Longest distance of a non-uniform dispersion process on the infinite line

Yilun Shang[1]

[1] *Department of Computer and Information Sciences, Northumbria University, Newcastle NE1 8ST, UK*
Email: `shylmath@hotmail.com`

Abstract

The non-uniform dispersion process on the infinite integer line is a synchronous process where n particles are placed at the origin initially, and any particle not exclusively occupying an integer site will move at the next time step to the right adjacent integer with probability p_n and to the left with probability $1 - p_n$ independently. We characterize the longest distance from the origin when the dispersion process stops, which is shown to be $\Theta(n)$ with high probability for fairly general p_n.

Keywords: Random process, particle, dispersion, distance.
2000 MSC: 60G50, 60C05

---

1. Introduction

Let $G = (V, E)$ be the integer line graph, where the vertex set $V = \mathbb{Z}$ and the edge set $E = \{\{i, i+1\} : i \in \mathbb{Z}\}$. Let $n \in \mathbb{Z}^+ := \{0, 1, 2, \cdots\}$ be a non-negative integer. The *dispersion process* [1] on G with parameter $p_n \in [0, 1]$ is defined as follows. At time $t = 0$, n particles are placed at the origin. At any time step $t \in \mathbb{Z}^+$, if two or more particles sit on the same vertex, all these particles move independently to the right neighbor with probability p_n and to the left neighbor with probability $1 - p_n$ at time step $t + 1$. The process stops when all vertices in G are occupied by at most one particle.

The dispersion process has a couple of unique features. Firstly, all n particles move away from each other to find a private site in a synchronous manner following the identical rule. Secondly, a particle may (by chance) stop at a vertex for some time, move, and then stop again following an irregular time pattern. Using the standard Landau asymptotic notations and random graph convention [4], the authors of [1] show that (amongst other results) w.h.p. the longest distance from the origin is $O(n \ln n)$ for $p_n \equiv 1/2$ when the dispersion finishes. With a refined argument, the upper bound is further reduced to $O(n)$ in [2] still for the uniform version with $p_n \equiv 1/2$.

Along this direction of investigation, we in this letter study the longest distance for the non-uniform dispersion process with general transition probability $p_n \in [0, 1]$. The results are stated and discussed in Section 2, and the proofs are given in Section 3.

2. Results

Theorem 1. *Consider the dispersion process on G with parameter $p_n \in [0, 1]$.*

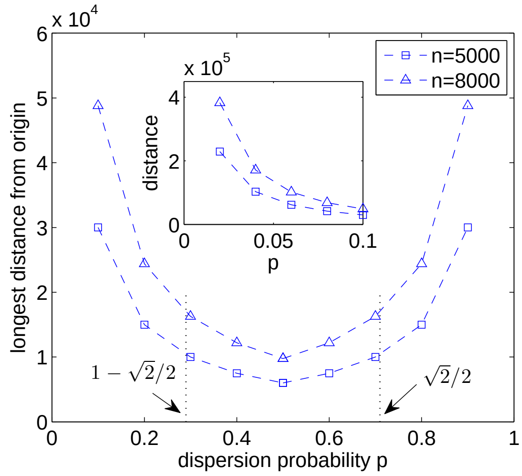

Figure 1: The longest distance from the origin as a function of dispersion probability p for $n = 5000$ (squares) and $n = 8000$ (triangles) particles. Inset shows the distance versus smaller p. Each data point is obtained by means of an ensemble averaging of 20 independently-computed processes.

(i) If there exist two constants $\underline{p}$ and $\overline{p}$ such that $1 - \frac{\sqrt{2}}{2} < \underline{p} \le p_n \le \overline{p} < \frac{\sqrt{2}}{2}$ for all n, then w.h.p. the longest distance from the origin is $\Theta(n)$ when the dispersion process stops. The dispersion process stops w.h.p. within $O(n^8)$ steps.

(ii) If $p_n = 0$ or $p_n = 1$, w.h.p. the longest distance from the origin is infinity, i.e., the dispersion process never stops. If $0 < p_n < 1$, w.h.p. the dispersion process stops within T_n steps for some $T_n < \infty$.

Recall that $f_n = \Theta(g_n)$ as $n \to \infty$ if there exists constants $c_1, c_2 > 0$ and n_0 such that $c_1 g_n \le f_n \le c_2 g_n$ for $n \ge n_0$ [4]. We will prove the result by providing an upper bound and a lower bound for the longest distance.

An omission from Theorem 1 is the lack of consideration of $0 < \underline{p} \le 1 - \frac{\sqrt{2}}{2} \approx 0.29$ (or $0.71 \approx \frac{\sqrt{2}}{2} \le \overline{p} < 1$). The restriction of $\underline{p}$ and $\overline{p}$ is mainly due to the estimate of the expectation of $\tilde{y}_j(t)$ in (2), which is essential in bounding the tail probability (3). Obviously, if there is such a $\underline{p}$ or $\overline{p}$, the dispersion will eventually end w.h.p.; c.f. Theorem 1(ii). In Fig. 1, we show the longest distance as a function for different constant probability $p_n \equiv p$ for $n = 5000$ and $n = 8000$ particles. The distance increases rapidly as p deviates from the middle value 0.5. But to determine the order of magnitude of the longest distance from the origin when $p \le 1 - \frac{\sqrt{2}}{2}$ (or equivalently $p \ge \frac{\sqrt{2}}{2}$) remains open.

3. Proofs

We start with the statement (ii). Then we prove the lower bound and the upper bound for statement (i). The following probabilistic inequality bounds the upper tail of the sum of a sequence of independent random variables, which has varied versions widely applied in random structures; c.f. [3, 4].

Lemma 1. ([2, Lemma 2.1]) *Suppose that* $x_1, x_2, \cdots, x_r$ *are independent non-negative integer random variables such that* $\mathrm{P}(x_j \geq k) \leq \sigma^k$ *for all* j *and some* $\sigma \in (0,1)$. *Let* $x = \sum_{j=1}^{r} x_j$ *and* $\mu = (1-\sigma)^{-1}$. *Then* $\mathrm{P}(x \geq (1+\varepsilon)\mu r) \leq e^{-c\varepsilon^2 r}$ *for any* $\varepsilon \in [0,1]$ *and some* $c = c(\sigma) > 0$.

Proof of Theorem 1 (ii). Firstly, it is obvious that if $p_n = 0$ or $p_n = 1$, the dispersion process never stops. Next, recall that the infinite integer line G is an infinite Abelian Cayley graph with the underlying group being the cyclic group $\mathbb{Z}$. For $0 < p_n < 1$, the termination of the dispersion process follows exactly the proof of [1, Theorem 8] by replacing $1/2$ with p_n. □

Proof of Theorem 1 (i). We claim that the longest distance is at least $\lfloor \frac{n}{2} \rfloor$, where $\lfloor\cdot\rfloor$ is the round down operator. This can be shown by contradiction. If this is not true, there are at most $f_n := 2(\lfloor \frac{n}{2} \rfloor - 1) + 1$ vertices in G which are occupied by particles. It is direct to show that the inequality $\lfloor \frac{n}{2} \rfloor < \frac{n+1}{2}$ for any $n \in \mathbb{Z}^+$ holds by considering odd and even n. Therefore, $f_n < n$. By the pigeonhole principle, there are two particles sitting on the same vertex, which contradicts the fact that the dispersion has stopped.

For the upper bound, we will adopt the proof technique of [2] by considering the ordered dispersion process where each particle is given a label $j \in [n] := \{1, 2, \cdots, n\}$. Denote by $\xi_j(t)$ the position of particle j at time step t and let $\eta_i(t)$ count the number of particles at vertex $i \in \mathbb{Z}$ at time t. At each time step t, all particles are relabeled such that $j_1 < j_2$ implies $\xi_{j_1}(t) \leq \xi_{j_2}(t)$. This means the labels of particles are always increasing along the real axis and all particles sitting on the same vertex can essentially be arbitrarily labeled. Graphically, we can think of the particles as piled in stacks located at vertices of G. We agree that the lowest labelled particle in the left to right ordering of the particles is at the bottom of the stack and the highest at the top; see Fig. 2.

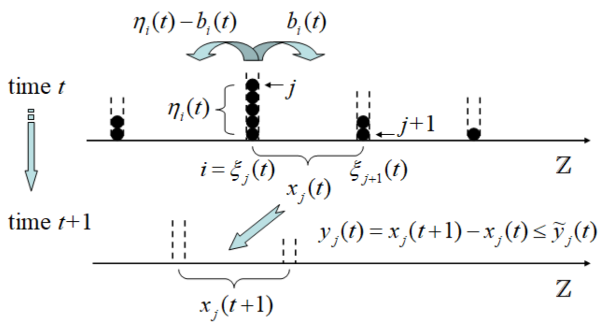

Figure 2: Schematic illustration of main parameters.

For $j \in [n-1]$, let $x_j(t) = \xi_{j+1}(t) - \xi_j(t)$ be the distance between particles $j+1$ and j at time t. Clearly, $\xi_n(t) - \xi_1(t) = \sum_{j=1}^{n-1} x_j(t)$ and $x_j(t) \in \mathbb{Z}^+$. Set the increment $y_j(t) = x_j(t+1) - x_j(t)$ and the index subset $\theta(t) = \{j : x_j(t) \geq 2\} \subseteq [n-1]$. For each vertex $i \in \mathbb{Z}$ satisfying $\eta_i(t) \geq 2$, we define independently a binomial random variable $b_i(t) \sim \mathrm{Binominal}(\eta_i(t), p_n)$, which represents the number of particles that will move from i (at time t) to $i+1$ (at time $t+1$). Recall that $\eta_{\xi_j(t)}(t)$ means the number of particles staying at the same vertex as j at time t, and $b_{\xi_j(t)}(t)$ of them will move to the right in the next time step. We define a random variable $\tilde{y}_j(t)$ for $j \in [n-1]$ and

$t \in \mathbb{Z}^+$ as follows.

$$\tilde{y}_j(t) = \begin{cases} 2, & \eta_{\xi_j(t)}(t) \ge 2, \eta_{\xi_{j+1}(t)}(t) \ge 2, j \in \theta(t), b_{\xi_j(t)}(t) = 0, b_{\xi_{j+1}(t)}(t) = \eta_{\xi_{j+1}(t)}(t); \\ -2, & \eta_{\xi_j(t)}(t) \ge 2, \eta_{\xi_{j+1}(t)}(t) \ge 2, j \in \theta(t), b_{\xi_j(t)}(t) > 0, b_{\xi_{j+1}(t)}(t) < \eta_{\xi_{j+1}(t)}(t); \\ 1, & \eta_{\xi_j(t)}(t) \ge 2, \eta_{\xi_{j+1}(t)}(t) = 1, j \in \theta(t), b_{\xi_j(t)}(t) = 0; \\ -1, & \eta_{\xi_j(t)}(t) \ge 2, \eta_{\xi_{j+1}(t)}(t) = 1, j \in \theta(t), b_{\xi_j(t)}(t) > 0; \\ 1, & \eta_{\xi_j(t)}(t) = 1, \eta_{\xi_{j+1}(t)}(t) \ge 2, j \in \theta(t), b_{\xi_{j+1}(t)}(t) = \eta_{\xi_{j+1}(t)}(t); \\ -1, & \eta_{\xi_j(t)}(t) = 1, \eta_{\xi_{j+1}(t)}(t) \ge 2, j \in \theta(t), b_{\xi_{j+1}(t)}(t) < \eta_{\xi_{j+1}(t)}(t); \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Intuitively, $\tilde{y}_j(t)$ governs the amount of change for adjacent stacks of particles in one step; c.f. Fig. 2. By straightforwardly checking all situations of the numbers of particles staying with particles j and $j+1$ as well as the distance between particles j and $j+1$ following [2], we see that $y_j(t) \le \tilde{y}_j(t)$ for all $j \in \theta(t)$. Given the configuration at time t, we have the following estimates for the conditional probabilities: $P(\tilde{y}_j(t) = 2) = (1-p_n)^{\eta_{\xi_j(t)}(t)} p_n^{\eta_{\xi_{j+1}(t)}(t)} \le (1-\underline{p})^2 \overline{p}^2$, $P(\tilde{y}_j(t) = -2) = (1 - p_n^{\eta_{\xi_j(t)}(t)})(1-(1-p_n)^{\eta_{\xi_{j+1}(t)}(t)}) \ge (1-\overline{p}^2)(1-(1-\underline{p})^2)$, $P(\tilde{y}_j(t) = 1$ w.r.t. the 3rd line of (1)) $= (1-p_n)^{\eta_{\xi_j(t)}(t)} \le (1-\underline{p})^2$, $P(\tilde{y}_j(t) = -1$ w.r.t. the 4th line of (1)) $= 1-(1-p_n)^{\eta_{\xi_j(t)}(t)} \ge 1-(1-\underline{p})^2$, $P(\tilde{y}_j(t) = 1$ w.r.t. the 5th line of (1)) $= p_n^{\eta_{\xi_j(t)}(t)} \le \overline{p}^2$, and $P(\tilde{y}_j(t) = -1$ w.r.t. the 6th line of (1)) $= 1-p_n^{\eta_{\xi_j(t)}(t)} \ge 1-\overline{p}^2$. Therefore, we have for any $j \in \theta(t)$, $|\tilde{y}_j(t)| \le 2$ and the conditional expectation of $\tilde{y}_j(t)$ is upper bounded by the three quantities $(1-\underline{p})^2 - 1 + (1-\underline{p})^2 = 2(1-\underline{p})^2 - 1 := \delta_1 < 0$, $\overline{p}^2 - (1-\overline{p}^2) = 2\overline{p}^2 - 1 := \delta_2 < 0$, and $2(1-\underline{p})^2\overline{p}^2 - 2(1-\overline{p}^2)(1-(1-\underline{p})^2) = \frac{1}{2}((1+\delta_1)(1+\delta_2) - (1-\delta_1)(1-\delta_2)) := \delta_3 < 0$. In other words,

$$E(\tilde{y}_j(t)) \le \max\{\delta_1, \delta_2, \delta_3\} := \delta < 0, \tag{2}$$

where, similarly as $P(\cdot)$, $E(\cdot)$ represents the conditional expectation given the configuration at time t throughout the paper.

Set $\tilde{x}_j(0) = 3 > x_j(0) = 0$ for any $j \in [n-1]$, and define $\tilde{x}_j(t+1) = \max\{3, \tilde{x}_j(t) + \tilde{y}_j(t)\}$ if $j \in \theta(t)$ and $\tilde{x}_j(t+1) = 3$ if $j \notin \theta(t)$. An application of mathematical induction yields $x_j(t) \le \tilde{x}_j(t)$ for all $t \in \mathbb{Z}^+$, $j \in [n-1]$, and $\tilde{y}_j(t) = 3 - 3 = \tilde{x}_j(t+1) - \tilde{x}_j(t)$ for $j \notin \theta(t)$. Noting that $\xi_n(t) - \xi_1(t) = \sum_{j=1}^{n-1} x_j(t) \le \sum_{j=1}^{n-1} \tilde{x}_j(t)$, we may bound the sum by modeling each $\tilde{x}_j(t)$ ($t \in \mathbb{Z}^+$) as a random walk having a wall at 3 and moving a distance bounded by $\tilde{y}_j(t)$ satisfying (2) when $j \in \theta(t)$ or staying put at 3 when $j \notin \theta(t)$. As in [2], there is $\sigma \in (0,1)$ such that $P(\tilde{x}_j(t) \ge 3+k) \le \sigma^k$ for any $k > 0$ and $t \in \mathbb{Z}^+$. In fact, in view of the boundedness of $\tilde{y}_j(t)$ and (2), we have $0 \le \frac{\tilde{y}_i(t)+2}{4} \le 1$, $E\left(\frac{\tilde{y}_j(t)+2}{4}\right) \le \frac{\delta+2}{4}$, and $k - \delta t > 0$ for any $t \in \mathbb{Z}^+$. Drawing on e.g. [3, Theorem 21.6], we have

$$\begin{aligned} P(\tilde{x}_j(t) \ge 3+k) &\le P(\tilde{y}_j(t) + \cdots + \tilde{y}_j(0) \ge k) \\ &= P\Big(\frac{\tilde{y}_j(t)+2}{4} + \cdots + \frac{\tilde{y}_j(0)+2}{4} \ge \frac{k+2t+2}{4}\Big) \\ &\le \exp\left(-\frac{\left(\frac{k+2t+2}{4} - \frac{(\delta+2)t}{4}\right)^2}{2\left[\frac{(\delta+2)t}{4} + \frac{1}{3}\left(\frac{k+2t+2}{4} - \frac{(\delta+2)t}{4}\right)\right]}\right) \\ &\le e^{-c(k+t)} \le e^{-ck} := \sigma^k. \end{aligned} \tag{3}$$

for some constant $c > 0$.

Consider the vertices in G that hold at least one particle at time t. We label these vertices sequentially from left to right and denote by $I_j(t)$ the label of stack holding particle j at time t. In view of the above comments and definition of $\tilde{y}_j(t)$, we know that for $j, j' \in [n-1]$ and $|I_j(t) - I_{j'}(t)| \geq 3$, the two random variables $\tilde{y}_j(t)$ and $\tilde{y}_{j'}(t)$ are independent. Divide the particles $[n-1]$ into $\lfloor (\ln n)^2 \rfloor$ sets $J_i := \{j \in [n-1] : j(\mod \lfloor (\ln n)^2 \rfloor) = i\}$, where $i = 1, 2, \cdots, \lfloor (\ln n)^2 \rfloor$. Clearly, $|J_i| = \lfloor \frac{n-1}{(\ln n)^2} \rfloor$. Set $\gamma_i(t) := \sum_{j \in J_i} \tilde{x}_j(t)$ for $i = 1, 2, \cdots, \lfloor (\ln n)^2 \rfloor$. Define A be the event that at some time $t \leq n^8$ there exists a vertex in G holding at least $\lfloor (\ln n)^2 \rfloor / 2$ particles such that all of them move to the same direction. Therefore,

$$\begin{aligned} \mathrm{P}(A) \leq & n^8 \cdot n \cdot 2\Big(p_n^{\frac{\lfloor (\ln n)^2 \rfloor}{2}} + (1-p_n)^{\frac{\lfloor (\ln n)^2 \rfloor}{2}}\Big) \\ \leq & 2n^9 \Big(\overline{p}^{\frac{\lfloor (\ln n)^2 \rfloor}{2}} + (1-\underline{p})^{\frac{\lfloor (\ln n)^2 \rfloor}{2}}\Big) = o(1), \end{aligned} \tag{4}$$

as $n \to \infty$. By considering separately the cases $|I_j(t) - I_{j'}(t)| = 1$ and $|I_j(t) - I_{j'}(t)| = 2$ following [2], we obtain that conditional on the event A not occurring, $\tilde{y}_j(t)$ and $\tilde{y}_{j'}(t)$ are independent, where $j, j' \in J_i$ for $i = 1, 2, \cdots, \lfloor (\ln n)^2 \rfloor$. Since $\gamma_i(t) - 3|J_i| = \sum_{j \in J_i} (\tilde{x}_j(t) - 3)$, we have

$$\mathrm{P}\Big(\gamma_i(t) - 3|J_i| \geq \frac{2|J_i|}{1-\sigma}\Big) \leq e^{-c|J_i|}, \tag{5}$$

for some constant $c > 0$, capitalizing on the analysis in the previous paragraph and Lemma 1. Therefore,

$$\begin{aligned} & \mathrm{P}\Big(\exists i \in [\lfloor (\ln n)^2 \rfloor], t \leq n^8 : \gamma_i(t) \geq |J_i|\Big(3 + \frac{2}{1-\sigma}\Big)\Big) \\ \leq & n^8 \lfloor (\ln n)^2 \rfloor e^{-c|J_i|} \\ = & e^{-\Omega\left(\frac{n}{(\ln n)^2}\right)}. \end{aligned} \tag{6}$$

Hence, with probability at least $1 - e^{-\Omega(\frac{n}{(\ln n)^2})} = 1 - o(1)$, for any fixed $t \leq n^8$ we have

$$\sum_{j=1}^{n-1} \tilde{x}_j(t) = \sum_{i=1}^{\lfloor (\ln n)^2 \rfloor} \gamma_i(t) \leq \lfloor (\ln n)^2 \rfloor |J_1| \Big(3 + \frac{2}{1-\sigma}\Big) = O(n), \tag{7}$$

as $n \to \infty$.

Next, we will show that the dispersion process comes to an end w.h.p. within time $t \leq n^8$. We have the following three pieces of argument.

(S1) Denote by $\alpha(t)$ the closest particle from the origin and $l(t)$ its distance to the origin. Suppose $l(t) > 0$ and let $z(t)$ be the number of particles sitting together with $\alpha(t)$. If $\alpha(t)$ is on the left-hand side of the origin, then

$$l(t+1) - l(t) = \begin{cases} 0, & \text{with prob. 1 if } z(t) = 1; \\ -1, & \text{with prob. } 1 - (1-p_n)^{z(t)} \geq 1 - (1-\underline{p})^2 \text{ if } z(t) \geq 2; \\ 1, & \text{with prob. } (1-p_n)^{z(t)} \leq (1-\underline{p})^2 \text{ if } z(t) \geq 2. \end{cases} \tag{8}$$

Note that $l(0) = 0$, $l(1) = 1$, and $\mathrm{E}(l(t+1) - l(t)) = -(1-(1-p_n)^{z(t)}) + (1-p_n)^{z(t)} \leq -1 + 2(1-\underline{p})^2 = \delta_1 < 0$. Recall that δ_1 is defined above (2). Hence, $\mathrm{E}\, l(t) \leq 1 + (t-1)\delta_1$ for $t \geq 2$. By using the

Azuma-Hoeffding bound (e.g. [3, Theorem 21.15]), we have for $t \le n^8$,

$$\begin{aligned} P(l(t) \ge \ln n) =& P(l(t) \ge 1 + (t-1)\delta_1 - 1 - (t-1)\delta_1 + \ln n) \\ \le& P(l(t) \ge E\, l(t) - 1 - (t-1)\delta_1 + \ln n) \\ \le& e^{1 - \frac{(-1-(t-1)\delta_1 + \ln n)^2}{2t}} = o(1), \end{aligned} \tag{9}$$

as $n \to \infty$. Therefore, w.h.p. we have $l(t) = O(\ln n)$ for any $t \le n^8$. On the other hand, if $\alpha(t)$ is on the right-hand side of the origin, then

$$l(t+1) - l(t) = \begin{cases} 0, & \text{with prob. } 1 \text{ if } z(t) = 1; \\ -1, & \text{with prob. } 1 - p_n^{z(t)} \ge 1 - \overline{p}^2 \text{ if } z(t) \ge 2; \\ 1, & \text{with prob. } p_n^{z(t)} \le \overline{p}^2 \text{ if } z(t) \ge 2. \end{cases} \tag{10}$$

Likewise, $E(l(t+1) - l(t)) = -(1 - p_n^{z(t)}) + p_n^{z(t)} \le -1 + 2\overline{p}^2 = \delta_2 < 0$, where δ_2 is defined above (2). Hence, $E\, l(t) \le 1 + (t-1)\delta_2$ for $t \ge 2$. Similarly, using the Azuma-Hoeffding bound we have for $t \le n^8$,

$$P(l(t) \ge \ln n) \le e^{1 - \frac{(-1-(t-1)\delta_2 + \ln n)^2}{2t}} = o(1), \tag{11}$$

as $n \to \infty$. Hence, w.h.p. again we have $l(t) = O(\ln n)$ for any $t \le n^8$.

(S2) Assume that the dispersion process lasts for at leat n^8 time steps. If the labels of particles are not re-arranged after each step, a give particle performs a random walk R interrupted by some periods when it stays put. After k steps, the distance of the particle to the origin can be approximated by the normal distribution Normal$(0, k(1-p_n)p_n)$. Choosing $k = n^7$, we have

$$P(|R| \ge n^2) = P\left(\left| \frac{R}{\sqrt{n^7(1-p_n)p_n}} \right| \ge \frac{1}{\sqrt{n^3(1-p_n)p_n}} \right) = 1 - o(1), \tag{12}$$

as $n \to \infty$. Thus, the random walk will reach w.h.p. a distance of n^2 away from the origin after n^7 steps.

(S3) For $t \le n^8$ and $j \in [n-1]$, we have

$$\begin{aligned} P(x_j(t) \ge (\ln n)^2) \le& P(\tilde{x}_j(t) \ge (\ln n)^2) \\ \le& \sigma^{(\ln n)^2 - 3} = n^{-\Omega(\ln n)} = o(1), \end{aligned} \tag{13}$$

as $n \to \infty$ based on the previous analysis.

If the dispersion process does not stop within time $t \le n^8$, it follows from (S1) and (S2) that w.h.p. there exists a particle, say $\beta(t)$, at some time $t \le n^8$ which is at distance at least $n^2 - O(\ln n)$ away from $\alpha(t)$. Thereby, there is a gap $x_{j_0}(t) = \Omega(n)$ between the positions of two particles j_0 and $j_0 + 1$. [In fact, if this is not true, then the distance between $\beta(t)$ and $\alpha(t)$ is $o(n^2)$ since there are n particles in total. It contradicts the fact that the distance in question being $n^2 - O(\ln n)$.] Now, this contradicts (S3). Consequently, we have shown that the dispersion process must come to an end w.h.p. within time $t \le n^8$. Combining (7) and the conclusion of (S1), we finally complete the proof of Theorem 1. □

References

[1] C. Cooper, A. McDowell, T. Radzik, N. Rivera, T. Shiraga, Dispersion processes. *Rand. Struct. Alg.*, 53(2018) 561–585.

[2] A. Frieze, W. Pegden, A note on dispersing particles on a line. *Rand. Struct. Alg.*, 53(2018) 586–591.

[3] A. Frieze, M. Karoński, *Introduction to Random Graphs*. Cambridge University, Cambridge, 2016.

[4] S. Janson, T. Luczak, A. Ruciński, *Random Graphs*. Wiley, New York, 2000.