A geometric Achlioptas process

Tobias Müller*
Utrecht University
t.muller@uu.nl

Reto Spöhel†
Berner Fachhochschule
reto.spoehel@bfh.ch

April 13, 2013

Abstract

The *random geometric graph* is obtained by sampling n points from the unit square (uniformly at random and independently), and connecting two points whenever their distance is at most r, for some given $r = r(n)$. We consider the following variation on the random geometric graph: in each of n rounds in total, a player is offered *two* random points from the unit square, and has to select exactly one of these two points for inclusion in the evolving geometric graph.

We study the problem of avoiding a linear-sized (or "giant") component in this setting. Specifically, we show that for any $r \ll (n \log\log n)^{-1/3}$ there is a strategy that succeeds with high probability in keeping all component sizes sublinear. We also show that this is tight in the following sense: for any $r \gg (n \log\log n)^{-1/3}$, with high probabiliy the player will be forced to create a component of size $(1 - o(1))n$, no matter how he plays. We also prove that the corresponding offline problem exhibits a similar threshold behavior at $r(n) = \Theta(n^{-1/3})$.

These findings should be compared to the existing results for the (ordinary) random geometric graph: there a giant component arises with high probability once r is of order $n^{-1/2}$. Thus our results show, in particular, that in the geometric setting the power of choices can be exploited to a much larger extent than in the classical Erdős-Rényi random graph, where the appearance of a giant component can only be delayed by a constant factor.

1 Introduction

The *random geometric graph* with parameters n and r is obtained by sampling n points from the unit square (uniformly at random and independently), and connecting two points whenever their distance is at most r. The study of this model essentially goes back to Gilbert [11] who defined a very similar model in 1961; for this reason it is sometimes also called the *Gilbert random graph*. Random geometric graphs have been the subject of considerable research effort over the past decades, and quite precise results are now known for this model on aspects such as connectivity, Hamilton cycles, the chromatic number and random walks on the graph. (See for instance [21, 4, 18, 9].) A comprehensive overview of the results prior to 2003 can be found in the monograph [22].

*The research in this paper was done while the author was partially supported by a VENI grant from Netherlands Organisation for Scientific Research (NWO).

†Part of the research in this paper was done while the author was at MPI Saarbrücken, and while he was supported by a grant of the Swiss National Science Foundation.

By the results in Chapter 10 of [22] (which build on the work of several previous authors, including Gilbert [11]), there is a constant λ_{crit} such that if $r = \sqrt{\lambda/n}$ with $\lambda \leq \lambda_{\text{crit}}$ then the largest component contains a sublinear proportion of all vertices, while if $\lambda > \lambda_{\text{crit}}$ then the largest component contains a linear fraction of all vertices. Phrased differently, a "giant" component suddenly emerges when the average degree exceeds a certain constant ($\pi\lambda_{\text{crit}}$ to be precise). An interesting detail is that the precise value of λ_{crit} remains unknown to this date.

1.1 Our results

We consider a *power of choices version* of the random geometric graph. By this we mean the following probabilistic process: There are n *rounds* in total, and in each round a player is offered *two* random points from the unit square, of which he has to select exactly one for inclusion in the evolving geometric graph.

The objective of the player is to keep the size of the largest component as small as possible. (In Section 5 we briefly discuss the setup when the player wants to maximize the size of the largest component.) In particular, we are interested in the question for which functions $r = r(n)$ the player can avoid the formation of a *linear-sized* ("giant") component with high probability.[1] This question is answered by the following theorem:

Theorem 1 *Consider the geometric power of choices process defined above. There exist functions $f, g : (0, \infty) \to (0, 1)$, $g < f$, such that the following holds:*
If $r = \sqrt[3]{\frac{c}{n \cdot \log \log n}}$ for some fixed $c > 0$ then

(i) *There exists a strategy such that, if the player follows this strategy, then w.h.p. in round n all components are smaller than $f(c)n$;*

(ii) *W.h.p., no matter which strategy the player employs, the largest component in round n has order at least $g(c)n$.*

Moreover, $f(c) \to 0$ as $c \downarrow 0$ and $g(c) \to 1$ as $c \to \infty$.

Here and in the rest of the paper, $\log(.)$ denotes the *base 2* logarithm (i.e. the inverse of 2^x).

Theorem 1 extends to the setting with an arbitrary fixed number $d \geq 2$ of choices per step; the expression for r then needs to be replaced by $r = \sqrt[d+1]{c/(n \cdot (\log \log n)^{d-1})}$. We will come back to this at the end of the paper.

Note that Theorem 1 implies in particular that for any $r \ll (n \log \log n)^{-1/3}$, w.h.p. the largest component is sublinear in size, and that for any $r \gg (n \log \log n)^{-1/3}$, w.h.p. the largest component will be of size $(1 - o(1))n$ no matter how the player plays. Thus Theorem 1 establishes a "threshold" of $\Theta((n \log \log n)^{-1/3})$ for the appearance of a giant component in the geometric power of choices process. Note that this threshold is higher than the threshold for the original geometric random graph by a power of n. This is in stark contrast with the power of choices version of the well-known Erdős-Rényi process (see Section 1.2 below); there, the apearance of a giant component can only be delayed by a constant factor.

[1] We say that a sequence of events $(A_n)_n$ holds *with high probability* (abbreviated: w.h.p.) if $\mathbb{P}(A_n) = 1 - o(1)$ as $n \to \infty$.

Observe also that the threshold behaviour is very different from that of the original random geometric graph setting: Theorem 1 states that qualitatively the behaviour of the process is the same for all $c > 0$. In the standard geometric graph on the other hand, the size of the largest component jumps from $\Theta(\log n)$ just below the threshold $r_{\text{crit}} := \sqrt{\lambda_{\text{crit}}/n}$ to $\Theta(n)$ just above r_{crit}. (In the well-known Erdős-Rényi graph a similar phenomenon occurs, see for instance [13].) In the process considered in Theorem 1, however, the order of the largest component under optimal play is $\Theta(n)$ for every $c > 0$. So in particular there is no c_{crit} analogous to λ_{crit}. See Section 5 for some additional discussion on the order of the largest component when r is below the threshold $\Theta((n \log\log n)^{-1/3})$.

The offline setting In the process we discussed so far, the player does not know which points will arrive in future rounds. It is interesting to consider what would happen if the player were *clairvoyant*, i.e. if he already knew from the start of the game where all the points in all the rounds will fall. Put differently, he is given n pairs of random points all at once, and needs to select one point from each of these pairs. This is often called the *offline* version of the game, and the original version is called the *online* version.

Intuitively, the additional advantage of being clairvoyant should allow the player to delay the onset of a giant component even further. The next theorem shows that this is indeed true, but that the advantage is rather modest – the threshold only increases by a factor of $(\log\log n)^{1/3}$.

Theorem 2 *Consider the geometric offline power of choices setting defined above. There exist functions $f, g : (0,\infty) \to (0,1)$, $g < f$, such that the following holds:*
If $r = \sqrt[3]{\frac{c}{n}}$ for some fixed $c > 0$ then

- **(i)** *W.h.p. it is possible to choose n points out of n pairs of random points such that all components are smaller than $f(c)n$;*
- **(ii)** *W.h.p., for every choice of n points out of n pairs of random points, the largest component has order at least $g(c)n$.*

Moreover, $f(c) \to 0$ as $c \downarrow 0$ and $g(c) \to 1$ as $c \to \infty$.

Again the result extends to the setting with $d \geq 2$ choices; the expression for r then is $r = \sqrt[d+1]{c/n}$.

1.2 Background and related work

The notion of the "power of choices" essentially dates back to a 1994 STOC paper by Azar, Broder, Karlin, and Upfal [2, 3]. In informal computer science terms, their result states that if one allocates a large number of jobs to a large number of servers by assigning each job to the currently less busy of *two* randomly chosen servers, one observes a dramatic improvement in load balancing over a completely random assignment. This result marked the beginning of the development of the power of choices as a powerful new paradigm in computer science, with applications to load balancing, hashing, distributed computing, network routing, and other areas (see [19] for a comprehensive survey).

The mathematical model for this setting is usually given in terms of balls and bins. In the standard balls and bins experiment, there are n balls and n bins, and each ball is dropped into a random bin (chosen uniformly at random, independently of the choices for the other balls). Denoting by M_n the number of balls in the fullest bin, also called the *maximum load*, it is well-known (and can be proved by the first and second moment methods) that w.h.p. $M_n = (1 + o(1)) \ln n / \ln \ln n$. (An even more precise result is given in [12]).

In the power of choices version of this setup, the n balls arrive sequentially, and for each ball *two* random bins are sampled (uniformly at random and independently from each other). The goal now is to devise a strategy for choosing between the bins that keeps the maximum load as small as possible. An obvious choice of a strategy is the *greedy* strategy where we always choose the least full bin (in case of a tie we can choose arbitrarily). Azar et al. [3] showed the following remarkable result. Recall that a random variable X *stochastically dominates* the random variable Y if $\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x)$ for all $x \in \mathbb{R}$.

Theorem 3 ([3]) *Consider the power of choices balls and bins process with n bins and n rounds, and let M_n denote the maximum load after the process ends if we employ the greedy strategy. Then*

$$M_n = \log \log n + O(1) \;\; w.h.p.$$

Moreover, the maximum load under any other strategy stochastically dominates the maximum load under the greedy strategy.

This result shows that in the power of two choices ball and bins process, the maximum load under the greedy strategy is *exponentially smaller* than in the ordinary non-power-of-choices version, and that moreover it is very strongly concentrated. It also shows that the greedy strategy is optimal in a very strong sense.

We should remark that Theorem 3 is in a fact a slight simplication of Theorem 1 of [3]. Among other things, it was also shown in [3] that allowing more than two choices per step further decreases the maximum load, but only by a constant factor. Note that the behaviour of the geometric power of choices process is in contrast with this: As mentioned in the previous section, in our setting a choice of d options in each round results in a threshold at $r = \Theta(\sqrt[d+1]{1/(n \cdot (\log \log n)^{d-1})})$. Thus, every additional choice per step increases the threshold by a power of n.

Theorem 3 plays an important role in our proof of Theorem 1, and may give some intuition why a power of $\log \log n$ appears in the threshold for our geometric power of choices process.

The Achlioptas process The power of choices version of the classical Erdős-Rényi graph process is usually called the *Achlioptas process* after Dimitris Achlioptas, who first suggested it. The Achlioptas process starts with an empty graph on n vertices. In each round, two random vertex pairs are presented, and the player needs to select exactly one of them for inclusion as an edge in the evolving graph. His goal is to delay or accelerate the occurence of some monotone graph property, such as containing a giant component, containing a triangle, or containing a Hamilton cycle.

Bohman and Frieze [5] were the first to study the Achlioptas process. They showed that by an appropriate edge-selection strategy, the emergence of a giant component can be delayed by a constant factor. Several authors subsequently improved on their bounds, and also showed that no improvement beyond a constant factor is possible [6, 24]. The opposite problem of creating a giant component as quickly as possible was studied in [10, 8], and an exact threshold for the offline problem was determined in [7]. All these results show that the power of choices offered by the Achlioptas process affects the threshold for the appearance of a giant component only by a constant factor.

More recently, the precise nature of the phase transition in the Achlioptas process received much attention: countering "strong numerical evidence" presented in *Science* [1], Riordan and Warnke [23] showed that for a large number of natural player strategies, the Achlioptas phase transition is in fact continuous.

Other properties that have been studied for the Achlioptas process include Hamiltonicity [16] and the appearance of copies of a given fixed graph F [15, 20, 17].

The vertex Achlioptas process The reader might wonder whether it is reasonable to compare our geometric power of choices process to the Achlioptas process – after all, we are selecting vertices in the former and edges in the latter. Let us therefore consider the following process: the n vertices of an Erdős-Rényi random graph $G(n,m)$ (i.e. the random graph sampled uniformly from all graphs on n vertices and m edges) are revealed two at a time, along with all edges induced by the vertices revealed so far, and we need to select one of of the two vertices for inclusion in a subgraph. Our goal is to avoid a linear-sized component in the subgraph induced by the vertices we select. This process indeed seems to be a better reference for our comparison, but its phase transition has not been studied explicitly in the literature. As it turns out, also in this *"vertex Achlioptas process"*, w.h.p. a giant component cannot be avoided as soon as the average degree of the underlying random graph exceeds a certain constant. We give a proof for this in Appendix A.

1.3 About the proofs

In the following we informally outline some of the main ideas used in the proofs of Theorems 1 and 2. Our goal here is not to give detailed proof sketches, but to give some impression of our overall proof strategies and of the type of arguments used. We will describe our proof strategies in more detail later where appropriate.

In all our proofs we consider a discretized version of the random geometric graph as follows: We divide the unit square into $\Theta(r^{-2})$ many small squares (called *boxes*) in such a way that, essentially, we no longer need to worry about the precise locations of the random points, but only need to know which of the boxes are occupied by at least one point. In this way, our analysis of the process reduces to an analysis of appropriate random subgraphs of a large finite square grid (or king's move grid, see Section 3.1 below). With some technical work, the results for this grid then translate back to the original geometric setting to give the desired results.

Lower bound proofs The key idea in the proofs of the lower bound parts of both Theorem 1 and 2 is the following: To avoid the formation of large connected components of occupied boxes (and thus also of large connected components in the original power of choices random geometric graph setting), we designate a subset of the boxes as the "moat". This moat separates the unit square into "small" parts (see Figure 1 below). The player's goal is to prevent the appearance of paths of occupied boxes crossing the moat. In both settings, he selects points outside the moat whenever possible. His main worry are the pairs of points that both fall into the moat, as these force him to select a point in the moat.

In the offline scenario, it is not too hard to show that even if the moat is only a (moderately large) constant number of boxes wide, then w.h.p. there is a choice of points for which the moat is not crossed. The proof of this relies on an (approximate) analogy of our setting with the Erdős-Rényi random graph in its subcritical stage, where the boxes in the moat play the role of the vertices and the pairs of points that both fall into the moat yield the edges.

In the online scenario, we can only ensure that the moat will not be crossed if it is at least $\Theta(\log\log n)$ boxes wide. Our strategy here is more involved; the key part is to set things up in such a way that, essentially, we can invoke Theorem 3 to argue that w.h.p. the moat will not be crossed. To that end, we divide the moat into "blocks" consisting of $\Theta(\log\log n) \times \Theta(\log\log n)$ boxes, and pay special attention to the blocks that become "dangerous" because, oversimplifying slightly, only relatively few more occupied boxes are required to create a (potentially crossing) long path inside them.

Upper bound proofs A key ingredient in both our upper bounds proofs is a simple isoperimetric inequality for subgraphs of the square grid. It allows us to conclude from the fact that there are relatively few unoccupied boxes that a significant proportion of the occupied boxes must belong to relatively large connected components of the graph induced by all occupied boxes.

For the offline case, we use this as part of a combinatorial counting argument. Essentially, we count the number of sets of boxes whose removal decomposes the grid graph into "small" components. More precisely, in order to keep the total number of sets to be considered sufficiently low, we will only count appropriate *subsets* of decomposing sets as described. By an expectation argument, we will then show that w.h.p. the player will not be able to avoid a single one of these sets, and therefore will be forced to create a "large" (linear-sized) component.

For the online case, we use a two-round approach and analyze the process after $n/2$ points rounds as an intermediate step. To this end, we divide the grid graph into "megablocks" consisting of $O(1/(r\log\log n)) \times O(1/(r\log\log n))$ boxes, and define an appropriate notion of "good" megablocks (see Figure 3(b) on page 22 below). We show that at time $n/2$, w.h.p. most boxes of the original grid graph are occupied, and that as a consequence most megablocks are good. It follows with the mentioned isoperimetric inequality that a bounded number of connected components of the graph of megablocks (defined in the obvious way – megablocks are adjacent if they share a side) covers most of the unit square. Conditional on that, we then show that in the remaining $n/2$ rounds of the process, w.h.p. at least one of these components will evolve into a linear-sized component of the original grid graph. To show that the player cannot avoid this, we apply a slight variation of Theorem 3 in an appropriate geometric setup.

2 Preliminaries

For the sake of readability and clarity of exposition we will mostly ignore rounding. I.e. we will usually omit floor and ceiling signs. In all cases it is a routine matter to check that all computations and proofs also work if floors and ceilings are added, and we leave this to the reader.

Throughout the paper, $\ln x$ will denote the *natural logarithm* and $\log x$ will denote the logarithm base 2 (i.e. $\ln x$ is the inverse of e^x and $\log x$ is the inverse of 2^x). For $n \in \mathbb{N}$ we will denote $[n] := \{1, \ldots, n\}$.

By a slight abuse of notation we will write $\mathbb{Z}^2$ to denote the *graph of the integer lattice*, i.e. the infinite graph with vertex set $\mathbb{Z}^2$ and an edge between two vertices if and only if their distance is exactly one. Similarly, we will also identify $A \subseteq \mathbb{Z}^2$ with the subgraph of $\mathbb{Z}^2$ it induces. Thus, $[s]^2$ in particular denotes an $s \times s$ grid.

We will use the notation $\mathrm{Bi}(n, p)$ to denote the binomial distribution with parameters n and p; we will use $\mathrm{Po}(\mu)$ to denote the Poisson distribution with mean μ; and we will use $\mathrm{Geom}(p)$ to denote the geometric distribution with parameter p.

We will use the following incarnation of the Chernoff bound. A proof can for instance be found in [22], c.f. Lemma 1.1 and Lemma 1.2.

Lemma 4 *Let X be a random variable with a binomial or Poisson distribution. Then, letting $\mu := \mathbb{E}X$, we have*

(i) *For $k \geq \mu$ we have $\mathbb{P}(X \geq k) \leq \exp[-\mu H(k/\mu)]$, and*

(ii) *For $k \leq \mu$ we have $\mathbb{P}(X \leq k) \leq \exp[-\mu H(k/\mu)]$,*

where $H(x) := x \ln x - x + 1$.

3 Lower bound proofs

3.1 Proof of part (i) of Theorem 1

We will consider a suitable discretization of the geometric graph. Let $\mathcal{D}_r$ be the obvious dissection into squares of side length r. That is,

$$\mathcal{D}_r := \{[ir, (i+1)r] \times [jr, (j+1)r] : 0 \leq i, j < 1/r\}. \tag{1}$$

(We assume that $1/r$ is an integer thoughout the section[2].) We refer to the elements of $\mathcal{D}_r$ as *boxes*, and we say that a box is *occupied* if is contains a point of our process and *empty* otherwise.

The *king's move grid* $\mathcal{K}_s$ on $s \times s$ vertices is the graph with vertex set $[s]^2$ and an edge between two vertices if and only if their distance is at most $\sqrt{2}$. This way, we can also move diagonally, which explains the name "king's move grid" – at least to those familiar with the rules of chess.

We will identify the boxes of the dissection $\mathcal{D}_r$ with the vertices of the king's move grid $\mathcal{K}_{(1/r)}$ – that is, we consider two boxes adjacent if they share a side or a corner. Note that if two points are adjacent in the original geometric graph, then they must lie in boxes that are adjacent in $\mathcal{K}_{(1/r)}$. Thus, denoting by Occ_r the subgraph of $\mathcal{K}_{(1/r)}$ induced by by the occupied boxes, the following holds: If C_1, C_2 are distinct components of Occ_r, then the points in C_1 and the points in C_2 belong to different components of the geometric graph.

We further group the boxes into "blocks" consisting of $h \times h$ boxes each, where $h = 100 \cdot \log\log n$. (For convenience we assume that $1/r$ and $1/hr$ are both integers.) Again it is useful to consider the blocks as vertices of a king's move grid $\mathcal{K}_{(1/hr)}$.

Our strategy can be described as follows. At the start of the game we will pick a constant $K = K(c)$, to be made explicit later on, and we pick a set of

$$N := K \cdot (1/hr),$$

[2]The concerned reader may check that, if we take $\tilde{h} := \lfloor 100 \log\log n \rfloor$ instead of $h = 100 \log\log n$ and $\tilde{r} := 1/\tilde{h}\lfloor 1/r\tilde{h} \rfloor$ instead of r, then $1/r$ and $1/hr$ are both integers, and $\tilde{h} = (1+o(1))h, \tilde{r} = (1+o(1))r$, and all proofs and computations in this section carry through. Since $\tilde{r} \geq r$ this also establishes the result for r.

blocks that forms a "moat" as in Figure 1 that we will "defend". The next lemma formally captures the essential properties of our "moat". Its proof is indicated in Figure 1; the details are left to the reader.

Lemma 5 *For every fixed $K > 0$, the following holds for all sufficiently large $s \in \mathbb{N}$. There is a subgraph $H \subseteq \mathcal{K}_s$ of the king's move grid with $v(H) \geq s^2 - Ks$ vertices such that*

$$v(H_{\max}) \leq (a(K) + o_s(1)) \cdot s^2,$$

where $H_{\max}$ denotes the largest component of H, and

$$a(K) = \begin{cases} 1/(\lfloor K \rfloor + 1) & \text{if } K > 1, \\ 1 - (K/2)^2 & \text{if } K \leq 1. \end{cases}$$

In particular, $0 < a(K) < 1$ for all $K > 0$ and $a(K) \to 0$ as $K \to \infty$. ∎

Let us remark that while it is clearly possible to improve on the expression for $a(K)$ given in Lemma 5 we have made no attempt to do so as the current version of the lemma suffices for our purposes.

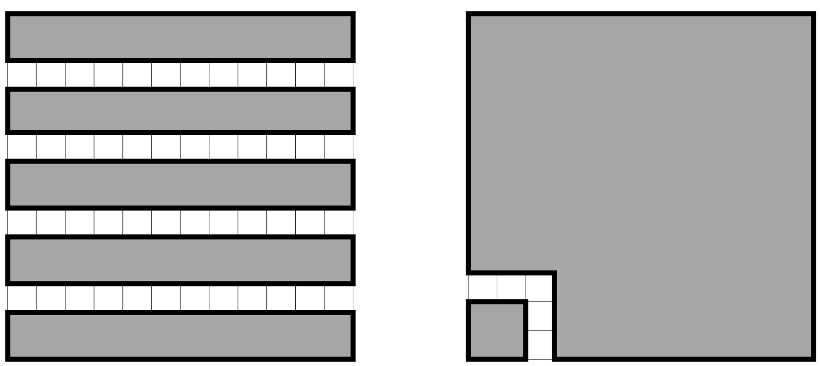


Figure 1: Dividing the unit square using a moat consisting of $K \cdot (1/hr)$ blocks of dimensions $hr \times hr$, when K is large (left) and when K is small (right).

Hence, our strategy begins by selecting N blocks according to Lemma 5 so that they form a moat $\mathcal{M}$ that divides the unit square $[0,1]^2$ into pieces of area at most $a(K)$. During the game we will now attempt to prevent the formation of (king's move) paths of occupied boxes crossing the moat. We will devise a strategy that will succeed in defending the moat in this sense (w.h.p.). This then implies that the area of the occupied boxes corresponding to the largest component of the geometric graph is $a(K) + o(1)$. (This may include some boxes from the moat; note however that the entire moat has area $o(1)$.) With an additional argument showing that (w.h.p.) no union of boxes with area $A \geq \varepsilon$ ($\varepsilon > 0$ arbitrary but fixed) will have received more than $(1 + \varepsilon) \cdot A \cdot n$ points, it then follows that the number of vertices of the largest component of the geometric graph is bounded by $(a(K) + o(1)) \cdot n$, which concludes the proof.

Let us now describe our strategy in more detail. We say that a block B is *bad* if there is a path of occupied boxes of length h that uses one of the boxes of B and uses only boxes of the moat. Observe that such a path may use boxes in blocks other than B, but it may not use

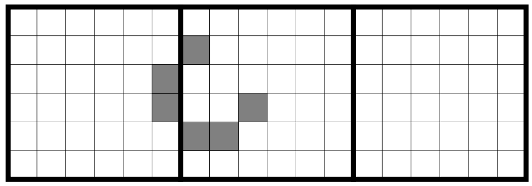


Figure 2: The left and middle blocks are bad. (The gray cells represent occupied cells.)

boxes in blocks that are either not adjacent to B or do not belong to the moat. See Figure 2 for a depiction.

Clearly, if there are no bad blocks then there cannot be any path crossing the moat.

We will say that a block B is *dangerous* if occupying up to $2 \log \log n$ additional boxes can render it bad. In other words, B is dangerous if there is some path P of length h that uses only boxes of the moat, and at least one box of B, such that at least $h - 2 \log \log n$ boxes of P are occupied. If B is not dangerous we will call it *safe*. Let $N_{\mathrm{bad}}(t), N_{\mathrm{dang}}(t)$ denote the number of bad resp. dangerous blocks at the end of round $t = 1, \ldots, n$.

We will keep an ordered list of blocks $L(t) = (L_1(t), \ldots, L_{2^{-h}N}(t))$ containing exactly $2^{-h}N$ blocks, which will be updated at the end of each round. We will make sure that if $N_{\mathrm{dang}}(t) \leq 2^{-h}N$ then $L(t)$ contains all dangerous blocks by applying the following update rule. If $N_{\mathrm{dang}}(t) < 2^{-h}N$ and some previously safe block $B \notin L(t)$ becomes dangerous during round $t+1$, then we replace an arbitrary safe block in the list with the new dangerous block. (If for some t, it happens that $N_{\mathrm{dang}}(t) > 2^{-h}N - k$ and k blocks become dangerous during round $t+1$, then our strategy will have failed.)

We will call the blocks in $L(t)$ *pseudodangerous*, and the blocks not in $L(t)$ *pseudosafe* (with respect to round t).

Our strategy can now be described as follows:

(STR-1) We always pick a point outside of the moat M if we can. If both points are outside the moat, we choose randomly.

(STR-2) If both points fall inside the moat and both are in pseudosafe blocks, then we play randomly.

(STR-3) If both fall inside the moat, one in a pseudodangerous, and one in a pseudosafe block, then we choose the pseudosafe block.

(STR-4) If both fall in the moat, both in pseudodangeous blocks, say in $L_i(t)$ and $L_j(t)$, then we do the following. Set

$$k_i := |\{t' < t : \text{ we played in } L_i(t') \text{ in round } t'\}|,$$

and define k_j similarly. If $k_i < k_j$ then we play in $L_i(t)$ and if $k_i > k_j$ then we play in $L_j(t)$. In case of a draw we play randomly.

Remark: Recall that which block $L_i(t)$ points to may change between rounds. Let us stress that in **(STR-4)** we compare the number of times we played in the i-th resp. the j-th pseudorandom block as opposed to the number of times we played in the specific blocks that $L_i(t)$ resp. $L_j(t)$ represent in round t. On the other hand, once $L_i(t)$ points to a dangerous block, the value of $L_i(t')$ will remain the same for all $t' \geq t$. Thus, if $L_i(t)$ points to a dangerous block B, the number of times we played in the i-th pseudorandom block until round t is an upper bound on the number of times we played in B *since it became dangerous*. This subtle, but very important point shows that – provided the number of dangerous blocks remains below $2^{-h}N$ – if the number of rounds in which we play in the i-th pseudorandom block stays below $2\log\log n$ for every index i then no bad blocks will be formed.

As indicated before, we will show that our strategy works by eventually proving that, w.h.p., no bad blocks will be formed (which implies the moat will not get crossed), and then showing that for each of the regions that the moat divides the unit square into, the number of points that fell into the region is proportional to its area (w.h.p).

We start by proving the following lemma which shows that, w.h.p., we never have more dangerous blocks than entries in the list $L(t)$. Clearly, this implies that we will be able to keep the dangerous blocks a subset of the pseudodangerous blocks.

Lemma 6 *There is an absolute constant $C_0 > 0$ such that if $K \leq C_0/c$, then $N_{dang}(t) \leq 2^{-h}N$ for all $t = 1, \ldots, n$, w.h.p.*

Proof: Observe that the area of the moat is

$$\text{area}(\mathcal{M}) := N \cdot (hr)^2 = Khr.$$

Let R denote the number of rounds in which we (are forced to) play inside the moat $\mathcal{M}$. Clearly $R \stackrel{d}{=} \text{Bi}(n, \text{area}^2(\mathcal{M}))$. Note that

$$\mathbb{E}R = n \cdot K^2 \cdot h^2 \cdot r^2 = n^{\frac{1}{3}+o(1)} \tag{2}$$

by choice of r.

To facilitate the analysis it is helpful to consider what would happen under a slightly different setup. Suppose that whenever both points fall inside the moat during some round we add both points and otherwise we pick a point outside the moat. Clearly we will end up with $2R$ points in the moat, distributed uniformly. Moreover, the set of occupied boxes inside the moat under this setup will be a superset of the set of occupied boxes in the original setup.

Let $Z \stackrel{d}{=} \text{Po}(8\mathbb{E}R)$ be a Poisson variable with mean $8\mathbb{E}R$. By applying the Chernoff bound, Lemma 4, together with (2) we can easily see that

$$\begin{aligned} \mathbb{P}(Z < 2R) &\leq \mathbb{P}(Z \leq \mathbb{E}Z/2) + \mathbb{P}(R \geq 2\mathbb{E}R) \\ &\leq \exp[-\mathbb{E}Z \cdot H(\tfrac{1}{2})] + \exp[-\mathbb{E}R \cdot H(2)] \\ &= \exp[-\Omega(n^{\frac{1}{3}+o(1)})]. \end{aligned}$$

Let us once more modify the setup slightly, and just drop Z points on the moat (their locations chosen uniformly at random and independent of Z and the locations of the other points). This way, the points in the moat will form a *Poisson process*, which has the convenient consequence that the events that different boxes in the moat are occupied become independent (see for instance [14]). Note that, by choosing a suitable coupling, we can ensure that if $Z \geq 2R$ then this setup dominates our previous setup in the sense that the set of points in the moat under the old setup is a subset of the set of points in the moat under the new setup.

This allows us to bound the probability that a block becomes dangerous by considering the new setup. Observe that the expected number of points in a given box is

$$\begin{aligned}\mu &:= \mathbb{E}Z \cdot \left(\tfrac{r^2}{\operatorname{area}(\mathcal{M})}\right) \\ &= 8 \cdot n \cdot K^2 \cdot h^2 \cdot r^2 \cdot \left(\tfrac{r^2}{K \cdot h \cdot r}\right) \\ &= 8 \cdot K \cdot h \cdot n \cdot r^3 \\ &= 800 \cdot K \cdot c.\end{aligned}$$

Let us write $p := 1 - e^{-\mu}$. Then a box in the moat is occupied with probability p, independently of all other boxes.

Next, observe that the number of paths of length h starting either in a given block or in one of the neighbouring blocks is at most $9h^2 8^{h-1}$ (such a path starts in one of the $9h^2$ boxes belonging to the block and adjacent blocks, and there are always at most 8 choices for the next box of the path). Let p_{dang} denote the probability that a particular block is dangerous under this new setup. The union bound gives

$$p_{\text{dang}} \le 9 \cdot h^2 \cdot 8^{h-1} \cdot \mathbb{P}\left(\operatorname{Bi}(h,p) \ge h - 2\log\log n\right).$$

Using the Chernoff bound (Lemma 4), we see that for all $k \ge h$

$$\begin{aligned}\mathbb{P}\left(\operatorname{Bi}(h,p) \ge h - 2\log\log n\right) &\le \exp\left[-hp \cdot H\left(\tfrac{h-2\log\log n}{hp}\right)\right] \\ &= \exp\left[-hp \cdot H(\tfrac{98}{100 \cdot p})\right],\end{aligned}$$

where as usual $H(x) = x\ln x - x + 1$.

Observe that $pH\left(\frac{98}{100p}\right) \to \infty$ as $p \downarrow 0$, since

$$pH\left(\frac{98}{100 \cdot p}\right) = p \cdot \left(\left(\frac{98}{100p}\right)\ln\left(\frac{98}{100p}\right) - \frac{98}{100p} + 1\right) = \frac{98}{100}\ln\left(\frac{98}{100p}\right) + \frac{98}{100} + p.$$

Hence, there exists a universal constant p_0 such that

$$\exp\left[-pH\left(\frac{98}{100 \cdot p}\right)\right] \le \frac{1}{100} \text{ for all } p \le p_0. \tag{3}$$

Let us set $C_0 := -\ln(1-p_0)/800$, so that $K = K(c) \le C_0/c$ implies that $p = 1 - \exp[-800Kc] \le p_0$.

With this choice of K we have

$$\begin{aligned}p_{\text{dang}} &\le 9 \cdot h^2 \cdot 8^{h-1} \cdot \exp\left[-hp \cdot H(\tfrac{98}{100 \cdot p})\right] \\ &= \tfrac{9}{8} \cdot h^2 \cdot \left(8 \cdot \exp[-pH(\tfrac{98}{100 \cdot p})]\right)^h \\ &\le \tfrac{9}{8} \cdot h^2 \left(\tfrac{8}{100}\right)^h \\ &\le 10^{-h},\end{aligned}$$

where the last line holds for n sufficiently large (recall $h \to \infty$ as $n \to \infty$).

Let D denote the number of dangerous blocks in our modified process where we simply drop Z points uniformly at random on the moat. Then $\mathbb{E}D = N \cdot p_{\text{dang}}$ and hence

$$\begin{aligned}\mathbb{P}(D > 2^{-h}N) &= \mathbb{P}(D > (2^{-h}/p_{\text{dang}}) \cdot \mathbb{E}D) \\ &\le \mathbb{P}(D > 5^h \cdot \mathbb{E}D) \\ &\le 5^{-h} \\ &= o(1),\end{aligned}$$

where we used Markov's inequality for the third line.

Putting everything together, we see that

$$\begin{array}{rcl}\mathbb{P}(N_{\text{dang}} > 2^{-h}N) & \leq & \mathbb{P}(Z < 2R) + \mathbb{P}(D > 2^{-h}N) \\ & = & o(1),\end{array}$$

which concludes the proof of the lemma. ■

Let R_{psd} denote the number of rounds in which we (are forced to) play in a pseudodangerous block.

Lemma 7 *$R_{psd} \leq 2^{-h}N$ w.h.p.*

Proof: The probability that both points land in pseudodangerous blocks in round $t = 1, \dots, n$ is

$$p := \left(2^{-h}Nh^2r^2\right)^2.$$

Thus we have

$$\begin{array}{rcl}\mathbb{E}R_{\text{psd}}/(2^{-h}N) & = & (np)/(2^{-h}N) \\ & = & (n2^{-2h}N^2h^4r^4)/(2^{-h}N) \\ & = & n2^{-h}Nh^4r^4 \\ & = & O\left(n2^{-h}h^3r^3\right) \\ & = & O\left(h^2 2^{-h}\right) \\ & = & o(1),\end{array}$$

where we used that $N = O(1/hr)$ in the fourth line, and that $nr^3 = c/\log\log n = \Theta(1/h)$ in the fifth line. It follows that

$$\mathbb{P}(R_{\text{psd}} > 2^{-h}N) \leq \frac{\mathbb{E}R_{\text{psd}}}{2^{-h}N} = o(1),$$

by Markov's inequality. ■

Lemma 8 *If $K \leq C_0/c$ with C_0 as in Lemma 6, then $N_{bad}(n) = 0$ w.h.p.*

Proof: If some bad block got created even though we stuck to our strategy, then it must be the case that either **a)** there were more dangerous blocks than places in our list of pseudodangerous blocks, or **b)** there is some index $1 \leq i \leq 2^{-h}N$ such that there were more than $2\log\log n$ rounds t when we played the i-th pseudodangerous block in our list. (Recall the remark immediately following the description of the strategy **(STR-1)**–**(STR-4)**.)

Let E denote the event that **b)** happens. Clearly

$$\mathbb{P}(N_{\text{bad}}(n) > 0) \leq \mathbb{P}(N_{\text{dang}}(n) > 2^{-h}N) + \mathbb{P}(R_{\text{psd}} > 2^{-h}N) + \mathbb{P}(E \text{ and } R_{\text{psd}} \leq 2^{-h}N).$$

By Lemmas 6 and 7 the first two terms of the right hand side are $o(1)$. Now notice that

$$\begin{array}{rcl}\mathbb{P}(E \text{ and } R_{\text{psd}} \leq 2^{-h}N) & \leq & \mathbb{P}(E|R_{\text{psd}} \leq 2^{h}N) \\ & \leq & \mathbb{P}(E|R_{\text{psd}} = 2^{-h}N),\end{array}$$

where the second inequality holds by obvious monotonicity. Now notice that, by part **(STR-4)** of our strategy, the event that E holds, given that $R_{\text{psd}} = 2^{-h}N$, can be viewed as the event that, in the standard power of choices balls and bins setup with $\tilde{n} := 2^{-h}N$ balls and $\tilde{n}$ bins, the maximum load is at least $2\log\log n$. Since $\tilde{n} = 2^{-h} \cdot K \cdot (1/hr) = n^{\frac{1}{3}+o(1)}$ we have that

$$\log\log\tilde{n} = (1+o(1))\log\log n.$$

It therefore follows immediately from Theorem 3 that $\mathbb{P}(E|R_{\text{psd}} = 2^{-h}N) = o(1)$. We see that $\mathbb{P}(N_{\text{bad}}(n) > 0) = o(1)$, as required. ■

By this last lemma, our strategy succeeds (w.h.p.) in confining the components of the evolving random geometric graph to subsets of the unit square of area bounded by $a(K)+o(1)$. The finishing touch of the proof of part **(i)** of Theorem 1 comes in the form of the following lemma.

Lemma 9 *For every $\varepsilon > 0$ the following holds if we follow the strategy set out above. W.h.p. every $A \subseteq [0,1]^2$ that is the union of boxes of the dissection $\mathcal{D}_r$ and with* $\text{area}(A) \geq \varepsilon$ *contains at most $(1+\varepsilon) \cdot \text{area}(A) \cdot n$ points.*

Proof: For $A \subseteq [0,1]^2$, let $\mathcal{N}(A)$ denote the number of points in A (in round n). Let us first recall that the moat satisfies $\text{area}(\mathcal{M}) = Khr = o(1)$. If R denotes the number of rounds in which we (are forced to) take a point in the moat then clearly $R \overset{d}{=} \text{Bi}(n, \text{area}^2(\mathcal{M}))$, and in particular $\mathbb{E}R = n \cdot \text{area}^2(\mathcal{M}) = o(n)$. Hence, by Markov's inequality

$$\mathbb{P}(\mathcal{N}(\mathcal{M}) \geq (\varepsilon^2/2)n) = \mathbb{P}(R \geq (\varepsilon^2/2)n) = o(1).$$

Let $\mathcal{A}$ denote all the subsets of $[0,1]^2$ under consideration, i.e. all A that are unions of boxes of $\mathcal{D}_r$ and have area at least ε.

Pick an arbitrary $A \in \mathcal{A}$ and set $A' := A \setminus \mathcal{M}$. Then $\text{area}(A') = \text{area}(A) - o(1)$. Note that in every round $1 \leq t \leq n$, a point is added to A' with probability

$$p := (1 - \text{area}^2(\mathcal{M})) \cdot \frac{\text{area}(A')}{1 - \text{area}(\mathcal{M})} = (1+o(1)) \cdot \text{area}(A).$$

(This is because if both points fall in the moat, we obviously add a point outside of A', and otherwise we add a point drawn according to the uniform distribution on $[0,1]^2 \setminus \mathcal{M}$.) We have $\mathcal{N}(A') \overset{d}{=} \text{Bi}(n,p)$ so that by the Chernoff bound (Lemma 4) we have

$$\begin{aligned} \mathbb{P}(\mathcal{N}(A') > (1+\varepsilon/2) \cdot \text{area}(A) \cdot n) &\leq \exp[-n \cdot p \cdot H\left(\tfrac{(1+\varepsilon/2)\cdot\text{area}(A)\cdot n}{np}\right)] \\ &= \exp[-\Omega(n)], \end{aligned}$$

since $p = \Omega(1)$ and $(1+\varepsilon/2) \cdot \text{area}(A) \cdot n/(np) = 1 + \varepsilon/2 + o(1)$ is bounded away from one. Let E denote the event that there exists $A \in \mathcal{A}$ with $\mathcal{N}(A) > (1+\varepsilon)\,\text{area}(A)n$. Noting that $(\varepsilon^2/2)n \leq (\varepsilon/2) \cdot \text{area}(A) \cdot n$ for all $A \in \mathcal{A}$, we obtain that

$$\begin{aligned} \mathbb{P}(E) &\leq \mathbb{P}(\mathcal{N}(\mathcal{M}) > (\varepsilon^2/2)n) + \textstyle\sum_{A\in\mathcal{A}} \mathbb{P}(\mathcal{N}(A \setminus \mathcal{M}) > (1+\varepsilon/2) \cdot \text{area}(A) \cdot n) \\ &\leq o(1) + 2^{(1/r)^2} \cdot \exp[-\Omega(n)] \\ &= o(1) + \exp[n^{\frac{2}{3}+o(1)} - \Omega(n)] \\ &= o(1), \end{aligned}$$

where in the second line we used that $|\mathcal{A}| \leq 2^{(1/r)^2}$ as sets in $\mathcal{A}$ are unions of the boxes of our dissection $\mathcal{D}_r$, and in the third line we used the specific form of r. ■

Lemmas 6–9 together imply part **(i)** of Theorem 1. For completeness we spell out the details:

Proof of part (i) of Theorem 1: If we take $K = C_0/c$ with C_0 as provided by Lemma 6, and follow the strategy described by **(STR-1)**–**(STR-4)** above, then by Lemma 8, w.h.p. every connected component of the resulting geometric graph will lie inside a set of boxes of area at most $a(K) + o(1)$ with $a(.)$ as in Lemma 5. Therefore, by Lemma 9, w.h.p., every component of the geometric graph will have at most $n \cdot (a(K) + o(1))$ vertices. Thus the claim follows for, say, $f(c) := \sqrt{a(C_0/c)}$. (Recall that $0 < a(K) < 1$ for all K.) Since $K = C_0/c \to \infty$ as $c \downarrow 0$ and $a(K) \to 0$ as $K \to \infty$, we also have $f(c) \to 0$ as $c \downarrow 0$. ■

3.2 Proof of part (i) of Theorem 2

Our proof strategy is similar to the one for part **(i)** of Theorem 1 used in the preceding section. We will make use of a standard result for the Erdős-Rényi random graph $G(n, m)$. Recall that $G(n, m)$ is obtained by taking a set of n vertices, and selecting a set of m edges uniformly at random from all possible sets of m edges. A graph is *1-orientable* if its edges can be oriented in such a way that every vertex has indegree at most 1.

The following result is a special case of Theorem 5.5 in the standard reference [13].

Theorem 10 *If $m \le cn$ with $c < \frac{1}{2}$ then $G(n, m)$ consists only of trees and unicyclic components, w.h.p. In particular, $G(n, m)$ is 1-orientable w.h.p.*

Let $\mathcal{D}_r$ again be defined by (1). Again we will consider blocks, i.e. $h \times h$ groups of boxes. However, this time we simply set $h := 100$. Again we pick a constant $K = K(c)$, and build a moat $\mathcal{M}$ consisting of $N := K \cdot (1/hr)$ blocks, in such a way that the moat divides the unit square into parts of area no more than $a(K)$ with $a(.)$ as in Lemma 5. As before, we select points outside the moat whenever possible, breaking ties randomly. As we will see, Theorem 10 will then allow us to deal relatively easily with pairs of points that both fall into the moat. Let R denote the number of rounds in which both points fall into the moat.

Lemma 11 *There is an absolute constant C_0 such that if $K \le C_0/c$ then $R < N/100$ w.h.p.*

Proof: Observe that $R \stackrel{d}{=} \mathrm{Bi}(n, \mathrm{area}^2(\mathcal{M}))$, and $\mathrm{area}(\mathcal{M}) = N \cdot (hr)^2 = Khr$. Hence

$$\mathbb{E}R = nK^2h^2r^2 = \frac{cK^2h^2}{r},$$

using that $nr^3 = c$. This is less than $N/100 = K/(100hr)$ for $K < 1/(100h^3c) = 1/(10^8c)$. Hence the claim follows for $C_0 := 10^{-8}$.

For such a choice of K, the Chernoff bound (Lemma 4) gives

$$\begin{aligned}\mathbb{P}(R > N/100) &\le \mathbb{P}(R > 10 \cdot \mathbb{E}R) \\ &\le \exp[-\Omega(\mathbb{E}R)] \\ &= \exp[-\Omega(n^{1/3})] \\ &= o(1),\end{aligned}$$

where we used the fact that $\mathbb{E}R = \Theta(1/r) = \Theta(n^{1/3})$. ■

Let us say that a block gets *doubly hit* in some round if both balls fall into the block in that round.

Lemma 12 *W.h.p., no block of the moat gets doubly hit in more than three rounds.*

Proof: Let us fix a block B, and let Z denote the number of rounds in which B gets doubly hit. Clearly $Z \stackrel{d}{=} \mathrm{Bi}(n, (hr)^4)$. Then we have

$$\mathbb{P}(Z \geq 4) = O(n^4 r^{16}) = O(n^{-4/3}).$$

Thus, the expected number of blocks that get doubly hit in at least four rounds is $O(N \cdot n^{-4/3}) = O(n^{-1}) = o(1)$. ■

Let us now define an auxiliary (random) graph $\tilde{G}$, whose vertices are the N blocks of the moat and where for every pair of blocks $B \neq B'$ there is an edge between them if in some round one of the points landed in B while the other landed in B'.

Lemma 13 *Provided $K \leq C_0/c$ with C_0 as in Lemma 11, the graph $\tilde{G}$ is 1-orientable, w.h.p.*

Proof: Let E denote the event that $\tilde{G}$ consists of trees and unicyclic components (and hence is 1-orientable), and let R' denote the number of rounds in which the points fell into two different blocks of the moat. Observe that if we condition on $R' = m$ then $\tilde{G}$ is just a copy of the Erdős-Rényi random graph $G(N, m)$. Let F denote the event that the Erdős-Rényi random graph $G(N, N/100)$ consists of trees and unicyclic components.

Then we have that

$$\mathbb{P}(E^c) \leq \mathbb{P}(R' > N/100) + \mathbb{P}(F^c) = o(1),$$

by Lemma 11 and Theorem 10. ■

As mentioned, our strategy for the offline process will always select a point outside the moat if possible, choosing randomly if both points fall outside the moat. (Which point we select when both points fall in the moat will be specified shortly.) Let us note that the proof of Lemma 9 in the previous section only used part **(STR-1)** of our strategy for the online setting, and therefore carries over to our offline strategy:

Lemma 14 *Consider the offline process, with the dissection $\mathcal{D}_r$, the moat $\mathcal{M}$, etc., as above. Assume that we always select a point outside the moat if we can, choosing randomly if both points fall outside the moat. Then for every $\varepsilon > 0$ the following holds w.h.p.: every $A \subseteq [0,1]^2$ that is the union of boxes of the dissection $\mathcal{D}_r$ and with $\mathrm{area}(A) \geq \varepsilon$ contains at most $(1+\varepsilon) \cdot \mathrm{area}(A) \cdot n$ points.* ■

We are now ready for the proof of part **(i)** of Theorem 2.

Proof of part (i) of Theorem 2: We show that, w.h.p., we can select one point from each pair in such a way that no block of the moat will contain more than four points. Clearly, this then implies that the moat will not be crossed.

To see this, note first that, by Lemma 12, w.h.p., no block will contain more than three points coming from rounds when it was doubly hit. Observe also that, by Lemma 13, w.h.p., the auxiliary graph $\tilde{G}$ is 1-orientable. Hence it is possible to select one point from each pair of points that both fall into the moat in such a way that this contributes at most one point to each block. Thus we can ensure that, in total, each block will indeed contain at most four points (w.h.p.).

This shows that the player succeeds (w.h.p.) in stopping the moat from getting crossed. The result now follows with Lemma 14 exactly as in the online case. ■

4 Upper bound proofs

4.1 Preliminaries

We need a number of auxiliary results for our upper bound proofs. We collect these in the next few subsections.

4.1.1 Isoperimetric inequalities

Recall that we identify subsets of $\mathbb{Z}^2$ or $[s]^2$ with the subgraphs of the infinite integer grid induced by them.

Lemma 15 *Suppose $H \subseteq \mathbb{Z}^2$ is a finite induced subgraph of the integer lattice. Then*

$$e(H, H^c) \geq 4\sqrt{v(H)}.$$

Proof: Let H be an arbitrary finite induced subgraph of $\mathbb{Z}^2$, let H_x denote the projection of H on the x-axis, and let H_y denote the projection on the y-axis. Let us write $\ell_x := |H_x|, \ell_y = |H_y|$. Note that every vertical line that intersects H contributes at least two vertical edges to $e(H, H^c)$, and that analogously every horizontal line that intersects H contributes at least two horizontal edges. Thus

$$e(H, H^c) \geq 2\ell_x + 2\ell_y.$$

On the other hand, it is clear that

$$v(H) \leq \ell_x \cdot \ell_y.$$

Thus we obtain with straightforward calculus that

$$v(H) \leq \max_{0 \leq z \leq \frac{e(H,H^c)}{2}} z \cdot \left(\frac{e(H, H^c)}{2} - z\right) = \left(\frac{e(H, H^c)}{4}\right)^2, \tag{4}$$

which is equivalent to the claim. ■

We will need the following strengthening of Lemma 15.

Lemma 16 *Let $x > 0$ be given, and suppose $H \subseteq \mathbb{Z}^2$ is a finite induced subgraph of the integer lattice such that all connected components of H have at most x vertices. Then*

$$e(H, H^c) \geq 4\frac{v(H)}{\sqrt{x}}.$$

Proof: Let $\mathcal{C}$ denote the set of components of H. By Lemma 15, each component C satisfies $\sqrt{v(C)} \leq \frac{e(C,C^c)}{4}$. Hence

$$v(H) = \sum_{C \in \mathcal{C}} v(C) \leq \sqrt{x} \sum_{C \in \mathcal{C}} \sqrt{v(C)} \leq \sqrt{x} \sum_{C \in \mathcal{C}} \frac{e(C, C^c)}{4} = \sqrt{x} \cdot \frac{e(H, H^c)}{4},$$

which is equivalent to the claim. ■

Lemma 17 *Let $\alpha, \beta > 0$ and $s \in \mathbb{N}$ be given. Suppose $H \subseteq [s]^2$ is an induced subgraph of the $s \times s$ grid with $v(H) \geq s^2 - \alpha s$. Moreover, let $H_\beta \subseteq H$ denote*

$$H_\beta := \bigcup \{C \subseteq H : C \text{ is a component of } H, \text{ and } v(C) \leq \beta s^2\},$$

i.e. H_β is the union of of all components of H with at most βs^2 vertices. Then $v(H_\beta) \leq \sqrt{\beta} \cdot (1+\alpha) \cdot s^2$.

Proof: Let H^c denote the complement of H in $\mathbb{Z}^2$ (not $[s]^2$), and observe that every edge of $e(H, H^c)$ connects a vertex of H either to a vertex of $A := (\{0, s+1\} \times [s]) \cup ([s] \times \{0, s+1\})$ or to one of the at most αs vertices of $B := [s]^2 \setminus H$. Observe also that every vertex of A can be adjacent to at most one vertex of H, while a vertex of B can be adjacent to at most 4 vertices of H. Hence we have

$$e(H, H^c) \leq |A| + 4|B| \leq 4(1+\alpha)s. \tag{5}$$

The claim follows by applying Lemma 16 (in the form $v(H) \leq \sqrt{x} \cdot e(H, H^c)/4$) to H_β. ∎

Lemma 18 *Let $\alpha > 0$ and $k \in \mathbb{N}$ be fixed. Then the following holds as $s \to \infty$. If $H \subseteq [s]^2$ is an induced subgraph with $v(H) \geq s^2 - \alpha s$ vertices, and $C_1, \ldots, C_k \subseteq H$ denote the k largest components of H (ties broken arbitrarily) then*

$$v(C_1) + \cdots + v(C_k) \geq (1 - o_s(1)) \cdot \lambda_k \cdot s^2,$$

where $\lambda_k = \lambda_k(\alpha)$ is given by

$$\lambda_1 = \left(\frac{1}{1+\alpha}\right)^2 \quad \text{and} \quad \lambda_{k+1} = \lambda_k + \left(\frac{1-\lambda_k}{1+\alpha}\right)^2 \text{ for } k \geq 1. \tag{6}$$

Proof: Let us first point out that $0 < \lambda_k < 1$ for all k, as can easily be seen from the definition. The proof is by induction on k. We start with the base case, $k = 1$. Set $\beta := (1-\varepsilon) \cdot \lambda_1 = (1-\varepsilon) \cdot \left(\frac{1}{1+\alpha}\right)^2$, with $0 < \varepsilon < 1$ arbitrary but fixed. By Lemma 17, the union of all components of order at most βs^2 contains no more than $\sqrt{1-\varepsilon} \cdot s^2$ vertices. Since the union of *all* components must clearly have $s^2 - \alpha s = (1 - o_s(1))s^2$ vertices, there must exist a component of order $> \beta s^2$. As $\varepsilon > 0$ can be chosen arbitrarily small, it follows that $v(C_1) \geq (1 - o_s(1))\lambda_1 s^2$, which establishes the base case.

Now suppose that $v(C_1) + \cdots + v(C_k) = \tilde{\lambda} s^2$ with $\tilde{\lambda} \geq (1 - o_s(1))\lambda_k$. If $\tilde{\lambda} > \lambda_{k+1}$ then we are done, so we can assume this is not the case. Aiming for a contradiction, suppose that $v(C_{k+1}) < \beta s^2$, where $\beta = (1-\varepsilon)\left(\frac{1-\tilde{\lambda}}{1+\alpha}\right)^2$ for some fixed $\varepsilon > 0$. Lemma 17 would then give that

$$v(H \setminus (C_1 \cup \cdots \cup C_k)) \leq \sqrt{\beta}(1+\alpha)s^2 = \sqrt{1-\varepsilon} \cdot (1 - \tilde{\lambda})s^2,$$

which is impossible as we must have

$$\begin{aligned} (1 - o_s(1))s^2 &= v(C_1 \cup \cdots \cup C_k) + v(H \setminus (C_1 \cup \cdots \cup C_k)) \\ &= \tilde{\lambda} s^2 + v(H \setminus (C_1 \cup \cdots \cup C_k)). \end{aligned}$$

It follows that $v(C_{k+1}) \geq (1-o_s(1))\left(\frac{1-\tilde{\lambda}}{1+\alpha}\right)^2$, so that

$$v(C_1)+\cdots+v(C_{k+1}) \geq (1-o_s(1))\cdot\left(\tilde{\lambda}+\left(\frac{1-\tilde{\lambda}}{1+\alpha}\right)^2\right).$$

By differentiating $f(x) := x+\left(\frac{1-x}{1+\alpha}\right)^2$ with respect to x it is easily seen that f is strictly increasing in x for $x \geq \lambda_1 = \left(\frac{1}{1+\alpha}\right)^2$. Since $\tilde{\lambda} \geq (1-o_s(1))\lambda_k$ by the inductive hypothesis, and $\lambda_k > \lambda_{k-1} > \cdots > \lambda_1$, it now follows that

$$\begin{aligned} v(C_1)+\cdots+v(C_{k+1}) &\geq (1-o_s(1))\cdot\left(\lambda_k+\left(\tfrac{1-\lambda_k}{1+\alpha}\right)^2\right)\cdot s^2 \\ &= (1-o_s(1))\cdot\lambda_{k+1}\cdot s^2, \end{aligned}$$

as required. ■

Corollary 19 *Fix $0<\varepsilon<1, \alpha>0$. Then there exists $k=k(\varepsilon,\alpha)$ such that the following holds for all large enough s. If $H \subseteq [s]^2$ is an induced subgraph with $v(H) \geq s^2-\alpha s$ then*

$$v(C_1)+\cdots+v(C_k) \geq (1-\varepsilon)s^2,$$

where C_i denotes the i-th largest component (ties broken arbitrarily). Moreover, for all ε we have $k(\varepsilon,\alpha)=1$ if $\alpha \leq \varepsilon/2$.

Proof: Let the numbers $\lambda_k=\lambda_k(\alpha)$ be as defined by (6).

The "moreover" part of Corollary 19 follows immediately from Lemma 18 since $\lambda_1 = \left(\frac{1}{1+\alpha}\right)^2 > 1-\varepsilon$ if $\alpha \leq \varepsilon/2$.

To see that the rest of the corollary also holds, notice that the numbers λ_k form an increasing sequence that is bounded above by one, and that the limit of the sequence must be a fixed point of the equation

$$\lambda=\lambda+\left(\frac{1-\lambda}{1+\alpha}\right)^2.$$

Since the only fixed point is $\lambda=1$, we must have $\lim_{k\to\infty}\lambda_k=1$. Hence there is a $k=k(\varepsilon,\alpha)$ such that $\lambda_k>1-\varepsilon$. The lemma follows. ■

4.1.2 Balls and bins

In our proof of part **(ii)** of Theorem 1, we will need a minor extension of the lower bound part of Theorem 3 that concerns the scenario where there are slightly fewer balls than bins. This case does not seem to have been treated explicitly in the literature. The proof is very similar to the original lower bound proof given in [3]; we include it here for completeness.

Lemma 20 *Let $\varepsilon>0$ be arbitrary, but fixed. Consider the power of two choices balls and bins process, with n bins and m rounds where $n/\ln n \leq m \leq n$, and let M_n denote the maximum load in round m. No matter what strategy the player utilizes, we have*

$$\mathbb{P}(M_n < (1-\varepsilon)\log\log n) \leq \exp[-n^{1+o(1)}].$$

Proof: By obvious monotonicity properties, it suffices to prove the lemma for the case when m, the number of rounds, is exactly equal to $n/\ln n$. Let us thus assume that $m = n/\ln n$. Our approach for the proof will be to bound from below, for each i, the number of rounds in which the player *is forced to* create a bin with i balls in it.

We denote by $N_i(t)$ the number of bins with *at least* i balls in them after round t. (Note that $N_0(t) = n$ for all t.) Furthermore, we set

$$k := \lceil (1-\varepsilon)\log\log n \rceil,$$

and write $\underline{N}(t) = (N_0(t), \ldots, N_k(t))$. Let

$$\alpha := \left(\frac{m}{4 \cdot k \cdot n}\right), \qquad t_i := m \cdot (i/k) \quad \text{for } i = 1, \ldots, k,$$

and define c_i by

$$c_i := \alpha^{2^i - 1}, \qquad i = 1, \ldots, k.$$

Observe that the c_i satisfy the recurrence relation

$$c_{i+1} = \alpha \cdot c_i^2. \tag{7}$$

Before proceeding, let us make some further observations about the c_i. Note that $\alpha = (1+\varepsilon)/(4 \log\log n \cdot \log n)$, so certainly $\alpha < 1$ and the c_i are thus decreasing. Moreover, we have

$$0 > \ln \alpha \geq -(1 + o(1)) \ln\ln n. \tag{8}$$

Also note that

$$\begin{aligned} c_k &= \alpha^{2^k - 1} \\ &\geq \alpha^{(\log n)^{1-\varepsilon} - 1} \\ &= \exp\left[\ln \alpha \cdot ((\log n)^{1-\varepsilon} - 1)\right] \\ &\geq \exp\left[-(1 + o(1)) \cdot \ln\ln n \cdot (\log n)^{1-\varepsilon}\right] \\ &= \exp[-o(\ln n)] \\ &= n^{-o(1)}, \end{aligned} \tag{9}$$

where we have used (8) in the fourth line.

Another key observation is that if in some round both balls fall in bins with *exactly* $i-1$ balls in them then a new bin with i balls in it will be created, regardless of the strategy of the player. In other words, for all i, t we have

$$\mathbb{P}(N_i(t+1) = N_i(t) + 1 | \underline{N}(t)) \geq \left(\frac{N_{i-1}(t) - N_i(t)}{n}\right)^2. \tag{10}$$

If for some t we have $N_1(t) < c_1 n$, then the observation (10) shows that

$$\mathbb{P}(N_1(t+1) = N_1(t) + 1 | N_1(t) < c_1 n) \geq (1 - c_1)^2 \geq \frac{1}{2}.$$

It follows that

$$\begin{aligned} \mathbb{P}(N_1(t_1) < c_1 n) &\leq \mathbb{P}(\mathrm{Bi}(t_1, \tfrac{1}{2}) < c_1 n) \\ &\leq \exp[-(t_1/2) \cdot H\left(\tfrac{c_1 n}{t_1/2}\right)] \\ &= \exp[-\tfrac{m}{2k} \cdot H(\tfrac{1}{2})] \\ &= \exp[-n^{1-o(1)}], \end{aligned} \tag{11}$$

where we used the Chernoff bound (Lemma 4), together with the facts that $t_1 = m/k = n^{1-o(1)}$ and $c_1 = \alpha = m/(4kn)$.

Let E_i denote the event

$$E_i := \{N_j(t_j) \geq c_j n \text{ for all } 1 \leq j \leq i\}.$$

Again using the observation (10) we see that, for all $t \geq t_i$:

$$\mathbb{P}(N_{i+1}(t+1) = N_{i+1}(t) + 1 | E_i, N_{i+1}(t) < c_{i+1}n) \geq (c_i - c_{i+1})^2 \geq \frac{1}{2}c_i^2.$$

(Here we use that $N_i(t) \geq N_i(t_i)$ for $t \geq t_i$ by obvious monotonicity.) It thus follows that

$$\begin{array}{rcl} \mathbb{P}(N_{i+1}(t_{i+1}) < c_{i+1}n | E_i) & \leq & \mathbb{P}(\mathrm{Bi}(t_{i+1} - t_i, c_i^2/2) < c_{i+1}n) \\ & \leq & \exp[-(m/k) \cdot (c_i^2/2) \cdot H(\frac{c_{i+1}n}{(m/k)\cdot(c_i^2/2)})] \\ & = & \exp[-(m/k) \cdot (c_i^2/2) \cdot H(\frac{1}{2})] \\ & = & \exp[-n^{1-o(1)}], \end{array} \tag{12}$$

where we used that $c_{i+1} = \alpha c_i^2 = (m/4kn) \cdot c_i^2$ in the third line, and that $c_i \geq c_k = n^{-o(1)}$ by (9) in the fourth line. It follows that

$$\begin{array}{rcl} \mathbb{P}(N_k(t_k) < c_k n) & \leq & \mathbb{P}(\text{there is some } 1 \leq i \leq k \text{ such that } N_i(t_i) < c_i n) \\ & = & \mathbb{P}(E_1^c) + \mathbb{P}(E_2^c | E_1)\mathbb{P}(E_1) + \cdots + \mathbb{P}(E_k^c | E_{k-1})\mathbb{P}(E_{k-1}) \\ & \leq & \mathbb{P}(E_1^c) + \mathbb{P}(E_2^c | E_1) + \cdots + \mathbb{P}(E_k^c | E_{k-1}) \\ & = & \mathbb{P}(N_1(t_1) < c_1 n) + \sum_{i=1}^{k-1} \mathbb{P}(N_{i+1}(t_{i+1}) < c_{i+1}n | E_i) \\ & = & k \cdot \exp[-n^{1-o(1)}] \\ & = & \exp[-n^{1-o(1)}], \end{array}$$

where we used (11) and (12) to get the fifth line. This concludes the proof of the lemma. ■

4.1.3 The Crazy Coupon Collector

In the classical coupon collector problem, each box of some product, say cereals, contains one of N types of coupons, sampled uniformly at random. There is a collector who keeps buying new boxes until he has collected at least one coupon of each type.

For our purposes it is useful to consider a variant of the coupon collector problem that we call the *crazy coupon collector*. Suppose again that there are N types of coupons, but now each box of cereals contains *two* coupons (whose types are chosen independently and uniformly at random from all N possible types). There is a crazy coupon collector (CCC, henceforth) who enjoys the process of collecting the coupons more than he does having a complete collection. Therefore, everytime he opens a box of cereals, he only adds a coupon to his collection if *both* coupons in the box are of types he does not have in his collection yet. Even in that case, he only adds one of the two coupons to his collection.

Lemma 21 *For any $s = s(N)$ that with $1 \ll s \ll N$ the following is true: w.h.p. the CCC needs to buy at most $2N^2/s$ boxes to collect all but s coupons.*

Proof: Let T denote the number of rounds it takes the CCC to collect exactly $N - s$ coupons. We need to show that

$$\mathbb{P}(T > 2N^2/s) = o(1). \tag{13}$$

Observe that T is a sum of independent geometrically distributed random variables. More precisely,

$$T = Z_1 + \cdots + Z_{N-s},$$

where $Z_i \stackrel{d}{=} \mathrm{Geom}(p_i)$ with $p_i := \left(\frac{N-i+1}{N}\right)^2$. Thus

$$\begin{aligned}
\mathbb{E}T &= \sum_{i=1}^{N-s} \frac{1}{p_i} = \sum_{i=1}^{N-s} \left(\frac{N}{N-i+1}\right)^2 \\
&= N^2 \sum_{j=s+1}^{N} j^{-2} \\
&= N^2 \cdot (1+o(1)) \int_s^\infty x^{-2}\mathrm{d}x \\
&= (1+o(1))\frac{N^2}{s},
\end{aligned}$$

where the integral approximation holds due to our assumptions on s. Similarly we have

$$\begin{aligned}
\operatorname{Var} T &= \sum_{i=1}^{N-s} \frac{1-p_i}{p_i^2} \le \sum_{i=1}^{N-s} \frac{1}{p_i^2} \\
&= \sum_{i=1}^{N-s} \left(\frac{N}{N-i+1}\right)^4 \\
&= N^4 \sum_{j=s+1}^{N} j^{-4} \\
&= N^4 \cdot (1+o(1)) \int_s^\infty x^{-4}\mathrm{d}x \\
&= (1+o(1))\frac{N^4}{3s^3}.
\end{aligned} \tag{14}$$

It follows with Chebyshev's inequality that

$$\mathbb{P}(T > 2N^2/s) \le \mathbb{P}(T > 1.5\,\mathbb{E}T) \le \frac{\operatorname{Var} T}{(0.5\,\mathbb{E}T)^2} = O(1/s) = o(1),$$

as desired.

■

4.2 Proof of part (ii) of Theorem 1

We now are ready to give the main argument for our upper bound on the online threshold. Throughout this section we will consider the boxes of the dissection $\mathcal{D}_\varrho$ as defined in (1), where $\varrho := r/\sqrt{5}$. This time however, we treat these boxes as vertices of the ordinary grid graph $[(1/\varrho)]^2$. (Again we assume $(1/\varrho)$ is an integer.) We will denote by $\mathrm{Occ}_\varrho(t)$ the subgraph induced by the occupied boxes after t rounds.

Note that if two points fall anywhere inside two adjacent boxes of $\mathcal{D}_\varrho$, by our choice of ϱ they are within distance r from each other. Hence, if a set of occupied boxes induces a connected component of $\mathrm{Occ}_\varrho(t)$, then all points inside these boxes belong to the same connected component of the geometric graph.

Lemma 22 *After $n/2$ rounds, w.h.p. all but $\left(\frac{100}{c}\right) \cdot \log\log n \cdot (1/\varrho)$ boxes are occupied, no matter how the player plays.*

Proof: Note that to minimize the number of occupied boxes, the player should play exactly like the crazy coupon collector from Section 4.1.3, with the $N := (1/\varrho)^2 = n^{2/3+o(1)}$ boxes playing the role of the coupons. Let $s := \left(\frac{100}{c}\right) \cdot \log\log n \cdot (1/\varrho)$, and note that $s = n^{1/3+o(1)} = o(N)$. It follows with Lemma 21 that w.h.p. after at most

$$\frac{2N^2}{s} = \frac{2c}{100 \cdot \varrho^3 \log\log n} = \frac{c \cdot 5^{3/2}}{50 \cdot r^3 \log\log n} = \frac{5^{3/2}}{50} \cdot n < n/2$$

rounds, all but s many boxes are occupied. ■

Let us now fix $0 < \varepsilon < 1/7$, to be determined later. We dissect the square into *megablocks* consisting of $b \times b$ boxes, where

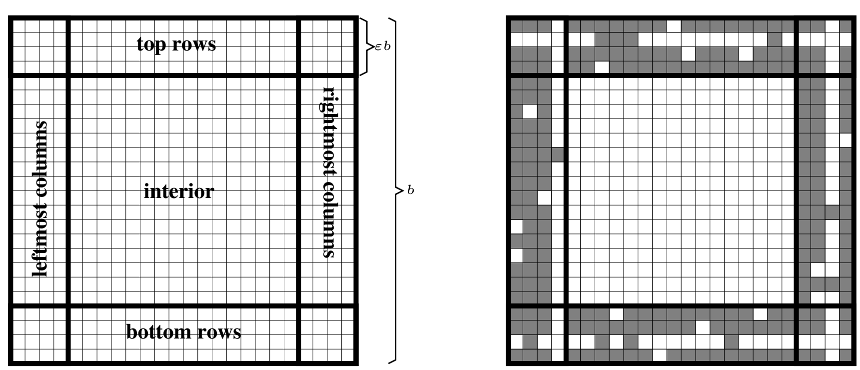
$$b := \left(\frac{\varepsilon \cdot c}{1000}\right) \cdot \frac{(1/\varrho)}{\log\log n}. \tag{15}$$

(Again we assume for convenience that $(1/\varrho)$ and $(1/b\varrho)$ are integers.) Thus, there are z^2 megablocks where

$$z = (1/b\varrho) = \left(\frac{1000}{\varepsilon \cdot c}\right) \cdot \log\log n. \tag{16}$$

It is convenient to consider the megablocks as vertices of the (ordinary) grid $[z]^2$. So a megablock is adjacent to megablocks that share a side with it, but not with megablocks that share only a corner with it.

We shall refer to the top εb rows of a megablock B simply as the *top rows* of B. Similarly, we call the bottom εb rows the bottom rows, the leftmost εb columns the leftmost columns and the rightmost εb the rightmost columns. Those boxes of a megablock B that belong to neither the top or bottom rows nor to the leftmost or rightmost columns will be called the *interior* of B. See Figure 3(a) for a depiction.



(a) A megablock with its top, bottom, leftmost and rightmost columns labelled

(b) A good megablock. Three of the four top rows have only few empty boxes, and similarly for the bottom rows, left columns and right columns.

Figure 3: Two megablocks.

Let us call a row of a megablock *good* if no more than $\frac{1}{3}\log\log n$ of its boxes are empty, and similarly we call a column of a megablock good if no more than $\frac{1}{3}\log\log n$ of its boxes are empty.

We call a megablock *good* if at least three quarters of the top rows are good, at least three quarters of the bottom rows are good, at least three quarters of the leftmost columns are good, and at least three quarters of the rightmost columns are good. See Figure 3(b) for a depiction.

If a megablock is not good we will call it *bad*. Let us denote by $\mathrm{Mega}_\varrho(t)$ the subgraph of the grid of megablocks (recall that we treat it like the ordinary $z \times z$ grid) induced by the megablocks that are good in round t.

Lemma 23 *W.h.p. in round $n/2$ at most $\left(\frac{10^5}{\varepsilon c}\right) \cdot z$ megablocks are bad, no matter what the player does.*

Proof: By Lemma 22 we can assume that in round $n/2$ at most $\left(\frac{100}{c}\right) \cdot \log\log n \cdot (1/\varrho)$ boxes are empty. Each bad megablock contains at least

$$\frac{1}{4} \cdot \varepsilon \cdot b \cdot \frac{1}{3}\log\log n = \left(\frac{\varepsilon^2 c}{12000}\right)(1/\varrho),$$

empty boxes, because a quarter of either the top rows or the bottom rows or the leftmost columns or the rightmost columns is not good. Hence, the number of bad megablocks can not be larger than

$$\frac{\left(\frac{100}{c}\right) \cdot \log\log n \cdot (1/\varrho)}{\left(\frac{\varepsilon^2 c}{12000}\right)(1/\varrho)} < \left(\frac{10^8}{\varepsilon^2 c^2}\right)\log\log n = \left(\frac{10^5}{\varepsilon c}\right) \cdot z.$$

■

We shall also need the following consequence of Lemma 22:

Corollary 24 *W.h.p. in round $n/2$ at least $(1-\varepsilon)(1/\varrho)^2$ boxes are contained in components of $\mathrm{Occ}_\varrho(n/2)$ of order strictly larger than b^2, no matter what the player does.*

Proof: By Lemma 22 we can assume that in round $n/2$ there are at most $\left(\frac{100}{c}\right)\log\log n(1/\varrho)$ empty boxes. Set

$$s := (1/\varrho), \quad \alpha := \left(\frac{100}{c}\right) \cdot \log\log n, \quad \beta := (b/s)^2 = \left(\frac{\varepsilon \cdot c}{1000\log\log n}\right)^2.$$

By Lemma 17, w.h.p. the number of occupied boxes in components of $\mathrm{Occ}_\varrho(n/2)$ of order at most b^2 is at most $\sqrt{\beta} \cdot (1+\alpha) \cdot s^2 \leq \sqrt{\beta} \cdot 2\alpha \cdot s^2 = \varepsilon/5 \cdot (1/\varrho)^2$. As only an $o(1)$-fraction of the boxes is empty, the claim follows. ■

Let us say that a row or a column of a megablock is *full* is it contains no empty boxes. We will us say that a megablock is *framed* if among the top rows there is one that is full, among the bottom rows there is one that is full, among the leftmost columns there is one that is full and among the righmost columns there is one that is full. If a megablock B is framed then we refer to the union of the full rows among the top rows, bottom rows, leftmost columns and rightmost columns as the *frame* of B. The choice of the name should be clear from the depiction in Figure 4.

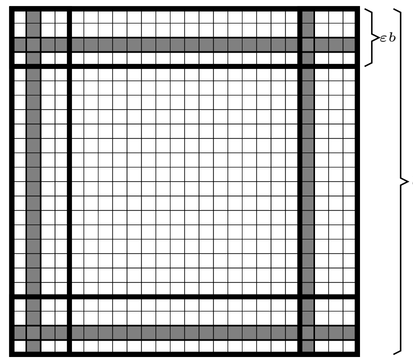


Figure 4: A framed megablock.

Lemma 25 *W.h.p. every megablock that is good in round $n/2$ is framed in round n, no matter what the player does.*

Proof: We will first compute the probability that a given megablock contains a full row among the top εb rows in round n, given that it was good in round $n/2$.

Let us thus fix a megablock B, condition on it being good in round $n/2$, and consider what happens to it in the rounds $t > n/2$. In round $n/2$, at least $M := \frac{3}{4}\varepsilon b$ of the top rows of B are good. Let us fix exactly M of these good rows $r_1, \ldots, r_M$.

We now consider the following balls and bins type process for the remaining $n/2$ rounds. In each round $n/2 < t \leq n$, as long as none of the rows $r_1, \ldots, r_M$ is full, we have a list $\underline{e}(t) = (e_1(t), \ldots, e_M(t))$ of empty boxes, where $e_i(t) \in r_i$. If at some round t the player plays in some box in our list, then we replace it with another box in the same row that is still empty (as long as this is possible; if it is not possible then evidently a full row has been created). Otherwise we keep the list the same. In other words, if in round t the player picks a point in $e_i(t)$ then we set $e_j(t+1) = e_j(t)$ for all $j \neq i$ and $e_i(t+1)$ is set to some box in row i that is still empty; and if the player does not play in any box of $\underline{e}(t)$ then we set $\underline{e}(t+1) = \underline{e}(t)$.

We want to compute the probability that there is some index i such that the player is forced to play more than $\frac{1}{3}\log\log n$ times in $e_i(t)$. Let R denote the number of rounds $n/2 < t \leq n$ in which both points fall into $e_1(t) \cup \cdots \cup e_M(t)$ – so that the player is forced to play in one of the e_is. Then

$$R \stackrel{d}{=} \mathrm{Bi}(n/2, (M \cdot \varrho^2)^2).$$

Observe that

$$\mathbb{E}R = \Theta(nM^2\varrho^4) = \Theta(n \cdot \varrho^2/(\log\log n)^2) = n^{\frac{1}{3}-o(1)}.$$

Hence, the Chernoff bound (Lemma 4) yields

$$\mathbb{P}(R < \mathbb{E}R/2) \leq e^{-n^{\frac{1}{3}-o(1)}}.$$

Set $N := \frac{1}{2}\mathbb{E}R$. If we condition on $R > N$, the probability that the player can achieve a situation where none of $r_1, \ldots, r_M$ is full by round n is upper bounded by the probability that in the player version of the two choices balls and bins process with N rounds and M bins, the player can achieve a maximum load of less than $\frac{1}{3}\log\log n$. Observe that

$$M = \frac{3}{4}\varepsilon b = \Theta((1/\varrho)/\log\log n) = n^{\frac{1}{3}-o(1)}.$$

Thus, we have

$$\log\log n = (1+o(1))\log\log M.$$

Also observe that

$$\begin{aligned} \tfrac{N}{M} &= \Theta(nM\varrho^4) \\ &= \Theta((n\varrho^3)M\varrho) \\ &= \Theta((\tfrac{c}{\log\log n})\cdot(\tfrac{1}{\log\log n})) \\ &= \Theta((\log\log M)^{-2}). \end{aligned}$$

Hence, by Lemma 20 we have:

$$\begin{aligned} & \mathbb{P}(\text{ none of } r_1,\ldots,r_M \text{ is full in round } n \mid r_1,\ldots,r_M \text{ were good in round } n/2) \\ \leq\ & \mathbb{P}(R < \tfrac{1}{2}\mathbb{E}R) + \exp[-M^{1-o(1)}] \\ \leq\ & e^{-n^{\frac{1}{3}-o(1)}} + e^{-n^{\frac{1}{3}-o(1)}} \\ =\ & e^{-n^{\frac{1}{3}-o(1)}}. \end{aligned}$$

The same argument and computations apply to the bottom εb rows, the leftmost εb columns and the rightmost εb columns. Since there are $z^2 = O\left((\log\log n)^2\right)$ megablocks in total, the union bound gives us

$$\begin{aligned} & \mathbb{P}(\text{There is a megablock which is good in round } n/2 \text{ and not framed in round } n) \\ \leq\ & z^2 \cdot 4 \cdot \exp[-n^{\frac{1}{3}-o(1)}] \\ =\ & o(1), \end{aligned}$$

as required. ■

Let us say that two megablocks B_1, B_2 that share a vertical side are *skewered* if there is a row that is full in both B_1 and B_2. Similarly we say that two megablocks B_1, B_2 that share a horizontal side are skewered if there is a common column that is full in both. See Figure 5 for a depiction.

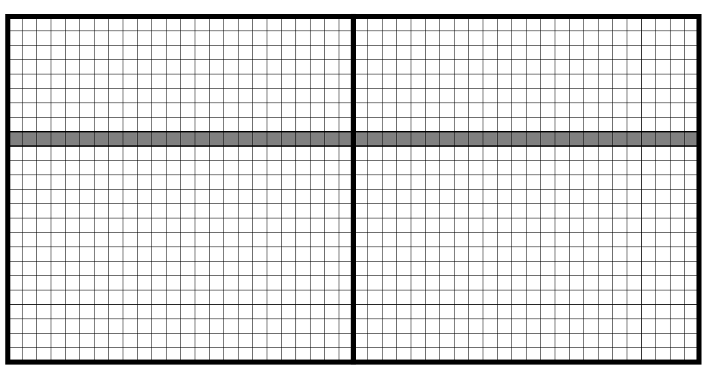


Figure 5: Two adjacent, skewered megablocks.

Lemma 26 *W.h.p. every two adjacent megablocks that are both good in round $n/2$ are skewered in round n, no matter what the player does.*

Proof: Let B_1, B_2 be two adjacent megablocks (without loss of generality we can assume they share a vertical side). If we condition on both being good in round $n/2$ then there must be at least $M = 2 \cdot \frac{1}{2}\varepsilon b = \varepsilon b$ rows that are good in both B_1, B_2. A row that is good in both has at most $\frac{2}{3}\log\log n$ empty boxes. We can thus follow the same reasoning as in the proof of Lemma 25 and the same computations with only very minor adaptations to prove that, with probability $1 - \exp[-n^{\frac{1}{3}-o(1)}]$ at least one of these rows will be full in both B_1 and B_2. Since there are only $O(z) = O(\log\log n)$ pairs of adjacent megablocks, the union bound again finishes the the proof. ■

Observe that, if two adjacent megablocks B_1, B_2 are both framed and if they are also skewered, then their frames will belong to the same component of Occ_ϱ. Hence, Lemma's 25 and 26 together immediately imply the following.

Corollary 27 *W.h.p. the folllowing holds, no matter what the player does. If $\mathcal{C}$ is a connected component of $\mathrm{Mega}_\varrho(n/2)$, then every megablock of $B \in \mathcal{C}$ will be framed in round n and the frames will all belong to the same component of the boxes graph $\mathrm{Occ}_\varrho(n)$.* ■

Lemma 28 *W.h.p., no matter what the player does, in round n there will be a component $\mathfrak{C}$ of $\mathrm{Occ}_\varrho(n)$ that contains at least $v(\mathfrak{C}) \geq a(\varepsilon, c) \cdot (1/\varrho)^2$ boxes, where*

$$a(\varepsilon, c) = \frac{1 - 7\varepsilon}{k\left(\varepsilon, \frac{10^5}{\varepsilon c}\right)}, \tag{17}$$

with $k(.,.)$ as provided by Corollary 19.

Proof: Let $A_1 \subseteq [0,1]^2$ denote the union of all boxes that belong to a component of $\mathrm{Occ}_\varrho(n/2)$ of order larger than b^2 (in round $n/2$). By Corollary 24, w.h.p., we have

$$\mathrm{area}(A_1) \geq 1 - \varepsilon.$$

Let A_2 denote the intersection of A_1 with the union of all boxes that lie in the interior of some good megablock $\in \mathrm{Mega}_\varrho(n/2)$. (The reason for these definitions will become clear later.) Then

$$\begin{aligned} \mathrm{area}(A_2) &\geq \mathrm{area}(A_1) - 4\varepsilon(b\varrho)^2 N_{\mathrm{good}} - (b\varrho)^2 N_{\mathrm{bad}} \\ &\geq \mathrm{area}(A_1) - 4\varepsilon - (b\varrho)^2 N_{\mathrm{bad}}, \end{aligned}$$

where $N_{\mathrm{good}}, N_{\mathrm{bad}}$ denote the number of good resp. bad megablocks in round $n/2$. Trivially we have $N_{\mathrm{good}} \leq z^2 = (1/b\varrho)^2$. Moreover, by Lemma 23 we have that $N_{\mathrm{bad}} = O(z) = o(z^2)$ w.h.p. Hence, w.h.p. it holds that

$$\mathrm{area}(A_2) \geq \mathrm{area}(A_1) - 5\varepsilon \geq 1 - 6\varepsilon.$$

Recall that, by Lemma 23, $\mathrm{Mega}_\varrho(n/2)$ contains at least $z^2 - \alpha z$ megablocks, where $\alpha := 10^5/(\varepsilon c)$. By Corollary 19 the $k = k(\varepsilon, \alpha)$ largest components of $\mathrm{Mega}_\varrho(n/2)$ together cover a fraction of at least $1 - \varepsilon$ of the unit square. Let A_3 denote the intersection of A_2 with the megablocks belonging to the k largest components of $\mathrm{Mega}_\varrho(n/2)$. Then

$$\mathrm{area}(A_3) \geq \mathrm{area}(A_2) - \varepsilon \geq 1 - 7\varepsilon.$$

Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ denote the k largest components of $\mathrm{Mega}_\varrho(n/2)$. Denoting by B_i the union of the megablocks of $\mathcal{C}_i$ for each $i = 1, \ldots, n$, there is an index $1 \leq i \leq k$ such that

$$\mathrm{area}(A_3 \cap B_i) \geq \mathrm{area}(A_3)/k \geq (1 - 7\varepsilon)/k.$$

Now recall that A_3 is a union of boxes that belong to components of $\mathrm{Occ}_\varrho(n/2)$ of order at least b^2, and that all these boxes belong to the interior of a megablock belonging to $\mathcal{C}_i$. Observe that, if $B \in \mathrm{Mega}_\varrho(n)$ is a framed megablock, and $\mathcal{C} \subseteq \mathrm{Occ}_\varrho(n)$ is a component with more than b^2 boxes that intersects the interior of B, then $\mathcal{C}$ also intersects the frame of B (as a megablock consists of exactly b^2 boxes).

By Corollary 27, we can assume that the frames of the megablocks in $\mathcal{C}_i$ all belong to the same component of the box graph. Hence, all boxes of $A_3 \cap B_i$ belong to the same component of $\mathrm{Occ}_\varrho(n)$. Consequently, $\mathrm{Occ}_\varrho(n)$ has a component consisting of at least $((1-7\varepsilon)/k) \cdot (1/\varrho)^2$ boxes, as required. ■

To transfer the result back to the original random geometric graph setting and conclude the proof, we need the next lemma, which is similar in spirit to Lemma 9.

Lemma 29 *For every $\varepsilon > 0$ the following holds w.h.p. Every $A \subseteq [0,1]^2$ that is the union of boxes of the dissection $\mathcal{D}_\varrho$ and with $\mathrm{area}(A) \geq \varepsilon$ contains at least $(1-\varepsilon)\cdot \mathrm{area}^2(A) \cdot n$ points in round n, no matter what the player does.*

Proof: Let $\mathcal{A}$ denote the set of all sets $A \subseteq [0,1]^2$ under consideration. Since every $A \in \mathcal{A}$ is a union of boxes, we have $|\mathcal{A}| \leq 2^{(1/\varrho)^2} = 2^{n^{2/3+o(1)}}$.

Fix a set $A \in \mathcal{A}$, and let Z the number of rounds in which the player cannot avoid playing in A because both points fall inside it. Then $Z \overset{d}{=} \mathrm{Bi}(n, \mathrm{area}^2(A))$. By the Chernoff bound (Lemma 4) we have

$$\mathbb{P}(Z < (1-\varepsilon)\cdot \mathrm{area}^2(A) \cdot n) \leq \exp\left[-n \cdot \mathrm{area}^2(A) \cdot H(1-\varepsilon)\right] = \exp[-\Omega(n)].$$

This holds for every set A under consideration. Hence, by the union bound we have

$$\begin{aligned} & \mathbb{P}(\text{ there is a set } A \in \mathcal{A} \text{ that receives less than } (1-\varepsilon)\cdot \mathrm{area}^2(A)\cdot n \text{ points }) \\ \leq\ & 2^{n^{\frac{2}{3}+o(1)}} \cdot \exp[-\Omega(n)] \\ =\ & \exp[n^{\frac{2}{3}+o(1)} - \Omega(n)] \\ =\ & o(1), \end{aligned}$$

which gives the lemma. ■

With Lemmas 21–29 in hand, it is easy to prove part **(ii)** of Theorem 1:

Proof of part (ii) of Theorem 1: For given c, set $\varepsilon := \sqrt{2 \cdot 10^5/c} > 0$ if $c \geq 10^8$, and $\varepsilon = 0.01$ otherwise. Let $\widetilde{a}(c) := a(\varepsilon, c)$ for $a(\varepsilon, c)$ as defined in (17). Note that in both cases $0 < \widetilde{a}(c) < 1$. Further, by the "moreover" part of Corollary 19, for $c \geq 10^8$ we have $\widetilde{a}(c) = 1 - 7\varepsilon = 1 - O(1/\sqrt{c})$. Hence we have $\widetilde{a}(c) \to 1$ as $c \to \infty$.

By Lemma 28, w.h.p. the boxes graph $\mathrm{Occ}_\varrho(n)$ will have a connected component of area at least $\widetilde{a}(c)$. By Lemma 29 this will give us a component of order at least $(1-o(1))\cdot \widetilde{a}^2(c) \cdot n$ in the resulting geometric graph. Hence the claim follows for, say, $g(c) := \widetilde{a}^3(c)$. Note that $g(c) \to 1$ as $c \to \infty$. ■

4.3 Proof of part (ii) of Theorem 2

As in the previous proof we divide the unit square into boxes of sidelength $\varrho := r/\sqrt{5}$. We set $s := 1/\varrho$ as before and assume, also as before, that s is an integer. Again we denote by Occ_ϱ the subgraph of the $s \times s$ grid $[s]^2$ induced by the occupied boxes.

For given $c > 0$, define $a = a(c)$ as the solution of

$$\frac{480\sqrt{a}}{(1-\sqrt{a})^2} = c. \tag{18}$$

Note that $0 < a(c) < 1$ with $a(c) \to 1$ as $c \to \infty$.

We will show:

Claim 30 *W.h.p. every possible choice of points is such that* Occ_ϱ *has a component with more than* $a(c) \cdot s^2$ *vertices.*

Observing that Lemma 29 carries over to the offline setting, Claim 30 implies part **(ii)** of Theorem 2 as in the argument just given for the online case (for, say, $g(c) := a^2(c)$).

It therefore remains to prove Claim 30. To do so, we proceed by combinatorial counting in the grid $[s]^2$. Let $\mathcal{X}$ denote the family of all subsets $X \subseteq [s]^2$ for which all components of the graph $[s]^2 \setminus X$ are of order at most as^2. Note that a choice of points for which Occ_ϱ has only components of order at most as^2 exists if and only if there is a set $X \in \mathcal{X}$ that can be completely avoided by the player; i.e., if and only if there is a set $X \in \mathcal{X}$ such that in each of the n pairs, at most one point falls into one of the boxes of X.

Naively speaking, we would wish to show that the expected number of such sets X is $o(1)$. Then Claim 18 would follow with Markov's inequality. Unfortunately, the number of sets $X \in \mathcal{X}$ is too large for this. We therefore refine our basic idea by defining a more manageable family $\mathcal{X}^*$ of subsets of $[s]^2$ with the crucial property that each $X \in \mathcal{X}$ has a subset $X^* \subseteq X$ with $X^* \in \mathcal{X}^*$. For this family $\mathcal{X}^*$ we will indeed be able to show that the expected number of sets $X^* \in \mathcal{X}^*$ that can be avoided in the sense discussed above is $o(1)$. Once this is established, it follows with Markov's inequality that w.h.p. no set from $\mathcal{X}^*$ can be avoided, which in turn implies that also no set from $\mathcal{X}$ can be avoided (recall that each set $X \in \mathcal{X}$ contains a subset $X^* \in \mathcal{X}^*$). To avoid confusion, let us point out explicitly that $\mathcal{X}^*$ will *not* be a subfamily of $\mathcal{X}$.

In the following we proceed with the construction of $\mathcal{X}^*$. For a given set $X \in \mathcal{X}$ we denote by $\mathcal{C}(X)$ the set of all components of $[s]^2 \setminus X$. Set

$$\delta := \frac{1-\sqrt{a}}{4}. \tag{19}$$

For $X \in \mathcal{X}$ given, let

$$t(|X|) := \frac{\delta^2 s^4}{(|X|+s)^2}, \tag{20}$$

and let $k = k(X)$ denote the number of components in $\mathcal{C}(X)$ that are of size at least $t(|X|)$. We shall refer to these components as *large* components, and denote them by $C_1, \dots, C_k$. We call the remaining components *small*. Note that the notions of large and small components are not absolute, but depend on the size of the set X considered.

For the following definitions it is convenient to go back to a geometric viewpoint of the $s \times s$ grid. Each component C of the graph $[s]^2 \setminus X$ corresponds to a connected subset of the unit square (with area $v(C) \cdot \varrho^2$), and has a geometrical boundary ∂C that is the union of one or several closed (rectilinear) walks in the unit square. Note that the length of this geometrical boundary is $e(C, C^c) \cdot \varrho$, where C^c denotes the complement of C in $\mathbb{Z}^2$ (not in $[s]^2$). For brevity we write, with slight abuse of notation, $|\partial C|$ for $e(C, C^c)$, and $|C|$ for $v(C)$ in the following.

For $i = 1, \ldots, k$, let C'_i denote the maximal superset of C_i whose geometrical boundary $\partial C'_i$ is contained in ∂C_i. Note that $\partial C'_i$ is a single closed walk in the unit square. (Informally speaking, C'_i is obtained from C_i by 'filling the holes' in C_i.) Note that $C'_1, \ldots, C'_k$ are not necessarily pairwise disjoint (think e.g. of $C_1, \ldots, C_k$ as concentric rings).

Going back to the combinatorial viewpoint, it is not hard to see that the neighborhood of C'_i is contained in the neighborhood of C_i for each $i = 1, \ldots, k$. Let $X' = X'(X)$ denote the union of the neighborhoods of $C'_1, \ldots, C'_k$, and note that $X' \subseteq X$.

For $X \in \mathcal{X}$ we now define

$$X^* = X^*(X) := \begin{cases} X & \text{if } |X| \geq s^{1.01}, \\ X'(X) & \text{otherwise.} \end{cases}$$

Note that $X^*(X) \subseteq X$ in both cases – as explained above, this is crucial for our argument. Finally, we define $\mathcal{X}^*$ to be the family of all sets $X^* \subseteq [s]^2$ that can arise in this way from some set $X \in \mathcal{X}$.

By our explanations above, it remains to show the following:

Claim 31 *The expected number of sets $X^* \in \mathcal{X}^*$ that contain no two points from the same random point pair (i.e., the expected number of sets $X^* \in \mathcal{X}^*$ that can be avoided by the player) is $o(1)$.*

Let $\mathcal{X}^*_m$ denote the family of all sets $X^* \in \mathcal{X}^*$ of size exactly m. We will bound the number of sets in $\mathcal{X}^*_m$ by combinatorial counting. We begin by showing that $\mathcal{X}^*_m$ is in fact empty for values of m smaller than

$$m_{\min} := \frac{(1-\sqrt{a})s}{2\sqrt{a}}. \tag{21}$$

Lemma 32 *For s large and $X \in \mathcal{X}$ with $|X| < s^{1.01}$ the set $X' = X'(X)$ satisfies $|X'| \geq m_{\min}$. Consequently, for $m < m_{\min}$ we have $\mathcal{X}^*_m = \emptyset$.*

Proof: Let X as in the lemma be given, and let H_{small} denote the union of the small components in $\mathcal{C}(X)$ (recall the definitions after (20)). Applying Lemma 17 with $\beta = (t(|X|)/s)^2 = \frac{\delta^2 s^2}{(|X|+s)^2}$ and $\alpha = |X|/s$ gives that $v(H_{\text{small}}) \leq \delta s^2$. The remaining occupied boxes must be in the $k = k(X)$ large components. Consequently we have

$$\sum_{i=1}^{k} |C_i| \geq (1-\delta)s^2 - |X| \geq (1-2\delta)s^2, \tag{22}$$

where the last inequality follows from $|X| \leq s^{1.01} = o(s^2)$.

The next argument is similar in spirit to the proofs of Lemmas 16 and 17; however, we have to deal with the subtlety that we want to bound $\sum_{i=1}^k |\partial C'_i| = \sum_{i=1}^k e(C'_i, (C'_i)^c)$ from below but we only have an upper bound on $|C_i| = v(C_i)$ (not $|C'_i| = v(C'_i)$) for all i.

Note first that

$$\sum_{i=1}^{k} |\partial C'_i| \leq 4|X'| + 4s, \tag{23}$$

where the inequality follows from the observation that each vertex of X' contributes at most 4 to the sum, and the boundary of the unit square contributes at most $4s$ in total.

On the other hand, by Lemma 15 the total circumference of $C'_1, \ldots, C'_k$ satisfies

$$\sum_{i=1}^{k} |\partial C'_i| \geq \sum_{i=1}^{k} 4\sqrt{|C'_i|} \geq 4\sum_{i=1}^{k} \sqrt{|C_i|} \geq 4\sum_{i=1}^{k} \frac{|C_i|}{\sqrt{as^2}} \stackrel{(22)}{\geq} \frac{4(1-2\delta)s}{\sqrt{a}},$$

where in the third inequality we used that $|C_i| \leq as^2$ (because $X \in \mathcal{X}$).

Together with (23) it follows that

$$|X'| \geq \frac{(1-2\delta)s}{\sqrt{a}} - s \stackrel{(19)}{=} \frac{(1-\sqrt{a})s}{2\sqrt{a}} = m_{\min}.$$

■

Next we bound the size of $\mathcal{X}^*_m$ for the intermediate values of m.

Lemma 33 *For s large and $m_{\min} \leq m < s^{1.01}$ we have $|\mathcal{X}^*_m| \leq e^{10m/(1-\sqrt{a})}$.*

Proof: To specify a set $X^* \in \mathcal{X}^*_m$ for m as in the lemma, it suffices to specify the boundaries of $C'_1, \ldots, C'_k$ (i.e., the *outer* boundaries of $C_1, \ldots, C_k$). For a given such component C'_i, we encode its geometric boundary $\partial C'_i$ by specifying, say, the leftmost point of the topmost horizontal line intersecting with $\partial C'_i$ as a starting point, and by specifying the direction of each of the $|\partial C'_i| = \ell_i$ steps along the boundary (say in clockwise direction). There are at most $s^2 \cdot 3^{\ell_i}$ ways of specifying a boundary $\partial C'_i$ (and thus a component C'_i) in this way.

For each set $X^* \in \mathcal{X}^*_m$ with m as in the lemma there exists, by definition, a set $X \in \mathcal{X}$ with $X'(X) = X^*$ and $|X| < s^{1.01}$. Thus the number $k = k(X)$ of large components can be bounded as

$$k \leq \frac{s^2}{t(|X|)} \stackrel{(20)}{=} \delta^{-2}\left(\frac{|X|}{s} + 1\right)^2 \leq \delta^{-2}(s^{0.01}+1)^2 =: x.$$

Note that $x = O(s^{0.02})$. Recall also from (23) that $\sum_{i=1}^{k} \ell_i \leq 4m + 4s$.

It follows that for $m_{\min} \leq m < s^{1.01}$ we have

$$|\mathcal{X}^*_m| \leq \sum_{k=1}^{x} \sum_{\substack{\ell_1,\ldots,\ell_k:\\ \ell_1+\cdots+\ell_k \leq 4m+4s}} \prod_{i=1}^{k} (s^2 3^{\ell_i}) \leq x(4m+4s)^x (s^2)^x 3^{4m+4s}.$$

Observe that, for s large enough and $m_{\min} \leq m < s^{1.01}$, the factor 3^{4m+4s} is much larger than x, $(4m+4s)^x$ and $(s^2)^x$ (which are all $\exp[O(s^{0.02}\log s)]$). It follows that, for s large enough:

$$|\mathcal{X}^*_m| \leq (3.01)^{4(m+s)} \leq e^{5(m+s)} \leq e^{10m/(1-\sqrt{a})},$$

where in the last step we used that

$$s \stackrel{(21)}{\leq} \frac{2\sqrt{a}}{1-\sqrt{a}} \cdot m_{\min} \leq \frac{2\sqrt{a}}{1-\sqrt{a}} \cdot m$$

and consequently

$$m + s \leq \frac{1+\sqrt{a}}{1-\sqrt{a}} \cdot m \leq \frac{2m}{1-\sqrt{a}}.$$

■

With the preceding lemmas in hand, Claim 31 follows with a routine calculation

Proof of Claim 31: Using Lemmas 32 and 33, and using the trivial bound

$$|\mathcal{X}^*_m| \leq \binom{s^2}{m} \leq s^{2m} = e^{2m\log s}$$

for $m \geq s^{1.01}$, we obtain that the expected number of sets $X^* \in \mathcal{X}^*$ that contain no two points from the same random point pair is

$$\begin{aligned}\sum_m |\mathcal{X}^*_m| \cdot \left(1 - \left(\frac{m}{s^2}\right)^2\right)^n &\leq \sum_{m_{\min} \leq m < s^{1.01}} e^{10m/(1-\sqrt{a}) - m^2 n/s^4} + \sum_{m \geq s^{1.01}} e^{2m\log s - m^2 n/s^4} \\ &\leq \sum_{m \geq m_{\min}} \left(e^{10/(1-\sqrt{a}) - mn/s^4}\right)^m + \sum_{m \geq s^{1.01}} \left(e^{2\log s - mn/s^4}\right)^m.\end{aligned} \tag{24}$$

By our choice of constants, we have

$$n/s^3 = n\varrho^3 = nr^3 \cdot 5^{-3/2} \geq c/12. \tag{25}$$

and consequently the exponents of the terms in parentheses are uniformly bounded by

$$\frac{10}{1-\sqrt{a}} - m_{\min} \cdot n/s^4 \overset{(21),(25)}{\leq} \frac{10}{1-\sqrt{a}} - \frac{c(1-\sqrt{a})}{24\sqrt{a}} \overset{(18)}{=} -\frac{10}{1-\sqrt{a}} < 0.$$

and

$$2\log s - s^{1.01} \cdot n/s^4 \overset{(25)}{\leq} 2\log s - c/12 \cdot s^{0.01} = -\omega(1),$$

respectively. It follows that the right hand side of (24) is $o(1)$. ■

As explained above, Claim 31 implies Claim 30, which in turn implies part **(ii)** of Theorem 2.

5 Concluding remarks

In this paper we have shown that in the power of choices version of the random geometric graph, the onset of a giant component can be delayed until the average degree is of order $n^{1/3}(\log\log n)^{2/3}$. This is an improvement by a power of n over the standard random geometric graph, where a giant appears as soon as the average degree exceeds a certain constant. As pointed out in the introduction, this behavior is in stark contrast to what happens in the (vertex) Achlioptas process, where the power of choices only yields a constant factor improvement.

We have also shown that in the *offline* version of our process a giant can be delayed just a little longer, until the average degree is of order $n^{1/3}$.

We offer the following two natural conjectures:

Conjecture 34 *There is a function $f : (0,\infty) \to (0,1)$ such that the following holds. Consider the* online *power of choices geometric graph process, where $r = \sqrt[3]{\frac{c}{n\log\log n}}$. Assuming optimal play, the largest component will have size $(1+o(1)) \cdot f(c) \cdot n$ w.h.p.*

Conjecture 35 *There is a function $f : (0, \infty) \to (0, 1)$ such that the following holds. Consider the* offline *power of choices geometric graph setting, where $r = \sqrt[3]{\frac{c}{n}}$. Assuming optimal play, the largest component will have size $(1 + o(1)) \cdot f(c) \cdot n$ w.h.p.*

Many steps in our proofs have been rather crude and we have made no attempt to optimize the expressions for $f(c), g(c)$ in Theorems 1 and 2. The main reason for this is that we believe that it will not be possible to prove the above two conjectures without significant new ideas.

The largest component just before the threshold Our proofs also give some insight into the behaviour before the threshold. For the following discussion, we let $r_0 = (n \log \log n)^{-1/3}$ for the online case, and $r_0 = n^{-1/3}$ for the offline case. Furthermore, we assume that r is asymptotically smaller than r_0 but only slightly so (say $n^{-1/3-0.01} \ll r \ll r_0$).

The strategies described in Sections 3.1 and Sections 3.2 guarantee that w.h.p. the largest component is of order $O((r/r_0)^6 \cdot n)$ vertices in both settings. To see this, note that in Lemma 5 the constant $a(K)$ can be improved to $a(K) = \Theta(1/K^2)$. Thus, $a(K) = \Theta(c^2)$ as $c = (r/r_0)^3 \to 0$, which translates to the claimed bound by (a slightly adapted version of) Lemma 9.

On the other hand, the upper bound proof given in Section 4.3 for the offline setting shows that w.h.p. the player will be forced to create a component with $\Omega((r/r_0)^{12} \cdot n)$ vertices: We have $a(c) = \Theta(c^2)$ as $c = (r/r_0)^3 \to 0$ in (18), and the resulting factor of $(r/r_0)^6$ is squared when (a slightly adapted version of) Lemma 29 is applied.

Similarly, the upper bound proof given in Section 4.2 for the online setting shows that w.h.p. the player will be forced to create a component with at least $\Theta((r/r_0)^{12} \cdot n)$ vertices: For ε fixed and $\alpha \to \infty$, we have $k(\varepsilon, \alpha) = O(\alpha^2)$ in Corollary 19, as $\lambda_i = O(i/\alpha^2)$. It follows that for $\varepsilon = 0.01$ fixed, $a(c, \varepsilon)$ in Lemma 28 is $\Omega(c^2)$ as $c = (r/r_0)^3 \to 0$. As for the online setting, the resulting factor of $(r/r_0)^6$ is squared when Lemma 29 is applied.

To summarize, in both the online and the offline power of choices setting, the size of the largest component in optimal play is between $\Theta((r/r_0)^{12} \cdot n)$ and $\Theta((r/r_0)^6 \cdot n)$, where r_0 denotes the respective threshold. Note that this behavior is again very different from what happens in the standard geometric and Erdős-Rényi random graphs, where the size of the largest component jumps fron $\Theta(\log n)$ to $\Theta(n)$ at the threshold. (Such a jump is also observed in the Achlioptas process when played with certain natural –but most likely not optimal– player strategies, see e.g. [24].)

It would be interesting to close the gap between our bounds for the moderately subcritical regime.

Question 36 *Both for the online and the offline setting, what is the order of the largest component in optimal play (w.h.p.) when r is slightly below the respective threshold?*

More choices Let us now sketch how our results generalize to the scenario with an arbitrary fixed number $d \geq 2$ of choices per step. As stated in the introduction, the resulting thresholds then are $n^{-1/(d+1)}(\log \log n)^{-(d-1)/(d+1)}$ for the online setting, and $n^{-1/(d+1)}$ for the offline case. This can be shown with only minor modifications to the proofs we gave for $d = 2$.

To give some intuition for these formulas, let us point out the following: In both scenarios, the threshold corresponds to the point where the *number of points that are forced to be in the moat* (as defined in our lower bound proofs) equals the *number of boxes of the moat* in order of magnitude. In the online scenario, the moat has an area of $\Theta(r \log \log n)$, and thus the (expected) number of points that we need to choose in the moat is of order $nr^d(\log\log n)^d$. On the other hand, the number of boxes in the moat is of order $r^{-1}\log\log n$. It is not hard to see that these terms are equal for r as stated. The threshold for the offline case can be motivated with a very similar calculation.

Creating a giant Another interesting related question is what happens if the player attempts to speed up the onset of a giant component instead of delaying it. For this setup one can quite easily derive the following result:

Theorem 37 *Suppose that $r = \sqrt{\lambda/n}$ for some constant $\lambda > 0$. Then the following holds, where λ_{crit} is the critical constant for the emergence of a giant component in the ordinary random geometric graph:*

(i) *If $\lambda \leq \lambda_{crit}/2$ then w.h.p. the largest component of the graph will be $o(n)$, no matter what the player does;*

(ii) *If $\lambda > \lambda_{crit}/2$ then the player has a strategy that will result in a component of order $\Omega(n)$, w.h.p.*

To see that part **(i)** holds, we just need to note that even if we allow the player to keep *both* points in each round he will just have a subcritical or critical random geometric graph. The proof of part **(ii)** is only slightly more involved. A sketch of the argument is as follows: The player fixes a small square A of area $\varepsilon = \varepsilon(\lambda)$ inside the unit square, and he always selects a point inside A if he can (if both fall in A he chooses randomly). If $\varepsilon > 0$ was chosen sufficiently small, then the graph induced by the points in A will be a *supercritical* random geometric graph, containing a linear proportion of all vertices that fall in A, and hence also a linear proportion of all n vertices. For completeness we spell out this argument in more detail in Appendix B.

Our strategy for $\lambda > \lambda_{\text{crit}}/2$ case is rather simple, and in a sense it might be suboptimal. While it does deliver a component of linear size, a more sophisticated strategy might achieve an even larger largest component.

Question 38 *If $\lambda > \lambda_{crit}$ is fixed and $r = \sqrt{\lambda/n}$, what is the order of the largest component the player can (w.h.p.) achieve?*

References

[1] D. Achlioptas, R.M. D'Souza, and J. Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009.

[2] Y. Azar, A.Z. Broder, A.R. Karlin, and E. Upfal. Balanced allocations. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, pages 593 – 602. ACM, 1994.

[3] Y. Azar, A.Z. Broder, A.R. Karlin, and E. Upfal. Balanced allocations. *SIAM Journal on Computing*, 29:180–200, 1999. A preliminary version of this paper appeared in Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing, 1994.

[4] J. Balogh, B. Bollobás, M. Krivelevich, T. Müller, and M. Walters. Hamilton cycles in random geometric graphs. *Annals of Applied Probability*, 21(3):1053–1072, 2011.

[5] T. Bohman and Alan F. Avoiding a giant component. *Random Structures Algorithms*, 19(1):75–85, 2001.

[6] T. Bohman, A. Frieze, and N. Wormald. Avoidance of a giant component in half the edge set of a random graph. *Random Structures Algorithms*, 25(4):432–449, 2004.

[7] T. Bohman and J.H. Kim. A phase transition for avoiding a giant component. *Random Structures Algorithms*, 28(2):195–214, 2006.

[8] T. Bohman and D. Kravitz. Creating a giant component. *Combin. Probab. Comput.*, 15(4):489–511, 2006.

[9] C. Cooper and A. Frieze. The cover time of random geometric graphs. In *SODA '09: Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 48–57, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[10] A.D. Flaxman, D. Gamarnik, and G.B. Sorkin. Embracing the giant component. *Random Structures Algorithms*, 27(3):277–289, 2005.

[11] E. N. Gilbert. Random plane networks. *J. Soc. Indust. Appl. Math.*, 9:533–543, 1961.

[12] G.H. Gonnet. Expected length of the longest probe sequence in hash code searching. *Journal of the ACM*, 28(2):289–304, 1981.

[13] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[14] J. Kingman. *Poisson Processes*. Oxford University Press, Oxford, 1993.

[15] M. Krivelevich, P.-S. Loh, and B. Sudakov. Avoiding small subgraphs in Achlioptas processes. *Random Structures Algorithms*, 34(1):165–195, 2009.

[16] M. Krivelevich, E. Lubetzky, and B. Sudakov. Hamiltonicity thresholds in Achlioptas processes. *Random Structures Algorithms*, 37(1):1–24, 2010.

[17] M. Krivelevich and R. Spöhel. Creating small subgraphs in Achlioptas processes with growing parameter. *SIAM J. Discrete Math.*, 26(2):670–686, 2012.

[18] C.J.H. McDiarmid and T. Müller. On the chromatic number of random geometric graphs. *Combinatorica*, 31(4):423–488, 2011.

[19] M. Mitzenmacher, A.W. Richa, and R. Sitaraman. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing, Vol. I, II*, volume 9 of *Comb. Optim.*, pages 255–312. Kluwer Acad. Publ., Dordrecht, 2001.

[20] T. Mütze, R. Spöhel, and H. Thomas. Small subgraphs in random graphs and the power of multiple choices. *J. Combin. Theory Ser. B*, 101(4):237–268, 2011.

[21] M. D. Penrose. On k-connectivity for a geometric random graph. *Random Structures Algorithms*, 15(2):145–164, 1999.

[22] M. D. Penrose. *Random Geometric Graphs.* Oxford University Press, Oxford, 2003.

[23] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.*, 22(4):1450–1464, 2012.

[24] J. Spencer and N. Wormald. Birth control for giants. *Combinatorica*, 27(5):587–628, 2007.

A An upper bound for the vertex Achlioptas process

.

In this section we show that, as claimed at the end of Section 1.2, in the *vertex* Achlioptas process the player is also forced to create a linear-sized component as soon as the average degree of the underlying random graph exceeds a certain constant.

We will use the following lemma, which is a straightforward generalization of Lemma 2 in [6]. We denote by $G(n, m)$ an ("Erdős-Rényi") random graph sampled uniformly from all graphs on n vertices and m edges. For a graph G and a set $S \subseteq V(G)$, we denote the graph induced by S in G by $G[S]$.

Lemma 39 ([6]) *Let $c > 0$. For every $\varepsilon > 0$ there exists $\delta = \delta(c, \varepsilon) > 0$ such that a.a.s. the random graph $G := G(n, cn)$ has the property that for every $S \subseteq V(G)$ for which $G[S]$ contains more than $(1 + \varepsilon)|S|$ edges we have $|S| \geq \delta n$.*

We are now able to deduce:

Theorem 40 *There is a constant $c > 0$ such that if $m \geq cn$ then w.h.p. a component of linear size will be formed in the vertex Achlioptas process, no matter what the player does.*

Proof: We will show that for c large enough and $m := cn$, w.h.p. $G(n, m)$ is such that every set of $n/2$ vertices induces a graph that contains a linear-sized component. Clearly, this then proves the claim. (In fact, our argument gives an upper bound for the *offline* problem corresponding to the vertex Achlioptas process.)

Note that the expected number of edges in a fixed set of $n/2$ vertices is

$$m \cdot \frac{\binom{n/2}{2}}{\binom{n}{2}} = (1 + o(1))m/4$$

By a Chernoff type bound (Theorem 2.10 of [13]) the probability that this number of edges is less than $m/8$ is $e^{-\Omega(m)}$. A union bound over all (trivially at most 2^n) sets of $n/2$ vertices thus yields that with probabiliy $1 - 2^n e^{-\Omega(m)}$, each such set contains at least $m/8$ edges. Clearly for c chosen large enough the last probability is $1 - o(1)$.

Note that the ratio of edges to vertices in each such set is $(m/8)/(n/2) = m/(4n) = c/4$, which we can ensure to be at least 2, say, by choosing $c \geq 8$. Moreover, by averaging, also at least one of the *components* of the graph induced by such a set has a ratio of edges to vertices of at least 2. By Lemma 39, w.h.p. each such subgraph of $G(n, m)$ is of order at least $\delta(c, 1)n$.

To summarize, w.h.p. $G(n, m)$ is such that each set of $n/2$ vertices has a ratio of edges to vertices of at least 2, and as a consequence of this induces a graph which contains a linear-sized component. ■

B Proof of part (ii) of Theorem 37

In this section we fill in the details of the proof sketch provided just after the statement of Theorem 37.

Proof of part (ii) of Theorem 37: We take $\varepsilon = \varepsilon(\lambda)$ sufficiently small, to be made more precise later, and we let $A \subseteq [0,1]^2$ be a square with $\text{area}(A) = \varepsilon$.

In every round, the player will always picks a point in A if he can. If it happens that both points fall in A the he chooses randomly. Observe that the probability that, in a given round, the player is able to select a point of A equals $1 - (1-\varepsilon)^2 = 2\varepsilon - \varepsilon^2$.

Let R denote the number of rounds when he succeeded to pick a point of A. Clearly $R \stackrel{d}{=} \text{Bi}(n, 2\varepsilon - \varepsilon^2)$. By the Chernoff bound (Lemma 4) we have that

$$\mathbb{P}(R < (1-\varepsilon) \cdot \mathbb{E}R) \leq \exp[-\Omega(n)] = o(1).$$

Let $\tilde{G}$ denote the subgraph of the player's graph induced by the points in A. Observe that we can rescale A by a factor of $1/\sqrt{\varepsilon}$ and translate it to map it to the unit square $[0,1]^2$. Thus, by stopping the process the instant $n' := (1-\varepsilon)\mathbb{E}R = (1-\varepsilon) \cdot (2\varepsilon - \varepsilon^2) \cdot n$ points have been selected inside A, we see that (w.h.p.) $\tilde{G}$ contains a copy of the ordinary random geometric graph with parameters n' and $r' := r/\sqrt{\varepsilon}$.

Let us now observe that we can rewrite r' as

$$r' = \frac{r}{\sqrt{\varepsilon}} = \sqrt{\frac{\lambda}{\varepsilon n}} = \sqrt{\frac{2 \cdot (1-\varepsilon) \cdot (1-\varepsilon/2) \cdot \lambda}{n'}} =: \sqrt{\frac{\lambda'}{n'}}.$$

As $\lambda > \lambda_{\text{crit}}/2$, we can choose $\varepsilon > 0$ small enough for $\lambda' > \lambda_{\text{crit}}$ to hold. Hence, in that case $\tilde{G}$ will (w.h.p.) contain a component spanning $\Omega(n') = \Omega(n)$ points. ■