While line percolation cannot be described by associating a *finite* family of neighbourhoods with each point of the lattice, there do exist, as we shall see, finite sets from which the infection can spread to the whole lattice in the line percolation model and our results about the critical probabilities of the line percolation model are in agreement with the general bounds for the critical probabilities of supercritical models proved in [9]. For some related work concerning subcritical models, see the paper of Balister, Bollobás, Przykucki and Smith [5].

2. OUR RESULTS

In this note, our main aim is to investigate what happens in the line percolation model when the initial set $A = A_p \subset [n]^d$ of infected points is determined by randomly selecting points from $[n]^d$, each independently with probability p . It would be natural to determine the values of p for which percolation is likely to occur. Let $\vartheta_p(n, r, d)$ denote the probability that such a randomly chosen initial set A_p percolates. We define the *critical probability* $p_c(n, r, d)$ by setting

$$p_c(n, r, d) = \inf\{p : \vartheta_p(n, r, d) \geq 1/2\}.$$

The primary question of interest is to determine the asymptotic behaviour of $p_c(n, r, d)$ for every $d, r \in \mathbb{N}$ as $n \rightarrow \infty$. Note that when infection parameter equals one, a set A of initially infected lattice points percolates if and only if $|A| > 0$ which implies that $p_c(n, 1, d) = \Theta(n^{-d})$; we restrict our attention to $r \geq 2$.

Before we state our results, a few remarks about asymptotic notation are in order. We shall make use of standard asymptotic notation; the variable tending to infinity will always be n unless we explicitly specify otherwise. When convenient, we shall also make use of some notation (of Vinogradov) that might be considered non-standard: given functions $f(n)$ and $g(n)$, we write $f \ll g$ if $f = O(g)$, $f \gg g$ if $g = O(f)$, and $f \sim g$ if $f = (1 + o(1))g$. Constants suppressed by the asymptotic notation are allowed to depend on the fixed infection parameter r , but of course, not on n or p .

In two dimensions, we are able to estimate the probability of percolation $\vartheta_p(n, r, 2)$ up to constant factors for all $0 \leq p \leq 1$. We also determine $p_c(n, r, 2)$ up to a factor of $1 + o(1)$ as $n \rightarrow \infty$.

Theorem 2.1. Fix $r \in \mathbb{N}$, with $r \geq 2$. Suppose that for some $s \in \mathbb{N}$, $0 \leq s \leq r - q$, we have $n^{-1-\frac{1}{r-s-1}} \ll p = o(n^{-1-\frac{1}{r-s}})$. Then as $n \rightarrow \infty$,

$$\vartheta_p(n, r, 2) = \Theta\left(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\right) \quad (1)$$

Alternatively, suppose that $p \gg n^{-1-\frac{1}{r}}$. Then $\vartheta_p(n, r, 2) = \Theta(1)$, and furthermore:

$$p_c(n, r, 2) \sim \lambda n^{-1-\frac{1}{r}}, \quad \lambda = \left(\frac{r! \log 2}{2}\right)^{1/r}$$

The techniques used to obtain the above formula for $\vartheta_p(n, r, 2)$ allow us to prove the following result about the critical probability in three dimensions, which is the main result of this paper.

Theorem 2.2. Fix $r \in \mathbb{N}$, with $r \geq 2$, and let

$$s = \left\lfloor \sqrt{r+1/4} - 1/2 \right\rfloor, \quad \gamma = \frac{r+s^2+s}{2(s+1)}.$$

Then as $n \rightarrow \infty$,

$$p_c(n, r, 3) = \Theta\left(n^{-1-\frac{1}{r-\gamma}}\right)$$

The nature of the threshold at the critical probability is also worth investigating. We say that the model exhibits a *sharp threshold* at $p_c = p_c(n, r, d)$ if for any fixed $\varepsilon > 0$, $\vartheta_{(1+\varepsilon)p_c}(n, r, d) = 1 - o(1)$ and $\vartheta_{(1-\varepsilon)p_c}(n, r, d) = o(1)$. It is not difficult to see from our proofs of Theorems 2.1 and 2.2 that in stark contrast to the classical r -neighbour bootstrap percolation model on the grid, there is *no sharp threshold* at p_c when $d = 2, 3$. We expect similar behaviour in higher dimensions but we do not have a proof of such an assertion.

It is also an interesting question to determine the size of a minimal percolating set for r -neighbour line percolation on $[n]^d$ for any $d, r \in \mathbb{N}$ and $n \geq r$. It is easy to check that the set $[r]^d$ percolates (see Figure 1). We shall demonstrate that this is in fact optimal.

Theorem 2.3. Let $d, r, n \in \mathbb{N}$, with $n \geq r$. Then the minimum size of a percolating set in the r -neighbour line percolation process on $[n]^d$ is r^d .

Establishing this fact turns out to be harder than it appears at first glance. The result is trivial when $d = 1$. When $d = 2$, it is not hard to demonstrate that any percolating set has size at least r^2 . To do this, we consider a generalised two-dimensional line percolation model on $[n]^2$ where the infection thresholds for horizontal and vertical lines are r_h and r_v respectively; indeed, we recover the r -neighbour line percolation model when $r_h = r_v = r$. Let $M(r_h, r_v)$ denote the size of a minimal percolating set in this generalised model.

Proposition 2.4. $M(r_h, r_v) \geq r_h r_v$

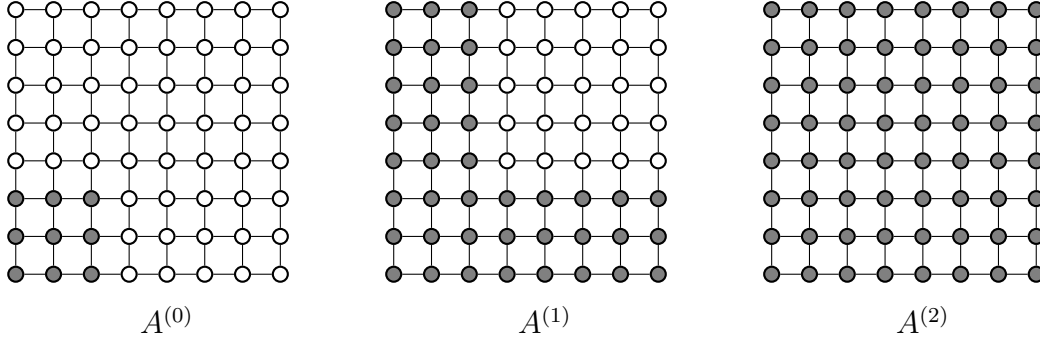


FIGURE 1. The spread of infection from $A = [3]^2$ in the 3-neighbour line percolation process on $[8]^2$.

Proof. Consider the first line \mathcal{L} to be infected: if \mathcal{L} is horizontal, then \mathcal{L} must contain r_h initially infected points and furthermore, if the set of initially infected points is a percolating set, then the set of initially infected points not on \mathcal{L} must constitute a percolating set for the generalised process with infection parameters r_h and $r_v - 1$. An analogous statement holds if \mathcal{L} is vertical. It follows that

$$M(r_h, r_v) \geq \min(r_v + M(r_h - 1, r_v), r_h + M(r_h, r_v - 1)).$$

We obtain by induction that $M(r_h, r_v) \geq r_h r_v$ which implies in particular that $M(r, r) \geq r^2$. \square

Note that the argument depends crucially on the fact that a line has codimension one in a two-dimensional space. Crucially, the incidence geometry of a collection of lines in the plane is essentially straightforward; this is no longer the case in higher dimensions and we need more delicate arguments to prove Theorem 2.3.s

The rest of this paper is organised as follows. We collect together some useful facts about binomial random variables in Section 3. We consider line percolation in two dimensions in Section 4, and prove Theorem 2.1. In Section 5, we turn to line percolation in three dimensions and prove Theorem 2.2, thus obtaining an estimate for the critical probability which is tight up to multiplicative constants. In Section 6, we determine the size of minimal percolating sets for r -neighbour line percolation on $[n]^d$ and prove Theorem 2.3. We conclude the paper in Section 7 with some discussion.

3. BINOMIAL RANDOM VARIABLES

We will need some standard facts about binomial random variables. We collect these here for the sake of convenience. As is usual, for a random variable with distribution $\text{Bin}(N, p)$, we write $\mu = Np$ for its mean.

The first proposition we shall require is an easy consequence of the fact that $e^{-2x} \leq 1 - x \leq e^{-x}$ for all $0 \leq x \leq 1/2$.

Proposition 3.1. *Let X be a random variable with distribution $\text{Bin}(N, p)$, with $p \leq 1/2$. Then for any $k \geq 1$,*

$$\exp(-2\mu) \left(\frac{\mu}{k}\right)^k \leq \mathbb{P}(X = k) \leq \exp(-\mu) \left(\frac{2e\mu}{k}\right)^k.$$

Also, $\exp(-2\mu) \leq \mathbb{P}(X = 0) \leq \exp(-\mu)$. □

We shall make use of the following standard concentration result; see [3] for example.

Proposition 3.2. *Let X be a random variable with distribution $\text{Bin}(N, p)$. Then for any $0 < \delta < 1$,*

$$\mathbb{P}(|X - \mu| > \delta\mu) \leq \exp\left(\frac{-\delta^2\mu}{3}\right). \quad \square$$

Finally, we shall make use of the following easy proposition.

Proposition 3.3. *Let X be a random variable with distribution $\text{Bin}(N, p)$ and suppose $\mu \ll 1$ as $N \rightarrow \infty$. Then for any $k \geq 0$,*

$$\mathbb{P}(X \geq k) = \Theta(\mathbb{P}(X = k)). \quad \square$$

4. LINE PERCOLATION IN TWO DIMENSIONS

The proof of the following proposition is essentially identical to the proof of Theorem 2.1 in [15]; we reproduce it here for completeness.

Proposition 4.1. *Fix $r \in \mathbb{N}$, with $r \geq 2$, and let $C > 0$ be a positive constant. If $p = Cn^{-1-1/r}$, then*

$$\vartheta_p(n, r, 2) \sim 1 - \exp(-2C^r/r!).$$

Proof. The probability that a given line has $r + 1$ or more initially infected points on it is bounded above by $\binom{n}{r+1} p^{r+1}$. Hence the probability that *any* line has $r + 1$ or more initially infected points on it is bounded above by

$$2n \binom{n}{r+1} p^{r+1} = O(n^{r+2} p^{r+1}) = O(n^{-1/r}).$$

Consequently, with high probability, no line has $r + 1$ or more initially infected points on it.

Let E_h denote the event that some horizontal line contains r initially infected points and define E_v analogously. Clearly, the process terminates on the first step if neither E_h nor E_v hold; so $\vartheta_p \leq \mathbb{P}(E_h \cup E_v)$.

Further, the probability that a line has $r - 1$ initially infected points is $\Theta((np)^{r-1}) = \Theta(n^{-1+1/r})$. Hence the number of (say) vertical lines with $r - 1$ initially infected points follows a binomial random variable with mean $\mu = \Omega(n^{1/r})$. Since $\mu \rightarrow \infty$ as $n \rightarrow \infty$, by Proposition 3.2, the probability that there exist at least $r + 1$ such lines is $1 - o(1)$. In turn, the probability that all the $r^2 - 1$ points involved are on different horizontal lines is at least $1 - r^2 n^{-1} = 1 - o(1)$. Hence with high probability, every horizontal line has at least r vertical lines with $r - 1$ initially infected points not on the given horizontal line. Similarly, with high probability every vertical line has at least r horizontal lines with $r - 1$ initially infected points not on the given vertical line. Let us denote these events by P_v and P_h .

Note that P_v, P_h, E_h and E_v are all increasing events. Hence, from the Fortuin-Kasteleyn-Ginibre inequality, conditional on the event $E_h \cup E_v$, with high probability P_v and P_h hold. Hence with probability $1 - o(1)$, any line which becomes infected has r perpendicular lines with $r - 1$ initially infected points away from the newly infected line. Hence, we deduce that $\vartheta_p \sim \mathbb{P}(E_h \cup E_v)$.

The number of horizontal lines with r initially infected points is binomially distributed and it is easily seen to converge in distribution to a Poisson random variable with mean $C^r/r!$. Thus $\mathbb{P}(E_h) \sim 1 - \exp(-C^r/r!)$; similarly, $\mathbb{P}(E_v) \sim 1 - \exp(-C^r/r!)$.

We now estimate $\mathbb{P}(E_h \cap E_v)$. Let $E_h \circ E_v$ denote the event that E_h and E_v occur disjointly. Now, E_h and E_v are increasing events, and so it follows from the Fortuin-Kasteleyn-Ginibre and the van den Berg-Kesten-Reimer inequalities that $\mathbb{P}(E_h \cap E_v) \geq \mathbb{P}(E_h)\mathbb{P}(E_v) \geq \mathbb{P}(E_h \circ E_v)$. Observe that $(E_h \cap E_v) \setminus (E_h \circ E_v)$ happens only if some lattice point v is initially infected and each of the two axis parallel lines through v contain $r - 1$ initially infected points. It follows that

$$\mathbb{P}((E_h \cap E_v) \setminus (E_h \circ E_v)) = O(n^2 p (np)^{2r-2}) = O(n^{-1+1/r})$$

and so $\mathbb{P}((E_h \cap E_v) \setminus (E_h \circ E_v)) = o(1)$. Consequently, we see that $\mathbb{P}(E_h \cap E_v) \sim \mathbb{P}(E_h)\mathbb{P}(E_v)$. Thus, $\mathbb{P}(E_h \cup E_v) \sim \mathbb{P}(E_h) + \mathbb{P}(E_v) - \mathbb{P}(E_h)\mathbb{P}(E_v)$ and the result follows. \square

We shall now prove Theorem 2.1.

Proof of Theorem 2.1. It follows from Proposition 4.1 that $p_c(n, r, 2) \sim \lambda n^{-1-1/r}$ where λ satisfies $\exp(-2\lambda^r/r!) = 1/2$, in particular that $\lambda = (\frac{1}{2}r! \log 2)^{1/r}$.

Furthermore, when $p \gg n^{-1-1/r}$, the probability that there exist r horizontal lines each containing r initially infected points is easily seen to be $\Theta(1)$. So $\vartheta_p(n, r, 2) = \Theta(1)$ when $p \gg n^{-1-1/r}$.

We now turn to estimating $\vartheta_p(n, r, 2)$ for smaller values of p . To do so, it will be convenient to work with a modified two-dimensional line percolation process

$$A_p = G^{(0)} \subset G^{(1)} \subset \dots G^{(t)} \subset \dots$$

where $G^{(2t+1)}$ is obtained from $G^{(2t)}$ by spreading the infection (only) along horizontal lines and $G^{(2t+2)}$ is obtained from $G^{(2t+1)}$ by spreading the infection along vertical lines. Since $G^{(t)} \subset A^{(t)}$ and $A^{(t)} \subset G^{(2t)}$, percolation occurs in the original process if and only if it occurs in the modified process.

Note that A_p percolates if and only if some $G^{(t)}$ contains r or more parallel lines; indeed, in this case $G^{(t+2)} = [n]^2$. We stop the process as soon it produces r or more parallel fully infected lines (or reaches termination). Note that if percolation occurs, then it does so in at most $2r$ steps in the original process, and consequently, in at most $4r$ steps in the modified process.

Let h_t and v_t be the number of horizontal and vertical lines infected when going from $G^{(2t)}$ and $G^{(2t+1)}$ and from $G^{(2t+1)}$ to $G^{(2t+2)}$ respectively; the pair $(\mathbf{h} = (h_t)_{t \geq 0}, \mathbf{v} = (v_t)_{t \geq 0})$ is called the *line-count* of the modified percolation process. Note that the line count of the process accounts for every line that is fully infected in the process except for those lines that are fully infected to begin with. Observe, however, that the probability of finding a fully infected line initially is at most $2np^n$; this probability is exponentially small in n , whilst the estimate of the percolation probability we are attempting to establish is merely polynomial in n . As a corollary, we can absorb this probability into the multiplicative constant and assume that no line is initially fully infected.

Given two sequences $\mathbf{h} = (h_t)_{t=0}^k$ and $\mathbf{v} = (v_t)_{t=0}^k$, we say that (\mathbf{h}, \mathbf{v}) is a *vertical line-count* if (\mathbf{h}, \mathbf{v}) is the line-count of a process which generates r fully infected vertical lines before it generates r fully infected horizontal lines, and does so in exactly $2k + 2$ steps; in other words (assuming no line is fully infected initially), if

- (1) $\sum_{t < k} v_t < r$,
- (2) $\sum_{t \leq k} h_t < r$, and
- (3) $\sum_{t \leq k} v_t \geq r$.

The definition of a *horizontal line-count* $(\mathbf{h} = (h_t)_{t=0}^{k+1}, \mathbf{v} = (v_t)_{t=0}^k)$ is analogous. Given a vertical line-count $(\mathbf{h} = (h_t)_{t=0}^k, \mathbf{v} = (v_t)_{t=0}^k)$, let us define its (*vertical*) *preface* to be the pair $(\mathbf{h}, \mathbf{v}')$ where $\mathbf{v}' = (v_t)_{t=0}^{k-1}$. Similarly, the (*horizontal*) *preface* of a horizontal line-count $(\mathbf{h} = (h_t)_{t=0}^{k+1}, \mathbf{v} = (v_t)_{t=0}^k)$ is the pair $(\mathbf{h}', \mathbf{v})$ where $\mathbf{h}' = (h_t)_{t=0}^k$.

Given a vertical preface $(\mathbf{h}, \mathbf{v}')$, let $E_v(\mathbf{h}, \mathbf{v}')$ be the event that the process generates r fully infected vertical lines before it generates r fully infected horizontal lines and furthermore, the (vertical) line-count of the process has preface $(\mathbf{h}, \mathbf{v}')$. For a horizontal preface $(\mathbf{h}', \mathbf{v})$, define $E_h(\mathbf{h}', \mathbf{v})$ analogously. We then note that

$$\vartheta_p(n, r, 2) = \sum_{(\mathbf{h}, \mathbf{v}')} \mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) + \sum_{(\mathbf{h}', \mathbf{v})} \mathbb{P}(E_h(\mathbf{h}', \mathbf{v}))$$

where the two sums are over all valid vertical and horizontal prefates respectively.

To specify a valid preface, we need to specify at most $2r$ distinct positive integers, each of which is at most r . So the number of valid prefices is at most r^{2r} ; consequently, to estimate the probability of percolation up to constant factors, it suffices to estimate the largest of the probabilities $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$ and $\mathbb{P}(E_h(\mathbf{h}', \mathbf{v}))$.

Now let us fix $s \in \{0, 1, \dots, r-1\}$ to be the least natural number for which $n(np)^{r-(s+1)} \gg 1$; since $p = o(n^{-1})$, we also have $n(np)^{r-s} = o(1)$. Having fixed s , we say that a vertical preface $(\mathbf{h} = (h_t)_{t=0}^k, \mathbf{v}' = (v_t)_{t=0}^{k-1})$ is *slow* if

- (1) $\sum_{t < k} v_t \leq s$ and
- (2) $\sum_{t < k} h_t \leq s$.

Similarly, let us say that a horizontal preface $(\mathbf{h}' = (h_t)_{t=0}^k, \mathbf{v} = (v_t)_{t=0}^k)$ is *slow* if $\sum_{t < k} v_t \leq s$ and $\sum_{t \leq k} h_t \leq s$.

The notion of a slow preface is motivated by observing that once we have generated $s+1$ parallel fully infected lines, then the most efficient way to percolate is to generate r parallel fully infected lines perpendicular to these $s+1$ lines in the next step. Suppose that at some stage in the process, we have l parallel fully infected lines; what is the probability that we generate l' parallel fully infected lines perpendicular to these l lines in the next step? The answer changes depending on whether $n(np)^{r-l} \gg 1$ or $n(np)^{r-l} \ll 1$. Intuitively, one expects that since the lines we have used are exceptional, the unused lines should still look “essentially” as if each site was independently infected with probability p .

Under this heuristic, if $n(np)^{r-l} \ll 1$, then the probability that the process generates exactly l' fully infected lines perpendicular to these l lines in the next step is seen, using Propositions 3.1, to be

$$\Theta\left(\binom{n}{l'} \left((np)^{r-l}\right)^{l'} \left(1 - (np)^{r-l}\right)^{n-l'}\right) = \Theta\left(\left(n(np)^{r-l}\right)^{l'}\right).$$

Similarly if $n(np)^{r-l} \gg 1$, then conditional on having generated l lines already, Proposition 3.2 would seem to imply that with probability $\Theta(1)$, there exist r lines perpendicular to these l lines, each containing $r-l$ initially infected points none of which lie on lines infected earlier (of which there are at most $2r$). These r perpendicular lines become infected in the next step; consequently, we would have percolation with probability $\Theta(1)$.

Formally, we have the following lemma which we will prove later:

Lemma 4.2. *Fix $2 \leq t \leq 2r$ and $1 \leq h, v, v' \leq r-1$ such that $n(np)^{r-v} = o(1)$ and $n(np)^{r-(h-1)} = o(1)$. Let $\mathcal{E}(h, v)$ be the event that the number of fully infected horizontal and vertical lines after the first $2t-1$ steps are h and v respectively (and that the process has not terminated within the first $2t-1$ steps). Also, let $\mathcal{F}(v')$ denote the event that*

the process generates v' vertical lines on step $2t$. Then

$$\mathbb{P}(\mathcal{F}(v') \mid \mathcal{E}(h, v)) = \begin{cases} \Theta\left(\left(n(np)^{r-h}\right)^{v'}\right) & \text{if } n(np)^{r-h} \ll 1, \\ \Theta(1) & \text{if } n(np)^{r-h} \gg 1. \end{cases}$$

By construction a slow preface constructs more than s parallel lines, and so conditional on a slow preface we have percolation with probability $\Theta(1)$. Furthermore, given (say) a vertical preface (\mathbf{x}, \mathbf{y}) , there exists a slow (possibly vertical or horizontal) preface $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$, such that $\mathbb{P}(E_v(\mathbf{x}, \mathbf{y})) = O(\mathbb{P}(E_{v/h}(\hat{\mathbf{x}}, \hat{\mathbf{y}})))$. Thus, to estimate ϑ_p , it suffices to restrict our attention to slow prefaces.

We shall demonstrate how to deal with slow, vertical prefaces; slow, horizontal prefaces can be handled analogously. Given a slow, vertical preface $(\mathbf{h}, \mathbf{v}')$, we would like to estimate $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$. We distinguish two cases depending on the value of $h = \sum_{t \leq k} h_t$.

Case 1: $h \leq s$. It follows from Lemma 4.2 and the fact that $n(np)^{r-h} = o(1)$ that

$$\mathbb{P}\left(v_k \geq r - \sum_{t < k} v_t\right) = \Theta\left(\mathbb{P}\left(v_k = r - \sum_{t < k} v_t\right)\right).$$

So in this case, we may assume that $\sum_{t \leq k} v_t = r$. We see from Lemma 4.2 that $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$, up to constant factors, is given by

$$(n(np)^r)^{h_0} \times (n(np)^{r-h_0})^{v_0} \times (n(np)^{r-v_0})^{h_1} \times \dots \\ \dots \times (n(np)^{r-\sum_{t < k} v_t})^{h_k} \times (n(np)^{r-h})^{v_k}$$

which, on algebraic simplification, is seen to be $\Theta(n^{r+h}(np)^{r^2})$. This is maximised when $h = s$. Thus, in this case, we see that

$$\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) = \Theta\left(n^{r+s}(np)^{r^2}\right).$$

Case 2: $h > s$. If $h > s$ on the other hand, $n(np)^{r-h} \gg 1$ and so the estimate for $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$, up to constant factors, becomes

$$(n(np)^r)^{h_0} \times (n(np)^{r-h_0})^{v_0} \times (n(np)^{r-v_0})^{h_1} \times \dots \times (n(np)^{r-\sum_{t < k} v_t})^{h_k} \times 1$$

which in turn, on simplification, is seen to be $\Theta(n^{r+h}(np)^{r^2}(n(np)^{r-h})^{\sum_{t < k} v_t - r})$. Since $n(np)^{r-h} \gg 1$ and we disregard constant factors, the probability of $E_v(\mathbf{h}, \mathbf{v}')$ is maximised when $\sum_{t < k} v_t$ is maximal. Recall that $\sum_{t < k} v_t \leq s$, and so we may assume that $\sum_{t < k} v_t = s$. It follows that

$$\begin{aligned} \mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) &= \Theta\left(n^{r+h}(np)^{r^2}\left(n(np)^{r-h}\right)^{s-r}\right) \\ &= \Theta\left((n(np)^r)^s (n(np)^{r-s})^h\right) \end{aligned}$$

Since $n(np)^{r-s} = o(1)$ and we seek to find the most likely slow prefaces, we may assume that $h = s + 1$. Hence in this case

$$\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) = \Theta\left(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\right).$$

can be achieved.

We claim that the main contributions to ϑ_p come from Case 2. To show this, first observe that $(n(np)^{r-s}) \ll 1$ and $s + 1 - r \leq 0$, and so $(n(np)^{r-s})^{s+1-r} \gg 1$. Hence we can compare the contribution from Case 2 to that from Case 1:

$$n^{2s+1}(np)^{r(2s+1)-s(s+1)} = n^{r+s}(np)^{r^2} (n(np)^{r-s})^{s+1-r} \gg n^{r+s}(np)^{r^2}$$

and so the contributions from Case 2 dominate. Recall that s was chosen to be least such that $n(np)^{r-(s+1)} \gg 1$, which is plainly seen to be equivalent to:

$$n^{-1-\frac{1}{r-s-1}} \ll p = o(n^{-1-\frac{1}{r-s}})$$

Thus, we conclude that

$$\vartheta_p(n, r, 2) = \Theta\left(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\right) \text{ when}$$

as required. □

We now prove Lemma 4.2, which is an exercise in conditioning once we have conditioned on the right set of events. This is to be expected, as at each alternate stage we are interested in finding successively fewer points on any line, and so the fact that an excess of points was not found earlier does not substantively distort the distribution of the number of points found on any given line. Furthermore, the number of points is small by comparison to n , and so distortions arising from a perpendicular line infecting an already infected point are negligible. Hence the bulk of the proof is in fact merely setting up the correct notation; the actual core calculation is simply observing that there are a fixed set of events, each of probability $1 - o(1)$, whose occurrence guarantees the naïve calculation goes through.

Proof. First, consider any subset \mathcal{H} of h horizontal lines and any subset \mathcal{V} of v vertical lines. Also fix a non-empty subset $\mathcal{H}' \subset \mathcal{H}$ of cardinality $h' > 0$. We start by conditioning on the event that the sets of fully infected horizontal and vertical lines generated by the process after $2t - 1$ steps are precisely the lines in \mathcal{H} and \mathcal{V} respectively. In addition to conditioning on the sets of horizontal and vertical lines generated by the process after $2t - 1$ steps, let us further assume that $\mathcal{H}' \subset \mathcal{H}$ is the set of horizontal lines generated by the process on step $2t - 1$. We also condition on the points in \mathcal{H}' and partition the vertical lines outside \mathcal{V} into sets $\mathcal{V}_0, \mathcal{V}_1, \dots, \mathcal{V}_{h'}$ where we add a vertical line to \mathcal{V}_i if it contains i initially infected points in its intersection with \mathcal{H}' . Note that as $t \geq 2$ and since we have assumed that no line of \mathcal{H}' was infected in the first step of the

process, each line of \mathcal{H}' must necessarily contain strictly fewer than r initially infected points. It follows by double counting that for each $1 \leq i \leq h'$, the set \mathcal{V}_i has size at most $h'(r-1)/i = O(1)$. Let us write $\mathcal{E}^* = \mathcal{E}^*(\mathcal{H}, \mathcal{V}, \mathcal{H}', \mathcal{V}_0, \dots, \mathcal{V}_{h'})$ for the event that all of the above happens, namely that the lines in \mathcal{H} and \mathcal{V} are those that are fully infected in the first $2t-1$ steps, the set of lines infected on step $2t-1$ is \mathcal{H}' , and that the lines outside \mathcal{V} with precisely i initially infected points in \mathcal{H}' are precisely those in \mathcal{V}_i .

Let $\mathcal{F}^* = \mathcal{F}^*(\mathcal{H}, \mathcal{V})$ denote the event that there are v' vertical lines outside \mathcal{V} with $r-h$ initially infected points outside \mathcal{H} . Since $\mathbb{P}(\mathcal{F}(v') | \mathcal{E}^*) = \mathbb{P}(\mathcal{F}^* | \mathcal{E}^*)$, it is clearly enough to show that

$$\mathbb{P}(\mathcal{F}^* | \mathcal{E}^*) = \begin{cases} \Theta\left(\left(n(np)^{r-h}\right)^{v'}\right) & \text{if } n(np)^{r-h} \ll 1, \\ \Theta(1) & \text{if } n(np)^{r-h} \gg 1. \end{cases}$$

It is easy to check using Propositions 3.1 and 3.2 that

$$\mathbb{P}(\mathcal{F}^*) = \begin{cases} \Theta\left(\left(n(np)^{r-h}\right)^{v'}\right) & \text{if } n(np)^{r-h} \ll 1, \\ \Theta(1) & \text{if } n(np)^{r-h} \gg 1. \end{cases}$$

So we need only check that conditioning on \mathcal{E}^* essentially changes nothing. To do this, we write $\mathbb{P}(\mathcal{F}^* | \mathcal{E}^*)$ in terms of events that depend only on points outside both \mathcal{H} and \mathcal{V} . For $0 \leq i \leq h'$, let us denote by \mathcal{B}_i the event that some vertical line in \mathcal{V}_i contains $r-h+(h'-i)$ or more initially infected points outside \mathcal{H} . Also, let us write \mathcal{B} for the event that some horizontal line outside \mathcal{H} contains $r-v$ or more initially infected points outside \mathcal{V} .

The lemma is straightforward to verify once we observe (see Figure 2) that

$$\mathbb{P}(\mathcal{F}^* | \mathcal{E}^*) = \mathbb{P}(\mathcal{F}^* | \mathcal{B}^c \cap \mathcal{B}_0^c \cap \mathcal{B}_1^c \cap \dots \cap \mathcal{B}_{h'}^c).$$

The facts that $p = o(n^{-1})$ and $|\mathcal{V}_i| = O(1)$ for $i \geq 1$ are sufficient to check that $\mathbb{P}(\mathcal{F}^* \cap \mathcal{B}_i) = o(\mathbb{P}(\mathcal{F}^*))$ for each $1 \leq i \leq h'$. Using the fact that $n(np)^{r-(h-1)} = o(1)$, we see that $\mathbb{P}(\mathcal{F}^* \cap \mathcal{B}_0) = o(\mathbb{P}(\mathcal{F}^*))$. Finally, it follows from the fact that $n(np)^{r-v} = o(1)$ that $\mathbb{P}(\mathcal{F}^* \cap \mathcal{B}) = o(\mathbb{P}(\mathcal{F}^*))$. Observe that each of the events $\mathcal{B}, \mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_{h'}$ occurs with probability $o(1)$; it is then clear that

$$\mathbb{P}(\mathcal{F}^* | \mathcal{E}^*) = \mathbb{P}(\mathcal{F}^* | \mathcal{B}^c \cap \mathcal{B}_0^c \cap \mathcal{B}_1^c \cap \dots \cap \mathcal{B}_{h'}^c) = (1 + o(1))\mathbb{P}(\mathcal{F}^*)$$

and the lemma follows. \square

planes with probability at least $1/2$ if C is greater than some sufficiently large constant. So it follows that $p_c(n, r, 3) = O(n^{-1-1/(r-\gamma)})$.

Case 2: $C \ll 1$. We first deal with the case $r = 2$, which is somewhat degenerate as $\gamma = 1$. In particular we note that there are $3n^2$ lines, and the probability that any of them becomes infected at the first step is bounded by $\binom{n}{2}p^2 < \frac{1}{2}C^2n^{-2}$. Hence the probability that any line becomes infected is bounded by $\frac{3}{2}C^2$ which tends to 0 as $C \rightarrow 0$, as required. Hence we may assume that $\gamma > 1$.

We claim that the probability of percolation is at most $1/2$, provided C is less than some sufficiently small constant. We shall demonstrate this by proving something much stronger. We shall track, as the infection spreads, the number of planes with k or more parallel fully infected lines for each $1 \leq k \leq s+1$ and show that, with probability at least $1/2$, these numbers are not too large when the process terminates; in particular, we shall show that there are no planes with $s+1$ or more parallel fully infected lines when the process reaches termination and consequently, that there is no percolation.

We shall work with a modified three-dimensional line percolation process in which the infection spreads one line at a time. Let $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_{3n^2}$ be an ordering of the $3n^2$ lines of the three-dimensional grid. In this modified process, we have a sequence of subsets $A_p = H^{(0)} \subset H^{(1)} \subset \dots \subset H^{(t)} \subset \dots$ of $[n]^3$ such that

$$H^{(t+1)} = \begin{cases} H^{(t)} \cup \mathcal{L}_k & \text{if } |\mathcal{L}_k \cap H^{(t)}| \geq r, \text{ where } k = t+1 \pmod{3n^2}, \\ H^{(t)} & \text{otherwise.} \end{cases}$$

Clearly, $H^{(t)} \subset A^{(t)} \subset H^{(3n^2t)}$ and so A_p percolates in the original process if and only if it percolates in this modified process.

We run the modified three-dimensional process starting from A_p and if we find at time t that

- (A) the number of planes containing k or more parallel fully infected lines will exceed $n^{1-k\gamma/(r-\gamma)}$ for some $1 \leq k \leq s+1$ at time $t+1$, or
- (B) $H^{(t)} = H^{(t+3n^2)}$ (ie. the process has terminated).

then we *stop the modified process at time t* . Let E_A be the event that we stop the modified process on account of condition ((A)). We state the following lemma which we will prove later.

Lemma 5.1. *In the modified process, and working with the definitions of Theorem 2.2*

$$\mathbb{P}(E_A) = O\left(\sum_{1 \leq k \leq s} C^{rk} + C^{r(2s+1)-s(s+1)}\right).$$

We note that the required lower bound on p_c follows immediately. First, observe that $rk > 0$ and $r(2s+1) > s(s+1)$. Then Lemma 5.1 implies that for any fixed $r \geq 2$, we

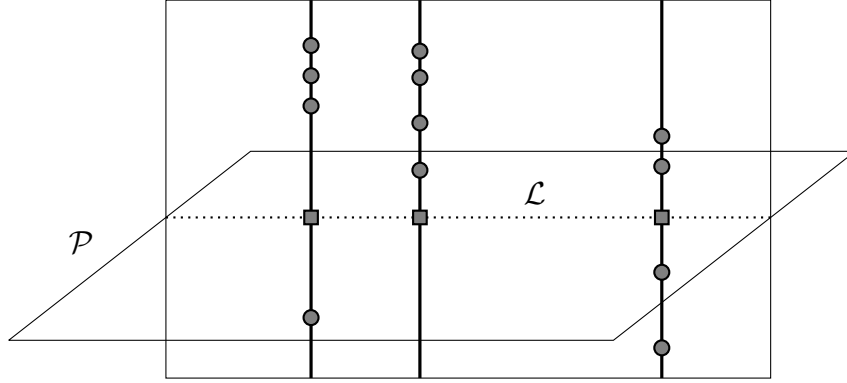


FIGURE 3. Boosted points on \mathcal{L} in \mathcal{P} .

have $\mathbb{P}(E_A) \rightarrow 0$ as $C \rightarrow 0$. Hence, if C is less than a suitably small constant (dependent on r), the probability that the three-dimensional r -neighbour line percolation process with $p = Cn^{-1-1/(r-\gamma)}$ generates a plane with $s + 1$ parallel fully infected lines before reaching termination is less than $1/2$. Consequently, the probability of percolation is also less than $1/2$. This implies that $p_c(n, r, 3) = \Omega(n^{-1-1/(r-\gamma)})$ as required.

Hence we have completed the proof of Theorem 2.2. \square

Proof of Lemma 5.1. Recall that we work with the definitions of the proof of Theorem 2.2. In particular we work with the modified process and seek to bound the number of planes containing many parallel fully infected lines.

For $k \geq 1$, let us write N_k for the number of planes containing k or more parallel fully infected lines when we stop the modified three-dimensional process and define $N = 3n - \sum_{k \geq 1} N_k$. Note that the modified process is defined in such a way that when it is stopped, $N_k \leq n^{1-k\gamma/(r-\gamma)}$ for each $1 \leq k \leq s$.

Observe that $r \in [s(s+1), (s+1)(s+2))$ by construction, and that as a corollary $\gamma \in [s, s+1)$. Hence $(s+1)\gamma/(r-\gamma) > 1$, which implies that $n^{1-(s+1)\gamma/(r-\gamma)} < 1$. As a corollary, $N_k = 0$ for $k \geq s+1$. It trivially follows that $N_0 = (1 - o(1))n$.

We shall prove Lemma 5.1 by estimating the probability that a given plane contains k or more parallel fully infected lines when we stop the process. Let us fix a plane \mathcal{P} . Suppose that a point v of \mathcal{P} gets infected before we stop the process and suppose further that v is not initially infected. Then v is either

- (1) infected when a line perpendicular to \mathcal{P} containing v has r other previously infected points on it (we call such points *boosted points*), or
- (2) infected when a line in \mathcal{P} containing v has r other previously infected points on it.

Let $A_{\mathcal{P}}$ denote the union of the boosted points and the initially infected points of \mathcal{P} . Observe that if we run the two-dimensional r -neighbour line percolation process on \mathcal{P} starting from $A_{\mathcal{P}}$, we infect all the points of \mathcal{P} that were infected in the modified three-dimensional process before it was stopped. Thus, the probability that \mathcal{P} contains k or more parallel fully infected lines when we stop the modified three-dimensional process is bounded above by the probability that we generate k parallel fully infected lines in the two-dimensional r -neighbour line percolation process on \mathcal{P} starting from $A_{\mathcal{P}}$.

Fix any arrangement of the boosted points in \mathcal{P} . Note that if we have k boosted points on a line \mathcal{L} in \mathcal{P} , then this implies that the plane perpendicular to \mathcal{P} which intersects \mathcal{P} in \mathcal{L} generated k parallel fully infected lines in the modified three-dimensional process before it was stopped (see Figure 3); consequently, the number of such lines \mathcal{L} in \mathcal{P} is at most N_k .

For $1 \leq k \leq s+1$, let E_k denote the event that the two-dimensional r -neighbour line percolation process on \mathcal{P} starting from $A_{\mathcal{P}}$ generates k parallel fully infected lines. We state the following lemma which we will prove later.

Lemma 5.2. *Conditional on any arrangement in \mathcal{P} of the boosted points satisfying the condition that at most N_k lines contain k boosted points,*

$$\begin{aligned}\mathbb{P}(E_k) &= O(C^{rk} n^{-k\gamma/(r-\gamma)}) \quad \forall 1 \leq k \leq s, \\ \mathbb{P}(E_{s+1}) &= O(C^{r(2s+1)-s(s+1)} n^{-1}).\end{aligned}$$

Recall that E_A is the event that we stop modified three-dimensional process on account of the number of planes containing k parallel fully infected lines exceeding $n^{1-k\gamma/(r-\gamma)}$ for some $1 \leq k \leq s+1$.

From Lemma 5.2, we see that expected number of planes with k parallel fully infected lines when we stop the modified process in three dimensions is $O(C^{(r-s)k} n^{1-k\gamma/(r-\gamma)})$ when $1 \leq k \leq s$ and $O(C^{r(2s+1)-2s(s+1)})$ when $k = s+1$. By Markov's inequality, the probability that the number of planes containing k parallel fully infected lines exceeds $n^{1-k\gamma/(r-\gamma)}$ is $O(C^{(r-s)k})$ when $1 \leq k \leq s$ and $O(C^{r(2s+1)-2s(s+1)})$ when $k = s+1$ since $\lfloor n^{1-(s+1)\gamma/(r-\gamma)} \rfloor = 0$. Applying the union bound, we get

$$\mathbb{P}(E_A) = O\left(\sum_{1 \leq k \leq s} C^{rk} + C^{r(2s+1)-s(s+1)}\right).$$

as required. \square

We are now in a position to prove Lemma 5.2 in a fashion essentially analogous to that of Theorem 2.1. In particular, we will eventually show that there are not enough lines with boosted points on to substantively alter the percolation, and thus the percolation is dominated (up to constants) by the percolation with no boosted points.

Proof of Lemma 5.2. As in the proof of Theorem 2.1, we consider the modified two-dimensional percolation process $A_{\mathcal{P}} = G^{(0)} \subset G^{(1)} \subset \dots$ on \mathcal{P} where in going from $G^{(2t)}$ to $G^{(2t+1)}$, only horizontal lines are infected, and in going from $G^{(2t+1)}$ to $G^{(2t+2)}$, only vertical lines are infected. Let us stop this modified two-dimensional process on \mathcal{P} as soon as it generates k or more parallel fully infected lines (or reaches termination).

For $0 \leq i \leq s$, let $h_{t,i}$ denote the number of horizontal lines containing i boosted points which are infected when going from $G^{(2t)}$ and $G^{(2t+1)}$ and define $v_{t,i}$ analogously. We say that $(\mathbf{h} = (h_{t,i}), \mathbf{v} = (v_{t,i}))$ is the *full line-count* of the modified two-dimensional process. Given (\mathbf{h}, \mathbf{v}) , let $(\mathbf{h}^*, \mathbf{v}^*)$ be defined by setting $h_{t,0}^* = \sum_{i \geq 0} h_{t,i}$, $v_{t,0}^* = \sum_{i \geq 0} v_{t,i}$, and $h_{t,i}^* = v_{t,i}^* = 0$ for $1 \leq i \leq s$.

Let $E_k(\mathbf{h}, \mathbf{v})$ denote the event that the modified two-dimensional process on \mathcal{P} generates k or more parallel fully infected lines and furthermore, the full line-count of the modified two-dimensional process on \mathcal{P} is given by (\mathbf{h}, \mathbf{v}) .

For any (\mathbf{h}, \mathbf{v}) , we shall show that $\mathbb{P}(E_k(\mathbf{h}, \mathbf{v})) = O(C^{-sk} \mathbb{P}(E_k(\mathbf{h}^*, \mathbf{v}^*)))$; in other words, we show that ‘long range effects’ are essentially negligible and that we may restrict our attention to the case where we never use any of the boosted points.

Having generated l parallel fully infected lines, let us consider the probability that the modified two-dimensional process generates exactly l' new fully infected lines perpendicular to these l lines in the next step. As in the proof of Theorem 2.1, we state a lemma which permits us to treat lines generated in the past as being fixed deterministically.

Lemma 5.3. *Fix $2 \leq t \leq 2r$ and $1 \leq h, v, v'_{i=0} \leq r-1$ such that $n(np)^{r-v} = o(1)$ and $n(np)^{r-(h-1)} = o(1)$. Let $\mathcal{E}(h, v)$ be the event that the number of fully infected horizontal and vertical lines after the first $2t-1$ steps are h and v respectively (and that the process has not terminated within the first $2t-1$ steps). Also, let $\mathcal{F}(v')$ denote the event that the process generates v'_i vertical lines containing i boosted points on step $2t$. Then:*

$$\mathbb{P}(\mathcal{F}(v') \mid \mathcal{E}(h, v)) = \begin{cases} \Theta\left(\left(n(np)^{r-h}\right)^{v'_0}\right) & \text{if } n(np)^{r-h} = o(1), \forall i > 0, v'_i = 0 \\ \Theta(1) & \text{if } n(np)^{r-h} \gg 1, \forall i > 0, v'_i = 0 \\ o\left(\left(n(np)^{r-h}\right)^{\sum_i v'_i}\right) & \text{if } n(np)^{r-h} = o(1), \exists i > 0, v'_i > 0 \end{cases}$$

Hence by induction on the steps of the line-count, we have:

$$\mathbb{P}(E_k(\mathbf{h}, \mathbf{v})) = O(\mathbb{P}(E_k(\mathbf{h}^*, \mathbf{v}^*))).$$

We now restrict our attention to the events $E_k(\mathbf{h}^*, \mathbf{v}^*)$. As in the proof of Theorem 2.1, we may suppose that $\sum_t v_{t,0}^* = k \leq s+1$ and that $\sum_t h_{t,0}^* = l < k$. Recall that s is the greatest natural number such that $s(s+1) \leq r$ and that $p = Cn^{-1-1/(r-\gamma)}$ satisfies $n(np)^{r-s} = o(1)$ and $n(np)^{r-s-1} \gg 1$.

We shall mimic the proof of Theorem 2.1. Since $l \leq s$ and $n(np)^{r-l} \ll 1$, we see that the probability of $E_k(\mathbf{h}^*, \mathbf{v}^*)$, up to constant factors, is given by

$$(n(np)^r)^{h_{0,0}^*} \times (n(np)^{r-h_{0,0}^*})^{v_{0,0}^*} \times (n(np)^{r-v_{0,0}^*})^{h_{1,0}^*} \times \dots \\ \dots \times (n(np)^{r-\sum_{t < m} v_{t,0}^*})^{h_{m,0}^*} \times (n(np)^{r-\sum_{t \leq m} h_{t,0}^*})^{v_{m,0}^*}.$$

We see on algebraic simplification that

$$\mathbb{P}(E_k(\mathbf{h}^*, \mathbf{v}^*)) = \Theta\left(n^{k+l}(np)^{rk+rl-kl}\right) = \Theta\left((n(np)^r)^k \left(n(np)^{r-k}\right)^l\right). \quad (3)$$

Note that $n(np)^r = C^r n^{-\frac{\gamma}{r-\gamma}}$ and recall that $n(np)^{r-k}$ is $\gg 1$ for $k \leq s$ and is $o(1)$ for $k = s+1$.

When $k \leq s$, we see that the estimate for the probability of $E_k(\mathbf{h}^*, \mathbf{v}^*)$ in (3) is maximised by taking $l = 0$, from which we conclude that

$$\mathbb{P}(E_k) = O\left((n(np)^r)^k\right) = O\left(C^{rk} n^{-\frac{k\gamma}{r-\gamma}}\right).$$

On the other hand, when $k = s+1$, the estimate for the probability of $E_k(\mathbf{h}^*, \mathbf{v}^*)$ in (3) is maximised by taking $l = s$, from which we conclude that

$$\mathbb{P}(E_{s+1}) = O\left((n(np)^r)^{s+1} (n(np)^{r-(s+1)})^s\right) \\ = O\left(n^{2s+1} (np)^{r(2s+1)-s(s+1)}\right)$$

Note that $2\gamma(s+1) = r + s(s+1)$ and hence $r(2s+1) - s(s+1) = 2(s+1)(r-\gamma)$. Hence we have:

$$\mathbb{P}(E_{s+1}) = O\left(C^{r(2s+1)-s(s+1)} n^{-1}\right).$$

This completes the proof of Lemma 5.2. \square

It remains to prove Lemma 5.3. In particular, we again find that we need to show that points away from fully infected lines are still essentially independently infected with probability p . As in Lemma 4.2, this proceeds essentially by

Proof of Lemma 5.3. We note that the first two statements follow immediately from Lemma 5.3, as in these cases no vertical lines with boosted points are permitted to be used at this stage, and $N_0 \sim n$.

To prove the final statement, we fix \mathcal{H}, \mathcal{V} and \mathcal{H}' as in the proof of Lemma 5.3. We partition the vertical lines outside \mathcal{V} into sets $\mathcal{V}_{i,j}$ for $i \in \{0, \dots, h'\}$ and $j \in \{0, \dots, s\}$, where a vertical line is in $\mathcal{V}_{i,j}$ if it contains i initially infected points in its intersection with \mathcal{H}' and j boosted points. As before $|V_{i,j}| \leq h'(r-1)/i = O(1)$ for $i \geq 1$. We condition on $\mathcal{E}^* = \mathcal{E}^*(\mathcal{H}, \mathcal{V}, \mathcal{H}', \mathcal{V}_{0,0}, \dots, \mathcal{V}_{h',s})$ for the event that all of the above happens.

Now, we let $\mathcal{F}_i^* = \mathcal{F}^*(\mathcal{H}, \mathcal{V})$ denote the event that there are v'_i vertical lines outside \mathcal{V} with $r - h$ initially infected points outside \mathcal{H} and i boosted points. We have that $\mathbb{P}(\mathcal{F}(v')|\mathcal{E}^*) = \mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*|\mathcal{E}^*)$. Now, it is immediate that:

$$\mathbb{P}(\mathcal{F}_0^*) = \Theta\left(\binom{N_0}{v'_0} \left((np)^{r-h}\right)^{v'_0} \left(1 - (np)^{r-h}\right)^{n-v'_0}\right) = \Theta\left(\left(n(np)^{r-h}\right)^{v'_0}\right)$$

since $N_0 \sim n$. Furthermore, recall that $N_i \leq n^{1-i\gamma/(r-\gamma)}$, $r > \gamma > 1$ and $n^{1+\gamma/(r-\gamma)}p = Cn^{(\gamma-1)/(r-\gamma)}$. Hence we have

$$\begin{aligned} \mathbb{P}(\mathcal{F}_i^*) &= O\left(\binom{N_i}{v'_i} \left((np)^{r-h-i}\right)^{v'_i}\right) = O\left(\left(n(np)^{r-h}\right)^{v'_i} \left(n^{1+\gamma/(r-\gamma)}p\right)^{-iv'_i}\right) \\ &= o\left(\left(n(np)^{r-h}\right)^{v'_i}\right). \end{aligned}$$

Additionally, these events are independent, as they are dependent on distinct sets of vertical lines. So when $n(np)^{r-h} = o(1)$ and some $v'_i > 0$ for $i > 0$, we have:

$$\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*) = o\left(\left(n(np)^{r-h}\right)^{\sum_i v'_i}\right)$$

Now, we denote by $\mathcal{B}_{i,j}$ the event that some vertical line in $\mathcal{V}_{i,j}$ contains $r - h + (h' - i)$ or more initially infected points outside \mathcal{H} . We also write \mathcal{B} for the event that some horizontal line outside \mathcal{H} contains $r - v$ or more initially infected points outside \mathcal{V} . As before, we have that:

$$\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*|\mathcal{E}^*) = \mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*|\mathcal{B}^c \cap \mathcal{B}_{0,0}^c \cap \dots \cap \mathcal{B}_{h',s}^c)$$

Since $p = o(1)$ and $|\mathcal{V}_{i,j}| = O(1)$, we have $\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^* \cap \mathcal{B}_{i,j}) = o(\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*))$ for all $i \geq 1$ and all j , and $\mathbb{P}(\mathcal{B}_{i,j}) = o(1)$. Since $n(np)^{r-(h-1)} = o(1)$, we have that $\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^* \cap \mathcal{B}_{0,j}) = o(\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*))$ and that $\mathbb{P}(\mathcal{B}_{0,j}) = o(1)$ for all j . Finally, since $n(np)^{r-v} = o(1)$, we have $\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^* \cap \mathcal{B}) = o(\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^* \cap \mathcal{B}))$ and $\mathbb{P}(\mathcal{B}) = o(1)$. Hence we deduce that:

$$\mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*|\mathcal{E}^*) = \mathbb{P}(\cap_{i=0}^s \mathcal{F}_i^*)$$

as required. \square

6. MINIMAL PERCOLATING SETS

In this section, we prove Theorem 2.3 which tells us the size of a minimal percolating set. We shall make use of the polynomial method which has had many unexpected applications in combinatorics; see [16] for a survey of many of these surprising applications. While linear algebraic techniques have previously been used to study bootstrap percolation processes (see [8]), we believe that this application of the polynomial method is new to the field.

Proof of Theorem 2.3. Suppose for the sake of contradiction that there is a set $A \subset [n]^d$ which percolates with $|A| < r^d$. We shall derive a contradiction using the polynomial method.

Proposition 6.1. *There exists a non-zero polynomial $P_A \in \mathbb{R}[x_1, x_2, \dots, x_d]$ of degree at most $r - 1$ in each variable which vanishes on A .*

Proof. Let $V \subset \mathbb{R}[x_1, x_2, \dots, x_d]$ be the vector space of real polynomials in d variables of degree at most $r - 1$ in each variable. The dimension of V is clearly r^d . Consider the evaluation map from V to $\mathbb{R}^{|A|}$ which sends a polynomial P to $(P(v))_{v \in A}$. Clearly, this map is linear. Since we assumed that $|A| < r^d$, this map has a non-trivial kernel. The existence of P_A follows. \square

We shall use the polynomial P_A to follow the spread of infection. The following claim will yield a contradiction.

Proposition 6.2. *The polynomial P_A vanishes on $A^{(t)}$ for every $t \geq 0$.*

Proof. We proceed by induction on t . The claim is true when $t = 0$ since $A^{(0)} = A$. Now, assume P_A vanishes on $A^{(t)}$ and consider a line \mathcal{L} which gets infected when going from $A^{(t)}$ to $A^{(t+1)}$. It must be the case that $|\mathcal{L} \cap A^{(t)}| \geq r$. Since P_A vanishes on $A^{(t)}$, the restriction of P_A to \mathcal{L} disappears on $\mathcal{L} \cap A^{(t)}$. If the direction of \mathcal{L} is $i \in [d]$, then the restriction of P_A to \mathcal{L} is a univariate polynomial in the variable x_i of degree at most $r - 1$. Since a non-zero univariate polynomial of degree at most $r - 1$ has at most $r - 1$ roots, the restriction of P_A to \mathcal{L} has to be identically zero. Consequently, P_A vanishes on $A^{(t+1)}$. \square

Since A percolates, we conclude that P_A vanishes on $[n]^d$. On the other hand, using the following proposition, the proof of which may be found in [2], we conclude that P_A cannot vanish on $[r + 1]^d$.

Proposition 6.3. *Let $P = P(x_1, x_2, \dots, x_d)$ be a polynomial in d variables over an arbitrary field \mathcal{F} . Suppose that the degree of P as a polynomial in x_i is at most k_i for $1 \leq i \leq d$, and let $S_i \subset \mathcal{F}$ be a set of at least $k_i + 1$ distinct elements of \mathcal{F} . If $P(u_1, u_2, \dots, u_d) = 0$ for every d -tuple $(u_1, u_2, \dots, u_d) \in S_1 \times S_2 \times \dots \times S_d$, then P is identically zero.* \square

It follows from Proposition 6.3 that P_A is zero and we have a contradiction. This establishes Theorem 2.3. \square

Remark. It follows from Theorem 2.3 that the size of a minimal percolating set in the r -neighbour bootstrap percolation model on $[n]^d$ with edges induced by the Hamming torus is at least $(r/d)^d$. On the other hand, it is possible to construct sets of size about $r^d/2d!$ which percolate. It would be interesting to determine the size of a minimal

percolating set in this model exactly for all $d, r \in \mathbb{N}$; we suspect that the lower bound of $(r/d)^d$ is quite far from the truth.

7. CONCLUDING REMARKS

There remain many challenging and attractive open problems, chief amongst which is the determination of $p_c(n, r, d)$ for all $d, r \in \mathbb{N}$. To determine $p_c(n, r, 3)$, we used a careful estimate for $\vartheta_p(n, r, 2)$ which is valid for all $0 \leq p \leq 1$. This estimate for $\vartheta_p(n, r, 2)$ depends crucially on the fact that the two-dimensional process reaches termination in a constant (depending on r , but not on n) number of steps. We believe that to determine $p_c(n, r, 4)$, one will need to determine $\vartheta_p(n, r, 3)$ for all $0 \leq p \leq 1$ but since it is not at all obvious that the three-dimensional process reaches termination in a constant number of steps with high probability, we suspect different methods will be necessary.

As remarked earlier, it is easily read out of our proofs that the line percolation model does not exhibit a sharp threshold at p_c in two or three dimensions. It would be interesting to prove an analogous statement for every $d, r \in \mathbb{N}$.

ACKNOWLEDGEMENTS

The first and second authors were partially supported by NSF grant DMS-1301614 and the second author also wishes to acknowledge support from EU MULTIPLEX grant 317532.

Some of the research in this paper was carried out while the authors were visitors at Microsoft Research, Redmond. This research was continued while the third and fourth authors were visitors at the University of Memphis. The authors are grateful to Yuval Peres and the other members of the Theory Group at Microsoft Research for their hospitality, and the third and fourth authors are additionally grateful for the hospitality of the University of Memphis.

We would also like to thank Janko Gravner for bringing [15] to our attention and the anonymous referees for their helpful comments.

REFERENCES

1. J. Adler and U. Lev, *Bootstrap Percolation: visualizations and applications*, Braz. J. Phys. **33** (2003), 641–644. [2](#)
2. N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), 7–29. [20](#)
3. N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008. [6](#)

4. H. Amini, *Bootstrap percolation in living neural networks*, J. Stat. Phys. **141** (2010), 459–475. [1](#)
5. P. Balister, B. Bollobás, M. Przykucki, and P. Smith, *Subcritical neighbourhood family percolation models have non-trivial phase transitions*, Trans. Amer. Math. Soc., To appear. [3](#)
6. J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris, *The sharp threshold for bootstrap percolation in all dimensions*, Trans. Amer. Math. Soc. **364** (2012), 2667–2701. [2](#)
7. J. Balogh, B. Bollobás, and R. Morris, *Bootstrap percolation in three dimensions*, Ann. Probab. **37** (2009), 1329–1380. [2](#)
8. J. Balogh, B. Bollobás, R. Morris, and O. Riordan, *Linear algebra and bootstrap percolation*, J. Combin. Theory, Ser. A **119** (2012), 1328–1335. [19](#)
9. B. Bollobás, P. Smith, and A. Uzzell, *Monotone cellular automata in a random environment*, Combin. Probab. Comput. **24** (2015), 687–722. [2](#), [3](#)
10. J. Chalupa, P. L. Leath, and G. R. Reich, *Bootstrap percolation on a Bethe lattice*, J. Phys. C. **12** (1979), 31–35. [2](#)
11. H. Duminil-Copin and A. C. D. Van Enter, *Sharp metastability threshold for an anisotropic bootstrap percolation model*, Ann. Probab. **41** (2013), 1218–1242. [2](#)
12. A. Fey, L. Levine, and Y. Peres, *Growth rates and explosions in sandpiles*, J. Stat. Phys. **138** (2010), 143–159. [1](#)
13. L. R. Fontes, R. H. Schonmann, and V. Sidoravicius, *Stretched exponential fixation in stochastic Ising models at zero temperature*, Comm. Math. Phys. **228** (2002), 495–518. [2](#)
14. M. Granovetter, *Threshold models of collective behavior*, American J. Sociology **83** (1978), 1420–1443. [2](#)
15. J. Gravner, C. Hoffman, J. Pfeiffer, and D. Sivakoff, *Bootstrap percolation on the hamming torus*, Ann. Appl. Probab. **25** (2015), 287–323. [2](#), [6](#), [21](#)
16. L. Guth, *Unexpected applications of polynomials in combinatorics*, The Mathematics of Paul Erdős, Springer New York, 2013, pp. 493–522. [19](#)
17. A. E. Holroyd, *Sharp metastability threshold for two-dimensional bootstrap percolation*, Probab. Theory Related Fields **125** (2003), 195–224. [2](#)
18. A. E. Holroyd, T. M. Liggett, and D. Romik, *Integrals, partitions, and cellular automata*, Trans. Amer. Math. Soc. **356** (2004), 3349–3368. [2](#)
19. I. H. Lee and A. Valentinyi, *Noisy contagion without mutation*, Rev. Econom. Stud. **67** (2000), 47–56. [1](#)
20. D. J. Watts, *A simple model of global cascades on random networks*, Proc. Nat. Acad. Sci. **99** (2002), 5766–5771. [2](#)

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF MEMPHIS, MEMPHIS TN 38152, USA

Email address: pbalistr@memphis.edu

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UK; *and* DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF MEMPHIS, MEMPHIS TN 38152, USA; *and* LONDON INSTITUTE FOR MATHEMATICAL SCIENCES, 35A SOUTH ST., MAYFAIR, LONDON W1K 2XF, UK.

Email address: b.bollobas@dpms.cam.ac.uk

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UK

Email address: j.d.lee@dpms.cam.ac.uk

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UK

Email address: b.p.narayanan@dpms.cam.ac.uk