Barrier coverage

Paul Balister[*†], Béla Bollobás[‡§], Amites Sarkar[¶]

March 9, 2016

Abstract

Suppose sensors are deployed randomly in a long thin strip, and suppose each sensor can detect objects within a fixed distance. We say that the sensors achieve *barrier coverage* if there is no path across the strip that a small object can follow that avoids detection by the sensors. We give fairly precise results on the probability that barrier coverage is achieved as a function of the range of the sensors, the height and length of the strip, and the number of sensors deployed. In particular, we show that the most likely obstruction — a rectangular region crossing the strip which is devoid of sensors — does *not* in general dominate the probability of failure of barrier coverage.

1 Introduction

Imagine a collection of randomly placed sensors deployed in some region in the plane, each capable of detecting events or objects within a given fixed distance. A natural and well studied question is whether or not they cover the region of interest, in the sense that every point in the region is in the sensing range of some sensor (see for example [12, 6, 7, 9, 3]). In this paper we consider a somewhat different question. Instead of fully covering the

[*]Department of Mathematical Sciences, University of Memphis, USA. E-mail: pbalistr@memphis.edu

[†]Research supported in part by NSF grants CCF-0728928 and DMS-1301614.

[‡]DPMMS, Cambridge University, UK, and Department of Mathematical Sciences, University of Memphis, USA. E-mail: B.Bollobas@dpmms.cam.ac.uk

[§]Research supported in part by NSF grants DMS-0906634, CNS-0721983, CCF-0728928, and DMS-1301614, ARO grant W911NF-06-1-0076, and TAMOP-4.2.2/08/1/2008-0008 program of the Hungarian National Development Agency.

[¶]Department of Mathematics, Western Washington University, Bellingham, USA. E-mail: amites.sarkar@wwu.edu

region, one simply wishes to block the movement of an intruder say, across a long thin strip. In this case, fully covering the strip is not necessary. It is simply enough that there be no path across the strip avoiding the sensing regions. We call such coverage *barrier coverage* and give fairly precise results on the probability of barrier coverage as a function of the range of the sensors, the height and length of the strip, and the number of sensors deployed.

Consider therefore a horizontal strip $[a, b] \times [0, h]$ of height h and length $b - a \gg h$. We shall assume each sensor has sensing range $\frac{r}{2}$, so that two sensors will have overlapping sensing regions when they are within distance r of each other. (The reason for this choice of normalization will become clearer when we discuss the graphs $G_{h,r}$ and $G_{h,r}(a, b)$ defined below.) As scaling all distances by a constant factor does not change the model, we will assume without loss of generality that the scale is chosen so that there is on average one sensor per unit area. If the sensors are required to lie inside the strip, then a vertical crossing path at $x = a$ or $x = b$ is more likely to avoid the sensors than one near the middle of the strip. Thus the regions of the strip close to the ends are in some sense atypical. To avoid these "boundary problems", we shall extend the strip horizontally and allow sensors at points (x, y) with $x < a$ or $x > b$. We shall however always insist that $y \in [0, h]$. To do this, rather than fixing the total number of sensors, we shall instead place the sensors according to a Poisson point process with intensity 1 in the *infinite* strip $\mathbb{R} \times [0, h]$. This now ensures that any horizontal translate of a crossing path has the same chances of avoiding the sensors. Of course, if $b - a \to \infty$, then there will almost surely be a crossing path that avoids the sensors. Thus we shall mainly be concerned with the frequency of breaks along the strip where a crossing path can avoid sensors. Results for a finite strip will then be deduced from this infinite model. This therefore motivates the following formalized version of our problem.

We consider an infinite strip $S_h = \mathbb{R} \times [0, h]$ of width (or height) h and place sensors inside this strip randomly according to a Poisson point process $\mathcal{P}$ of intensity 1. We shall assume that each sensor has the ability to detect intruders strictly within a (Euclidean) distance $\frac{r}{2}$ and write $D_v = \{x \in \mathbb{R}^2 : \|x - v\|_2 < \frac{r}{2}\}$ for the sensing region of a sensor located at a point $v \in S_h$. Construct an infinite random geometric graph $G_{h,r}$ with vertex set given by the set of sensors $\mathcal{P}$, by joining every sensor to every other sensor that is strictly within distance r, that is, whenever their sensing regions intersect. For $a < b$, define $G_{h,r}(a, b)$ to be the subgraph of $G_{h,r}$ consisting of sensors with x-coordinate lying in the interval $(a - \frac{r}{2}, b + \frac{r}{2})$, and with two extra "virtual" sensors s and t, where s is joined to all sensors of $\mathcal{P}$ strictly within distance $\frac{r}{2}$ of the line $x = a$ and t is joined to all sensors of $\mathcal{P}$ strictly within distance $\frac{r}{2}$ of the line $x = b$ (see Figure 1). The sensors s and t are never joined directly to each other.

Define a *separating path* to be a continuous simple path in S_h starting at some point

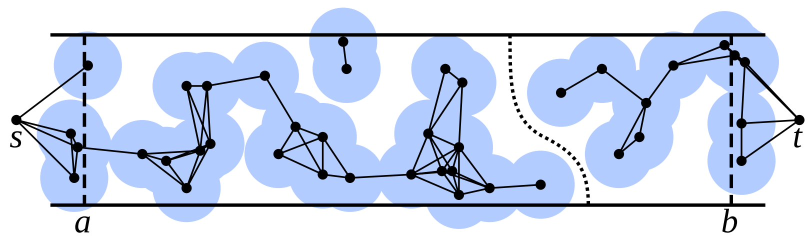

Figure 1: Model with sensor range $\frac{r}{2}$ (sensing regions indicated by shading). Dotted line indicates a separating path — either a possible path of an undetected intruder, or a path disconnecting s from t in the network.

on the line $y = 0$, ending at some point on the line $y = h$ and not passing strictly within distance $\frac{r}{2}$ of any sensor. The following is immediate from the above definitions and standard topological properties of the plane.

Lemma 1. *The graph $G_{h,r}(a, b)$ contains a path from s to t if and only if there is no separating path of the infinite graph $G_{h,r}$ that lies entirely within the rectangle $[a, b] \times [0, h]$.* □

Thus the questions of avoiding separating paths, and obtaining long paths along $G_{h,r}$ are effectively equivalent, and barrier coverage can also be thought of as a question of "percolation" in the Gilbert disk model (see for example [11, 13]) when confined to a thin strip. Of course, strictly speaking there can be no percolation in a strip of bounded width, as an infinite strip almost surely has a large gap somewhere with no sensors, and hence no infinite component can exist in $G_{h,r}$. However, for large r and h, components can exist that are exponentially large, and it is these components, and the gaps between them, that we shall be interested in.

The presence of a separating path indicates a "break" in the coverage. The aim will be to determine the frequency or *intensity* $I_{h,r}$ of these breaks, defined so that the number of breaks in $[a, b] \times [0, h]$ is on average $(b - a)(I_{h,r} + o(1))$ as $b - a \to \infty$ (the $o(1)$ error including, for example, end effects near $x = a$ and $x = b$). The precise definition of $I_{h,r}$ is given in Section 4. We also wish to estimate the probability that a break occurs in a finite interval, or more generally the probability distribution of the number of breaks. We aim to define the notion of a break so as to make breaks "almost independent". This will imply that the number of breaks in $G_{h,r}(a, b)$ is given approximately by a Poisson variable of mean $(b - a)I_{h,r}$, and in particular, the probability that no break occurs is about $\exp(-(b - a)I_{h,r})$.

The definition of a break is more delicate than it might appear at first. For example, in Figure 2, should the two separating paths be considered as defining the same break, or two different breaks? Since there may be several small connected "islands" in the break,

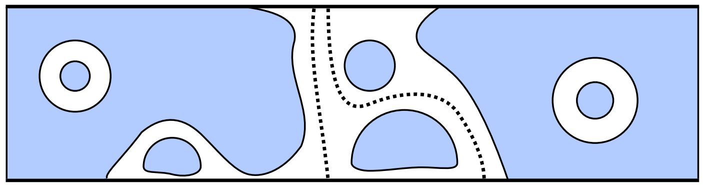

Figure 2: Ambiguity in counting breaks. Do the two separating paths indicate two separate breaks, or should we consider this as just one "compound" break? Also, there are several small components of $G_{h,r}$ that do not cause any breaks.

one may be able to construct very many homotopically distinct separating paths. The two paths in Figure 2 are not really independent since the existence of one makes the existence of the other much more likely. If we were to count these paths as distinct breaks, then the probability distribution of the number of breaks in $G_{h,r}(a, b)$ could be very far from Poisson, and the probability of a break could be much less than the expected number of breaks, even when the expected number of breaks is much less than 1. For this reason, we wish to consider the situation in Figure 2 as a *single* break. There are several suitable definitions of a break, which although different, give the same asymptotic frequencies when h and r are large. The following definition is chosen since it is easy to compute in simulations and is fairly convenient theoretically.

Define a *good component* as a (graph) component C of the infinite graph $G_{h,r}$ which contains a sensor strictly within distance $\frac{\sqrt{3}}{2}r$ of the top boundary ∂S_h^+ of S_h, and also contains a sensor strictly within distance $\frac{\sqrt{3}}{2}r$ of the bottom boundary ∂S_h^- of S_h. We shall show (see the comments before Lemma 8 below) that there is a natural left-to-right ordering of good components. Indeed, the $\frac{\sqrt{3}}{2}r$ in the definition was chosen to be the largest value such that good components cannot "jump" past one another.[1] Now define a *break* to be the gap between two consecutive good components. In other words, a break is a partition of the set of good components into two classes, those on the left of the break, and those on the right, which is compatible with the left-right ordering of the good components.

Our main results are the following.

Theorem 2. *Assume $r \geq 7$. Then the intensity of breaks almost surely exists and satisfies*

$$I_{h,r} = r^{1/3}\varepsilon(hr^{-1/3})e^{-hr+O(hr^{-5/3})} \tag{1}$$

where $\varepsilon(z)$ is the function defined in Section 6 below, and the constant implicit in the $O(\cdot)$ notation is independent of both r and h.

[1] One could use any distance that is at most $\frac{\sqrt{3}}{2}r$ here, however the choice $\frac{\sqrt{3}}{2}r$ is needed in our proof of Lemma 23.

Note that h is arbitrary in Theorem 2, so the intensity is determined up to a factor of $1 + o(1)$ when $h = o(r^{5/3})$ and $r \to \infty$. Even for larger h, the logarithm of the intensity is determined up to a factor of $1 + o(1)$ as $r \to \infty$.

The lower bound of 7 on r is not best possible. Indeed, we expect Theorem 2 to hold for all $r \geq r_0$ where r_0 is any value above the percolation threshold $r_c \approx 1.1984$ for the Gilbert model (see [11, 13, 14]). However, if $r < r_c$ then components are generally small, and so good components are very rare (for large h). The breaks are therefore few in number, but large in horizontal extent — it is easy to find crossing paths across the strip.

More details are given on the function $\varepsilon(z)$ in Section 6, but we just note here the following, which allows us to give a simple explicit expression for the intensity when $h = \omega(r^{1/3})$.

Theorem 3. *There exist constants $\alpha > 0$, β, and $\eta > 0$, such that*

$$\varepsilon(z) = \exp\left(\alpha z + \beta + O(e^{-\eta z^{1/9}})\right)$$

as $z \to \infty$.

Remark. The most "obvious" obstruction to barrier coverage is an $r \times h$ rectangle containing no sensor (see Section 2). Based on this one would expect a break intensity of order he^{-hr}. However, as $h \to \infty$ the actual break intensity is exponentially larger, so the most obvious obstructions do *not* dominate the break intensity. Indeed, the contribution given by the obvious obstructions corresponds to taking $\varepsilon(z) = z$. However, $\varepsilon(z) \gg z$ when $z \gg 1$ (see also (30)). Equivalently, non-obvious obstructions dominate when $h \gg r^{1/3}$.

Having estimated the intensity of breaks, we have the following result on the distribution of breaks. Here we say that $G_{h,r}(a, b)$ *contains* a break, if there is a separating path that lies entirely within $[a, b] \times [0, h]$ that induces this break (see Lemma 8). Note that by Lemma 1, $G_{h,r}(a, b)$ contains an s-t path if and only if it contains no break.

Theorem 4. *Fix $r \geq 7$, $x > 0$, and $k \geq 0$. Then the probability that $G_{h,r}(0, x/I_{h,r})$ contains exactly k breaks tends to $e^{-x}x^k/k!$ as $h \to \infty$. In particular, the probability of barrier coverage in $[0, x/I_{h,r}] \times [0, h]$ tends to e^{-x} as $h \to \infty$.*

For strips of length $o(1/I_{h,r})$, Theorem 4 implies that the probability of there being a break is $o(1)$. However, for this case the following much stronger result will be shown.

Theorem 5. *If $r \geq 7$ and $h \geq 1$ then*

$$\mathbb{P}(G_{h,r}(0, \ell) \text{ has no } s\text{-}t \text{ path}) \leq (\ell + 5h)I_{h,r}.$$

2 Heuristics

Before giving proofs and further results, we introduce some non-rigorous heuristics about the types of component of $G_{h,r}$ and their frequency. The purpose is to help provide some intuitive explanation for our results. We generally assume r is large, so that in most areas the graph $G_{h,r}$ is highly connected. The main idea is the concept of "excluded area", that is that rare configurations can be described by the absence of sensors within some nice region A. Outside A, the density of sensors will be assumed to be close to the expected density, which will result in a high degree of connectivity in the graph $G_{h,r}$. As r gets smaller, these approximations become less accurate. Indeed, for $r < r_c$ the graph breaks up into small components and the heuristics described here become totally inapplicable.

Given a region $A \subseteq S_h$, the probability that it contains no sensors is $e^{-|A|}$, where $|A|$ is the area of A. We consider the minimal regions A that can force a component, or a break, under the assumptions that there is a reasonable density of sensors outside of A. For example, an excluded $r \times h$ rectangle across the strip is very likely to cause a break (see Figure 3). On either side of this rectangle, the components are very likely to be good (for large r), but the region disconnects $G_{h,r}$. This region is the smallest such empty region that results in a break, so we might expect most breaks to be approximately rectangular, and the frequency of breaks $I_{h,r}$ to be about e^{-hr}. (More precisely he^{-hr}, since such a rectangle can be placed after any sensor, and there are about h sensors per unit distance along the strip.) Small components can form near the boundary of S_h with an excluded area of $\frac{1}{2}\pi r^2$, or in the interior of S_h with excluded area πr^2. Hence these are likely to occur with frequencies about $e^{-\pi r^2/2}$ and $e^{-\pi r^2}$ respectively. Note that if $h < \frac{\pi}{2}r$, then breaks should be more common than these small components, and so most components are likely to be good. On the other hand, if $h > \frac{\pi}{2}r$, then most components are likely to be small (bad) components that do not form breaks.

Compound breaks (breaks containing homotopically inequivalent separating paths) need an excluded area of at least $rh + (\frac{\pi}{3} - \frac{\sqrt{3}}{4})r^2 > rh$ (see fourth example in Figure 3), so at first sight these seem far rarer than simple breaks. However, we need to take care of combinatorial issues — how many ways such breaks can occur. It is possible for compound breaks to be more common than simple breaks if h is extremely large, in particular if $1 \ll r^2 \ll \log h$. This is because although we lose a factor of $e^{-O(r^2)}$ in the frequency of compound breaks due to the extra $O(r^2)$ excluded area, we gain a combinatorial factor of order h due to the choice of the vertical position of the small component inside the break. (The excluded area is still $hr + O(r^2)$ even if the small component is in the center of the break.)

Although vertical excluded rectangles are the most obvious breaks, diagonal breaks are also possible. If the top of the break is displaced by a distance d, the excluded area

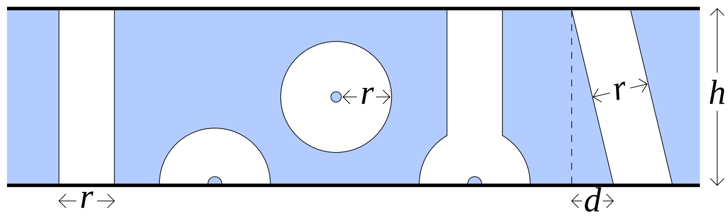

Figure 3: Minimal excluded areas for breaks, small components near the boundary and in the interior, compound breaks, and diagonal breaks.

becomes $r\sqrt{h^2+d^2} \approx rh + \frac{rd^2}{2h}$. If $rd^2 \sim h$ this does not impose a large penalty, so we may expect breaks to often deviate by about $O(\sqrt{h/r})$ from vertical.

These heuristics tend to be good for large r, however, corrections are needed for the effects of finite r. These corrections tend to make breaks more likely. Indeed, the main difference between the e^{-hr} heuristic and Theorem 2 is given by the $\varepsilon(hr^{-1/3})$ factor, which by Theorem 3 gives an extra positive $\Theta(hr^{-1/3})$ term in the exponent. Determining this extra term is rather difficult; in fact we believe that this is the most important contribution of this paper. This extra term arises as a trade off between combinatorial factors counting the many different types of break one can have, versus the penalty one pays in extra excluded area for non-rectangular breaks. Roughly speaking, one gains an exponential factor per unit length (h) of the break due to the extra "entropy" allowed by making the walls of the excluded area fluctuate. Increasing r reduces this effect as these walls cannot curve so much without increasing the excluded area significantly.

3 Good components and crossing sensor-paths

Recall that a *good component* is a component of $G_{h,r}$ that has sensors within distance $\frac{\sqrt{3}}{2}r$ of both the top and bottom boundaries of the strip $S_h = \mathbb{R} \times [0,h]$.

Define a *crossing sensor-path* to be a path from the bottom of S_h to the top of S_h consisting of line segments, with the first line segment being a vertical line segment bv_1 of length at most $\frac{\sqrt{3}}{2}r$ from some point $b \in \partial S_h^-$ to some sensor v_1 of $G_{h,r}$. Then line segments corresponding to a graph theoretic walk[2] $v_1, \ldots, v_n$ through $G_{h,r}$, and finally a vertical line segment v_nt of length at most $\frac{\sqrt{3}}{2}r$ to a point $t \in \partial S_h^+$ (see Figure 4). Note that a crossing sensor-path need not be a simple curve and there exists a crossing sensor-path going though some sensor if and only if that sensor is contained in a good component.

[2]A *walk* through a graph is a sequence of vertices $v_1, \ldots, v_n$ such that each v_iv_{i+1} is an edge. We allow both edges and vertices to be repeated.

More generally, define a *sensor-path* to be a walk through $G_{h,r}$ which may optionally also include an initial and/or final vertical segment of length at most $\frac{\sqrt{3}}{2}r$ to the boundary ∂S_h of S_h. Sensor-paths can of course exist in any component of $G_{h,r}$.

Notice that the concepts of good components and sensor-paths are defined in terms of the *infinite* graph $G_{h,r}$, and are not related to the particular interval $[a, b]$ along the x-axis that we are concerned with.

We should also note that from now on, when we refer to, for example, a point to the right of a separating path or crossing sensor-path γ, we mean right in the topological sense of belonging to the infinite component of $S_h \setminus \gamma$ that is unbounded to the right. Hence points to the left of such a path may in fact have a larger x-coordinate than another point that is to the right of the same path. We now give some simple consequences of these definitions.

Lemma 6. *If C is a component of $G_{h,r}$ and v is a sensor that lies within distance $\frac{r}{2}$ of a sensor-path γ_C through C, then $v \in C$.*

Proof. Let z be the closest point on γ_C to v, so $\|z - v\| < \frac{r}{2}$. If z is on some edge $v_i v_{i+1}$ of γ_C then $\|v_i - v_{i+1}\| < r$, so either $\|z - v_i\| < \frac{r}{2}$ or $\|z - v_{i+1}\| < \frac{r}{2}$. Thus either $\|v - v_i\| < r$ or $\|v - v_{i+1}\| < r$ implying that v is adjacent to either v_i or v_{i+1} in $G_{h,r}$, and so $v \in C$. Now assume z lies on a vertical segment bv_1, say, meeting ∂S_h^-. If $z = v_1$ we are done, since then v is within distance $\frac{r}{2}$ of the sensor $v_1 \in C$. Otherwise, the segment vz must be horizontal and $\|v - v_1\|^2 = \|v - z\|^2 + \|z - v_1\|^2 < (\frac{r}{2})^2 + (\frac{\sqrt{3}}{2}r)^2 = r^2$. Thus once again, v is within distance r of a vertex of C, so $v \in C$. □

Lemma 7. *If C and C' are distinct components, then the minimum distance between any two sensor-paths γ_C and $\gamma_{C'}$, through C and C' respectively, is at least $\frac{r}{2}$.*

Proof. Let z and z' be the closest pair of points with z on γ_C and z' on $\gamma_{C'}$, and assume $\|z - z'\| < \frac{r}{2}$. Since the paths are piecewise linear, we may assume one of z and z' (say z) is a corner or an endpoint of its path. If z is a corner, then it is one of the vertices of C. But then $z \in C'$ by Lemma 6, a contradiction since C and C' are distinct components. Thus we may assume z is either a top t or bottom b boundary point of S_h on γ_C, say $z = b \in \partial S_h^-$. But since z is the closest point on a vertical segment of γ_C to z', and $z' \in S_h$, z' must be a bottom boundary point b' of S_h on $\gamma_{C'}$. But then either v_1 or v_1' must be within distance $\frac{r}{2}$ of $\gamma_{C'}$ or γ_C respectively, where bv_1 and $b'v_1'$ are vertical segments of γ_C and $\gamma_{C'}$. Thus either $v_1 \in C'$ or $v_1' \in C$, a contradiction. □

By Lemma 7 we see that crossing sensor-paths of distinct components cannot intersect. On the other hand, we can find a crossing sensor-path of a good component C that intersects all other crossing sensor-paths of C (by, for example, having it pass through every vertex of C). Thus we can order the good components from left to right according to the order of their crossing sensor-paths. Recall that a *break* is a gap between two consecutive good components.

Lemma 8. *Any separating path γ partitions the good components into those that lie to the left of γ and those that lie to the right of γ, so in particular defines a break. Conversely, for any break there exists such a separating path γ. Indeed, if the break occurs between good components C and C' then we can choose γ to lie between every pair of crossing sensor-paths γ_C and $\gamma_{C'}$ through C and C' respectively.*

Proof. If some sensor x lies to the left of γ and some sensor y lies to the right of γ, then the line segment xy meets γ at some point z, say. Since $\|x - z\|, \|y - z\| \geq \frac{r}{2}$ we have $\|x-y\| \geq r$, and so x and y are not adjacent in $G_{h,r}$. Thus any separating path disconnects $G_{h,r}$, and so defines a break. For the converse, suppose C is a good component. Define $S_C := (\bigcup_{\gamma_C} \gamma_C) \cup (\bigcup_{v \in C} \overline{D}_v)$ to be the closed region consisting of the union of all crossing sensor-paths γ_C through C and the closures of the sensing regions of C. Note that S_C forms a connected subset of the plane. Indeed, $\bigcup_{v \in C} \overline{D}_v$ is connected, and all crossing sensor-paths γ_C meet this set. No sensor outside of C is within distance $\frac{r}{2}$ of any γ_C (by Lemma 6) or within distance r of any sensor in C. Thus no sensor outside of C is within distance $\frac{r}{2}$ of S_C. Since each γ_C intersects both the top and bottom boundaries of S_h, so does S_C. Let γ be the rightmost boundary of S_C in S_h, i.e., the intersection of S_C with the closure of the component of $S_h \setminus S_C$ that is unbounded to the right. Now γ is a separating path (made up from line segments and arcs of circles), no part of which is to the left of any γ_C. If C' is a good component that lies to the right of C and $\gamma_{C'}$ is a crossing sensor-path through C', then no point of $\gamma_{C'}$ can lie to the left of γ, otherwise some point of $\gamma_{C'}$ would be within distance $\frac{r}{2}$ of some sensor in C, contradicting Lemma 6, or would meet some γ_C, contradicting Lemma 7. Thus γ lies between any pair of crossing sensor-paths γ_C and $\gamma_{C'}$. □

4 Widths of good components and breaks

Call a separating path *good* if it is as in Lemma 8, i.e., it does not cross any crossing sensor-path.

For any good component C define x_C^+ to be the smallest x-coordinate of any point on any good separating path to the right of C. Similarly define x_C^- to be the largest x-coordinate of any point on any good separating path to the left of C (see Figure 4). Define the *width* of a good component C as $x_C^+ - x_C^-$ and define the *width* of a break between consecutive good components C and C' as $x_{C'}^- - x_C^+$.

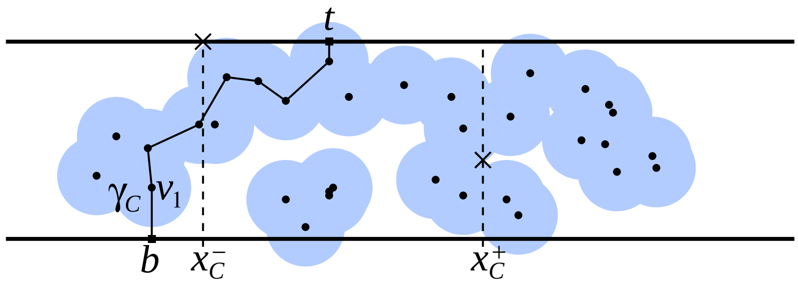

Figure 4: Crossing sensor-path γ_C (solid path) and $x_C^{\pm}$. Crosses indicate the rightmost point to the left of every γ_C, or leftmost point to the right of every γ_C that a good separating path can pass through. Note that *good* separating paths cannot pass beneath v_1.

Note that the width of a good component may be negative, as a good separating path to the right of C may extend further to the left than the rightmost point of a good separating path on the left of C. The width of a break, between C and C' say, is however always positive. In fact this width is (almost surely) the maximum horizontal extent of a good separating path γ between C and C'. Indeed, if there were good separating paths meeting (x_C^+, y_C^+) and $(x_{C'}^-, y_{C'}^-)$, say, but no good separating path meeting both, then there would be a collection C'' of sensors whose combined sensing region $\bigcup_{v\in C''} D_v$ separates these two points. But then there would be a good component between C and C'.

Lemma 9. *If γ is a separating path between good components C and C' that lies entirely in $[x,x']\times[0,h]$, then there is a good separating path between C and C' that also lies in $[x,x']\times[0,h]$. Also, some point of γ lies in $[x_C^+, x_{C'}^-]\times[0,h]$.*

Proof. Let γ^L be the good separating path constructed in Lemma 8 as the rightmost boundary path of $S_C = (\bigcup_{\gamma_C}\gamma_C)\cup(\bigcup_{v\in C}\overline{D}_v)$, where $\overline{D}_v$ denotes the closure of the sensing region D_v of v. The rightmost point (x_L, y_L) of γ^L lies on the boundary of some disk D_v, $v\in C$. As γ must pass to the right of v and cannot approach within distance $\frac{r}{2}$ of v, we must have $x'\geq x_L$. Thus replacing γ by the rightmost boundary of $\gamma\cup\gamma_L$ results in a separating path that does not cross any crossing sensor-path γ_C of C and still lies in $[x,x']\times[0,h]$. Repeating this process using the good separating path γ^R which is the leftmost boundary of $S_{C'} = (\bigcup_{\gamma_{C'}}\gamma_{C'})\cup(\bigcup_{v\in C'}\overline{D}_v)$ results in a good separating path in $[x,x']\times[0,h]$.

Now take $[x,x']$ minimal, so that $[x,x']$ is the projection onto the x-axis of γ. Then as any good separating path lies in $[x_C^+, x_{C'}^-]\times[0,h]$, $[x,x']$ must intersect $[x_C^+, x_{C'}^-]$, and so γ contains a point in $[x_C^+, x_{C'}^-]\times[0,h]$. □

As a consequence of Lemma 9, we see that Lemma 1 holds even if we insist on the separating path being good.

For all (fixed) $h, r > 0$, define the *intensity* $I_{h,r}$ of breaks as the limit

$$I_{h,r} = \lim_{\ell \to \infty} \frac{N_\ell}{\ell},$$

where N_ℓ is the number of breaks of $G_{h,r}$ containing separating paths in $[0, \ell] \times [0, h]$. It is easy to see that this limit exists. Indeed, translational invariance and long range independence imply that horizontal translation is an ergodic transformation on the probability space of this model. Thus almost surely, breaks (or any other event that can be defined in a translational invariant manner) occur with a well defined and deterministic frequency along the strip. In particular, $I_{h,r}$ is almost surely well defined and constant. It is clear that $I_{h,r}$ is also equal to the intensity of good components (defined in any reasonable way). Moreover, it makes sense to talk about widths of random good components and breaks. For example, if W is the width of a random break then $\mathbb{P}(W > x)$ is almost surely equal to the asymptotic proportion of breaks with width $W > x$, and is well defined for any $x \in \mathbb{R}$.

We now prove bounds on the average widths of breaks and good components. We shall need these to show that the average width of the breaks is relatively small compared with the average width of the good components when h is large. This is false when $r < r_c$, and is the main reason for the lower bound $r \geq 7$ assumed in our results. The bounds we obtain are relatively crude, but will be sufficient for our purposes.

Lemma 10. *Assume $r \geq 7$. Then the average width of a break is at most*

$$\max\{5h, 1/h + 2h\}.$$

Proof. We first deal with the case when $h < \frac{\sqrt{3}}{2}r$. In this case every vertex is within distance $\frac{\sqrt{3}}{2}r$ of both the top and the bottom of S_h, so in particular, every component is good. Consider two points of the Poisson process, (x_1, y_1) and (x_2, y_2), that are consecutive in the horizontal ordering of the points given by their x-coordinates. Counting breaks is equivalent to counting such pairs that lie in distinct components of $G_{h,r}$. Let $d = \max\{y_1, h - y_1\}$ and assume (x_i, y_i) lies in the component C_i, $i = 1, 2$. If $d \leq \frac{r}{2}$ then the sensing region of (x_1, y_1) covers the strip S_h up to x-coordinate $x_1 + (\frac{r^2}{4} - d^2)^{1/2} \geq x_1 + \frac{r}{2} - d \geq x_1 + \frac{r}{2} - h$. If $d > \frac{r}{2}$ then $h > \frac{r}{2}$, but there is still a vertical crossing sensor-path at $x = x_1$. Thus in all cases $x^+_{C_1} \geq x_1 + \frac{r}{2} - h$. Similarly $x^-_{C_2} \leq x_2 - \frac{r}{2} + h$, so the width of the break is at most $(x_2 - \frac{r}{2} + h) - (x_1 + \frac{r}{2} - h) = (x_2 - x_1 - r) + 2h$. Let $Z = x_2 - x_1 - r$. If $Z < 0$ a break may or may not be formed, but if it is then its width is at most $2h$. On the other hand, if $Z \geq 0$ then $x_2 - x_1 \geq r$, so there is always a break and its width is at most $Z + 2h$. Conditioned on $Z \geq 0$, Z is an exponential random variable with mean $1/h$. Thus the average width of a break is at most $1/h + 2h$ conditioned on $Z \geq 0$ and at most $2h$ otherwise. Thus the unconditioned average width of a break is also at most $1/h + 2h$.

We may now assume that $h \geq \frac{\sqrt{3}}{2}r$. We shall now bound the intensity of breaks with large width. Tile S_h with $a \times b$ rectangles, where a and b are chosen so that

(i) h/b is an integer;

(ii) the diameter $\sqrt{a^2+b^2}$ of these rectangles is $\frac{r}{2}$;

(iii) the area ab of these rectangles is maximized subject to (i) and (ii).

For large h/r, the rectangles are approximately square with side length $r/\sqrt{8}$, however to tile S_h we need h/b to be an integer, and this decreases the area ab slightly. It is easy to see that for any $h \geq \frac{\sqrt{3}}{2}r$,

$$h/b \geq 3, \qquad a, b \in \left[\tfrac{r}{\sqrt{12}}, \tfrac{r}{\sqrt{6}}\right], \qquad \text{and} \qquad ab \geq \tfrac{\sqrt{2}}{12}r^2. \tag{2}$$

The smallest area ab occurs when $h = \frac{\sqrt{3}}{2}r$, $h/b = 3$, and as one increases h, extreme values of a and b occur when the aspect ratio of the rectangles is $k : k+1$ for some $k \geq 3$.

Fix one of these tiling rectangles R adjacent to ∂S_h^-. We estimate the probability p that there exists a good separating path γ of width at least w meeting R, with R being the leftmost rectangle meeting both γ and ∂S_h^-. Take the set of rectangles that intersect γ and regard this set as the vertices of a graph G, rectangles being joined if they share a common edge. Thus G is a subgraph of the square lattice. The rectangles of G cannot contain any point of the Poisson process, since all the points of these rectangles lie within $\frac{r}{2}$ of γ and γ is a separating path. Ignoring probability zero events, it is clear that we can, without loss of generality, assume that γ goes through no corner of any rectangle, so that G is a connected graph that joins R to ∂S_h^+, and joins the leftmost rectangle to the rightmost rectangle meeting γ. It is easy to see that one can take a connected subset G' of these rectangles containing exactly $n = h/b + \lfloor w/a \rfloor$ rectangles, still meeting our original rectangle R, although G' may now fail to meet the top of S_h or the leftmost/rightmost rectangles. For example, any spanning tree of G rooted at R contains at least $\frac{h}{b} - 1$ vertical and at least $\lfloor w/a \rfloor$ horizontal edges, and so G contains at least n vertices. Now repeatedly prune leaves until G' contains exactly n vertices. Is is well known that the number of connected subgraphs of order n of a graph of maximum degree $\Delta \geq 3$ containing a specified vertex is at most $(e(\Delta-1))^{n-1}$, where e is Euler's constant (see, for example, [5, problem 45]). Thus the number of such choices of the subgraph G' of the square lattice rooted at R is at most $(3e)^{n-1}$. However, we can do a bit better. Using the results of Klarner and Rivest [10] (equation (7) and the preceding discussion[3]) the number of connected subsets

[3]Although they prove a slightly sharper asymptotic bound on the number of these "lattice animals", it is not so obvious what their stronger upper bound is for fixed n. Hence we use this slightly weaker result.

of size n of the square lattice up to translation is at most

$$\sum_k \binom{n}{k,k+1,n-2k-1}2^{n-k-1} = \frac{1}{2}\bigg(\sum_k \binom{n}{k,k+1,n-2k-1}2^{n-k-1} + \sum_k \binom{n}{k+1,k,n-2k-1}2^{n-k-1}\bigg)$$
$$\leq \frac{1}{2}\sum_{i+j+\ell=n} \binom{n}{i,j,\ell}2^{n-(i+j+1)/2}$$
$$= \frac{1}{2\sqrt{2}}2^n(1+\tfrac{1}{\sqrt{2}}+\tfrac{1}{\sqrt{2}})^n$$
$$< \frac{1}{2}(2+2\sqrt{2})^n.$$

As we shall use this repeatedly, we shall state it as a lemma.

Lemma 11. *For all $n \geq 1$, the number of equivalence classes of connected subsets of the square lattice of size n up to translation is at most $\frac{1}{2}\mu^n$, where $\mu = 2+2\sqrt{2}$.* □

Since we assumed that R is the leftmost rectangle on the lowest level of this subset, fixing R determines which translate of this set we have. Thus there are at most $\frac{1}{2}\mu^n$ possible choices for G', where $\mu = 2+2\sqrt{2}$. These rectangles form an empty region of area nab, and so the probability p that there exists such a separating path γ satisfies

$$p \leq \tfrac{1}{2}\mu^n e^{-nab} \leq \tfrac{1}{2}(\mu e^{-ab})^n \leq \tfrac{1}{2}(\mu e^{-ab})^{h/b+w/a-1},$$

where we are using the fact that $\lfloor w/a \rfloor \geq w/a-1$ and $\mu e^{-ab} < 1$. It is conceivable (if rather implausible) that two breaks may give rise to the same graph G'. However, it is clear that three breaks cannot all give rise to the same graph. Indeed, if γ, γ', γ'' are good separating paths giving rise to three successive breaks, then there are crossing sensor-paths γ_C and $\gamma_{C'}$ corresponding to good components C and C' such that $\gamma < \gamma_C < \gamma' < \gamma_{C'} < \gamma''$ in the left-right ordering of these paths (none of these paths cross). Since no point of γ_C is within $\frac{r}{2}$ of any point of $\gamma_{C'}$, it follows that no point of γ is within distance $\frac{r}{2}$ of any point of γ'', and thus the sets of rectangles these paths pass through are disjoint. Consequently, the intensity $I_{h,r}(w)$ of breaks of width at least w can be bounded above by $2p/a$, that is

$$I_{h,r}(w) \leq \tfrac{1}{a}(\mu e^{-ab})^{h/b+w/a-1}.$$

Now we need a lower bound on the total intensity of breaks. Consider a vertical strip of rectangles of the tiling, each containing at least one point of $\mathcal{P}$. The vertices in these rectangles will form (part of) a good component. If the $r \times h$ rectangle immediately to the right of this strip is empty, it will certainly force a break. The probability of this occurrence is

$$(1-e^{-ab})^{h/b}e^{-hr},$$

so the intensity $I_{h,r}$ of breaks can be bounded below by

$$I_{h,r} \geq \tfrac{1}{a}(1 - e^{-ab})^{h/b} e^{-hr}. \tag{3}$$

Hence, if W is the (positive-valued) random variable expressing the width of a break and $w \geq a$, then

$$\begin{aligned}
\mathbb{P}(W \geq w) &= \tfrac{I_{h,r}(w)}{I_{h,r}} \\
&\leq (\mu e^{-ab})^{h/b + w/a - 1} (1 - e^{-ab})^{-h/b} e^{hr} \\
&\leq e^{hr} \left(\frac{\mu e^{-ab}}{1 - e^{-ab}} \right)^{h/b + (w-a)/a} \\
&= e^{(r - \kappa a)h - \kappa b(w-a)}
\end{aligned} \tag{4}$$

where

$$\kappa := -\frac{1}{ab} \log \left(\frac{\mu e^{-ab}}{1 - e^{-ab}} \right) = 1 - \frac{1}{ab} \log \left(\frac{\mu}{1 - e^{-ab}} \right) \geq 0.72. \tag{5}$$

The last inequality here follows from the fact that $\mu = 2 + 2\sqrt{2}$ and, by (2), $ab \geq \frac{\sqrt{2}}{12} r^2 \geq 5.77$ as $r \geq 7$. Now, for all $x \geq a$ we have by (4)

$$\mathbb{E}(W) \leq x + \int_x^\infty \mathbb{P}(W \geq w)\, dw \leq x + \tfrac{1}{\kappa b} e^{(r - \kappa a)h - \kappa b(x-a)}.$$

To minimize the right hand side, we set $x = a + \frac{(r-\kappa a)h}{\kappa b} \geq a$, at which point the exponent in the second term is 0. Hence

$$\mathbb{E}(W) \leq a + \frac{(r - \kappa a)h}{\kappa b} + \frac{1}{\kappa b} = \frac{r}{\kappa b} h + \frac{\kappa ab - \kappa ah + 1}{\kappa b}.$$

Now $h/b \geq 3$ and $\kappa ab \geq (0.72)(5.77) > 1$, so $\kappa ab - \kappa ah + 1 \leq -2\kappa ab + 1 < 0$. Also, by (2), $b \geq r/\sqrt{12}$, so

$$\mathbb{E}(W) \leq (\sqrt{12}/\kappa) h \leq 5h.$$

□

Lemma 12. *Assume $r \geq 7$. Then the proportion of good components with width less than $w \geq 0$ is at most*

$$(w + c_W) e^{-h/3},$$

where c_W is some absolute constant.

Proof. As in the proof of Lemma 10, we consider the case $h < \frac{\sqrt{3}}{2}r$ first. In this case all components are good. The width of a good component is at least as large as the horizontal distance between its leftmost and rightmost points as we can take vertical crossing sensor-paths through these points and no good separating path can cross these. Any two points in S_h within horizontal distance $3 < \frac{r}{2}$ are within distance r of each other, and so are in the same component. Thus we can stochastically bound from below the width of a good component by a random variable $X = \sum_{i=1}^{T-1} X_i$, where X_i are i.i.d. exponential random variables of mean $1/h$ giving the horizontal distance between consecutive sensors, and $T = \min\{i : X_i \geq 3\}$. Consider $\mathbb{E}e^{-\alpha X}$ for $\alpha > 0$. Conditioning on the value of X_1 we obtain

$$\begin{aligned}\mathbb{E}(e^{-\alpha X}) &= \mathbb{E}\big(1_{\{X_1\geq 3\}} + e^{-\alpha(X_1+X')}1_{\{X_1<3\}}\big)\\ &= \mathbb{P}(X_1 \geq 3) + \int_0^3 \mathbb{E}(e^{-\alpha(z+X')})he^{-hz}\,dz\\ &= e^{-3h} + \mathbb{E}(e^{-\alpha X'})\int_0^3 he^{-(h+\alpha)z}dz\\ &\leq e^{-3h} + \mathbb{E}(e^{-\alpha X'})\tfrac{h}{h+\alpha},\end{aligned}$$

where $X' = \sum_{i=2}^{T'-1} X_i$, $T' = \min\{i > 1 : X_i \geq 3\}$, has the same distribution as X. Since $e^{-\alpha X}$ is bounded,

$$\mathbb{E}(e^{-\alpha X}) \leq \tfrac{h+\alpha}{\alpha}e^{-3h},$$

so

$$\mathbb{P}(X < w) \leq \mathbb{E}(e^{\alpha w-\alpha X}) \leq \tfrac{h+\alpha}{\alpha}e^{\alpha w}e^{-3h}.$$

Setting $\alpha = 1/w$ gives $\mathbb{P}(X < w) \leq (wh+1)e^{1-3h}$. But $he^{1-h} \leq 1$ for all h, so

$$\mathbb{P}(X < w) \leq we^{-2h} + ee^{-3h} \leq (w + c_W)e^{-h/3}$$

(with room to spare).

Turning to the heart of the proof, we now assume $h \geq \frac{\sqrt{3}}{2}r$. For each good component C we shall define a "leftmost" crossing sensor-path γ_C^L through C. Given any good separating path γ to the left of C, we "explore" the Poisson process $\mathcal{P}$ to the right of γ in order to find γ_C^L. Let U be the set of points $u \in S_h$ that are at distance at least $(\sqrt{3}-1)\frac{r}{2}$ from the boundary of S_h, are to the right of γ, and such that $D_u = \{x : \|x-u\| < \frac{r}{2}\}$ does not contain any vertex of $\mathcal{P}$. Let U_0 be the component of U (as a subset of $\mathbb{R}^2$) that is connected to γ. Note that one can determine U_0 by examining just the union of the regions $\overline{D}_u$ with $u \in U_0$ (not just $u \in U$). Indeed, one can imagine the points of $\mathcal{P}$ as pegs, and $\overline{D}_u$ a disk that can be moved around as long as it is not obstructed by these pegs, and does not get too close to ∂S_h. Then U_0 is the region that the center u is allowed to trace out. Let $S = \bigcup_{u\in U_0} \overline{D}_u$, then the rightmost boundary of S consists of a sequence of arcs of circles, these circles joining points of $\mathcal{P}$ or ∂S_h. The points of $\mathcal{P}$ determining this boundary form the path γ_C^L (see Figure 5). Indeed, each such point must be strictly within r of the next one, or within $\frac{\sqrt{3}}{2}r$ of the boundary of S_h. Thus they form a crossing sensor-path. Conversely, given any crossing sensor-path γ_C, U_0 must lie to the left of γ_C as the center u of the disk $\overline{D}_u$ cannot cross an edge between two vertices of γ_C and cannot cross the vertical segments of γ_C joining it to the boundary of S_h. This justifies our description of γ_C^L as the "leftmost" crossing sensor-path, although it is possible that γ_C^L contains redundant vertices that lie to the right of some shorter γ_C. It is also possible that γ_L^C fails to contain the leftmost vertex of C in the case when this vertex is close to ∂S_h. Note also that some vertices of γ_C^L may lie in the interior of S (see Figure 5).

Since there are good separating paths to the left of C that go through any given point of U_0, the region S (and hence γ_C^L) lies entirely to the left of the vertical line $x = x_C^- + \frac{r}{2}$.

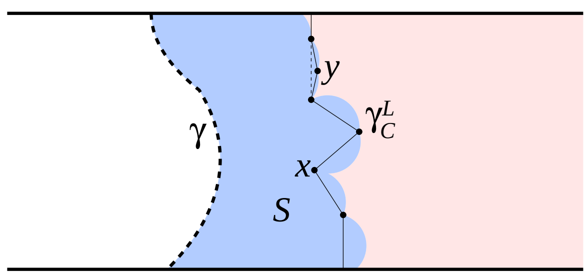

Figure 5: Leftmost crossing sensor-path γ_C^L of good component C. Region to the right is "unexplored". Dotted line is a separating path to the left of C. Note that x lies in the interior of S, and y could be removed to form a shorter path to the left of γ_C^L. Also, near the top of γ_C^L there is a region to the left of γ_C^L that is not in S. This could potentially contain some vertices of C, and even the leftmost vertex of C.

We now aim to bound the width of the good component conditioned on particular values of the sets S and $\mathcal{P} \cap S$ (and hence also on γ_C^L). It is important in what follows that we have not conditioned on anything to the right of S. Tile the strip S_h with $a \times b$ rectangles as in the proof of Lemma 10. For definiteness, align the rectangles horizontally so that the line $x = x_C^- + \frac{r}{2}$ lies on a vertical boundary between two rectangles. Recall that S lies entirely to the left of this line.

We now bound the probability that the width W of C is less than w in two stages. First we bound the probability that $\frac{r}{2} \le W < w$. Consider a good separating path γ to the right of C that meets the line $x = x_C^+$ and mark all the rectangles that intersect γ. Then no marked rectangle contains a sensor. If $W \ge \frac{r}{2}$ then $x_C^+ \ge x_C^- + \frac{r}{2}$, so γ must stay to the right of the line $x = x_C^- + \frac{r}{2}$, and thus none of the marked rectangles intersect the region S. Also, at least one of the marked rectangles lies in the column L of rectangles intersecting the line $x = x_C^+$.

As in the proof of Lemma 10, we take the connected subgraph G of marked rectangles, so that G joins the top to the bottom of S_h, consists entirely of empty rectangles, and also intersects L. We bound the probability that any such graph G exists meeting L (ignoring the requirement that L is in fact the leftmost column intersecting G). Let R be the leftmost rectangle of G that meets the bottom boundary of S_h. Suppose R is horizontally k rectangles to the right of L (or $-k$ rectangles to the left of L). Then there are at least $h/b + |k|$ rectangles in G. Moreover, R is uniquely determined by k and L. All these rectangles are empty, and are in regions not yet examined when constructing γ_C^L. Thus, by Lemma 11, we can bound the probability that such a γ exists with a given choice of R by

$$\sum_{n \geq h/b+|k|} \tfrac{1}{2}\mu^n e^{-nab} = \tfrac{1}{2}(\mu e^{-ab})^{h/b+|k|}/(1-\mu e^{-ab}),$$

where $\mu = 2 + 2\sqrt{2}$. Summing over $k \in \mathbb{Z}$, we have a bound of

$$p_0 = \tfrac{1}{2}(\mu e^{-ab})^{h/b}/(1-\mu e^{-ab})^3 \leq \tfrac{1}{2}\big(\tfrac{\mu e^{-ab}}{1-\mu e^{-ab}}\big)^{h/b}$$

on the probability that any γ exists intersecting L. (Here we have used $\sum x^{|k|} = 1 + 2x + 2x^2 + \cdots \leq 1/(1-x)^2$ and $h/b \geq 3$ from (2).) As in the proof of Lemma 10, we can write this last expression as $\frac{1}{2}e^{-\kappa ah}$, where $\kappa \geq 0.72$ by (5). But $\kappa a \geq \kappa \frac{r}{\sqrt{12}} \geq 1$, so

$$p_0 \leq e^{-h}. \tag{6}$$

As L contains the line $x = x_C^+$ and $x_C^- + \frac{r}{2} \leq x_C^+ < x_C^- + w$, there are at most $\lceil (w - \frac{r}{2})/a \rceil$ choices for the column L and so

$$\mathbb{P}(\tfrac{r}{2} \leq W < w) \leq \lceil (w - \tfrac{r}{2})/a \rceil p_0 \leq w e^{-h/3}, \tag{7}$$

where we have used the facts (from (2)) that $\frac{r}{2} \geq a$ and $a \geq \frac{r}{\sqrt{12}} \geq 1$.

Now we consider the probability that $W < \frac{r}{2}$. Unfortunately, in this case it is possible for some of the rectangles of G to intersect the set S, which we have already conditioned on. Indeed, if we are not careful, almost all of G may be covered by S. Dealing with this case is therefore significantly more complicated.

If $W < \frac{r}{2}$, the separating path γ must pass to the left of $x = x_C^- + \frac{r}{2}$ (as it meets $x = x_C^+$ and $x_C^+ - x_C^- = W < \frac{r}{2}$) and also to the right of $x = x_C^-$ (as it is to the right of γ_C^L). Inductively add to G any rectangles adjacent to G and (topologically) to the right of G that contain no point of $\mathcal{P}$, making G a maximal connected set of empty rectangles each of which either intersects γ or is to the right of γ. If one can then remove rectangles from G that intersect S and still connect the top and bottom of S_h, then do this. We may have removed all rectangles intersecting $x = x_C^+$ in this process, but at least one of the remaining rectangles must be adjacent to a rectangle that is to the left of $x = x_C^- + \frac{r}{2}$, either because it is adjacent to a rectangle meeting S that was removed, or because no rectangles were removed and γ passes to the left of $x = x_C^- + \frac{r}{2}$. In particular, in the case when G now does not meet S, the remaining set of rectangles meets one of three columns L, the one meeting $x = x_C^-$, $x = x_C^- + a$, or $x = x_C^- + 2a > x_C^- + \frac{r}{2}$. Thus we can bound the probability that such a G exists by $3p_0$ where p_0 is as above.

If every crossing path in G meets S, then by removing rectangles if necessary, we may assume G is a path from the bottom to the top of S_h. By taking G to be the shortest such path, we may assume that only one of our rectangles is adjacent to ∂S_h^-, and only one of our rectangles is adjacent to ∂S_h^+. By assumption, G still meets S.

If a large portion of G is covered by S then we will have difficulty proving a sufficiently strong bound on the probability of the existence of γ. Thus we need to add some extra area that we know to be devoid of sensors. Let B be the union of the rectangles of G and let T be the region topologically to the right of G (and hence to the right of γ) that lies within distance r of some vertex of γ_C^L. Then we know that T contains no sensors, as any such sensor would be to the right of γ but joined to the component C. Thus $B \cup T$ contains no sensors and $(B \cup T) \setminus S$ has not been examined. There are at most 3^{n-3} possible paths G of length n starting at a fixed rectangle R as there are at most 3 choices at each of $n - 1$ steps, but the first and last steps go in a known direction. Thus we can bound the probability that C is of width $\leq \frac{r}{2}$ by

$$p_1 = 3p_0 + \sum_{n \geq h/b} (2 + 2(n - h/b))3^{n-3} \sup_{|G|=n} e^{-|(B \cup T) \setminus S|}.$$

Here we have included $3p_0$ to take into account the case when we could have removed rectangles to obtain a configuration that avoids S, together with $2 + 2(n - h/b)$ times the probability of a path from a fixed R crossing S_h. The factor of $2 + 2(n - h/b)$ counts the possible choices of R, given that the path can make at most $n - h/b$ horizontal steps, and must intersect one of the two columns intersecting $x = x_C^-$ or $x = x_C^- + a$ if it is to meet S (which is to the left of $x = x_C^- + \frac{r}{2} < x_C^- + 2a$) and also pass to the right of γ_C^L (and hence to the right of $x = x_C^-$).

Using (6), and the bound (8) proved below in Lemma 13, we have

$$p_1 \le 3p_0 + \sum_{n \ge h/b} 2(1 + (n - h/b))3^{n-3} e^{-(\pi r^2/96)\max\{n-h/b-4,0\} - (rb/(4\sqrt{3}))(h/b-1)}$$
$$\le 3e^{-h} + 3^{h/b-3} e^{-(rb/(4\sqrt{3}))(h/b-1)} \sum_{i=0}^{\infty} 2(1+i)3^i e^{-(\pi r^2/96)\max\{i-4,0\}}.$$

Now $r \ge 7$, so $3e^{-\pi r^2/96} < 0.7 < 1$. Thus the last sum above converges and is bounded by an absolute constant $c > 0$. Also, $r \ge 4\sqrt{3}$, so

$$p_1 \le 3e^{-h} + 3^{h/b-3} e^{-b(h/b-1)} c = 3e^{-h} + 3^{-2} c e^{-(2b/3 - \log(3))(h/b-1) + b/3 - h/3}.$$

Now $b \ge r/\sqrt{12} \ge 2$, so $2b/3 - \log(3) \ge 4/3 - \log(3) > 0$. Thus we can replace h/b by its smallest value $h/b = 3$ to get

$$p_1 \le 3e^{-h} + c3^{-2} e^{-2(2b/3 - \log 3) + b/3 - h/3} \le 3e^{-h} + ce^{-b} e^{-h/3}.$$

But $b \ge 2$ and $h \ge \frac{\sqrt{3}}{2} r \ge 5$, so

$$\mathbb{P}(W < \tfrac{r}{2}) \le p_1 \le (3e^{-10/3} + ce^{-2}) e^{-h/3} = c_W e^{-h/3}.$$

The result then follows on adding (7). □

Lemma 13. *Let G, B, and T be as in the proof of Lemma 12. Then*

$$|(B \cup T) \setminus S| \ge \tfrac{\pi r^2}{96} \max\{n - \tfrac{h}{b} - 4, 0\} + \tfrac{rb}{4\sqrt{3}}(\tfrac{h}{b} - 1), \qquad (8)$$

where n is the number of rectangles in G.

Proof. Recall that G is a minimal path of empty rectangles joining ∂S_h^- to ∂S_h^+, B is the union of these rectangles, and T is the region topologically to the right of B that is within distance r of some vertex of γ_C^L.

We first show that any point P topologically to the right of $B \cup T$ that is not in the bottom or top row of rectangles must be at least distance $c := r/\sqrt{12}$ from S. Note that $c \le a, b$ by (2). Any point on the right boundary of S is either on an arc of a circle between two sensors of γ_C^L, or on an arc of a circle joining a sensor to the boundary of S_h. In both cases one can check using simple geometry that P is at distance at least c from S (see Figure 6). In the case when the closest point on ∂S to P is on an arc joining two vertices v_i and v_{i+1} of γ_C^L, the minimum distance of P from S is at least

$$x := \tfrac{\sqrt{3}}{2} r - \tfrac{r}{2} \ge \tfrac{r}{\sqrt{12}} = c.$$

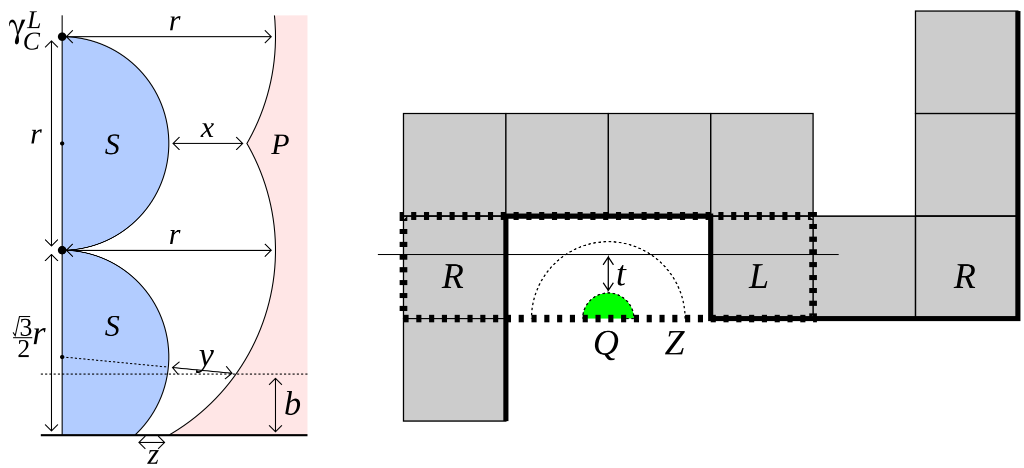

Figure 6: Left: Points P at distance at least r from any vertex of γ_C^L are at horizontal distance $z \geq \frac{c}{2}$ from S, and if P is at least b from ∂S_h then it is also at distance at least $\min\{x, y\} \geq c$ from S. Right: Horizontal line through Z that approaches within distance t of Q intersects $(B \cup T) \setminus S$ in an interval of length at least $2(c^2 - t^2)^{1/2}$.

The worst case however is near the boundary when the closest point to P on ∂S is on an arc meeting ∂S_h. In this case P could be within distance

$$y := \left(r^2 - (\tfrac{\sqrt{3}}{2}r - b)^2 + (\tfrac{\sqrt{3}}{2}r - \tfrac{r}{2} - b)^2\right)^{1/2} - \tfrac{r}{2} \geq \tfrac{r}{\sqrt{12}} = c$$

of S. If P is in the top or bottom rows of rectangles it can be closer to S, however in all cases it must be at least a horizontal distance of

$$z := \tfrac{r}{2} - \left(\tfrac{r^2}{4} - (\tfrac{\sqrt{3}}{2}r - \tfrac{r}{2})^2\right)^{1/2} \geq \tfrac{r}{2\sqrt{12}} = \tfrac{c}{2}$$

from S. It is also a vertical distance of at least c from S (since it is not vertically above or below the arc of ∂S adjacent to the corresponding boundary of S_h).

We now estimate the area $|(B \cup T) \setminus S|$ in two ways. First we estimate the area in a row of rectangles, and add up the contributions from the h/b rows, then we estimate by columns. Averaging these two estimates will then give the result.

First fix a horizontal row of rectangles. We call a rectangle R in this row *right-facing* if R lies in G, the rectangle to the right of R does not lie in G, and the rectangle to the right of R is in the component (topologically) to the right of the path G of rectangles. We call a rectangle L in this row *left-facing* if L lies in G, the rectangle to the left of L does not lie in G, but the rectangle to the left of L is in the component to the *right* of G. (This can occur if the path of rectangles G loops round and travels downwards across this row, see Figure 6.) Note that some rectangles of G may be neither left nor right-facing, but no rectangle can be both as the path G is of minimal length. Since the path begins at the bottom of S_h and ends at the top, there will be one more right-facing than left-facing rectangle in this row. Consider the rightmost rectangle R of G in this row. Clearly it is right-facing. The right boundary of the excluded area T is at horizontal distance at least c from S ($c/2$ if the row is adjacent to ∂S_h). Thus any horizontal line $y = y_0$ intersects $(B \cup T) \setminus S$ in an interval of length at least $\min\{a, c\} = c$ (or $\min\{a, c/2\} = c/2$) to the right of the left boundary of R. Integrating over y gives a contribution of bc (or $bc/2$) for this rectangle to the area in $(B \cup T) \setminus S$ that lies either in, or to the right, of R. The remaining left and right-facing rectangles can be paired up, with a right-facing rectangle R followed by some rectangles not in G, followed by a left-facing rectangle L. Note that rows adjacent to the boundary of S_h contain only a single right-facing rectangle, so we may assume that the row is not adjacent to ∂S_h. Let Z be the rectangular region consisting of R, L, and the rectangles between R and L on this row (see Figure 6). As the path G is minimal, not all of Z lies in $B \cup T$. (Otherwise adding Z to B and cutting off the "loop" between R and L would shorten the path of rectangles.) Let Q be the region in Z that does not lie in $B \cup T$. Then Q is at distance at least c from S. Consider the intersection I of a horizontal line $y = y_0$ with $Z \setminus (Q \cup S) = ((B \cup T) \setminus S) \cap Z$. If the line $y = y_0$ meets Q, then there is an interval of length at least $\min\{a, c\} = c$ in I to the left of Q, and an interval of length at least $\min\{a, c\} = c$ in I to the right of Q that does not meet S. If the line $y = y_0$ passes within distance t of Q, then there is an interval of length $2(c^2 - t^2)^{1/2}$ in I all of whose points are within distance c of Q and so does not meet S. Integrating over y thus gives an area in $((B \cup T) \setminus S) \cap Z$ of at least $\pi c^2/2$, the minimum area of intersection with Z of the disk of radius c about a point in $Z \setminus (R \cup L)$. Thus, summing over rows, we have

$$|(B \cup T) \setminus S| \geq (\tfrac{h}{b} - 2)bc + 2\tfrac{bc}{2} + \tfrac{1}{2}(n_h - \tfrac{h}{b})\tfrac{\pi c^2}{2}, \tag{9}$$

where n_h is the total number of left- or right-facing rectangles. A similar calculation can be performed for columns of rectangles, giving

$$|(B \cup T) \setminus S| \geq \tfrac{1}{2} \max\{n_v - 4, 0\}\tfrac{\pi c^2}{2}, \tag{10}$$

where n_v is the total number of up- or down-facing rectangles. Here pairs of up- and downward facing rectangles give a contribution of at least $\frac{\pi c^2}{2}$. Most unpaired "outward facing" rectangles R' contribute at least $\frac{\pi c^2}{4}$. Indeed, we can use the same argument as for paired rectangles, but limit the integration to the side of the region Q that faces the rectangle. There is however one case in which this argument can fail. Although rectangles adjacent to ∂S_h are never outward-facing, rectangles on the second row from the top or bottom of S_h can be. In this case, the adjacent rectangle R'' to R' borders ∂S_h, and so points in $Q = R' \setminus T$ may not be at distance at least c from S. This only affects the argument if R' is within horizontal distance $\frac{r}{2}$ of the vertical segments of γ_C joined to $\partial S_h^\pm$ as it is only in this case does a part of S within distance c of Q meet the vertical column through R'. However, the right hand side of B is to the right of γ_C, and $2a > \frac{r}{2}$, so this can affect at most two rectangles R' near ∂S_h^- and two rectangles R' near ∂S_h^+. Excluding these four possible rectangles, we have a total contribution at of least $(n_v - 4)\frac{\pi c^2}{4}$ as required.

As we travel along the path G, $n_v + n_h$ counts the number of segments of the right boundary of B. Thus each counterclockwise turn adds 2 to $n_v + n_h$, and each clockwise turn adds 0 to $n_v + n_h$. Each straight edge adds 1, so since there are as many clockwise as counterclockwise turns we have $n_v + n_h = n$, the total number of rectangles. Thus by averaging (9) and (10) we get

$$|(B \cup T) \setminus S| \geq \tfrac{\pi c^2}{8} \max\{n - \tfrac{h}{b} - 4, 0\} + \tfrac{bc}{2}(\tfrac{h}{b} - 1).$$

The result follows as $c = r/\sqrt{12}$. □

5 The distribution of breaks

To show that breaks have an approximately Poisson distribution, we use the Stein-Chen method, a particularly convenient form of which can be found in [1]. The following theorem is immediate consequence of Theorem 1 of [1].

Theorem 14. *Let $\{X_I : I \in \mathcal{I}\}$ be a collection of Bernoulli random variables indexed by a countable collection of subsets I of some ground set. Suppose that $\sum_I \mathbb{E}(X_I) = \lambda$, and for each I, X_I is independent of $\{X_J : I \cap J = \emptyset\}$. Let $b_1 = \sum_{I,J\colon I\cap J\neq\emptyset} \mathbb{E}(X_I)\mathbb{E}(X_J)$, $b_2 = \sum_{I,J\colon I\cap J\neq\emptyset,\, I\neq J} \mathbb{E}(X_I X_J)$, and write $X = \sum_I X_I$. Then for all k,*

$$|\,\mathbb{P}(X = k) - e^{-\lambda}\lambda^k/k!\,| \leq b_1 + b_2.$$

□

Theorem 1 in [1] includes another term b_3 that bounds dependency when $I \cap J = \emptyset$, but in our case $b_3 = 0$. Also, the sums in the definition of b_1 and b_2 are over *ordered* pairs (I, J).

Proof of Theorem 4. Let W_g be a random variable giving the width of a good component, and W_b be a random variable giving the width of the subsequent break. Then $W_b \geq 0$, $W_g + W_b \geq 0$, and

$$\mathbb{E}(W_g + W_b)I_{h,r} = 1.$$

By Lemma 12, $\mathbb{E}(W_g)$ grows exponentially with h, while by Lemma 10, $\mathbb{E}(W_b)$ grows at most linearly with h. Thus $\mathbb{E}(W_b)/\mathbb{E}(W_g + W_b) \to 0$, and so $\mathbb{E}(W_b)I_{h,r} \to 0$ as $h \to \infty$. Set

$$\ell = x/I_{h,r} = x\mathbb{E}(W_g + W_b),$$

so that $\ell \to \infty$ as $h \to \infty$.

Let $I \subseteq [0, \ell]$ be a half-open interval with integer endpoints and length $|I| > r$, say $I = [p, q)$ with $q - p > r$, $p, q \in \mathbb{Z}$. We denote by $\mathcal{I}$ the set of such intervals, and given $I = [p, q) \in \mathcal{I}$ write $I' = [p + \frac{r}{2}, q - \frac{r}{2}]$. Define B_I to be the event that there are two crossing sensor-paths, γ_1 and γ_2, that lie in $I' \times [0, h]$, and a separating path lies in $I' \times [0, h]$ between γ_1 and γ_2 (so γ_1 and γ_2 go through distinct good components). Clearly the event B_I depends only on the Poisson process in $I \times [0, h]$. Now let X_I be the indicator function of the event that B_I holds, but that B_J does not hold for any proper subinterval J of I with $J \in \mathcal{I}$. Clearly X_I also depends only on the Poisson process in $I \times [0, h]$ and if $X_I = 1$ then B_I holds, so there is a break. Moreover this break has break interval $[x_C^+, x_{C'}^-] \subseteq I' \subseteq I$ as no good separating path corresponding to this break can cross the crossing sensor-paths γ_1 and γ_2 in $I' \times [0, h]$.

From the construction of the leftmost crossing sensor-path in Lemma 12, there is always a crossing sensor-path through a good component C that lies entirely to the left of $x_C^- + \frac{r}{2}$. Similarly, there is a crossing sensor-path entirely to the right of $x_C^+ - \frac{r}{2}$. Thus, every break interval $[x_C^+, x_{C'}^-]$ that is a subset of $[r, \ell - r]$ results in an interval I with $X_I = 1$. Indeed, we have

$$I \subseteq [\lfloor x_C^+ - r \rfloor, \lceil x_{C'}^- + r \rceil), \tag{11}$$

so the width of I is at most $2r + 2$ more than the width of the break it contains.

If there are two breaks resulting in the event B_I holding, then there must be distinct proper subintervals $J_1, J_2 \subset I$, $J_1, J_2 \in \mathcal{I}$, such that B_{J_1} and B_{J_2} hold. This is because there would be three crossing sensor-paths, γ_1, γ_2, γ_3, in $I' \times [0, h]$ corresponding to three distinct good components. But these would have to be separated by a distance of at least $\frac{r}{2} > 3$ from each other. Hence B_J would hold with $J = [p, q - 3)$ and with $J = [p + 3, q)$. Thus each I with $X_I = 1$ corresponds to a single break. Conversely, each break can give rise to only one minimal interval $I \in \mathcal{I}$ with $X_I = 1$. Hence if no break interval intersects either $[0, r]$ or $[\ell - r, \ell]$, then $X = \sum_{I \in \mathcal{I}} X_I$ exactly counts the number of breaks in $G_{h,r}(0, x/I_{h,r})$, and is also the number of coordinates x_C^+ that lie in $[0, \ell]$. The probability that X is not equal to the number of breaks in $G_{h,r}(0, x/I_{h,r})$ is at most the probability that a break intersects either $[0, r]$ or $[\ell - r, \ell]$. But the probability that a break intersects $[0, r] \cup [\ell - r, \ell]$ is the probability that the break starts in $[-W_b, r] \cup [\ell - r - W_b, \ell]$, where W_b is its width. By translational invariance, this probability is is at most

$$b_0 := 2(\mathbb{E}(W_b) + r) I_{h,r} \leq (10h + 2r) I_{h,r}$$

by Lemma 10 when $h \geq 1$. But $I_{h,r}$ tends to 0 exponentially fast with h (by Lemma 12). Thus $b_0 \to 0$, and so $\mathbb{E}(X) \to x$, as $h \to \infty$.

From the definition of X_I it is clear that X_I is independent of all X_J with $I \cap J = \emptyset$. If $I \cap J \neq \emptyset$, $I \neq J$, and $X_I X_J = 1$, then by (11) we have $\lfloor x_C^+ - r \rfloor < \lceil x_C^- + r \rceil$ for any good component C between the breaks corresponding to I and J, and hence all these good components must have width

$$x_C^+ - x_C^- \leq \lfloor x_C^+ - r \rfloor - \lceil x_C^- + r \rceil + 2r + 2 \leq 2r + 1.$$

In particular, one of the two good components adjacent to the break given by $I \in \mathcal{I}$ must have width at most $2r+1$. Thus, by Lemma 12, the expected number of I's such that $X_I X_J = 1$ for some $J \neq I$, $I \cap J \neq \emptyset$, is at most $2(2r+1+c_W)e^{-h/3}\ell I_{h,r} = 2(2r+1+c_W)e^{-h/3}x$. Let $I = [p,q) \in \mathcal{I}$ and suppose there is such a J. There must be a crossing sensor-path γ_C in $I \times [0,h]$ intersecting $[q - \frac{r}{2} - 1, q] \times [0,h]$ or else $B_{[p,q-1)}$ would hold. Each $J = [p',q') \in \mathcal{I}$ with $X_J = 1$ corresponding to a break on the right of I must contain a crossing sensor-path meeting $[p' + \frac{r}{2}, p' + \frac{r}{2} + 1] \times [0,h]$, otherwise $B_{[p'+1,q')}$ would hold. As distinct J's give rise to distinct breaks, all but at most one of these crossing sensor-paths must lie to the right of γ_C. Thus if there are k intervals J intersecting I and belonging to a break on the right of I, then there are $k-1$ components with crossing sensor-paths intersecting $[q - \frac{r}{2} - 1, p' + \frac{r}{2} + 1] \times [0,h] \subseteq [q - \frac{r}{2} - 1, q + \frac{r}{2}] \times [0,h]$. However, we cannot have more than $O(h/r)$ crossing sensor-paths corresponding to distinct components intersecting this region as each is at distance at least $\frac{r}{2}$ from the others. Thus for each I with $X_I = 1$ there can be at most $O(h/r+1)$ J's such that $I \cap J \neq \emptyset$ and $X_I X_J = 1$. Thus

$$b_2 = O(h/r+1)(2r+1+c_W)e^{-h/3}x,$$

and so $b_2 \to 0$ as $h \to \infty$.

Fixing I, $\sum_{J \cap I \neq \emptyset} \mathbb{E}(X_J) \leq \sum_k (|I|+k-1)\mathbb{P}(X_{[0,k)} = 1)$ as there are $|I|+k-1$ translates of an interval of length k intersecting I and $\mathbb{P}(X_I = 1)$ depends only on $|I|$. But if $X_{[0,k)} = 1$ then $k \leq W_b + 2r + 2$ where W_b is the width of the break corresponding to this interval. Thus

$$\sum_{J \cap I \neq \emptyset} \mathbb{E}(X_J) \leq (|I| + \mathbb{E}(W_b) + 2r + 1) I_{h,r}.$$

Now $|I| \leq W_b' + 2r + 2$, where W_b' is the width of the break corresponding to I. Hence

$$b_1 \leq (\mathbb{E}(W_b') + \mathbb{E}(W_b) + 4r + 3)\ell I_{h,r}^2 \leq (10h + 4r + 3)x I_{h,r},$$

where the second inequality follows from Lemma 10 and the fact that we may assume $h \geq 1$. Since $I_{h,r}$ decreases exponentially with h, $b_1 \to 0$ as $h \to \infty$.

Applying Theorem 14 we deduce that

$$\left|\mathbb{P}(X = k) - e^{-\mathbb{E}(X)}(\mathbb{E}(X))^k/k!\right| \leq b_1 + b_2,$$

and hence

$$\left|\mathbb{P}(\#\text{breaks} = k) - e^{-\mathbb{E}(X)}(\mathbb{E}(X))^k/k!\right| \leq b_0 + b_1 + b_2.$$

The result follows as $\mathbb{E}(X) \to x$ and $b_0, b_1, b_2 \to 0$ as $h \to \infty$. □

For practical purposes one would aim to show that the existence of any break occurring in a short strip is unlikely. In this case we recall Theorem 5.

Theorem 5. *If $r \geq 7$ and $h \geq 1$ then*

$$\mathbb{P}(G_{h,r}(0,\ell) \text{ has no } s\text{-}t \text{ path}) \leq (\ell + 5h)I_{h,r}.$$

Proof. If $G_{h,r}(0,\ell)$ has no s-t path then, by Lemma 1 and Lemma 9, a good separating path exists in $[0,\ell] \times [0,h]$. Thus $[0,\ell]$ must intersect one of the intervals $[x^+_{C_i}, x^-_{C_{i+1}}]$ corresponding to the breaks in $G_{h,r}$. Equivalently, either $x^+_{C_i} \in [0,\ell]$, or $0 \in [x^+_{C_i}, x^-_{C_{i+1}}]$ for some i, where C_i, $i \in \mathbb{Z}$, are the good components ordered horizontally. The expected number of i's with the first property is $\ell I_{h,r}$ since the asymptotic density of points $x^+_{C_i}$ equals the density of breaks $I_{h,r}$. The expected number of i's with the second property is at most $5hI_{h,r}$ since, by Lemma 10, the expected width of a break is at most $5h$ (for $h \geq 1$) and their density is $I_{h,r}$. Thus the probability that $G_{h,r}(0,\ell)$ is not s-t connected is at most $(\ell + 5h)I_{h,r}$. □

6 Small h and the function $\varepsilon(z)$

Having shown in Theorem 4 that breaks occur with an approximately Poisson distribution, it remains to derive the break intensity. Once we have the break intensity, it is a simple matter to either estimate or bound the probability that no breaks exist in a strip of given length (which by Lemma 1 is equivalent to the strip being barrier covered) using either Theorem 4 or Theorem 5.

The most interesting case is when h is larger than r, however, to estimate $I_{h,r}$ for large h, we shall reduce to the case of small h. Thus we shall need to study the small h ($h < \frac{\sqrt{3}}{2}r$) case in some detail first. We count the number of vertices that are the rightmost point of some good component, since this is equivalent to counting good components, and hence breaks.

For $h < \frac{\sqrt{3}}{2}r$, all components are good, so the probability that a fixed sensor v is the rightmost sensor of a good component is given by the probability that there is no sensor w to the right of v that is adjacent in $G_{h,r}$ to v, or to any sensor u to the left of v (as in this case either u or w would be joined to v by Lemma 6 applied to the vertical crossing sensor-path through v). To calculate this probability, fix v, and place sensors to the left of v in S_h according to a Poisson point process. Let A be the region in S_h to the right of v that is within distance r of v or any sensor to the left of v (see Figure 7). Then v is the rightmost sensor of a component if and only if the region A is empty. Thus conditioned on the process to the left of v, the probability that v is the rightmost sensor of a component is $e^{-|A|}$. The overall probability that we are interested in is just $\mathbb{E}(e^{-|A|})$, where the expectation is over the position of v and the state of the Poisson process to the left of v. The intensity of such sensors, and hence of breaks, is then given by

$$I_{h,r} = h\,\mathbb{E}(e^{-|A|}), \tag{12}$$

since the intensity of sensors v per unit length of the strip is just h. It remains to calculate $\mathbb{E}(e^{-|A|})$.

The excluded area is the union of a number of disks of radius r. We first approximate these areas by parabolic regions, replacing the disk $(x-x_0)^2+(y-y_0)^2 < r^2$ about a sensor (x_0, y_0) by the parabolic region $x - x_0 < r - (y - y_0)^2/2r$. We then estimate the excluded area as $|A| \approx rh - |B|$, where B is the shaded region in Figure 7 which lies to the right of all these approximating parabolas and to the left of the vertical line which is at distance r to the right of v. The advantage of this approximation is that the two parameters r and h can be reduced to a single parameter $z = hr^{-1/3}$, making the analysis of the function $\varepsilon(z)$ defined below much easier. Indeed, if we rescale the strip by a factor $r^{1/3}$ along the x-axis and by a factor $r^{-1/3}$ along the y-axis, the density of the Poisson process and all areas unchanged, but the parabolas are now of the form $x = z_0 - (y - y_0)^2/2$, where the points (z_0, y_0) are the locations of the sensors, shifted a constant amount to the right so that they lie on the vertices of the parabolas.

For convenience we now swap the x and y-coordinates. Then B is defined by placing a Poisson point process with intensity 1 in the half infinite strip $[0, z] \times [0, \infty)$ plus one more point chosen uniformly at random on $[0, z] \times \{0\}$ and then taking the area below all the parabolas $y = y_0 + (x - x_0)^2/2$, where (x_0, y_0) ranges over all of these points. Define $\varepsilon(z)$ by

$$\varepsilon(z) = z\,\mathbb{E}(e^{|B|}), \tag{13}$$

so that

$$I_{h,r} \approx r^{1/3}\varepsilon(hr^{-1/3})e^{-hr}. \tag{14}$$

Note that taking a Poisson point process in $[0, z] \times [0, \infty)$ and then adding one random point in $[0, z] \times \{0\}$ is equivalent to taking a Poisson process in $[0, z] \times [0, \infty)$ and then shifting all these points down until the lowest point is on the x-axis. Let $\mathcal{P}$ be defined as the set of these points.

It is clear that $|B| \le z^3/6$, the worst case being when there is just one point in $\mathcal{P}$ that is located at the origin. Thus $\varepsilon(z) \le ze^{z^3/6}$. We wish to improve this bound to $e^{O(z)}$ first, before showing that $\varepsilon(z) = e^{\alpha z+\beta+o(1)}$. Moreover, we wish to show that in (13), one can, without too much error, restrict $\mathcal{P}$ so that the lowest parabola above any point $(x_0, 0)$ has its vertex (i.e., its lowest point) not too far from the line $x = x_0$. Note that the horizontal distance to the vertex of one of the parabolas is just the slope of the parabola at that point, so that this is equivalent to restricting the maximum slope of the upper boundary

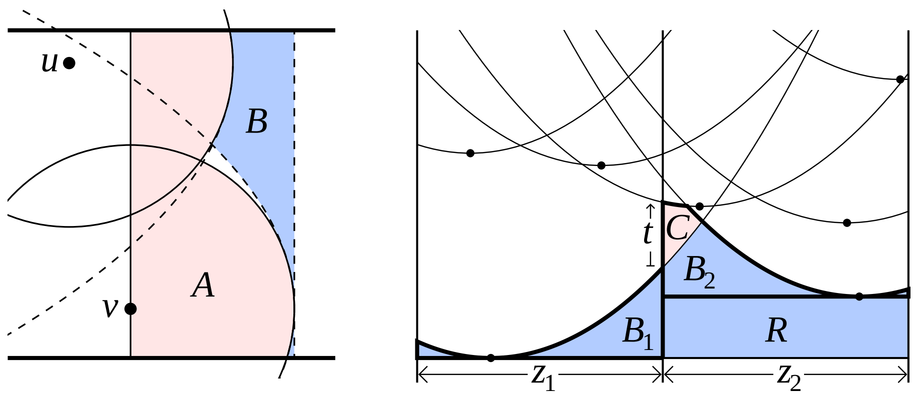

Figure 7: Approximating excluded disks by parabolas. The vertex v is the rightmost vertex of the good component, and u is another vertex to the left of v. The excluded area A consists of all points within the circles that lie to the right of v, so $|A| \approx hr - |B|$. For clarity, the picture is drawn with $h > r$. The diagram on the right is used in the proof of Lemma 17.

of the region B. In doing so, we shall also bound the error introduced in (14) as a result of approximating the excluded disks by parabolas. We now leave the task of estimating $I_{h,r}$ to Section 8. The rest of this section will be devoted to studying the function $\varepsilon(z)$.

Lemma 15. *Let $s(B)$ be the maximum absolute value of the slope of the upper boundary of B. Define $\varepsilon_s(z) = z\,\mathbb{E}(e^{|B|}1_{s(B)\geq s})$. Then there are absolute constants $c_1, c_2 > 0$ independent of z and s such that $\varepsilon_s(z) \leq e^{c_1 z - c_2 s^3}$.*

Note that $\varepsilon(z) = \varepsilon_0(z)$, so Lemma 15 implies that $\varepsilon(z) \leq e^{c_1 z}$.

Proof. First we note that $\varepsilon_s(z) \leq \varepsilon(z) \leq z e^{z^3/6}$ and $s(B) \leq z$, so by choosing c_1 sufficiently large, and c_2 sufficiently small, we may assume the result holds for $z \leq 1$.

Now assume $z > 1$ and tile $[0, z] \times [0, \infty]$ with rectangles of width α and height $\frac{\alpha^2}{2}$. Here we choose $\alpha = z/\lceil z \rceil$ so that $n = z/\alpha = \lceil z \rceil$ is an integer and $\frac{1}{2} \leq \alpha \leq 1$. Let a_i, $i = 0, \dots, n-1$, be the number of rectangles above the interval $[i\alpha, (i+1)\alpha]$ that intersect B. In other words, we bound B by a step function which has height $a_i \frac{\alpha^2}{2}$ on the interval $[i\alpha, (i+1)\alpha)$. Thus $|B| \leq \frac{\alpha^3}{2} \sum_i a_i$. Let b_i, $i = 0, \dots, n-1$, be defined as the maximum of

$$a_i - 2, \quad a_{i\pm 1} - 5, \quad a_{i\pm 2} - 10, \quad \dots \quad a_{i\pm k} - 1 - (k+1)^2, \quad \dots$$

Then there can be no point of $\mathcal{P}$ in the lowest b_i rectangles above $[i\alpha, (i+1)\alpha]$. Indeed, if say $b_i = a_{i+k} - 1 - (k+1)^2$ and there were a point $(x_0, y_0) \in \mathcal{P}$ with $y_0 \leq b_i \frac{\alpha^2}{2}$ and

$x_0 \in [i\alpha, (i+1)\alpha]$, then any point $(x, y) \in B$ with $x \in [(i+k)\alpha, (i+1+k)\alpha]$ would satisfy

$$y \le y_0 + \tfrac{1}{2}(x - x_0)^2 \le b_i \tfrac{\alpha^2}{2} + (k+1)^2 \tfrac{\alpha^2}{2} \le (a_{i+k} - 1)\tfrac{\alpha^2}{2}.$$

contradicting the definition of a_{i+k} as there should be no points in $B \cap [(i+k)\alpha, (i+1+k)\alpha] \times [0, (a_{i+k} - 1)\frac{\alpha^2}{2}]$. The probability of a configuration occurring with a particular sequence $(a_i)_i$ is thus at most $\exp(-\frac{\alpha^3}{2}\sum_i b_i)$. Since $|B| \le \frac{\alpha^3}{2}\sum_i a_i$, the contribution to $\varepsilon(z)$ from such configurations is at most $z\exp(\frac{\alpha^3}{2}\sum_i (a_i - b_i))$.

Let $\delta_i = a_i - a_{i-1}$, $i = 1, \dots, n-1$. Then the sequence $(\delta_i)_i$ determines $(a_i)_i$ up to the addition of a constant. However, $a_i = 1$ when the point of $\mathcal{P}$ on the x-axis lies in $[i\alpha, (i+1)\alpha)$. Thus the contribution to $\varepsilon(z)$ from configurations with a fixed sequence $(\delta_i)_i$ is at most

$$nz\exp\Big(\frac{\alpha^3}{2}\sum_i (a_i - b_i)\Big)$$

and the differences $a_i - b_i$ depend only the sequence $(\delta_i)_i$. Now $b_i \ge \max\{a_i - 2, a_{i\pm 1} - 5\}$, so $a_i - b_i \le \min\{2, 5 + \delta_i, 5 - \delta_{i+1}\}$, where we define $\delta_0 = \delta_n = 0$ when $i = 0$ or $n-1$. Thus

$$\begin{aligned} a_i - b_i &\le \tfrac{1}{2}\min\{5, 5+\delta_i\} + \tfrac{1}{2}\min\{5, 5-\delta_{i+1}\} \\ &= \tfrac{1}{2}(5 - \max\{-\delta_i, 0\}) + \tfrac{1}{2}(5 - \max\{\delta_{i+1}, 0\}) \\ &= 5 - \tfrac{1}{2}\delta_i^- - \tfrac{1}{2}\delta_{i+1}^+, \end{aligned} \tag{15}$$

where $\delta_i^+ = \max\{\delta_i, 0\}$ and $\delta_i^- = \max\{-\delta_i, 0\}$. Now $|\delta_i| = \delta_i^+ + \delta_i^-$, so

$$\sum_{i=0}^{n-1}(a_i - b_i) \le 5n - \tfrac{1}{2}(\delta_0^- + |\delta_1| + \dots + |\delta_{n-1}| + \delta_n^+) = 5n - \tfrac{1}{2}\sum_{i=1}^{n-1}|\delta_i|. \tag{16}$$

Thus

$$nz\exp\Big(\frac{\alpha^3}{2}\sum_{i=0}^{n-1}(a_i - b_i)\Big) \le nze^{5n\alpha^3/2}\prod_{i=1}^{n-1} e^{-|\delta_i|\alpha^3/4}.$$

Summing over all values of $\delta_1, \delta_2, \dots \delta_{n-1} \in \mathbb{Z}$ in turn gives

$$\sum_{(\delta_i)} nz\exp\Big(\frac{\alpha^3}{2}\sum_i (a_i - b_i)\Big) \le nze^{5n\alpha^3/2}\Big(\frac{2}{1 - e^{-\alpha^3/4}} - 1\Big)^{n-1} \le nze^{c_3 n}, \tag{17}$$

for some absolute constant $c_3 > 0$. Substituting $n = z/\alpha$ and recalling that $\alpha \ge \frac{1}{2}$ gives $\varepsilon(z) \le 2z^2 e^{2c_3 z}$. Thus the result for $s = 0$ (or even bounded s) follows.

Now suppose the upper boundary of B has maximum slope $s = s(B) > 0$ at the point $(x_0 + s, y_0 + \frac{s^2}{2})$, so that the slope is determined by the vertex $(x_0, y_0) \in \mathcal{P}$. (The case when $s(B)$ is the absolute value of the largest negative slope is similar.) Let $i, j \in \mathbb{Z}$ be such that

$$x_0 \in [i\alpha, (i+1)\alpha) \qquad \text{and} \qquad y_0 \in [(j-1)\tfrac{\alpha^2}{2}, j\tfrac{\alpha^2}{2}).$$

Then $a_{i+k} \le (k+1)^2+j$ for all $k \ge 0$. Now choose $t \in \mathbb{Z}$ so that $x_0+s \in [(i+t)\alpha, (i+t+1)\alpha)$. Note that $(t-1)\alpha \le s \le (t+1)\alpha$. Then $a_{i+t}\frac{\alpha^2}{2} \ge y_0 + \frac{s^2}{2} \ge (j-1+(t-1)^2)\frac{\alpha^2}{2}$. Hence for $0 < k < t$

$$b_{i+k} \ge (j-1+(t-1)^2) - 1 - (t-k+1)^2.$$

Thus

$$a_{i+k} - b_{i+k} \le (k+1)^2 + j - (j-1) - (t-1)^2 + 1 + (t-k+1)^2 = 4t + 3 - 2k(t-k).$$

Since the maximum slope of the upper boundary of B is $s \le (t+1)\alpha$, we have $|\delta_i| \le t+1$ for all i, and so

$$a_{i+k} - b_{i+k} \le (5 - \tfrac{1}{2}\delta^-_{i+k} - \tfrac{1}{2}\delta^+_{i+k+1}) + 5t - 2k(t-k).$$

Summing, and using (15) when $j \le i$ or $j \ge i+t$, gives

$$\sum_{j=0}^{n-1}(a_j - b_j) \le 5n - \tfrac{1}{2}\sum_{j=1}^{n-1}|\delta_j| - \sum_{k=1}^{t-1}(2k(t-k) - 5t).$$

This last sum is $\Theta(t^3) = \Theta(s^3)$, so summing over the $(\delta_i)_i$ as in (17) now gives the result for $\varepsilon_s(z)$. □

We observe that we only used the parabolic bound $y_0 + (x-x_0)^2/2$ on the upper boundary of B when $|x - x_0| < 2\alpha \le 2$ for $s = 0$, or $|x - x_0| < s + 2\alpha \le s+2$ for $s > 0$. Clearly modifying the parabola slightly does not affect the proof of Lemma 15, so we can generalize it to the following.

Lemma 16. *Assume $f\colon \mathbb{R}\times\mathbb{R} \to [0,\infty]$ is such that $f(x_0, x_0) = 0$ and*

$$\frac{\partial}{\partial x} f(x, x_0) = \eta(x, x_0)(x - x_0), \qquad 1 \le \eta(x, x_0) \le K,$$

for all x and x_0 with $|x - x_0| < s + 2$. Then the conclusion of Lemma 15 also holds for $\varepsilon_{f,s}(z)$, where we define

$$\varepsilon_{f,s}(z) = z\mathbb{E}(e^{|B|}) \tag{18}$$

as for $\varepsilon_s(z)$, but with the curves $y = y_0 + f(x, x_0)$ used instead of the parabolas $y = y_0 + (x-x_0)^2/2$ in the definition of the region B. The constants c_1 and c_2 do however depend on K.

Proof. By the monotonicity of the expression for $\varepsilon_{f,s}$ in terms of $e^{|B|}1_{s(B)\geq s}$, it is enough to prove the result when $f(x,x_0) = K(x-x_0)^2/2$ and the slope bound is replaced with $s(B) \geq Ks$. In this case the result follows from the proof of Lemma 15 by using $K\frac{\alpha^2}{2}$ for the height of the rectangles instead of $\frac{\alpha^2}{2}$. ☐

To show that $\varepsilon(z)$ is of the form $e^{\alpha z+\beta+o(1)}$ we now show that $\varepsilon(z_1+z_2) \approx c\varepsilon(z_1)\varepsilon(z_2)$ for some constant $c > 0$. Let $z = z_1 + z_2$ and decompose the area B into two pieces, the part above $[0, z_1]$ and the part above $[z_1, z_1+z_2]$. If y_1 (respectively y_2) is the height of the lowest vertex on the left (respectively right), then the set of vertices can be written as a union of $\mathcal{P}_1 + (0, y_1)$ and $\mathcal{P}_2 + (z_1, y_2)$, where $\mathcal{P}_1$ and $\mathcal{P}_2$ are corresponding configurations used in the definition of $\varepsilon(z_1)$ and $\varepsilon(z_2)$. The area $|B|$ can be calculated as $|B_1|+|B_2|+|R|-|C|$, where B_1 and B_2 are the corresponding areas for $\mathcal{P}_1$ and $\mathcal{P}_2$, R is either $[0, z_1] \times [0, y_1]$ or $[z_1, z_1 + z_2] \times [0, y_2]$ depending on whether y_2 or y_1 is zero, and C is the set of points that are below all the parabolas on their own side, but above a parabola on the opposite side (see Figure 7). Fix $\mathcal{P}_1$ and $\mathcal{P}_2$ and imagine sliding one of the regions $[0, z_1] \times \mathbb{R}$ or $[z_1, z_1 + z_2] \times \mathbb{R}$ either up or down. Let t be the y-coordinate of the top left point of B_2 minus the y-coordinate of the top right point of B_1. We consider the region $C = C(t)$ to be a function of t. Note in particular that $C(0) = \emptyset$ and that C is to the right (respectively left) of the vertical line $x = z_1$ when $t > 0$ (respectively $t < 0$), and C intersects the vertical line $x = z_1$ in an interval of length exactly $|t|$.

Lemma 17.

$$\varepsilon(z_1 + z_2) = z_1 z_2 \mathbb{E}_{\mathcal{P}_1} \mathbb{E}_{\mathcal{P}_2} e^{|B_1|} e^{|B_2|} \int_{-\infty}^{\infty} e^{-|C(t)|} dt, \tag{19}$$

where $\mathbb{E}_{\mathcal{P}_1}$ and $\mathbb{E}_{\mathcal{P}_2}$ are expectations over the corresponding configurations and, as above, t is the distance of the top left point of B_2 over the top right point of B_1.

Proof. The set $\mathcal{P}$ can be constructed with the correct distribution by setting $\mathcal{P} = (\mathcal{P}_1 + (0, y_1)) \cup (\mathcal{P}_2 + (z_1, y_2))$, where with probability z_1/z, $y_1 = 0$ and y_2 has an exponential distribution with mean $1/z_2$, while with probability z_2/z, $y_2 = 0$ and y_1 has an exponential distribution with mean $1/z_1$. Let t_0 be the value of t when $y_1 = y_2 = 0$. Then recalling that $|B| = |B_1| + |B_2| + |R| - |C|$,

$$\begin{aligned}
\varepsilon(z_1 + z_2) &= z\mathbb{E}_{\mathcal{P}_1}\mathbb{E}_{\mathcal{P}_2} e^{|B_1|+|B_2|} \left(\frac{z_1}{z} \int_0^{\infty} e^{\tau z_2 - |C(t_0+\tau)|} z_2 e^{-\tau z_2} d\tau + \frac{z_2}{z} \int_0^{\infty} e^{\tau z_1 - |C(t_0-\tau)|} z_1 e^{-\tau z_1} d\tau \right) \\
&= z_1 z_2 \mathbb{E}_{\mathcal{P}_1}\mathbb{E}_{\mathcal{P}_2} e^{|B_1|} e^{|B_2|} \int_{-\infty}^{\infty} e^{-|C(t)|} dt.
\end{aligned}$$

☐

The area $|C(t)|$ can perhaps be more easily visualized by translating the points (x, y) down vertically to $(x, y - (x - z_1)^2/2)$. As each point in any vertical segment is shifted down by the same amount, the area $|C(t)|$ remains constant. However, the parabolas used to define the regions B, B_1, B_2, and C now become straight lines. The upper boundaries of B, B_1, B_2 become polygonal paths that are concave down, and C becomes a polygonal region. Lines bounding B_1, B_2, or B that slope upwards arise from points to the left of $x = z_1$, and lines that slope downwards arise from points to the right of $x = z_1$. Indeed, the parabola with vertex $(z_1 - s, y_1) \in \mathcal{P}$ is transformed into a line of slope s. Note that the slopes of the upper boundaries of B, B_1, and B_2 at $x = z_1$ are unaffected by this transformation (see Figure 8).

Define

$$Q_{\mathcal{P}_1,\mathcal{P}_2} = \int_{-\infty}^{\infty} e^{-|C(t)|}\, dt.$$

The next result shows that $Q_{\mathcal{P}_1,\mathcal{P}_2}$ can "almost" be factored as a function of $\mathcal{P}_1$ times a function of $\mathcal{P}_2$. This will be the key step in the proof of Theorem 3.

Lemma 18. *For any choice of $\mathcal{P}_1$, $\mathcal{P}_1'$, $\mathcal{P}_2$, $\mathcal{P}_2'$, we have*

$$\frac{Q_{\mathcal{P}_1,\mathcal{P}_2} Q_{\mathcal{P}_1',\mathcal{P}_2'}}{Q_{\mathcal{P}_1,\mathcal{P}_2'} Q_{\mathcal{P}_1',\mathcal{P}_2}} \le 8\sqrt{2\pi}\big(\min\{z_1, z_2\}\big)^{3/2} + 16.$$

Proof. Write $Q^L_{\mathcal{P}_1,\mathcal{P}_2}$ (respectively $Q^R_{\mathcal{P}_1,\mathcal{P}_2}$) for the integral of $e^{-|C(t)|}$ over $t < 0$ (respectively $t > 0$), i.e., over t such that the region $C(t) = C_{\mathcal{P}_1,\mathcal{P}_2}(t)$ is to the left (respectively right) of the line $x = z_1$. Write $\mathcal{P}_1^0$ for the single point $(z_1, 0)$ and $\mathcal{P}_2^0$ for the single point $(0, 0)$, so that both represent a single point on the x-axis at the intersection of the two strips. For $t < 0$ and any $\mathcal{P}_2$ we have

$$|C_{\mathcal{P}_1,\mathcal{P}_2}(t)| \le |C_{\mathcal{P}_1,\mathcal{P}_2^0}(t)|.$$

Indeed, the lowest upper boundary of B to the left of $x = z_1$ we can force by placing points to the right of $x = z_1$ and with a given value of t is obtained when we place points on $x = z_1$ as this has slope 0 while any point further to the right would give a boundary with negative slope. Hence by integrating we have

$$Q^L_{\mathcal{P}_1,\mathcal{P}_2} \ge Q^L_{\mathcal{P}_1,\mathcal{P}_2^0}.$$

On the other hand, if $R := C_{\mathcal{P}_1,\mathcal{P}_2^0}(t/2)$ is not a subset of $C(t) = C_{\mathcal{P}_1,\mathcal{P}_2}(t)$ (keeping the vertical alignment of $\mathcal{P}_1$ fixed), then $T := C(t) \setminus R$ has no boundary determined by points in $\mathcal{P}_1$ (see Figure 8). However, in this case, convexity of the boundary determined by $\mathcal{P}_2$ implies that $|T| \ge |S|$ where $S = |C_{\mathcal{P}_1^0,\mathcal{P}_2}(-t/2)|$. Hence if $R \not\subseteq C(t)$ then $|C(t)| \ge |T| \ge |S|$, while if $R \subseteq C(t)$ then $|C(t)| \ge |R|$. Hence

$$e^{-|C_{\mathcal{P}_1,\mathcal{P}_2}(t)|} = e^{-|C(t)|} \le e^{-|R|} + e^{-|S|} = e^{-|C_{\mathcal{P}_1,\mathcal{P}_2^0}(t/2)|} + e^{-|C_{\mathcal{P}_1^0,\mathcal{P}_2}(-t/2)|}.$$

Integrating from $t=-\infty$ to $t=0$, we deduce that

$$Q^L_{\mathcal{P}_1,\mathcal{P}_2} \le 2(Q^L_{\mathcal{P}_1,\mathcal{P}_2^0} + Q^R_{\mathcal{P}_1^0,\mathcal{P}_2}).$$

Hence, writing $c_L = Q^L_{\mathcal{P}_1,\mathcal{P}_2^0}$ and $c_R = Q^R_{\mathcal{P}_1^0,\mathcal{P}_2}$, we have

$$c_L \le Q^L_{\mathcal{P}_1,\mathcal{P}_2} \le 2(c_L + c_R),$$

and by symmetry

$$c_R \le Q^R_{\mathcal{P}_1,\mathcal{P}_2} \le 2(c_L + c_R).$$

Now

$$Q_{\mathcal{P}_1,\mathcal{P}_2} = Q^L_{\mathcal{P}_1,\mathcal{P}_2} + Q^R_{\mathcal{P}_1,\mathcal{P}_2},$$

so

$$c_L + c_R \le Q_{\mathcal{P}_1,\mathcal{P}_2} \le 4(c_L + c_R).$$

Thus the quotient in the statement of the lemma is at most

$$\frac{4(c_L+c_R)4(c'_L+c'_R)}{(c_L+c'_R)(c'_L+c_R)},$$

where c'_L and c'_R are defined using $\mathcal{P}'_i$ instead of $\mathcal{P}_i$. Without loss of generality we may assume $z_1 \le z_2$ and $c'_L \ge c_L$. Then

$$\frac{4(c_L+c_R)4(c'_L+c'_R)}{(c_L+c'_R)(c'_L+c_R)} \le 16\frac{c'_L+c'_R}{c_L+c'_R} \le 16\frac{c'_L}{c_L}$$

The smallest value of c_L occurs when $\mathcal{P}_1 = \mathcal{P}_1^0$ as in this case $|C(t)|$ is as large as possible. Indeed, $|C(t)| = |t|z_1$ as $C(t)$ is of constant height $|t|$ and width z_1. Thus $c_L = \int_{\tau=0}^\infty e^{-\tau z_1}\,d\tau = 1/z_1$. The largest value of c'_L occurs when $\mathcal{P}'_1 = \{(0,0)\}$ as this causes the height of $C(t)$ to decrease as rapidly as possible away from the line $x = z_1$ and hence minimizes the area $|C(t)|$. In this case the height of $C(t)$ decreases linearly away from $x = z_1$ with slope z_1 and so

$$|C(t)| = \begin{cases} |t|^2/(2z_1) & \text{if } t \le z_1^2; \\ (|t| - z_1^2/2)z_1 & \text{otherwise.} \end{cases}$$

(The first expression occurring when the shifted $C(t)$ is a triangular region and the second when it forms a trapezoid that extends all the way to the line $x = 0$.) Hence in this case we have

$$c'_L = \int_0^{z_1^2} e^{-t^2/(2z_1)}dt + \int_{z_1^2}^\infty e^{-(t-z_1^2/2)z_1}dt \le \sqrt{\pi z_1/2} + 1/z_1$$

Thus

$$\frac{Q_{\mathcal{P}_1,\mathcal{P}_2} Q_{\mathcal{P}_1',\mathcal{P}_2'}}{Q_{\mathcal{P}_1,\mathcal{P}_2'} Q_{\mathcal{P}_1',\mathcal{P}_2}} \leq 16 \frac{\sqrt{\pi z_1/2} + 1/z_1}{1/z_1} = 8\sqrt{2\pi} z_1^{3/2} + 16.$$

□

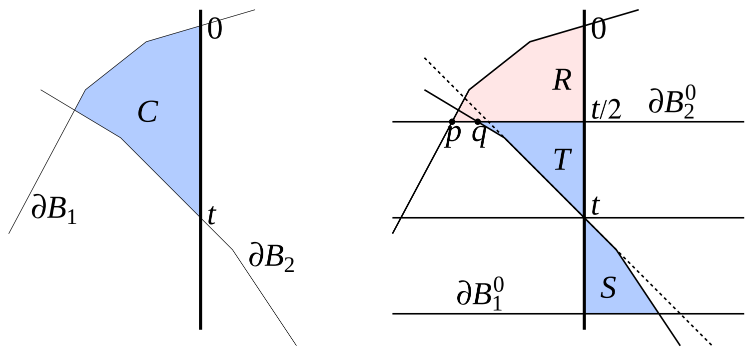

Figure 8: $|R| = |C_{\mathcal{P}_1,\mathcal{P}_2^0}(t/2)|$ and $|T| \geq |S| = |C_{\mathcal{P}_1^0,\mathcal{P}_2}(-t/2)|$. Either R or T is contained within $C = C(t) = C_{\mathcal{P}_1,\mathcal{P}_2}$ depending on whether the intersection point p is to the right or left of q. Note slopes of all lines from $\mathcal{P}_1$ are positive while slopes of all lines from $\mathcal{P}_2$ are negative. Lines ∂B_1^0 and ∂B_2^0 defined by $\mathcal{P}_1^0$ and $\mathcal{P}_2^0$ are horizontal.

To derive Theorem 3 from Lemma 18, we require some results on finite real matrices, which we shall now present.

Let $A = (a_{ij})$ be a square matrix with non-negative real entries. For any $K \geq 1$, we say A is K-*nearly rank* 1 if for all i, j, k, l,

$$a_{ij} a_{kl} \leq K a_{il} a_{kj}. \tag{20}$$

Note that a matrix that is 1-nearly rank 1 is in fact of rank 1.

Lemma 19. *If A is K-nearly rank 1 and B has non-negative entries, then the matrix product AB (or BA) is K-nearly rank 1 (with the same value of K).*

Proof. Writing $A = (a_{ij})$, $B = (b_{ij})$, and $AB = (c_{ij})$, we have

$$a_{ij} a_{kl} b_{jp} b_{lq} \leq K a_{il} a_{kj} b_{jp} b_{lq}.$$

Summing over j and l gives

$$c_{ip} c_{kq} \leq K c_{iq} c_{kp}$$

as required. □

The following lemma is a quantitative version of Perron's Theorem on eigenvalues of strictly positive matrices (see [8]).

Lemma 20. *Suppose A is a square matrix with strictly positive entries which is K-nearly rank 1. Then for any ε, $0 < \varepsilon < \frac{1}{2}$, and any $N \geq 1 + K \log(3K/\varepsilon)$, the matrix A^N is $(1+\varepsilon)$-nearly rank 1.*

Note that the bound on N is independent of the dimensions of the matrix A.

Proof. By Perron's Theorem, there exists a real and strictly positive eigenvector v, $Av = \lambda v$, with λ the (unique and real) maximal eigenvalue. From (20), and writing $A = (a_{ij})$,

$$\lambda a_{ij} v_k = \sum_l a_{ij} a_{kl} v_l \leq \sum_l K a_{il} a_{kj} v_l = K \lambda v_i a_{kj}.$$

Define u_j for each $j > 0$ to be the maximal u_j such that $a_{ij} \geq v_i u_j$ for all i. Then for each j there exists an k such that $a_{kj} = v_k u_j$. Thus $\lambda a_{ij} v_k \leq K \lambda v_i v_k u_j$, and so $a_{ij} \leq K v_i u_j$ for all i and j. Let $B = (v_i u_j)$. Then B is a rank 1 matrix that approximates A in the sense that $B \leq A \leq KB$ (inequality holding entry-wise). Now v is an eigenvector for B with eigenvalue $\lambda' = u^T v$. But $Bv \leq Av \leq KBv$, so $\lambda' \leq \lambda \leq K\lambda'$. Now fix a basis vector $\delta_k = (\dots, 0, 1, 0, \dots)$ and consider $A^N \delta_k$. Write

$$A^N \delta_k = ((A-B)+B)^{N-1} A \delta_k = (A-B)^{N-1} A \delta_k + c_k v = w + c_k v,$$

where c_k is some scalar. (For any vector x, $M_1 M_2 M_3 \dots M_{N-1} x$ is proportional to v when all the $M_i \in \{A-B, B\}$ and at least one $M_i = B$.) We aim to show that w is small compared with $c_k v$. Now $A\delta_k \leq KB\delta_k = K u_k v$, so

$$w = (A-B)^{N-1} A \delta_k \leq K(\lambda - \lambda')^{N-1} u_k v.$$

But $\lambda' \geq \lambda / K$, so

$$K(\lambda - \lambda')^{N-1} \leq K(1 - 1/K)^{N-1} \lambda^{N-1} \leq K e^{-(N-1)/K} \lambda^{N-1} \leq \varepsilon' \lambda^{N-1},$$

where $\varepsilon' = \varepsilon/3$. Hence $w \leq \varepsilon' \lambda^{N-1} u_k v$. Now $A^N \delta_k \geq A^{N-1} B \delta_k = \lambda^{N-1} u_k v$, so $c_k v = A^N \delta_k - w \geq (1-\varepsilon') A^N \delta_k$, and hence

$$c_k v \leq w + c_k v = A^N \delta_k \leq (1-\varepsilon')^{-1} c_k v.$$

Thus $v_i c_j \leq (A^N)_{ij} \leq (1-\varepsilon')^{-1} v_i c_j$. The result now follows since $(A^N)_{ij}(A^N)_{kl}$ and $(A^N)_{il}(A^N)_{kj}$ both lie between $v_i v_k c_j c_l$ and $(1-\varepsilon')^{-2} v_i v_k c_j c_l$, and $(1-\varepsilon')^{-2} \leq 1+\varepsilon$. $\square$

Proof of Theorem 3. Our aim is to use Lemma 17 to estimate $\varepsilon(z)$ for large z by relating it to the values of $\varepsilon(z)$ for smaller z. To do this, we would like the integral, and in particular $C(t)$, in Lemma 17 to depend only on points close to common boundary of the two strips $[0, z_1] \times \mathbb{R}$ and $[z_1, z_1 + z_2] \times \mathbb{R}$. Fix $z_0 \geq 4$ and define $f(x, x_0)$ by

$$f(x, x_0) = \begin{cases} (x - x_0)^2/2, & \text{if } (2\lfloor \frac{x_0}{2z_0} \rfloor - 1)z_0 \leq x < (2\lfloor \frac{x_0}{2z_0} \rfloor + 3)z_0; \\ \infty, & \text{otherwise.} \end{cases}$$

Now define $\varepsilon_f(z) = \varepsilon_{f,0}(z)$ as in Lemma 16. The rather bizarre definition of f is chosen so that the upper boundary of B is genuinely independent of horizontally distant points, and it will ensure that equation (24) below is an exact equality, rather than just an approximation. Indeed, f is defined so that if $x_0 \in [2kz_0, (2k+2)z_0)$ then f is a quadratic precisely when $x \in [(2k-1)z_0, (2k+3)z_0)$. In particular $f(x, x_0) = (x - x_0)^2/2$ for $|x - x_0| \leq z_0$, so the regions B used in the definition of $\varepsilon_f(z)$ and $\varepsilon(z)$ are the same unless there is a point on the upper boundary of (either version of) B with slope at least z_0. Moreover, $f(x, x_0) \geq (x - x_0)^2/2$ so $\varepsilon_f(z) \geq \varepsilon(z)$. Together with Lemma 16, we have

$$0 \leq \varepsilon_f(z) - \varepsilon(z) \leq \varepsilon_{f,z_0-2}(z) \leq e^{c_1 z - c_2(z_0-2)^3} \leq e^{c_1 z - c_2 z_0^3/8}, \tag{21}$$

where the last inequality follows from our assumption that $z_0 \geq 4$. Now assume that $z \leq cz_0^3$ where $c > 0$ is chosen so that $c < c_2/(8(c_1+1))$ and hence $c_1 z - c_2 z_0^3/8 < c_1 z - (c_1+1)z = -z$. Since $\varepsilon(z) \geq z$ for all $z \geq 0$ and $\varepsilon(z) = \varepsilon_f(z)$ for $0 < z < 4 \leq z_0$, we have

$$1 \leq \varepsilon_f(z)/\varepsilon(z) \leq 1 + e^{-z} \qquad \text{for } 0 < z \leq cz_0^3. \tag{22}$$

Divide $[0, z]$ into three intervals, $[0, 2z_0]$, $[2z_0, z - 2z_0]$, and $[z - 2z_0, z]$. Let y_1, y_2, y_3 be the height of the lowest point of $\mathcal{P}$ above each interval, so that $\min\{y_1, y_2, y_3\} = 0$, and conditioned on $y_i = 0$, the other y_j are independent exponential random variables. Write $\mathcal{P} = (\mathcal{P}_1 + (0, y_1)) \cup (\mathcal{P}_2 + (2z_0, y_2)) \cup (\mathcal{P}_3 + (z - 2z_0, y_3))$ as in the proof of Lemma 17. Define for $z \geq 4z_0$ a function $P^{(z)}$ of a pair of distributions of points by

$$P^{(z)}_{\mathcal{P}_1,\mathcal{P}_3} = z\, \mathbb{E}_{\mathcal{P}_2, y_1, y_2, y_3} e^{|B|},$$

where the upper boundary of B is defined using the function f. In other words, $P^{(z)}_{\mathcal{P}_1,\mathcal{P}_3}$ is given by the same formula as for $\varepsilon_f(z)$ except conditioned on $\mathcal{P}_1$ and $\mathcal{P}_3$. In particular

$$\varepsilon_f(z) = \mathbb{E}_{\mathcal{P}_1} \mathbb{E}_{\mathcal{P}_3} P^{(z)}_{\mathcal{P}_1,\mathcal{P}_3}. \tag{23}$$

Now consider the interval $[0, z]$ with $z = 4z_0$. Split $\mathcal{P} = (\mathcal{P}_1 + (0, y_1)) \cup (\mathcal{P}_2 + (2z_0, y_2))$ as in the proof of Lemma 17 and define a function Q of a pair of distributions of points as above by

$$Q_{\mathcal{P}_1,\mathcal{P}_2} = \int_{-\infty}^{\infty} e^{-|C(t)|}\, dt,$$

where $C(t)$ is defined as in Lemma 17, except using f instead of parabolas. In this case $C(t)$ is restricted to lie above $[z_0, 3z_0]$, and would be unaffected by points of $\mathcal{P}$ lying to the left of $x = 0$ or to the right of $x = 4z_0$ if they were to exist.

Using the same argument as in Lemma 17 we have that

$$P^{(4z_0+z)}_{\mathcal{P}_1,\mathcal{P}_4} = \mathbb{E}_{\mathcal{P}_2,\mathcal{P}_3} P^{(4z_0)}_{\mathcal{P}_1,\mathcal{P}_2} Q_{\mathcal{P}_2,\mathcal{P}_3} P^{(z)}_{\mathcal{P}_3,\mathcal{P}_4}, \qquad z \geq 4z_0, \tag{24}$$

where we have split the strip into two strips of widths $4z_0$ and z respectively, $\mathcal{P}_1$, $\mathcal{P}_2$ represent the points in the leftmost and rightmost $2z_0$ of the first strip and $\mathcal{P}_3$, $\mathcal{P}_4$ represent the points in the leftmost and rightmost $2z_0$ of the second strip (up to vertical translations). Here we use the fact that only points of $\mathcal{P}_2$ and $\mathcal{P}_3$ can affect the area of the region $C(t)$ used in the definition of $Q_{\mathcal{P}_2,\mathcal{P}_3}$. We wish to use Lemma 20 to show that $P^{(z)}_{\mathcal{P}_1,\mathcal{P}_2}$ increases exponentially with z to a high degree of accuracy. First, since Lemma 20 considers only finite matrices, we discretize the probability space of possible values of the $\mathcal{P}_i$. Since the bounds in Lemma 20 are independent of the size of the matrix, this can be done to arbitrary accuracy. We now can regard $P = (P_{\mathcal{P},\mathcal{P}'})_{\mathcal{P},\mathcal{P}'}$ and $Q = (Q_{\mathcal{P},\mathcal{P}'})_{\mathcal{P},\mathcal{P}'}$ as finite matrices whose rows and columns are indexed by the possible configurations of $\mathcal{P}_i$. Note that all entries of P and Q are strictly positive. Then (24) becomes

$$P^{(4z_0+z)} = P^{(4z_0)} EQEP^{(z)}, \qquad z \geq 4z_0,$$

where E is a diagonal matrix with entry $E_{\mathcal{P},\mathcal{P}}$ being the probability of $\mathcal{P}$. Thus by induction

$$P^{(4kz_0+z)} = (P^{(4z_0)}EQE)^k P^{(z)} \qquad \text{for } k \geq 1,\ z \geq 4z_0. \tag{25}$$

By Lemma 18 we also know that Q is K-nearly rank 1 with $K = O(z_0^{3/2})$. (The proof of Lemma 18 applies to the slightly modified parabolas with only a slight change in the constants.) Thus by Lemma 19 we know that $P^{(4z_0)}EQE$ is also K-nearly rank 1. Thus, by Lemma 20, if k is larger than $z_0^{3/2+1/3} = z_0^{11/6}$ then $A = (P^{(4z_0)}EQE)^k$ is $(1+\varepsilon)$-nearly rank 1 with $\varepsilon = e^{-\Omega(z_0^{1/3})}$. Now by (23) and (25), $\varepsilon_f(4kz_0 + z)$ is of the form $u^T Av$ and $\varepsilon_f(8kz_0 + z)$ is of the form $u^T A^2 v$ for some positive vectors u and v in the discretized approximation. Using $B \leq A \leq (1+\varepsilon)B$ with B a matrix of rank 1, $B^2 = \lambda' B$, as in the proof of Lemma 20, we obtain

$$\begin{aligned} \lambda'(1+\varepsilon)^{-1}u^T Av \leq \lambda' u^T Bv = u^T B^2 v \leq u^T A^2 v \\ \leq (1+\varepsilon)^2 u^T B^2 v = \lambda'(1+\varepsilon)^2 u^T Bv \leq \lambda'(1+\varepsilon)^2 u^T Av, \end{aligned}$$

so by taking a sufficiently fine discretization we obtain

$$|\log \varepsilon_f(8kz_0 + z) - \log \varepsilon_f(4kz_0 + z) - \log \lambda'| \leq 2\varepsilon$$

for $z \geq 4z_0$, $k \geq z_0^{11/6}$. Now λ' depends on z_0 and k, but not on z, and $\varepsilon = e^{-\Omega(z_0^{1/3})}$, so setting $z_1 = 4kz_0$ and replacing $z + 4kz_0$ by z, we have

$$\log \varepsilon_f(z + z_1) - \log \varepsilon_f(z) = h(z_1) + O(e^{-\eta' z_0^{1/3}})$$

for $z \geq 2z_1$ and some fixed $\eta' > 0$ independent of z_0 or z_1. Using (22) we deduce that provided $2z_1 \leq z \leq cz_0^3 - z_1$,

$$\log \varepsilon(z + z_1) - \log \varepsilon(z) = h(z_1) + O(e^{-\eta'' z_1^{1/9}}), \tag{26}$$

for some $\eta'' > 0$ independent of z, z_0, and z_1. Here we have used the fact that $z_1 = O(z_0^3)$ so that $z_1^{1/9} = O(z_0^{1/3})$.

Now for sufficiently large z_1 we can find a $z_0 \geq 4$ and $k \geq z_0^{11/6}$ with $z_1 = 4kz_0$ and $10z_1 \leq cz_0^3 - z_1$, say. Indeed, it is enough to find z_0 with $z_0^{17/6} \ll z_1 \ll z_0^3$. We fix such a choice and now regard z_0, k, and hence $h(z_1)$, as functions of z_1 only. Thus (26) holds for all sufficiently large z_1 and $2z_1 \leq z \leq 10z_1$.

Assume $z_1 \leq z_2 \leq 2z_1$. Then applying (26) twice with $(z, z_1) = (2z_1 + 2z_2, z_1)$, $(3z_1 + 2z_2, z_2)$, adding, and comparing with the result for $(2z_1 + 2z_2, z_1 + z_2)$, gives

$$h(z_1 + z_2) = h(z_1) + h(z_2) + O(e^{-\eta'' z_1^{1/9}}). \tag{27}$$

Using $\{(\lceil \frac{n-1}{2} \rceil + 1)z_1, (\lfloor \frac{n-1}{2} \rfloor + \zeta)z_1\}$ for $\{z_1, z_2\}$ in (27), ordered so that $z_1 \leq z_2$, we can show by induction that $h((n+\zeta)z_1) = nh(z_1) + h(\zeta z_1) + nO(e^{-\eta'' z_1^{1/9}})$ for all $n \geq 0$ and $\zeta \in [1, 2]$. Letting $n \to \infty$ we see that both $\limsup h(z)/z$ and $\liminf h(z)/z$ are of the form $(h(z_1) + O(e^{-\eta'' z_1^{1/9}}))/z_1$ for every sufficiently large z_1. Therefore the limit $\alpha = \lim_{z\to\infty} h(z)/z$ exists, and $h(z_1) = \alpha z_1 + O(e^{-\eta'' z_1^{1/9}})$.

Set $g(z) = \log \varepsilon(z) - \alpha z$. Then by (26) we have

$$g(z + z_1) = g(z) + O(e^{-\eta'' z_1^{1/9}}) \tag{28}$$

for all z with $2z_1 \leq z \leq 10z_1$. From this one can deduce that the limit $\beta = \lim_{z\to\infty} g(z)$ exists. Indeed, setting $z = 2z_1$ and using induction gives

$$g((1.5)^k 2z_1) = g(2z_1) + \sum_{i=0}^{k} O(e^{-\eta''(1.5^k z_1)^{1/9}}) = g(2z_1) + O(e^{-\eta'' z_1^{1/9}}).$$

One more application of (28) with $\{z - (1.5)^k 2z_1, (1.5)^k 2z_1\}$ in place of $\{z, z_1\}$ gives $g(z) = g(2z_1) + O(e^{-\eta'' z_1^{1/9}})$ for all z with $3(1.5)^k 2z_1 \leq z \leq 11(1.5)^k 2z_1$, and hence for all sufficiently large z. As this holds for all sufficiently large z_1, $\beta = \lim_{z\to\infty} g(z)$ exists and $g(2z_1) = \beta + O(e^{-\eta'' z_1^{1/9}})$ for all sufficiently large z_1. Thus

$$\log \varepsilon(z) = \alpha z + g(z) = \alpha z + \beta + O(e^{-\eta''(z/2)^{1/9}}) = \alpha z + \beta + O(e^{-\eta z^{1/9}}).$$

Finally we note that $\varepsilon(z) = z\mathbb{E}(e^{|B|}) \geq z \to \infty$ as $z \to \infty$, so $\alpha > 0$. □

Monte Carlo computer simulations were performed to estimate $\varepsilon(z)$ for various values of z up to $z = 8$ (see Figure 9 for a plot up to $z = 5$). Using these results, the constants in Theorem 3 were estimated as

$$\begin{aligned} \alpha &= \quad 1.12794 \pm 0.00001 \\ \beta &= -1.05116 \pm 0.00005 \end{aligned} \tag{29}$$

(errors are ± 1 standard deviation). The $O(e^{-\eta z^{1/9}})$ error term in Theorem 3 appears to be conservative as it actually seems to tend to zero extremely rapidly as $z \to \infty$. Indeed, the approximation $\varepsilon(z) \approx e^{\alpha z + \beta}$ is within 2% of the correct value when $z > 0.85$, and for $z > 3$ the error is insignificant.

For small z, one can expand $\varepsilon(z)$ as a power series in z. One can show that the only non-zero terms are of the form cz^{1+3k} and the first few terms are

$$\varepsilon(z) = z + \tfrac{1}{12} z^4 + \tfrac{1}{64800} z^{10} - \tfrac{1}{2721600} z^{13} + \cdots \tag{30}$$

(there is no z^7 term). We obtained these coefficients by first expanding $\varepsilon(z) = z\mathbb{E}(e^{|B|}) = z + z\mathbb{E}|B| + z\mathbb{E}(|B|^2)/2 + \ldots$. If we write $R = [0, z] \times [0, z^2/2]$, then B depends only on the points in $\mathcal{P} \cap R$. Expanding $\mathbb{E}(|B|^i)$ according to the (Poisson distributed) number of points in $\mathcal{P} \cap R$, it is enough to calculate $\mathbb{E}(|B|^i)$ conditioned on $\mathcal{P} \cap R$ containing exactly k points for small values of i and k. These conditional expectations can then be represented as integrals over i-tuples of points in R of the probability that all i points lie in B, which is equivalent to all k Poisson points lying in some subset of R determined by these i points. The values of these integrals were evaluated via symbolic integration using Mathematica for small values of i and k so as to obtain the expansion (30).

7 Extending to large h

We now consider the issues that occur when h is large. The main problem is that the break may be topologically complicated, and may meander back and forth as it crosses the strip S_h. We wish to show an approximately exponential dependence of $I_{h,r}$ as a function of h, and so we aim to compare $I_{2h,r}$ with $I_{h,r}^2$. This will be done by considering S_{2h} as two

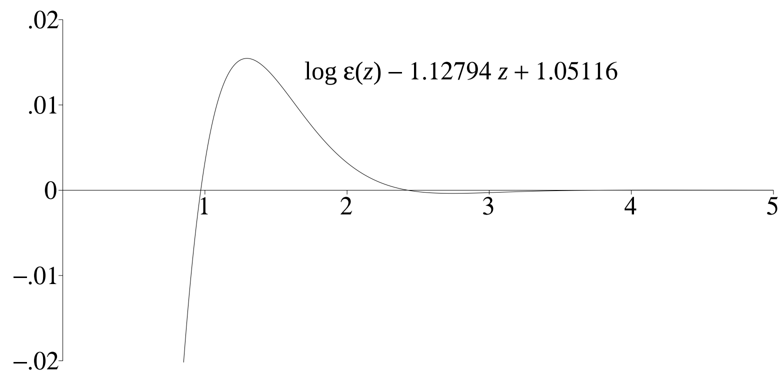

Figure 9: Plot of $\log \varepsilon(z) - 1.12794z + 1.05116$ against z.

copies of S_h, one on top of the other, and matching breaks up on these two independent strips. However, to combine breaks of two thin strips to form a break of a thicker strip needs these breaks to look "nice" at the point at which they meet.

We use a technique involving "surgery", that is, we take breaks that are "bad" and map them in a measure preserving way into breaks that are "good" by cutting and pasting certain regions in S_h. As long as this map is a "not too many"-to-1 map, we can then lower-bound the proportion of breaks that are good. We shall use the following general lemma, leaving the precise definition of a "good" break to later.

Lemma 21. *Suppose we have, for each bad break b, an at most k-to-1 map ψ_b on the probability space that converts b into a good break without destroying any break or converting any other good break into a bad break. Suppose further that it can only change bad breaks to good breaks if they are originally adjacent to b, and can generate at most one new break which is then immediately to the right of b. If ψ_b is measure preserving when restricted to any subset where it is injective, then the intensity of good breaks is at least $I_{h,r}/(12k+1)$.*

In Lemma 21 we assume that there is some consistent labeling of the breaks, for example, labeling by integers b from left to right with 0 being assigned to the first break ending to the right of $x = 0$. Also, the map ψ_b actually only needs to be defined on the subset of the probability space where the break b is bad.

Proof. Fix a large interval $[0, x]$ and let $T = \{b_1, \dots, b_n\}$ be a linearly ordered set of symbols representing the breaks in $[0, x] \times [0, h]$ in their left-to-right order. Let $p_{T,S}$ be the probability that there are $|T|$ breaks and the subset $S \subseteq T$ corresponds to those breaks that are good. The existence of ψ_b, $b \in T \setminus S$, implies that

$$p_{T,S} \leq k\Big(\sum_{A \subseteq \{b^-, b^+\}} p_{T, S \cup A \cup \{b\}} + \sum_{A \subseteq \{b^-, b^+, b'\}} p_{T \cup \{b'\}, S \cup A \cup \{b\}} \Big)$$

where $b^{\pm}$ are the breaks adjacent to b in T, and b' is a new break inserted immediately to the right of b in T (between b and b^+). Indeed, the right hand side gives at most k times the probability of the image of ψ_b restricted to the event where there are $|T|$ breaks and S is the set of good breaks. This image includes the cases when b is made good, $b^{\pm}$ are possibly made good, and either zero or one new break b' is generated, which may be either good or bad.

Summing over all pairs (S, b) with $|S| = r$ and $b \in T \setminus S$ and letting $p_{n,r}$ denote the probability that there are n breaks of which r are good, we have

$$(n-r)p_{n,r} \leq k\Big(\sum_{i=0}^{2}\tbinom{2}{i}(r+1+i)p_{n,r+1+i} + \sum_{i=0}^{3}\tbinom{3}{i}(r+1+i)p_{n+1,r+1+i}\Big).$$

The factor of $n-r$ gives the number of choices for b for each choice of S, and the factors of $r+1+i$ on the right hand side count the number of choices of b for each fixed choice of the set $S \cup A \cup \{b\}$ when $|A| = i$. Summing over r and n gives

$$E_B - E_G \leq k(4E_G + 8E_G),$$

where $E_B = \sum_{n,r} np_{n,r}$ is the expected number of breaks and $E_G = \sum_{n,r} rp_{n,r}$ the expected number of good breaks in $[0, x] \times [0, h]$. Thus $E_G \geq E_B/(12k+1)$, and so the result follows on letting $x \to \infty$. □

For any point $v \in \mathbb{R}^2$, write $B_r(v) = \{x \in \mathbb{R}^2 : \|x - v\| < r\}$ for the open disk of radius r about v. For $z \in \mathbb{R} \times [0, r]$, write $r_z = \max\{x : \|(x, 0) - z\| \leq r\}$ for the rightmost point on the x-axis that lies within distance r of z. We shall make use of the following simple lemma.

Lemma 22. *Let $v_i = (x_i, y_i) \in \mathcal{P}$, $i = 1, 2, 3$, be pairwise non-adjacent vertices of $G_{h,r}$ with $y_i \in [0, r)$ and $x_1 < x_2 < x_3$. Then $r_{v_3} > r_{v_1}$.*

Proof. If $y_1 > y_2$ then we must have $r_{v_1} \leq x_2$ as $(x_2, 0)$ is at least as far from v_1 as $v_2 = (x_2, y_2)$ and $\|v_1 - v_2\| \geq r$. Thus $r_{v_1} < x_3 < r_{v_3}$. Similarly, if $y_1 > y_3$ then $r_{v_1} \leq x_3 < r_{v_3}$. Thus we may assume $y_1 \leq \min\{y_2, y_3\}$. Set $v_4 = (x_4, y_4) := (r_{v_1}, 0)$, $\tilde{v}_1 := (x_1, r)$, and assume $r_{v_3} \leq r_{v_1}$. Then $x_3 < r_{v_3} \leq r_{v_1}$ and so $v_2, v_3 \in (x_1, x_4) \times [y_1, r)$. But $r_{v_3} \leq r_{v_1}$ implies that $v_3 \notin B_r(v_4)$. Also $|x_4 - x_1| \leq r$, so $(x_1, x_4) \times [y_1, r) \subseteq B_r(\tilde{v}_1) \cup B_r(v_4)$. Hence $v_3 \in B_r(\tilde{v}_1)$ and so $\tilde{v}_1 \in B_r(v_3)$. But then $(x_1, x_3) \times [y_3, r) \subseteq B_r(v_3)$. Also $(x_1, x_3) \times [y_1, y_3) \subseteq B_r(v_1) \cup B_r(v_3)$, so there is no possible location for the vertex $v_2 \in ((x_1, x_3) \times [y_1, r)) \setminus (B_r(v_1) \cup B_r(v_3)) = \emptyset$, a contradiction. □

A break between good components C and C' is called a *bottom left clean* break if the following holds. Assume C is to the left of C' and $v \in C$ is the rightmost vertex of C

that is within distance r of ∂S_h^-. Then there do not exist vertices u, u' in some component $C'' \neq C$ (possibly $C'' = C'$) and within distance r of ∂S_h^-, with u on the left of v, u' on the right of v, and $r_{u'} > r_v$ (see Figure 10). We define bottom right clean and top left/right clean similarly.

Lemma 23. *There exists $h_0 > 0$ such that for all $h \geq h_0$ and $r \geq 7$, the intensity of breaks that are bottom left clean and lie between two good components, each of width at least $e^{h/4}$, is at least $I_{h,r}/38$.*

Proof. If $h \leq \frac{\sqrt{3}}{2}r$ then all breaks are bottom left clean as no component C'' can cross the vertical crossing sensor-path through v. The result then follows from Lemma 12 for sufficiently large h. Thus we may assume $h > \frac{\sqrt{3}}{2}r$. Fix a large constant K and choose h_0 sufficiently large so that for $h \geq h_0$,

$$e^{h/4} \geq 2Kh > Kr.$$

We define a break b between two good components C and C' to be *bad* if C and C' are of width at least $e^{h/4}$, but the break is not bottom left clean. Hence a break is *good* if *either* it is bottom left clean, *or* it is adjacent to a good component of width less than $e^{h/4}$. Suppose b is bad and let $v \in C$ and $u, u' \in C''$ be as in the above definition, with u the leftmost vertex for which such a pair (u, u') exists. Let γ be a sensor-path in C'' joining u to u'. Then this path passes either above or below the vertex v (see Figure 10). In both cases we shall construct a map ψ_b as in Lemma 21 which converts the break into a bottom left clean (and hence good) break. In each case, the map ψ_b will only affect vertices within distance $O(r)$ of the break, and hence (for large enough K) will not affect components to the left of C or to the right of C'. In particular it will not affect any break that is not adjacent to b. Moreover, the only effect on C and C' relevant to adjacent breaks will be to possibly change their widths (by $O(r)$). This may cause an adjacent break to become good by reducing the width of C or C' below $e^{h/4}$. (As b is bad, both C and C' are originally of width at least $e^{h/4}$.)

Claim 0.1. No sensor-path γ_C of C can pass under a vertex w that is to the right of v and within distance r of ∂S_h^-.

If a line segment $v_i v_{i+1}$ of γ_C passes under w with v_i to the left and v_{i+1} to the right of w, then $v_{i+1} \in C$ must at least distance r from ∂S_h^- (and hence be above w) by choice of v as the rightmost vertex of C within distance r of ∂S_h^-. But then v_i must be below w. Hence w is adjacent to both v_i and v_{i+1} as both x and y-coordinates of w are between those of v_i and v_{i+1}, and $\|v_i - v_{i+1}\| < r$. Hence $w \in C$ and is to the right of v, contradicting the choice of v.

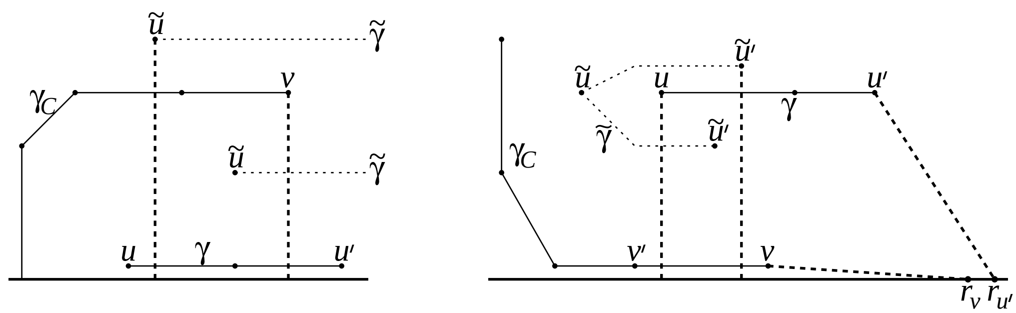

Figure 10: Two configurations giving an unclean break. The vertices v, v' lie in C and u, u' lie in C''. Path γ joining u to u' passes below v (left) or above v (right). Other paths $\tilde{\gamma}$ demonstrating the uncleanness of break cannot exist before (left) or after (right) surgery as sensor-paths in distinct components would then be too close.

Claim 0.1 implies that any crossing sensor-path of C must pass to the left of u as it cannot cross γ, cannot meet ∂S_h^- to the right of v, and cannot pass under u'. Hence x_C^+ is at least $\frac{r}{2}$ to the left of u. As the width of C is at least Kr, we also deduce that there is a sensor-path γ_C joining v to a point of ∂S_h^- that is at least Kr to the left of u (x_C^- is at most $\frac{r}{2}$ to the left of the leftmost vertex of C within $\frac{\sqrt{3}}{2}r$ of ∂S_h^-). Let γ_C be a such path. We may assume (by e.g., taking it to be of minimal length) that γ_C is a simple path.

Case 1. The sensor-path γ passes below v.

We first show that in this case there cannot be another pair $(\tilde{u}, \tilde{u}')$ in a component $\tilde{C}''$ demonstrating the uncleaness of the break unless $\tilde{C}'' = C''$ and any sensor-path $\tilde{\gamma}$ joining $\tilde{u}$ and $\tilde{u}'$ also passes below v. Indeed, if the path $\tilde{\gamma}$ went below v and $\tilde{C}'' \neq C''$, there would be two sensor-paths from different components passing below $w := v$ (see Figure 10, left), while if $\tilde{\gamma}$ went above v, then there would be two sensor-paths (γ_C and γ with γ_C above γ) from different components passing below $w := \tilde{u}$ (u is to the left of $\tilde{u}$ by choice of u; also γ_C must pass under $\tilde{u}$ as it cannot pass under $\tilde{u}'$ by Claim 0.1 and it cannot cross $\tilde{\gamma}$). In either case this would contradict either Lemma 6 or Lemma 7 as either the two sensor-paths would approach within distance $\frac{r}{2}$ of each other, or the higher one would approach within distance $\frac{r}{2}$ of w.

Define the point s to be vertically below v at distance r from v (so s will actually lie outside of the strip S_h). Let p be the point at distance r from v, but $\frac{\sqrt{3}}{2}r$ above s and to the right of s (see Figure 11). The point q is defined to be the nearest point to s that is at distance r from both v and p, so that vpq forms an equilateral triangle. The point s' is r to the left of s, $ws's$ is equilateral with w above $s's$, and g is at distance r from w, at the same height as v, and to the left of w.

Define the region $D^+ = (B_r(s') \cap B_r(v)) \setminus (B_r(g) \cup B_r(q))$. Define the region D^- to be those points that are outside and below $B_r(v)$ but above s, to the left of q, and either within horizontal distance r of q (if below q) or within distance r of q (if above q). Let L be the set of points in $\mathbb{R}^2$ that are above and to the left of s.

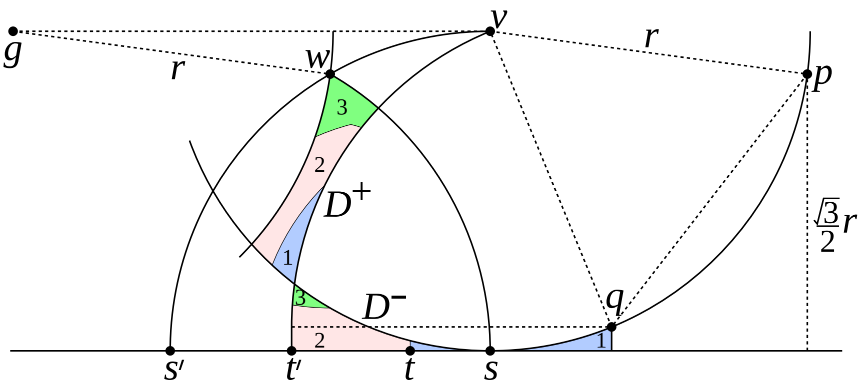

Figure 11: The regions $D^{\pm}$. The subsets $D_i^{\pm}$, $i = 1, 2, 3$, used in the proof of Lemma 24 are given by the shading. Note that the horizontal line through s is *below* ∂S_h^-.

Let $\phi\colon D^- \to D^+$ be the injective area-preserving map given by Lemma 24 below. Since γ joins u and u' and crosses vs, the last vertex u_L of γ lying in L must lie in $(B_r(s) \setminus B_r(v)) \cap L$. This vertex u_L lies in C'' and is within distance r of any point of D^+. But any point of D^+ is within distance r of $v \in C$. Thus $\mathcal{P} \cap D^+$ must be empty, otherwise the components C and C'' would be joined. Move all the vertices $x \in \mathcal{P} \cap D^-$ to their corresponding positions $\phi(x) \in D^+$. Some points of D^- may lie outside S_h, but if $D^- \cap S_h \neq \emptyset$ then D^+ must lie entirely within S_h. We define ψ_b in this case to be the induced map on (this subset of) the probability space of $G_{h,r}$.

Claim 1.1. This map neither creates nor destroys any break.

Equivalently, ψ_b does not create or destroy any good component. All vertices originally in D^- were in C'' as consecutive vertices on γ to the left and right of s are, between them, adjacent to every vertex in D^-. Thus C'' is the only component that can lose vertices. On the other hand, C gains the vertices that were moved since they are all now adjacent to v. Since C does not lose any vertices, C remains good.

The vertices that have been moved are now not adjacent to anything outside $B_r(v)$ to the right of q, and there are no longer any vertices in D^-, so they are no longer adjacent to anything outside of C to the right of s. Recall that γ_C joins v to ∂S_h^- at a point far to the left of v. Let $E \subseteq S_h$ be the region cordoned off by γ_C and vs. Suppose a vertex $\phi(x) \in D^+$ is adjacent (or equal) to a vertex z outside of E. Let $x' = x$ if $x \in L$, and

$x' = u_L$, the last vertex of γ in L, if $x \notin L$. Then $x' \in E$ as a path in C'' joining x' to u', say, cannot cross γ_C. By assumption, the vertical segment of γ_C meeting ∂S_h^- is far from $\phi(x)$, so either $\phi(x)z$ or $x'\phi(x)$ must cross an edge v_iv_{i+1} of γ_C. But $\|\phi(x)-x'\| \leq (\sqrt{3}-1)r$ by Lemma 24. Thus there is a curve joining $x' \in C''$ to z and crossing v_iv_{i+1} which is of length at most $\|z-\phi(x)\| + \|\phi(x)-x'\| \leq \sqrt{3}r$. Hence either x' or z is within distance $\frac{\sqrt{3}}{2}r$ of the line segment v_iv_{i+1}. But any such vertex must be adjacent to either v_i or v_{i+1}. As $x' \notin C$, we must have $z \in C$. Thus if $\phi(x)$ is adjacent to a vertex $z \notin C$, then z must be in a component of the new graph which is contained within E. This component cannot be good (and most importantly, it cannot meet C'), as a crossing sensor-path for this component would cross γ_C. In particular, we have not merged two good components.

Now consider the component C''. Suppose C'' is good (so that $C'' = C'$). Consider the segment of a crossing sensor-path $\gamma_{C'}$ from u to ∂S_h^+. This must go through some vertex in D^-. Let u_D be chosen so that it is the last such vertex in D^- on this path. The next vertex u_D^+ on $\gamma_{C'}$ must lie to the right of q, outside of $B_r(v)$, and within r of D^-. Thus u_D^+ lies below p and hence within $\frac{\sqrt{3}}{2}r$ of ∂S_h^-. Hence the subset $C''_\psi \subseteq C''$ of vertices still connected to u_D^+ is good.

Now suppose there is a second good component in $C'' \setminus C''_\psi$ and suppose $\gamma'_{C'}$ is a crossing sensor-path in this component joining some vertex of D^- to ∂S_h^+. Let $u_D'^+$ be the vertex after the last vertex of $\gamma'_{C'}$ in D^-. Without loss of generality $u_D'^+$ is to the left of u_D^+. A simple calculation shows that u_D^+ is within distance $\sqrt{3}r$ of v. But then $u_D'^+$ is separated from ∂S_h^+ by a region $B_r(u_D^+) \cup B_r(v) \cup D^-$ of width at least r, contradicting the existence of the path $\gamma'_{C'}$. Thus after removing the vertices in D^-, C'' contains a single good component C''_ψ containing the vertices of $\gamma_{C'}$ after u_D, and the remaining vertices of C'' lie in bad components.

If on the other hand C'' is bad, removing vertices from C'' will not change this fact, so all components of $C'' \setminus D^-$ are bad. Hence the resulting configuration is still a break between good components, and no other breaks have been created or destroyed.

Claim 1.2. The break b is transformed into a bottom left clean break.

From the proof of Claim 1.1 we know that all vertices of $C_\psi \setminus C$ lie in E. But if a vertex $v' \in E$ is to the right of v then γ_C (forming part of ∂E) must pass below v', contradicting Claim 0.1. Hence v is still the rightmost vertex of the new component $C_\psi \supseteq C$ that is within r of ∂S_h^-.

There is now no sensor-path outside of C passing below v as such a path would contain a vertex in D^-. If some pair $(\tilde{u}, \tilde{u}')$ demonstrated the uncleanness of this break, then the path joining $\tilde{u}$ and $\tilde{u}'$ would pass above v and not contain any vertices in D^+ (as these vertices have been absorbed into C). But then $\tilde{u}$ and $\tilde{u}'$ would have demonstrated the uncleanness of the original break, which we have already ruled out.

Claim 1.3. The map ψ_b is injective and measure preserving on Case 1 configurations.

As v is still the rightmost vertex of C within r of ∂S_h^-, the sets D^+ and D^- are determined by the resulting configuration. Hence this transformation can be reversed, and ψ_b is injective on this subset of the probability space. As we can regard ψ_b as just swapping regions in S_h, it is clearly measure-preserving. Thus the claim follows.

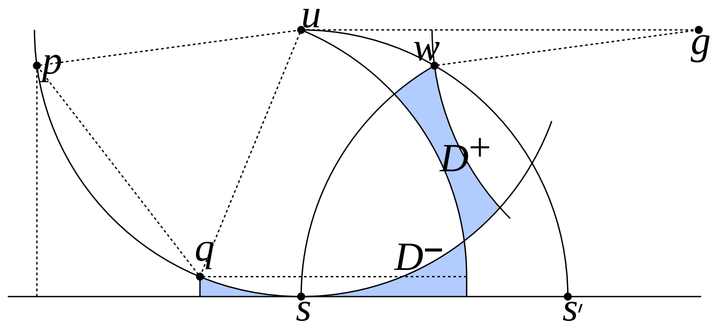

Figure 12: The case when γ passes above v.

Case 2. The path γ passes above v.

Recall that the path γ_C which joins v to ∂S_h^- must pass under the vertex u on γ. We now repeat the above argument with u in place of v and with left and right reversed (see Figure 12). However, there are a few differences in the proof. Note that the surgery is now being applied to the good component C as all vertices in D^- are adjacent to some vertex of γ_C. Let C_r be the set of vertices of C that are within distance r of ∂S_h^- and are to the right of the region D^-. Note that if $v \in D^-$ then $C_r = \emptyset$, otherwise v is the rightmost vertex of C_r. Let C_l be the set of vertices of C that are within distance r of ∂S_h^- and are to the left of the region D^-. Note that $C_l \neq \emptyset$ as, for example, it contains the first vertex of the sensor-path γ_C which is at least Kr to the left of u. Finally, let C^+ be the set of vertices $\phi(x)$, $x \in \mathcal{P} \cap D^-$, that have been moved by ψ_b.

Claim 2.1. If a sensor-path γ_y avoiding $C'' \cup D^-$ joins a vertex $y \in C^+ \cup C_r$ to a vertex which is to the left of u, then it must pass through a vertex $z \notin C^+$ which is to the right of u', within distance r of ∂S_h^-, and satisfies $r_z > r_v$.

Note that by replacing y by the last vertex of γ_y in $C^+ \cup C_r$, we can assume without loss of generality that y is the only vertex of γ_y in $C^+ \cup C_r$. In particular, all vertices of γ_y except possibly y are vertices in the original graph $G_{h,r}$.

As there is no sensor-path below u avoiding $C'' \cup D^-$, γ_y must pass to the right of u' and then above u. Indeed, γ_y cannot cross the sensor-path γ joining u and u' by the same argument as in Claim 1.1. Hence γ_y must cross the horizontal ray $u' + ([0, \infty) \times \{0\})$ from u'.

Suppose y is to the left of u'. Since the vertices of γ_y other than y cannot approach within distance r of $u' \in C''$, there must be some vertex z on γ_y which is below and to the right of u'. However, any such vertex z would satisfy $r_z > r_{u'} > r_v$ as required. Thus we can assume that y is already to the right of u', and as u' is to the right of v we must then have $y \notin C_r$ by choice of v. Hence $y = \phi(x) \in C^+$ and all other vertices of γ_y lie outside of $C^+ \cup C_r$.

Let z be the vertex adjacent to y on γ_y. Now $z \notin B_r(u) \cup D^-$ as $z \notin C'' \cup D^-$. Also $z \notin B_r(s) \cap B_r(s')$ as otherwise z would be adjacent to the first vertex of γ_C that is to the right of s, and hence z would lie in C_r. As z is also adjacent to $y \in D^+$, z must lie to the right of y and hence to the right of u'. If z is within distance r of ∂S_h^-, then $z \notin C$ as $z \notin C_r$. Applying Lemma 22 to $\{v, u', z\}$ then implies $r_z > r_v$ as required. Hence we can assume that z is not within distance r of ∂S_h^-. But that means that z must higher than the point g in Figure 12. But as $z \in B_r(y) \setminus B_r(u)$ and $y \in D^+$, $z \in B_r(w)$. However, $u' \notin B_r(z)$, $u' \notin B_r(s) \cap B_r(s')$ (as otherwise it would be adjacent to a vertex in γ_C), and u' is below z (as it is within distance r of ∂S_h^-). As u' is to the left of y and avoids $B_r(z) \cup (B_r(s) \cap B_r(s'))$ and as $w \in B_r(z)$, u' is in fact to the left of w. In particular, u' is at most $\frac{r}{2}$ to the right of v and $\|u' - v\| \geq r$. But this implies $r_{u'} \leq r_v$, contradicting the choice of u'.

Claim 2.2. This map does not destroy any break, and can generate at most one new break immediately to the right of b.

After surgery, there is a single component C_ψ containing C_l in the transformed graph, which is good (by the argument in Claim 1.1 showing that C'' remains good if it was originally so). As before, it is possible that the moved vertices $\phi(x) \in D^+$ may join components. However, to show that ψ_b does not destroy breaks, it is enough to show that there is no path from any $\phi(x) \in D^+$ to C_l. Taking the shortest such path, we can assume that such a path does not meet C'', as no path could join C'' and C_l without first going through a vertex in C^+. Claim 2.1 then implies C_l is joined in the original graph to a vertex z within distance r of ∂S_h^- and to the right of u' and hence to the right of v. But then $z \in C$ contradicts the choice of v.

It is possible that a new good component may be generated. This may happen if C'' becomes good by the addition of the vertices in C^+ and possibly by amalgamation with some other bad components. However, it is clear that at most one new good component is formed as all the new vertices are joined to a single component. Hence at most one new break is formed, and that this break is immediately to the right of b, which we now identify with the break immediately after the good component containing C_l.

Claim 2.3. The break b is transformed into a bottom left clean break.

Let C_ψ be the (good) component containing C_l in the transformed graph. Then $C_\psi \subseteq C$.

Let v' be the rightmost vertex of C_ψ that is within distance r of ∂S_h^-. Then Claim 2.1 implies that v' is to the left of u, as otherwise (as in Claim 2.2) we would have a vertex $z \in C_r$ to the right of v. Suppose there is a pair $(\tilde{u}, \tilde{u}')$ demonstrating the uncleanness of this break (see Figure 10, right) and let $\tilde{\gamma}$ be the path between them. As v' is to the left of u, $\tilde{u}$ must also be to the left of u.

Now $\tilde{u} \notin C''$, as otherwise we could have originally used the pair $(\tilde{u}, u')$ in place of (u, u'), contradicting the choice of u as the leftmost vertex for which such a pair (u, u') exists. Suppose then that $\tilde{\gamma}$ contains one of the moved vertices $\phi(x) \in C^+$ and consider the shortest subpath of $\tilde{\gamma}$ from $\tilde{u}$ to C^+. This path avoids C'' as otherwise $\tilde{u} \in C''$. Thus by Claim 2.1 $\tilde{\gamma}$ contains a vertex $z \notin C^+$ with $r_z > r_v$. But the segment of this path from $\tilde{u}$ to z does not meet C^+. Thus this segment exists in the original graph. But then $(\tilde{u}, z)$ demonstrates the uncleanness of the original break and $\tilde{u}$ is to the left of u, contradicting the choice of u. Hence we may assume $\tilde{\gamma}$ avoids C^+ and so is a path in the original graph $G_{h,r}$.

Now consider $\tilde{u}'$. As $\tilde{\gamma}$ is a sensor-path in the original graph and $\tilde{u} \notin C''$ we must have $\tilde{u}' \notin C''$. If $\tilde{u}'$ were to the right of u', then $\tilde{u}'$ could not be adjacent to v (by choice of v) and then applying Lemma 22 to the vertices $\{v, u', \tilde{u}'\}$ gives $r_{\tilde{u}'} > r_v$. Hence the pair $(\tilde{u}, \tilde{u}')$ demonstrates the uncleanness of the original break, again contradicting the choice of u. Thus $\tilde{u}'$ is to the left of u'. Suppose that $\tilde{u}'$ is to the right of u. The path $\tilde{\gamma}$ cannot pass below u as this would imply it would pass through D^- which is now empty. Thus $\tilde{\gamma}$ passes above u. Thus the line segment $v\tilde{u}'$ crosses γ. But both v and $\tilde{u}'$ must be at least $\frac{\sqrt{3}}{2}r$ from γ, so $\|\tilde{u}' - v\| \geq \sqrt{3}r$. But both v and $\tilde{u}'$ are within r of ∂S_h^- and so $\tilde{u}'$ must be at horizontal distance at least $\sqrt{2}r$ from v. If $\tilde{u}'$ is to the right of v then $r_{\tilde{u}'} > r_v$ and so $(\tilde{u}, \tilde{u}')$ demonstrate the uncleanness of the original break, contradicting the choice of u. Thus v is to the right of $\tilde{u}'$. But then we have a paths γ and γ_C passing below $\tilde{u}'$, with γ above γ_C. Thus γ passes within distance $\frac{r}{2}$ of either $\tilde{u}'$ or γ_C, contradicting the fact that γ is a path in a different component from $\tilde{u}'$ or γ_C. Thus $\tilde{u}'$ is to the left of u. But then $\tilde{u}'$ must be at least $\frac{\sqrt{3}}{2}r$ from ∂S_h^- as the path from v' to v in the original component C passes below $\tilde{u}'$. As $\tilde{u}'$ is to the left of u, within vertical distance $r - \frac{\sqrt{3}}{2}r$ of u, and $\|\tilde{u}' - u\| \geq r$, u' must be at least $(1 - (1 - \frac{\sqrt{3}}{2})^2)^{1/2}r$ to the left of u, and hence is to the left of p in Figure 12. By assumption $r_{\tilde{u}'} > r_{v'}$, so $v' \notin B_r(\tilde{u}') \cup B_r((r_{\tilde{u}'}, 0))$. Since v' is within distance r of ∂S_h^-, v' must be at least $\frac{r}{2}$ to the left of $\tilde{u}'$. But then it is at horizontal distance more than r from D^-. However, there is a path in C joining v' to some point in D^- and the last vertex on this path before entering D^- is within distance r of ∂S_h^-. This contradicts the fact that v' is the rightmost vertex of C_ψ within distance r of ∂S_h^-.

Claim 2.4 The map ψ_b is at most 2-to-1 on Case 2 configurations, and is measure-preserving when restricted to subsets where it is injective.

Consider the rightmost vertex v' of C_ψ that is within distance r of ∂S_h^-. The x-coordinate of v' must be within r of that of q, so determines the x-coordinate position of u up to an interval of length r. The y-coordinate of u must lie in $[\frac{\sqrt{3}}{2}r, r]$ as otherwise no path γ_C could pass beneath it. Thus, given v', u is restricted to lie in an r by $(\sqrt{3}-1)\frac{r}{2}$ rectangle R. As all points of D^+ are within distance $\frac{\sqrt{3}}{2}r$ of ∂S_h^-, all vertices in R lie in the original graph $G_{h,r}$. The subgraph of $G_{h,r}$ consisting of vertices in R can have at most two components. Once a component is selected, u is uniquely determined as the leftmost vertex in this component in R. Thus there are at most two choices for u. Thus the mapping is at most 2-to-1. It is clear that it is also measure-preserving on subsets where it is injective.

Putting both cases together, we have an (at most) 3-to-1 locally measure preserving mapping on an unclean break converting it into a clean break which can, at worst, generate a new break immediately to the right of b and convert neighboring breaks into good breaks. By Lemma 21, the intensity of good breaks is then at least $I_{h,r}/37$. The proportion of breaks that are not surrounded by good components of widths at least $e^{h/4}$ is at most $2(e^{h/4} + c_W)e^{-h/3}$ by Lemma 12. This is at most 10^{-4}, say, if $h \geq h_0$ and h_0 is chosen sufficiently large. Thus the intensity of bottom left clean breaks surrounded by wide good components is at least $I_{h,r}/37 - I_{h,r}/10^4 \geq I_{h,r}/38$. □

Lemma 24. *There exists an injective area-preserving map* $\phi\colon D^- \to D^+$ *such that*

(a) if $x \in D^- \cap L$ *then* $\|\phi(x) - x\| \leq (\sqrt{3}-1)r$,

(b) if $x \in D^- \setminus L$ *then* $\|\phi(x) - x'\| \leq (\sqrt{3}-1)r$ *for all* $x' \in (B_r(s) \setminus B_r(v)) \cap L$,

where $D^\pm$ *are as in Figure* 11 *and* L *is the set of points above and to the left of* s.

Proof. Define D_1^- to be the subset of D^- which is to the right of the point t, where t is $r/4$ to the left of s (see Figure 11). Let D_1^+ be the subset of points of D^+ that lie within $(\sqrt{3}-1)r$ of s. It can be shown that D_1^+ lies to the left of w, and hence all points of D_1^+ are also within distance $(\sqrt{3}-1)r$ of s'. It is then easy to check that all points of D_1^+ are within distance $(\sqrt{3}-1)r$ of any point $x' \in (B_r(s) \setminus B_r(v)) \cap L$.

Define D_3^- to be the set of points of D^- that are within distance $(\sqrt{3}-1)r$ of w, and D_3^+ to be the set of points of D^+ that are *not* within distance $(\sqrt{3}-1)r$ of either t or t', where t' is the bottom left corner of D^-. Let $D_2^\pm = D^\pm \setminus (D_1^\pm \cup D_3^\pm)$. Then all points of D_2^+ are within distance $(\sqrt{3}-1)r$ of any point in D_2^-, and all points of D_3^+ are within distance $(\sqrt{3}-1)r$ of any point of D_3^-.

A rather tedious calculation shows that $|D_i^-| < |D_i^+|$ for $i = 1, 2, 3$. Hence there exists an area-preserving injective function $\phi\colon D^- \to D^+$ so that $\phi(D_i^-) \subseteq D_i^+$. Any such map satisfies conditions (a) and (b), so the result follows. □

Lemma 25. *There exists $c > 0$ such that for all $h \geq \frac{\sqrt{3}}{4}r$, $r \geq 7$,*

$$cr^{-1}I_{h,r}^2 \leq I_{2h,r} \leq 100hI_{h,r}^2. \tag{31}$$

Proof. Cut the strip S_{2h} horizontally into two strips $S_h^{(t)}$ and $S_h^{(b)}$ with $S_h^{(t)}$ lying above $S_h^{(b)}$. If a (good) separating path γ exists for S_{2h} then its restrictions to $S_h^{(t)}$ and $S_h^{(b)}$ must contain separating paths $\gamma^{(t)}$ and $\gamma^{(b)}$ of $S_h^{(t)}$ and $S_h^{(b)}$ respectively. The path γ may cross the dividing line $y = h$ in more than one point, and hence may include segments that are not in either $\gamma^{(t)}$ or $\gamma^{(b)}$. Thus we cannot assume that $\gamma^{(t)}$ and $\gamma^{(b)}$ join up. Moreover, we cannot assume that $\gamma^{(t)}$ and $\gamma^{(b)}$ are good separating paths, even though γ is. However, the interval $J = [x_C^+, x_{C'}^-]$ corresponding to the break in $G_{2h,r}$ containing γ must intersect the intervals $J^{(t)} = [x_{C^{(t)}}^+, x_{C^{(t)\prime}}^-]$ and $J^{(b)} = [x_{C^{(b)}}^+, x_{C^{(b)\prime}}^-]$ corresponding to the breaks containing $\gamma^{(t)}$ and $\gamma^{(b)}$ in $S_h^{(t)}$ and $S_h^{(b)}$ respectively. Indeed, by Lemma 9, $\gamma^{(t)}$ and $\gamma^{(b)}$ intersect $J^{(t)} \times [h, 2h]$ and $J^{(b)} \times [0, h]$, but both paths are subpaths of γ which lies entirely within $J \times [0, 2h]$.

Write W, $W^{(t)}$, and $W^{(b)}$ for the random variables representing the widths of the breaks in S_{2h}, $S_h^{(t)}$, and $S_h^{(b)}$ respectively. Since $h \geq \frac{\sqrt{3}}{4} \cdot 7 > 1$ by hypothesis, we know from Lemma 10 that $\mathbb{E}(W) \leq 10h$ and that $\mathbb{E}(W^{(t)}) = \mathbb{E}(W^{(b)}) \leq 5h$. Therefore by Markov's inequality $\mathbb{P}(W \geq 20h) \leq \frac{1}{2}$. Fix a break in $S_h^{(t)}$ of width w. The probability that a break occurs within distance $20h$ of this break in $S_h^{(b)}$ is at most $\mathbb{E}(40h + w + W^{(b)})I_{h,r} \leq (45h + w)I_{h,r}$, as the break must start within an interval of length $40h + w + W^{(b)}$. Here the expectation is over the Poisson process in $S_h^{(b)}$, which is independent of the process in $S_h^{(t)}$. Thus the frequency of pairs of breaks lying within $20h$ of each other is at most $\mathbb{E}(45h + W^{(t)})I_{h,r}^2 \leq 50hI_{h,r}^2$. Since at least half of all breaks in S_{2h} give rise to such a pair, $I_{2h,r} \leq 100hI_{h,r}^2$.

For the lower bound, note that $I_{h,r} > 0$ and the region $\{(h, r) : h_0 \geq h \geq \frac{\sqrt{3}}{4}r, r \geq 7\} \subseteq \mathbb{R}^2$ is compact, so by making c smaller if necessary we may assume the result holds for all $h \leq h_0$. Thus in what follows we shall assume $h \geq h_0$ is sufficiently large.

Now fix the Poisson process in $S_h^{(t)}$ and $S_h^{(b)}$ and suppose that $C_L^{(t)}$ and $C_R^{(t)}$ are consecutive good components of widths at least $e^{h/4}$ separated by a bottom left clean break in $S_h^{(t)}$, while $C_L^{(b)}$ and $C_R^{(b)}$ are consecutive good components of widths at least $e^{h/4}$ separated by a top right clean break in $S_h^{(b)}$. Let $v^{(t)}$ be the rightmost vertex of $C_L^{(t)}$ that is within distance r of the bottom of $S_h^{(t)}$. Define $\tilde{C}_L^{(t)}$ as the union of all components C of $S_h^{(t)}$ which meet a vertex u which is to the left of $v^{(t)}$ and within distance r of the bottom of $S_h^{(t)}$. Define $\tilde{C}_R^{(t)}$ as the complement of $\tilde{C}_L^{(t)}$ in $G_{h,r}^{(t)}$. As the width of $C_R^{(t)}$ is $\gg r$, there is a vertex $u' \in C_R^{(t)}$ within distance $\frac{\sqrt{3}}{2}r < r$ of $\partial S_h^{(t)-}$ with $r_{u'} > r_{v^{(t)}}$. By the bottom left cleanness of the break, $C_R^{(t)}$ cannot then contain a vertex u within r of $\partial S_h^{(t)-}$ which is to the left of $v^{(t)}$. Thus $C_R^{(t)} \subseteq \tilde{C}_R^{(t)}$. Similarly define $\tilde{C}_R^{(b)}$ and its complement $\tilde{C}_L^{(b)}$, with $v^{(b)}$ the leftmost vertex of $C_R^{(b)}$ that is within distance r of the top of $S_h^{(b)}$. We now try to shift $S_h^{(t)}$ horizontally so that the breaks in $S_h^{(t)}$ and $S_h^{(b)}$ line up to form a break in S_{2h}.

Define $C_L(t)$ as the component of $G_{2h,r}$ containing the vertices of $C_L^{(t)}$ after $S_h^{(t)}$ has been shifted a distance t to the right relative to $S_h^{(b)}$. We claim that there exists $t_m \in \mathbb{R}$ and $\varepsilon > 0$ such that, for all $t \in (t_m - \varepsilon, t_m)$, $C_L(t)$ is a good component and there exists a break in $G_{2h,r}$ just to the right of this component and before $C_R^{(t)}$. We choose t_m to be the minimum t such that after the shift we have $d(\tilde{C}_L^{(t)}, \tilde{C}_R^{(b)}) = r$. Note that t_m almost surely exists, as $\tilde{C}_L^{(t)}$ contains all good components to the left of $C_L^{(t)}$ and $\tilde{C}_R^{(b)}$ contains all good components to the right of $C_R^{(b)}$, and both sets almost surely contain vertices within $\frac{r}{2}$ of the boundary between $S_h^{(t)}$ and $S_h^{(b)}$.

Let $x \in \tilde{C}_L^{(t)}$ and $y \in \tilde{C}_R^{(b)}$ be such that after a shift of t_m, $\|x - y\| = r$. Almost surely there is an $\varepsilon > 0$ such that no vertex of $\tilde{C}_R^{(t)}$ is within distance $r + \varepsilon$ of x and no vertex of $\tilde{C}_L^{(b)}$ is within distance $r + \varepsilon$ of y. Now fix $t \in (t_m - \varepsilon, t_m)$. Then with a shift of t we have $r < d(\tilde{C}_L^{(t)}, \tilde{C}_R^{(b)}) \leq \|x - y\| < r + \varepsilon$. Also, y is to the right of x (see Figure 13).

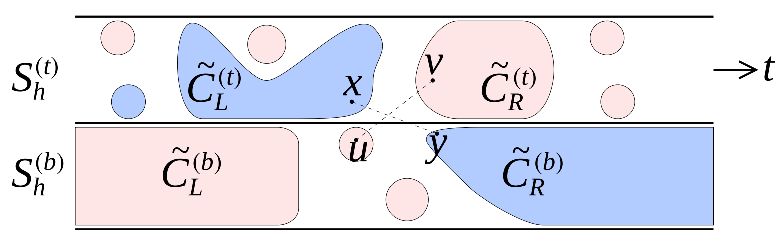

Figure 13: Aligning two clean breaks so as to form a break in S_{2h}.

Claim. If $u \in \tilde{C}_L^{(b)}$ and $v \in \tilde{C}_R^{(t)}$, then $\|u - v\| > r$.

Suppose otherwise. Then both u and v are within distance r of the line $y = h$. We first show that v is to the right of x. By the definition of $\tilde{C}_R^{(t)}$, v is to the right of $v^{(t)}$. Thus if x is to the right of v, then it is to the right of $v^{(t)}$ and hence $v \notin C_L^{(t)}$ by definition of $v^{(t)}$. Thus $x \in \tilde{C}_L^{(t)} \setminus C_L^{(t)}$ and so is in the same component as some point to the left of $v^{(t)}$ that is also within distance r of $\partial S_h^{(t)-}$. Applying Lemma 22 to $\{v^{(t)}, v, x\}$, we deduce that $r_x > r_{v^{(t)}}$, contradicting the bottom cleanness of the break. Thus v lies to the right of x. As $\|x - v\| \geq r + \varepsilon \geq \|x - y\|$, $v \in S_h^{(t)}$ cannot lie below the line segment xy. Also, if v lies above x, then as $\|x - v\| \geq r$, the line vu cannot pass to the left of x. Thus the ray from v through u crosses the ray from x though y. Similarly, the ray from u through v crosses the ray from y through x. Thus the finite line segments xy and uv intersect at some point z, say. Now $2r + 2\varepsilon \leq \|u - y\| + \|v - x\| \leq \|u - z\| + \|z - y\| + \|v - z\| + \|z - x\| =$

$\|x-y\|+\|u-v\|\le 2r+\varepsilon$, a contradiction. Hence $\|u-v\|>r$ as claimed.

The component in $G_{2h,r}$ containing $C_L^{(t)}$ must be contained within $\tilde{C}_L^{(t)}\cup\tilde{C}_L^{(b)}$ as there is no edge from this set to its complement $\tilde{C}_R^{(t)}\cup\tilde{C}_R^{(b)}$. It remains to estimate the frequency of these breaks for which ε is not too small and the resulting component $C_L(t)$ is good. The sets $\tilde{C}_L^{(t)}$ and $\tilde{C}_R^{(b)}$ can be found by exploring the Poisson process in $S_h^{(t)}$ and $S_h^{(b)}$ without encountering any vertex of $\tilde{C}_R^{(t)}$ or $\tilde{C}_L^{(b)}$. As the vertices x and y depend only on $\tilde{C}_L^{(t)}$ and $\tilde{C}_R^{(b)}$ we can condition on $\tilde{C}_L^{(t)}$ and $\tilde{C}_R^{(b)}$ and ask for the probability that no vertex of $\tilde{C}_R^{(t)}$ is within $r+\varepsilon$ of x and no vertex of $\tilde{C}_L^{(b)}$ is within $r+\varepsilon$ of y. The extra excluded area is at most $\pi(r+\varepsilon)^2-\pi r^2=2\pi r\varepsilon+\pi\varepsilon^2$ in each case. If $\varepsilon=\varepsilon_0/r$ this area is of order ε_0. Taking ε_0 to be sufficiently small, this occurs with probability at least $1-10^{-4}$, say. Thus by Lemma 23, the intensity of clean breaks where this occurs is at least $I_{h,r}/39$, say, independently in both $S_h^{(t)}$ and $S_h^{(b)}$.

Now since h is sufficiently large and the widths of the good component $C_L^{(t)}$ is at least $e^{h/4}$, we can show that with extremely high probability, $C_L^{(t)}$ joins with part of $\tilde{C}_L^{(b)}$ to form a good component $C_L(t)$ in S_{2h}. Indeed, all that is necessary is a column of non-empty $a\times b$ rectangles (a and b as in Lemma 10) spanning one of $a^{-1}e^{h/4}$ columns of S_{2h}. The probability that one column consists entirely of non-empty rectangles is $(1-e^{-ab})^{2h/b}\ge e^{-h/300}$ (as $ab\ge 5.77$ and $b\ge 2$). The probability that none of the $a^{-1}e^{h/4}$ columns have this property is $(1-e^{-h/300})^{a^{-1}e^{h/4}}$. But this is at most $\exp(-e^{h/5})$ for large enough h. Since this is far smaller than the intensity of breaks, we now have a lower bound on $I_{2h,r}$ of the form $(\varepsilon_0/r)(1-o(1))(I_{h,r}/39)^2\ge cr^{-1}I_{h,r}^2$. □

The following lemma shows that breaks are usually "almost rectangular" when $h\ll r^2$. We shall need this so that we can approximate the excluded areas by parabolic regions even when h is reasonably large, extending the argument in Section 6, which only applied when $h\le\frac{\sqrt{3}}{2}r$, to all $h\le\delta r^2$.

Lemma 26. *There is a constant $c>0$ such that, for all $r\ge 7$, the proportion of good components C for which the rightmost boundary of the excluded region to the right of C deviates more than an angle $\pm\theta$, $\theta\in[0,\pi/2]$, from vertical at any point is at most $e^{-(\theta-\sin\theta)r^2+ch}$.*

Proof. Let γ_b denote the boundary of the excluded region, that is, the rightmost boundary of $\bigcup_{v\in C}B_r(v)$ crossing S_h. We first note that for $h\ll r$, γ_b cannot deviate by more than an angle $O(h/r)$ from vertical and so $-(\theta-\sin\theta)r^2+ch\ge ch-O(h^3/r)$. Thus, by increasing c if necessary, the result is automatic for $h=O(1)$. Hence from now on we shall assume h is bounded away from zero.

Let A denote the set of points at distance at most r from γ_b and to the left of γ_b. The interior of A cannot contain any points of the Poisson process $\mathcal{P}$, although the leftmost boundary γ_l of A must contain the points of $\mathcal{P}$ giving rise to γ_b. The curves γ_b and γ_l are composed of sequences of arcs of circles of radii r centered on points on γ_l and γ_b respectively (see Figure 14). We can calculate the area of A by dividing A into sectors. Each sector has area $\frac{r}{2}$ times the length of the arc that gives rise to it (area $\phi\frac{r^2}{2}$ is $\frac{r}{2}$ times arc length ϕr when the angle of the sector is ϕ). Since the radii of the sectors start and end horizontally, the sum of the angles of the sectors giving rise to γ_l is the same as the sum of the angles of the sectors giving rise to γ_b. Hence $|\gamma_l| = |\gamma_b|$, where $|.|$ denotes the lengths of these paths. Thus, provided γ_l does not self-intersect (so that areas are not double counted), we obtain

$$|A| = \tfrac{r}{2}(|\gamma_l| + |\gamma_b|) = r|\gamma_b|.$$

It is clear from the definition of γ_b that γ_b does not self-intersect, however γ_l may self-intersect (see Figure 14). In this case, exclude sectors (equivalently sub-arcs of γ_l) that intersect previously encountered sectors. The remaining arcs of γ_l still have length at least h since the remaining arcs connect the top and bottom of S_h. Thus we still have

$$|A| \geq \tfrac{r}{2}(h + |\gamma_b|),$$

even when γ_l does self-intersect.

We now estimate the length $|\gamma_b|$ of γ_b. Let $\theta(s)$ be the (counterclockwise) angle of γ_b from vertical at the point on γ_b that is distance sr along γ_b from the point where γ_b meets ∂S_h^-. The function $\theta(s)$ is piecewise linear with derivative 1 except at the corners of γ_b where it discontinuously drops. Since γ_b goes a total vertical distance of h, we have

$$h = r\int_0^{|\gamma_b|/r} \cos\theta(s)\, ds.$$

Assume $\theta_0 = \theta(s_0)$ and $\theta_1 = \theta(s_1)$ are two values of $\theta(s)$ with $s_0 < s_1$ and $\theta_0 < \theta_1$. Then there are subintervals of values of s where $\theta(s)$ increases (with derivative 1) which between them cover all values from θ_0 to θ_1. Thus

$$h \leq |\gamma_b| + r\int_{\theta_0}^{\theta_1} (\cos\phi - 1)\, d\phi = |\gamma_b| - r(\theta_1 - \sin\theta_1) + r(\theta_0 - \sin\theta_0),$$

or equivalently

$$|\gamma_b| \geq h + r(\theta_1 - \sin\theta_1) - r(\theta_0 - \sin\theta_0).$$

In the case when γ_l does not self-intersect, we can take $\theta_0 = 0$, $\theta_1 = \theta$, or $\theta_0 = -\theta$, $\theta_1 = 0$ if θ measured clockwise. Note that $\theta(0) < 0 < \theta(|\gamma_b|/r)$ so 0 is always a value of $\theta(s)$ with

$0 < s < s_1$ (if $\theta_1 = \theta > 0$) or $s_0 < s < |\gamma_b|/r$ (if $\theta_0 = -\theta < 0$). Then $|\gamma_b| \geq h + r(\theta - \sin\theta)$, so for any fixed constant $c > 0$,

$$|A| - c|\gamma_b| = (r - c)|\gamma_b| \geq rh - ch + r^2(\theta - \sin\theta) - O(\min\{r, |\gamma_b|\}).$$

In the case when γ_l does self-intersect, we can choose $\theta_1 = \theta_0 + \pi$ for some θ_0 as $\theta(s)$ must increase by a total of more than π in order that γ_l self-intersects. Then $|\gamma_b| \geq h + \pi r - 2r$, so

$$\begin{aligned} |A| - c|\gamma_b| &\geq \tfrac{rh}{2} + \tfrac{r}{2}|\gamma_b| - c|\gamma_b| \\ &\geq rh - ch + r^2(\tfrac{\pi}{2} - 1) - O(r) \\ &\geq rh - ch + r^2(\theta - \sin\theta) - O(r). \end{aligned}$$

If $h \leq \frac{r}{2}$ then $|\gamma_b| = O(h)$ and γ_l does not self-intersect, otherwise $r = O(h)$. Thus for any h we have

$$|A| - c|\gamma_b| \geq rh + r^2(\theta - \sin\theta) - O(h) \tag{32}$$

regardless of whether or not γ_l self-intersects.

Now tile S_h with squares of side length $a \approx 1$ (which we will assume divides h). Let T be the set of squares of our tiling intersecting γ_b and assume T contains n squares. By Lemma 11, there are at most $\frac{1}{2}\mu^n$ choices for the set T starting at some fixed square R. We choose R to be the leftmost square on the bottom row of T adjacent to ∂S_h^-.

As γ_b passes through every square of T and there is no sensor within distance r of γ_b on the left of γ_b, there can be no sensor within distance $r - \sqrt{2}a$ of T on the left of T. This excludes an area of at least $|A| - \sqrt{2}a(|\gamma_l| + |\gamma_b|) \geq |A| - 2\sqrt{2}a|\gamma_b|$. Indeed, the excluded area contains all of A except those points within $\sqrt{2}a$ and to the left of γ_b (as these may be inside T), and those points within $\sqrt{2}a$ and to the right of γ_l (as from T we cannot guarantee that these are sufficiently close to γ_b). As all of T lies within distance $\sqrt{2}a$ of γ_b, the area of T is at most $2\sqrt{2}a|\gamma_b|$ and hence $n \leq 2\sqrt{2}a^{-1}|\gamma_b|$. Thus the intensity of such good components is bounded by

$$a^{-1} \sum_{n \leq 2\sqrt{2}a^{-1}|\gamma_b|} \tfrac{1}{2}\mu^n e^{-|A| + 2\sqrt{2}a|\gamma_b|} \leq e^{-|A| + c|\gamma_b|}$$

for some constant $c > 0$. Thus by (32), the intensity of such good components is

$$e^{-hr - r^2(\theta - \sin\theta) + O(h)}.$$

The total intensity of good components is at least $e^{-hr+O(h)}$ by (3) for $h \geq \frac{\sqrt{3}}{2}r$, and at least $he^{-hr} = e^{-hr+O(h)}$ for $h < \frac{\sqrt{3}}{2}r$, $h = \Omega(1)$, by (12). Thus the proportion of good components satisfying the conditions of the lemma is at most $e^{-r^2(\theta - \sin\theta) + O(h)}$, as claimed. $\square$

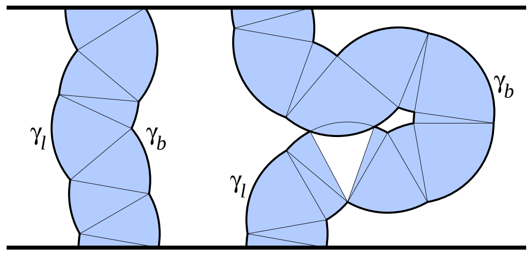

Figure 14: Estimating excluded area in Lemma 26.

8 Proof of Main Theorem

Proof of Theorem 2. The intensity of breaks is equal to the intensity of good components, or equivalently, the frequency of points of $\mathcal{P}$ along the strip that are the rightmost vertices of good components.

Assume first that $h \le \max\{\delta r^2, \frac{\sqrt{3}}{2}r\}$ with $\delta > 0$ a small constant. Using Lemma 26, there exists a $c > 0$ so that at most a proportion $e^{-(\theta-\sin\theta)r^2+ch}$ of breaks are such that γ_b makes an angle of more than $\theta = \frac{\pi}{3}$, say, with the vertical. This is at most $e^{-c'r^2}$ for some $c' > 0$ when $h \le \delta r^2$ and δ is sufficiently small. Moreover, this proportion is exactly zero if $h \le \frac{\sqrt{3}}{2}r$ as then it is impossible for the rightmost boundary of the excluded region to make an angle of more than $\frac{\pi}{3}$ with the vertical. Since $e^{-c'r^2} = O(hr^{-5/3})$ for $h \ge \frac{\sqrt{3}}{2}r$, we only need to consider the intensity of these "almost rectangular" breaks, as the inclusion of the non-rectangular breaks will only change the intensity by a factor of $e^{O(hr^{-5/3})}$, and this factor can be absorbed into the error term in (1).

We now show that the probability of a vertex v being the rightmost vertex of a good component followed by an almost rectangular break is at most $z^{-1}\varepsilon_{f_r}(z)e^{-hr}$, where $z = hr^{-1/3}$ and $f_r(x, x_0)$ is defined by

$$f_r(x,x_0) = \begin{cases} r^{4/3} - r^{2/3}(r^{4/3} - (x-x_0)^2)^{1/2}, & \text{if } |x - x_0| \le r^{2/3}; \\ \infty, & \text{otherwise.} \end{cases}$$

Condition first on the presence of a vertex v and the vertices of a crossing sensor-path γ_C^R that lies to the left of v (possibly including the vertex v), and consider the probability that γ_C^R is the "rightmost" crossing sensor-path of a good component whose rightmost vertex is v. Here the rightmost crossing path is constructed analogously to the leftmost crossing path from the proof of Lemma 12, simply swapping "left" and "right" throughout. Usually v will lie on γ_C^R, but this need not be the case when v is close to ∂S_h.

Let R be the subset of S_h that lies to the right of v, and let S be the region explored in the construction of γ_C^R as in the proof of Lemma 12. Let $\mathcal{P}'$ be the Poisson process

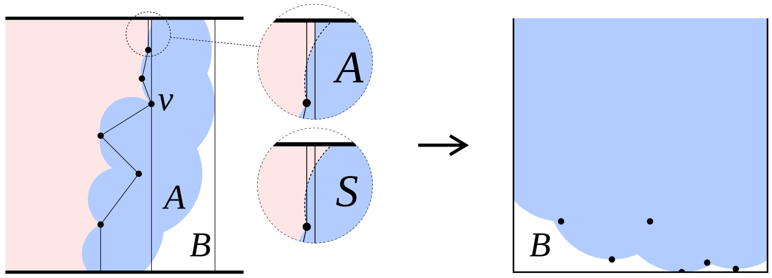

Figure 15: Rightmost crossing path and regions A and B defined in the proof of Theorem 2. Note that A includes points to the right of v, but to the left of the explored region S. Also, some points to the left of v can be to the right of A.

to the left of $S \cup R$, and let A be the subset of $S \cup R$ that is within distance r of the sensors of $\tilde{\mathcal{P}} := \gamma_C^R \cup \{v\} \cup \mathcal{P}'$ (see Figure 15). Note that no sensor-path outside of C can cross γ_C^R, so excluding points within distance r of $\tilde{\mathcal{P}}$ is the same as excluding points within distance r of $\tilde{\mathcal{P}} \cap C = C$. Also, any point in R that is to the left of S must necessarily be close to ∂S_h, and within distance r of some point of γ_C^R. Thus these points also need to be excluded if v is to be the *rightmost* vertex of C. Thus, conditioned on the event that the vertices of $\gamma_C^R \cup \{v\}$ exist, the event that γ_C^R is the rightmost crossing sensor-path of a good component whose rightmost vertex is v is precisely the event $A \cap \mathcal{P} = \emptyset$, and hence occurs with probability $\mathbb{E}_{\mathcal{P}'} e^{-|A|}$, where the expectation is over the choice of $\mathcal{P}'$.

Now rotate $\tilde{\mathcal{P}}$ clockwise $90°$, scale horizontally by $r^{-1/3}$ and vertically by $r^{1/3}$. Shift vertically so that v is on the x-axis. Provided the rightmost boundary of the original set A never becomes horizontal, the area of the region B defined in Section 6, using the function $f_r(x, x_0)$ above and the vertex set of $\tilde{\mathcal{P}}$, is equal to the area between A and the vertical line at distance r to the right of v. Hence $|S \setminus R| = |A| + |B| - hr$. The probability that a vertex is the rightmost vertex of a good component is obtained by integrating $\mathbb{E}_{\mathcal{P}'} e^{-|A|}$ over all valid choices of $\gamma_C \cup \{v\}$.

Now restrict these choices to those for which the right boundary of A stays within $\frac{\pi}{3}$ of vertical. (Note that whether or not the right boundary of A stays within $\frac{\pi}{3}$ of vertical is determined by γ_C^R. Indeed, the only points of A that can be affected by $\mathcal{P}' \cup \{v\} \setminus \gamma_C^R$ are those that are very close to ∂S_h, and ∂A is always within $\frac{\pi}{3}$ of vertical here.) Conditioned on the presence of vertices corresponding to those of the transformed $\gamma_C^R \cup \{v\}$ in $[0, z] \times [0, \infty]$, the probability that the region corresponding to $S \setminus R$ is empty is $\mathbb{E} e^{-|S \setminus R|} = \mathbb{E} e^{-|A| - |B| + hr}$. Now in the formula (18) defining $\varepsilon_{f_r}(x)$, this event contributes $z e^{|B|}$. Thus the contribution to $z^{-1} \varepsilon_{f_r}(z) e^{-hr}$ is just $\mathbb{E} e^{-|A|}$. Integrating over all choices of $\gamma_C^R \cup \{v\}$ for which the original excluded area A could have right boundary within $\frac{\pi}{3}$ of vertical gives that the probability that v is the rightmost vertex of a good component which is followed by an almost rectangular break is at most $z^{-1}\varepsilon_{f_r}(z)e^{-hr}$.

Conversely, consider a configuration $\mathcal{P}$ contributing to $z^{-1}\varepsilon_{f_r}(z)e^{-hr}$. Suppose that $\frac{\sqrt{3}}{2}r \le h \le \delta r^2$. We first show that we can restrict our attention to $\mathcal{P}$ for which the upper boundary of B has slope at most $r^{2/3}\tan\frac{\pi}{6}$. Indeed, for sufficiently small δ, $c_1 z - c_2(r^{2/3}\tan\frac{\pi}{6})^3 \le c_1\delta r^{5/3} - \frac{c_2}{\sqrt{27}}r^2 \le -c'r^2$, where c_1 and c_2 are the constants given by Lemma 15. Thus by Lemma 15, the remaining configurations contribute a fraction at most $e^{-c'r^2} = O(hr^{-5/3})$ to this expression. Scaling and rotating back to the S_h picture, the vertices of $\mathcal{P}$ defining the right boundary of A are within distance r of each other (or $\frac{r}{2}$ from the boundary of S_h). Thus they form a crossing sensor-path. As this is automatic when $h < \frac{\sqrt{3}}{2}r$, such a configuration is in the image of the mapping above. Thus the probability that a vertex is the rightmost vertex of a good component is

$$z^{-1}\varepsilon_{f_r}(z)e^{-hr}e^{O(hr^{-5/3})}. \tag{33}$$

A simple calculation shows that for $r \ge r_0$

$$\tfrac{1}{2}(x-x_0)^2 \le f_r(x,x_0) \le (1-(\tfrac{r_0}{r})^{4/3})\tfrac{1}{2}(x-x_0)^2 + (\tfrac{r_0}{r})^{4/3} f_{r_0}(x,x_0).$$

Indeed, the second inequality follows from the convexity of $\eta_r(x,x_0)$ (defined in (35) below) as a function of $r^{-4/3} \in [0, r_0^{-4/3}]$, followed by integration of (34).

Write B_r for the region B used in the definition of $\varepsilon_{f_r}(z)$, and B_∞ for the limit as $r \to \infty$, which gives the region B used in the definition of $\varepsilon(z)$. We have, using $r_0 = 7$,

$$|B_\infty| \le |B_r| \le (1-(\tfrac{7}{r})^{4/3})|B_\infty| + (\tfrac{7}{r})^{4/3}|B_7|.$$

Now in general for bounded random variables X and Y, Hölder's inequality gives that

$$\mathbb{E}(e^{(1-\varepsilon)X+\varepsilon Y}) \le (\mathbb{E}e^X)^{1-\varepsilon}(\mathbb{E}e^Y)^{\varepsilon}.$$

Thus, with $X = B_\infty$ and $Y = B_7$, we deduce that

$$\varepsilon(z) \le \varepsilon_{f_r}(z) \le \varepsilon(z)(\varepsilon_{f_7}(z)/\varepsilon(z))^{(7/r)^{4/3}}.$$

Now

$$\frac{\partial}{\partial x} f_r(x,x_0) = (x-x_0)\eta_r(x,x_0), \tag{34}$$

where

$$\eta_r(x,x_0) = (1-r^{-4/3}(x-x_0)^2)^{-1/2} \tag{35}$$

satisfies $1 \le \eta_r(x,x_0) \le 2$ for all $|x-x_0| \le 2$ and $r \ge 7$. Thus $\varepsilon_{f_7}(z)/\varepsilon(z) = e^{O(z)}$ by Lemma 16. Hence we have

$$\varepsilon_{f_r}(z) = \varepsilon(z)e^{O(z(7/r)^{4/3})} = \varepsilon(z)e^{O(hr^{-5/3})}.$$

The theorem follows for $h \le \max\{\delta r^2, \frac{\sqrt{3}}{2}r\}$, since there are h vertices per unit length along the strip and so $I_{h,r} = hz^{-1}\varepsilon_{f_r}(z)e^{-hr+O(hr^{-5/3})} = r^{1/3}\varepsilon(z)e^{-hr+O(hr^{-5/3})}$.

For $h > \max\{\delta r^2, \frac{\sqrt{3}}{2}r\}$, we use Lemma 25 to reduce inductively to the case $h \le \max\{\delta r^2, \frac{\sqrt{3}}{2}r\}$. Choose $k \in \mathbb{N}$ minimal so that $h_0 = h/2^k \le \max\{\delta r^2, \frac{\sqrt{3}}{2}r\}$. By induction on k, Lemma 25 implies that

$$(c/r)^{2^k-1} I_{h_0,r}^{2^k} \le I_{h,r} \le (200h_0)^{2^k-1} 2^{-k} I_{h_0,r}^{2^k}.$$

As $h_0 = \Theta(r^2)$ we have

$$I_{h,r} = I_{h_0,r}^{2^k} e^{O(2^k \log r)} = r^{2^k/3} \varepsilon(z/2^k)^{2^k} e^{-2^k h_0 r + O(2^k h_0 r^{-5/3}) + O(2^k \log r)}.$$

But $z/2^k = h_0 r^{-1/3} = \Theta(r^{5/3})$ is bounded away from zero. Thus by Theorem 3 we have $\varepsilon(z/2^k)^{2^k}/\varepsilon(z) = \exp(O(2^k))$. Hence

$$I_{h,r} = r^{1/3}\varepsilon(z)e^{-hr+O(hr^{-5/3})+O(2^k \log r)}.$$

However, the $O(hr^{-5/3})$ absorbs the $O(2^k \log r) = O(h(\log r)/r^2)$ term, giving the result. □

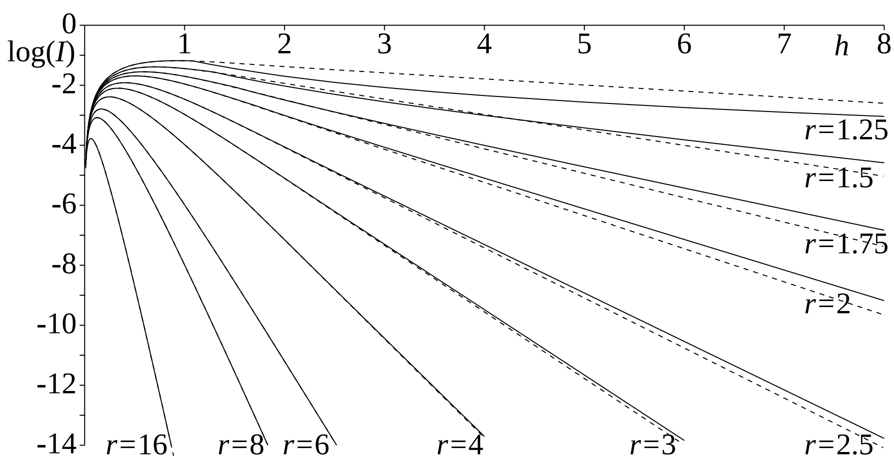

Figure 16: Plots of $\log(I_{h,r})$ against h for various values of r. Dotted lines indicate the approximation $r^{1/3}\varepsilon(hr^{-1/3})e^{-hr}$ from Theorem 2.

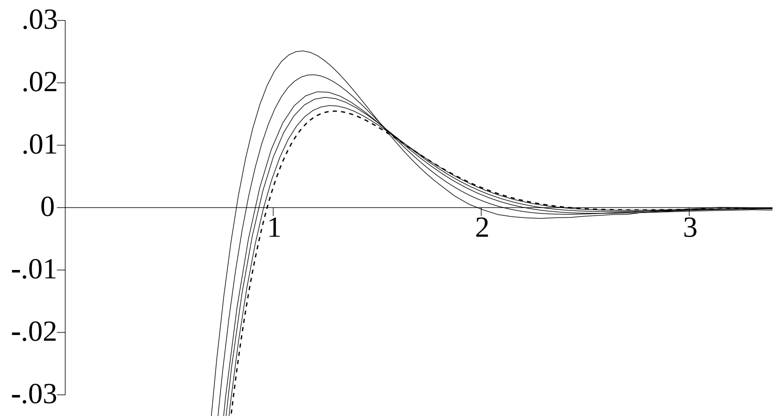

Figure 17: Plots of $\log(I_{h,r}) + \alpha_r h + \beta_r$ against $z = hr^{-1/3}$ for $r = 3, 4, 6, 8, 16$ (outer to inner solid lines respectively). Dotted line indicates the estimate $\log \varepsilon(z) - \alpha z - \beta$. Here α_r and β_r are as given in Table 1.

9 Simulation Results

In this section, we provide results of our simulations and compare them with our estimates. The main result here is that our estimates are almost indistinguishable from that observed in simulations for large r, and there is good agreement even for r significantly less than 7. All our simulations were sufficiently extensive so that statistical error bounds are not visible in the graphs.

To calculate the function $\varepsilon(z)$ in Figure 9, and to estimate the parameters α and β in (29), three different algorithms were employed. The first and simplest used just the definition of $\varepsilon(z)$: points were repeatedly placed in $[0, z] \times [0, z^2/2]$ according to a Poisson process, and the area of the region B was calculated. The average value of $ze^{|B|}$ was then computed.

The second algorithm was based on Lemma 17. Two instances of $\mathcal{P}$ were generated for strips of width $z/2$ and $\varepsilon(z)$ was calculated using (19). This algorithm was better at estimating $\varepsilon(z)$ than the first one for larger values of z. This is because the calculation of $ze^{|B|}$ for large z is significantly affected by low probability events giving rise to large values of $|B|$. This is difficult to estimate using simple Monte-Carlo methods. The formula in Lemma 17 captures most of this effect in the integral, which can be calculated with high accuracy. (If $|B|$ is large, it is probably due to one half-strip having its lowest vertex very high up.)

Finally, for small values of z one can use the Taylor expansion (30), which is quite accurate for $z < 1$.

For most parameter values, more than one algorithm was used so as to check the

r	α_r	β_r	α_r^{est}	β_r^{est}
1.25	0.0428[1]	3.6200[15]	0.0650	1.1773
1.5	0.3668[1]	1.6790[15]	0.4129	1.0732
1.75	0.7018[1]	1.2198[10]	0.7353	0.9927
2	1.0195[2]	1.0240[10]	1.0418	0.9273
2.5	1.6152[2]	0.8490[10]	1.6255	0.8253
3	2.1809[2]	0.7548[10]	2.1859	0.7474
4	3.2678[2]	0.6342[10]	3.2696	0.6316
6	5.3686[2]	0.4801[10]	5.3692	0.4787
8	7.4296[1]	0.3751[2]	7.4298	0.3749
16	15.5504[1]	0.1337[2]	15.5504	0.1337

Table 1: Estimates of α_r and β_r from simulations, together with the approximations given by (38). Numbers in square brackets indicate approximate 1 standard deviation errors in the last decimal place.

consistency of the different algorithms.

Our theoretical results strongly suggest that Theorem 2 can be strengthened to

$$I_{h,r} = e^{-\alpha_r h - \beta_r + o(1)}, \tag{36}$$

as $h \to \infty$, where

$$\begin{aligned} \alpha_r &= r - \alpha r^{-1/3} + O(r^{-5/3}), \\ \beta_r &= -\tfrac{1}{3}\log r - \beta + O(r^{-4/3}) \end{aligned} \tag{37}$$

as $r \to \infty$. This does not quite follow from the above, since we may still have a $O(\log h)$ error term in the exponent resulting from the inductive use of Lemma 25. We believe however that this is just an artifact of the proof. Indeed, a $O(\log h)$ term would arise from the factor of h on the right hand side of (31), which in turn comes from the $O(h)$ bound on the expected width of a break (Lemma 10) used in Lemma 25. In practice, one should use the amount by which the breaks in $S_h^{(t)}$ and $S_h^{(b)}$ can be moved relative to one another and still form a break in S_{2h}. This should be $O(1)$, and in particular should not depend noticeably on h.

Figure 16 plots the logarithm of the frequency of breaks observed in simulations against h for various values of r together with the estimate from Theorem 2. We observe that this approximation provides an extremely good fit to the simulation data except when r is close to the critical radius for the Gilbert model $r_c \approx 1.1984$ (see [14]). The network becomes highly disconnected in this case, and so the analysis breaks down.

We used two algorithms to estimate $I_{h,r}$. The first was to simply place points at random in a long strip and count good components. We generated the random points in order of their x-coordinate, and kept track of the changes in component structure as each point was added. It is only necessary to record points and components within horizontal distance r of the new point added, so the algorithm actually used very little memory, and the length of strip that could be simulated was limited only by the run-time.

The second algorithm estimated the probability that a fixed point v chosen uniformly at random in $\{0\} \times [0, h]$ is the rightmost vertex of a good component. To do this, points were generated further and further to the left of v until the status of the components containing points in $[-r, 0] \times [0, h]$ was determined (i.e., which vertices lay in the same component C as v, and whether C was good). If C was good, the area A to the right of v within distance r of any point of C was calculated. The probability that v is the rightmost point of C conditioned on the process to the left of v is then just $e^{-|A|}$.

The first method was effective when $I_{h,r}$ was not too small, while the second was effective provided r was not too close to r_c. For many choices of parameters both algorithms were used and results compared to check consistency between them.

The simulated values of α_r do indeed appear consistent with (37) and (29). Using simulations, one can estimate the error terms for α_r and β_r giving

$$\begin{aligned} \alpha_r &\approx \alpha_r^{\text{est}} := r - 1.12794r^{-1/3} - 0.20r^{-5/3} \\ \beta_r &\approx \beta_r^{\text{est}} := -\tfrac{1}{3}\log r + 1.05116 + 0.27r^{-4/3} \end{aligned} \tag{38}$$

Note that the constants 1.12794 and 1.05116 are the constants α and $-\beta$ from Theorem 3, and only the last coefficients (0.20 and 0.27) were estimated from the simulation estimates of $I_{h,r}$. From Table 1 one sees that the approximations in (38) are extremely good for $r \geq 3$, but get progressively less accurate for smaller values of r.

By comparison with Theorem 3, the $o(1)$ term in (36) should be approximately equal to $\log \varepsilon(z) - \alpha z - \beta$ where $z = hr^{-1/3}$. Figure 17 shows the values of this error term obtained from simulations for $r \geq 3$ (for $r < 3$ the error is much larger). Once again, the theoretical result is very close to the results from simulations.

References

[1] R. Arratia, L. Goldstein and L. Gordon, Two Moments Suffice for Poisson Approximations: The Chen-Stein Method, *Ann. Probab.* **17,** (1989), 9–25.

[2] P. Balister, B. Bollobás, S. Kumar, A. Sarkar, Reliable density estimates for deployment of sensors in thin strips, *ACM MobiCom 2007.*

[3] P. Balister, B. Bollobás, A. Sarkar, and M. Walters, Sentry selection in wireless networks, *Advances in Applied Probability* **42** (2010), 1-25.

[4] B. Bollobás, Random Graphs, 2nd edition. Cambridge studies in advanced mathematics 73, Cambridge University Press, (2001).

[5] B. Bollobás, The art of mathematics: coffee time in Memphis. Cambridge University Press, (2006).

[6] E.N. Gilbert, The probability of covering a sphere with N circular caps, *Biometrika* **56** (1965), 323–330.

[7] P. Hall, On the coverage of k-dimensional space by k-dimensional spheres, *Annals of Probability* **13** (1985), 991–1002.

[8] R. Horn and C. Johnson, Matrix Analysis, Cambridge University Press (1985).

[9] S. Janson, Random coverings in several dimensions, *Acta Mathematica* **156** (1986), 83–118.

[10] D.A. Klarner and R.L. Rivest, A proceedure for improve the upper bound for the number of n-ominos. Canadian J. Math. **25** (1973) 585–602.

[11] R. Meester, R. Roy, *Continuum percolation,* volume 119. Cambridge University Press (1996).

[12] P.A.P. Moran and S. Fazekas de St Groth, Random circles on a sphere, *Biometrika* **49** (1962), 389–396.

[13] M. Penrose, *Random Geometric Graphs,* Oxford Studies in Probability 5 (2003).

[14] J. Quintanilla, S. Torquato and R.M. Ziff, Efficient measurement of the percolation threshold for fully penetrable discs, *J. Phys.* A **33** (2000), L399–L407.