Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 139 (2021) 321–342
stochastic processes and their applications
www.elsevier.com/locate/spa

Two-type annihilating systems on the complete and star graph

Irina Cristali[a], Yufeng Jiang[b], Matthew Junge[c,*], Remy Kassem[d], David Sivakoff[e], Grayson York[f]

[a] *University of Chicago, United States*
[b] *University of Washington, United States*
[c] *Baruch College, Department of Mathematics, United States*
[d] *Columbia University, United States*
[e] *Ohio State University, United States*
[f] *Duke University, United States*

Received 19 September 2019; received in revised form 5 November 2020; accepted 14 May 2021
Available online 10 June 2021

Abstract

Red and blue particles are placed in equal proportion throughout either the complete or star graph and iteratively sampled to take simple random walk steps. Mutual annihilation occurs when particles with different colors meet. We compare the time it takes to extinguish every particle to the analogous time in the (simple to analyze) one-type setting. Additionally, we study the effect of asymmetric particle speeds.

1. Introduction

We introduce a discrete-time annihilating particle system and study the effects of multiple particle types and asymmetric speeds on the time to extinguish every particle. We consider such systems in two geometries: the complete graph on $2n$ vertices, K_{2n}, and the star graph with $2n$ leaves and a single non-leaf vertex, called the *core*, S_{2n}. Initially, one particle is placed at every site of K_{2n}, or at every leaf of S_{2n}. In the *one-type system*, at each step a particle is chosen

* Corresponding author.
E-mail addresses: icristali@statistics.uchicago.edu (I. Cristali), yj49@uw.edu (Y. Jiang), Matthew.Junge@baruch.cuny.edu (M. Junge), remy.kassem@duke.edu (R. Kassem), dsivakoff@stat.osu.edu (D. Sivakoff), grayson.york@duke.edu (G. York).

https://doi.org/10.1016/j.spa.2021.05.004
0304-4149/

uniformly at random and takes one step of a simple random walk. When any two particles meet, they mutually annihilate.

In our *two-type system*, half of the particles are colored blue and half are colored red. At each step, a blue particle is chosen uniformly at random with probability $p \in [1/2, 1]$, or else a red particle is chosen uniformly at random, and the chosen particle takes a random walk step. When two particles with different colors meet, they mutually annihilate; particles of the same color do not interact. Note that the incremental movement of particles corresponds to the embedded jump chain from the analogous process with particles performing continuous time random walks. Increasing p is equivalent to increasing the rate at which blue particles jump. Accordingly, we call the case $p = 1/2$ the *symmetric speeds case* and $p > 1/2$ the *asymmetric speeds case*.

The two-type system belongs to a family of processes that model systems with two compounds in which reactions neutralize both chemicals involved. These dynamics have been rigorously studied on infinite graphs, typically lattices. The main focus is on the asymptotic density of the different particle types. We initiate the study of such systems on finite graphs. Besides being natural for studying reactions with inherently limited space and material, the finite setting also introduces a new quantity: the time to neutralize all reactants. By working with simple geometries – complete and star graphs – we reveal complicated underlying features of the dynamics. For example, clustering of like compounds is a phenomenon that makes the two-type system more challenging to analyze than related one-type systems. Results for these dynamics on more "realistic" finite graphs, such as tori and random networks, or for general topologies would be natural next steps. However, as with the infinite setting, rigorous results appear difficult to obtain. More background and references are provided in Section 1.2.

1.1. Results

Let $T^1(G)$ and $T_p^2(G)$ be the number of steps it takes in the one-type and two-type systems for every particle to be annihilated on the graph G. We begin with an informal summary of our results. It is straightforward and elementary to compute the distributions of $T^1(K_{2n})$ and $T^1(S_{2n})$ exactly. This is done in Proposition 1, which we include for comparison with our quantitative bounds on $ET_p^2(G)$. In particular, we show that for $G = K_{2n}$ and $G = S_{2n}$ and for all $p \in [1/2, 1]$, we have $ET_p^2(G)$ is asymptotically larger than $ET^1(G)$. How much larger depends of course on the particular graph and the value of p. For the complete graph we prove that

$$2n \log n \le ET_p^2(K_{2n}) \le 20n(\log n)^2/\log\log n$$

for large n, and in particular, $\liminf ET_p^2(K_{2n})/ET^1(K_{2n}) \ge 2$. Our strongest results are for the star graph. For $p = 1/2$, we have

$$c\sqrt{n} \le ET_{1/2}^2(S_{2n}) - ET^1(S_{2n}) \le C\sqrt{n}\log n$$

for large n. For $p \in (1/2, 1)$, we have that $ET_p^2(S_{2n})/ET^1(S_{2n})$ is bounded away from 1 and ∞ as $n \to \infty$ and diverges like $\log(1/(1-p))$ as $p \uparrow 1$; and for $p = 1$, the ratio diverges like $2\log n$.

Throughout this article we let $X(p)$ denote a geometric random variable with distribution $P(X(p) = k) = (1-p)^{k-1}p$ for $k \ge 1$. We write $X \preceq Y$ to denote the usual notion of stochastic dominance $P(X \ge a) \le P(Y \ge a)$ for all $a \ge 0$. Or, equivalently, that there is a coupling so that $X \le Y$ almost surely. We say that $X \overset{d}{=} Y$ if X and Y have the same distribution. Our results make use of the standard asymptotic notation:

- $f = O(g)$ if $\limsup f/g < \infty$,
- $f = \Omega(g)$ if $\liminf f/g > 0$, and
- $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.
- We write $f \sim g$ if $\lim f/g = 1$.

One can exactly characterize how long it takes to go from having $2i$ to $2(i-1)$ particles in the system in terms of a geometric random variable. Though elementary, this gives us a baseline for comparing to the two-type system.

Proposition 1. *In both distributional equalities below the geometric random variables being summed are independent.*

(i) $T^1(K_{2n}) \stackrel{d}{=} \sum_{i=1}^n X(p_i)$ *with* $p_i = (2i-1)/2n$. *In particular,*

$$ET^1(K_{2n}) - (n\log n + \gamma n) = \Theta(1)$$

where $\gamma = \lim(-\log n + \sum_1^n i^{-1})$ *is the Euler–Mascheroni constant.*

(ii) $T^1(S_{2n}) \stackrel{d}{=} 2\sum_{i=1}^n X(q_i)$ *with* $q_i = 1 - (\frac{1}{2i})(\frac{2n-2i+1}{2n})$. *In particular,*

$$ET^1(S_{2n}) - (2n + 2\log n) = \Theta(1).$$

Precisely analyzing the two-type system appears to be much more difficult. The issue on the complete graph is that, as the process evolves, like-particles tend to cluster at the same sites. The clustering should not be too extreme. Namely, when there are $\Omega(n)$ particles, red and blue should occupy $\Omega(n)$ distinct sites at all times, and when there are $o(n)$ particles red and blue should be nearly perfectly spread out. However, there is dependence between which particles are removed and the number of particles at each site. This appears to make it difficult to prove that red and blue particles stay sufficiently spread out.

While we do not completely overcome the issues mentioned above, we are able to confirm that the two-type system survives longer than the one-type system. Below we prove that $ET_p^2(K_{2n}) \geq 2ET^1(K_{2n})(1 - o(1))$. This result should not be all that surprising since the two-type system in some sense has at least twice as many "safe" sites for particles to jump to among the occupied sites as the one-type setting. We also prove an upper bound that differs by a logarithmic factor.

Theorem 2. *For all* $p \in [1/2, 1]$ *it holds that*

$$T_p^2(K_{2n}) \succeq \sum_{i=1}^n X(i/2n) \tag{1}$$

with the $X(i/2n)$ *independent. Thus,* $ET_p^2(K_{2n}) - 2n\log n = \Omega(1)$. *Furthermore, the distributional inequality is an equality when* $p = 1$, *so* $ET_1^2(K_{2n}) - 2(n\log n + \gamma n) = \Theta(1)$. *As for an upper bound, it holds for any fixed* $p \in [1/2, 1]$ *that*

$$ET_p^2(K_{2n}) - \frac{20n(\log n)^2}{\log\log n} = O(1). \tag{2}$$

The proof of the lower bound uses a comparison to a process that has red and blue particles take up the maximal amount of space at each time step. Analogous to what occurs in Proposition 1, we show that $T_p^2(K_{2n})$ stochastically dominates a sum of geometric random variables. The upper bound goes by showing that it is overwhelmingly unlikely for any site

to host more than $C \log n / \log\log n$ particles in the first n^3 steps of the process. This gives a tractable way to lower bound the probability of a collision, but comes at the cost of the additional logarithmic factor.

We can say more about the two-type system on the star graph. The process in this setting has the same clustering issue at the leaves as what occurs globally on K_{2n}. Moreover, the number of like-particles grouped at the core introduces another hub for many like-particles to cluster. While, in principle, one could write down an explicit Markov chain for this process, to do this precisely would require keeping track of the number of particles at the core, as well as the number of red and blue leaves with $1, 2, \ldots$ particles at them. Analyzing this Markov chain exactly appears challenging since the state–space is of growing dimension, and there is significant dependence between the transition rate and the population size.

Despite these difficulties, we prove a fairly precise characterization for all $p \in [1/2, 1]$. For symmetric speeds we show that the second order term is different for the two-type case. The proof reveals that this is caused by clustering at the core. Recall that Proposition 1 shows $ET^1(S_{2n})$ has a logarithmic second order term. We show that $ET^2_{1/2}(S_{2n})$ has a second order term on an order between $\sqrt{n}$ and $\sqrt{n}\log n$. This demonstrates the effect of clustering at the core and, along with Proposition 1 (ii), also implies that $ET^2_{1/2}(S_{2n}) - ET^1(S_{2n}) = \Omega(\sqrt{n})$.

Theorem 3. *It holds that*

(i) $ET^2_{1/2}(S_{2n}) - (2n + C\sqrt{n}) = \Omega(1)$ *for any* $C < (32\pi)^{-1/2}$, *and*
(ii) $ET^2_{1/2}(S_{2n}) - (2n + c\sqrt{n}\log n) = O(1)$ *for some* $c > 0$.

The starting point for the lower bound is a "master formula" in Lemma 7 that equates the number of remaining particles to what has occurred up to that point at the core. We use this to make estimates on the number of particles in the system at time $2n$. This relies on a coupling to the simple random walk which tracks the discrepancy between the number of times red and blue have been sampled. The upper bound again uses the identity in Lemma 7, but this time couples to a different random walk to estimate the number of particles clustered at the core as the process evolves. The argument concludes by bounding the probability a particle is sampled at the core.

For asymmetric speeds, we focus on the leading order coefficient and provide universal upper and lower bounds. The lower bound implies that the asymmetric case has a strictly larger leading coefficient than the symmetric case.

Theorem 4. *Fix* $p \in (1/2, 1)$. *It holds for all* n *that*

$$\left(2 + \frac{2p-1}{2}\right) n - 1 \le ET^2_p(S_{2n}) \le \frac{2}{(1-p)} n.$$

The lower bound is proven in a similar manner as Theorem 3 (i), and the upper bound follows from the observation in Lemma 11 that from any configuration, after two steps, the probability of a collision is at least $1 - p$. So $T^2_p(S_{2n})$ is stochastically dominated by a sum of independent geometric random variables. While these bounds hold for all p, they become rather far from the truth for p near 1. The following theorem addresses what happens in this regime. We provide matching order upper and lower bounds for the rate the leading constant tends to infinity.

Theorem 5. *Given $c < 4$ there exists a value $p_* < 1$ (which depends on c) such that for any fixed $p \in (p_*, 1)$ it holds that*

$$ET_p^2(S_{2n}) - c\log\left(\frac{1}{1-p}\right)n = \Omega(1).$$

And, given $C > 12$ there exists $p^* < 1$ (which depends on C) such that for any fixed $p \in (p^*, 1)$ it holds that

$$ET_p^2(S_{2n}) - C\log\left(\frac{1}{1-p}\right)n = O(1).$$

These results use stochastic lower and upper bounds that relate $T_p^2(S_{2n})$ to a coupon collector process. This connection was not so obvious to make, and requires technical estimates on the time to collect a random subset of the coupons, as well as on the number of coupons collected after a random amount of time.

Finally, we handle the case $p = 1$, which corresponds to the setting in [15]. This setting is tractable because the now immobile red particles cannot cluster.

Theorem 6. *It holds that*

$$T_1^2(S_{2n}) \stackrel{d}{=} 2\sum_{i=1}^{n} X(p_i)$$

with the $X(p_i)$ independent and $p_i = i/2n$. In particular,

$$ET_1^2(S_{2n}) - (4n\log n + 4\gamma n) = \Theta(1)$$

with γ the Euler–Mascheroni constant.

1.2. Background

The study of annihilating particle systems dates back to the work of Erdős and Ney [17]. They considered a system of continuous time random walks started at each nonzero integer in which collisions cause both particles to annihilate and disappear from the process. In particular, they asked if the origin was visited infinitely often, and, more precisely, they studied the asymptotic decay of p_t, the probability the origin is occupied at time t.

The question of whether or not the origin is visited infinitely often was answered in the affirmative by Lootgieter in [25] in discrete time and by Schwartz in [28] in continuous time. Later, Arratia in [1,2] generalized the process to higher dimensions and more general initial configurations. One of his main findings was that

$$p_t \sim \begin{cases} 1/(2\sqrt{\pi t}), & d = 1 \\ \log t/(2\pi t), & d = 2 \\ 1/(2\gamma_d t), & d \geq 3 \end{cases} \tag{3}$$

where γ_d is the probability the simple random walk never returns to its starting position in $\mathbb{Z}^d$. Due to a parity relation observed by Arratia, p_t decays exactly twice as fast as what Bramson and Griffeath in [4] proved occurs for *coalescing random walk*. This is the system in which particles coalesce rather than annihilate upon colliding. The main proof technique in these systems is to analyze a dual process known as the voter model.

Two-type annihilating particle systems first garnered interest in the chemistry and physics literature [21,24,26,29]. Initially particles are assigned to be either of type A or B, and only collisions between different particle types result in annihilation. Unlike the one-type annihilating and coalescing systems, the two-type system has no known tractable dual process. Ovchinnikov and Zeldovich and later Toussaint and Wilczek predicted that in low dimensions the density of particles at the origin of $\mathbb{Z}^d$ is asymptotically much larger than in the one-type system [26,29] due to local clustering of like particles.

In a definitive series of papers, Bramson and Lebowitz [5–8] proved this (and more) for the two-type system on $\mathbb{Z}^d$, where initially each site has a Poi(μ_A)-distributed number of A particles and a Poi(μ_B)-distributed number of B particles. At time 0, pairs of A and B particles at the same site instantly annihilate. Particles then perform continuous time simple random walks at rates λ_A and λ_B, and annihilate when they meet a particle of opposite type. Since multiple particles can occupy a given site, the main quantity of interest is the expected number of particles at the origin at time t, which we denote by ρ_t. In the critical case, with particle types in balance ($\mu_A = \mu_B > 0$) and symmetric speeds ($\lambda_A = \lambda_B > 0$), Bramson and Lebowitz [7] proved that

$$\rho_t \approx \begin{cases} t^{-d/4}, & d \leq 3 \\ t^{-1}, & d \geq 4 \end{cases}.$$

Here $f \approx g$ if $0 < \liminf f/g \leq \limsup f/g < \infty$. Note that, in low dimension, this is asymptotically much larger than the formula for p_t at (3).

There has been recent interest in extending the results of Bramson and Lebowitz to asymmetric speeds. On lattices, physicists predicted that the asymptotic order of ρ_t does not change as the speeds are varied [21,24]. Cabezas, Rolla, and Sidoravicius in [10] considered the asymmetric speed case on a class of infinite transitive graphs and proved a universal lower bound $\rho_t = \Omega(t^{-1})$, and that the root is visited infinitely often when particle types are initially in balance. In a different work [9], Cabezas, Rolla, and Sidoravicius considered the case that red particles move and blue particles are stationary. They proved that there is a phase transition between transience and recurrence when the different particle types are in balance on a broad class of transitive graphs. An Abelian property ensures that the results hold in either discrete or continuous time. More recently Johnson, Junge, Lyu, and Sivakoff proved new upper and lower bounds for the particle density in two-type annihilating systems on lattices and bi-directed regular trees [19]. Bahl, Barnet, Johnson and Junge further explored how the volatility of the distributions of the initial particle counts impacts the total occupation time of the root [3].

Damron, Gravner, Junge, Lyu, and Sivakoff considered a similar problem as [9] in discrete time and proved transience/recurrence results along with more quantitative estimates on the number of visits to the origin when the particle densities are initially out of balance [15]. Very recently, Przykucki, Roberts, and Scott proved quantitative results in discrete time with B-particles stationary on the integers [27]. A slightly different, but related process was studied by Goldschmidt and Przykucki on Galton–Watson trees [18]. The papers [15,18,27] refer to the annihilating system as *parking* since they view A-particles as cars and B-particles as parking spots. Parking was introduced over fifty years ago in [22] and has attracted interest ever since. See [23] for an overview.

As for the finite setting, Cooper, Frieze, and Radzik studied similar quantities as us on random regular graphs [12,13]. They considered an "explosive" particle system with the same dynamics as our one-type system, but with the modification that all particles move simultaneously. They proved that the time it takes to remove all particles is $O(n \log n)$ when

there are sufficiently few particles initially. Additionally, the authors considered a two-type "predator–prey" dynamics, in which predators remove prey on contact, but predators persist, and they studied the expected time to remove all prey. These quantities are closely related to the *coalescence time*. This is the number of steps needed to reach a single particle when particles coalesce, rather than annihilate, upon colliding. There have been recent results for how this behaves on general finite graphs [11,20], as well as a result from Cox concerning the coalescence time on the torus [14]. Note that [11,14,20] only considered one-type systems. To the best of our knowledge, the quantity $T_p^2(G)$ for two-type systems has not been studied on any finite graph.

1.3. Further questions

It would be interesting to find the correct leading order coefficient for $ET_p^2(K_{2n})$ and for $ET_p^2(S_{2n})$. Note that currently we do not have a proof that $ET_p^2(K_{2n}) = O(n \log n)$. For the star graph, we conjecture that our asymptotic lower bound in Theorem 5 is sharp so that, for large enough p,

$$ET_p^2(S_{2n}) \sim 4 \log\left(\frac{1}{1-p}\right) n.$$

This is the answer one gets for the simplified model in which one assumes that the core is always occupied by blue particles and that every red step results in a collision. While the connection is difficult to make rigorous, it seems to be a reasonable approximation for large p.

We also would like to know the exact second order term for $ET_{1/2}(S_{2n})$. This is a more delicate question, but it would be interesting to decide if the logarithmic factor is needed, and if so, what is causing its appearance. We discuss this a bit more at Remark 15. Another future direction is to understand two-type annihilating systems on other finite graphs, such as Erdős–Rényi graphs, tori, and trees.

1.4. Organization

In Section 2 we analyze the one-type system and prove Proposition 1. In Section 3 we prove our lower and upper bounds for the two-type system on the complete graph from Theorem 2. In Section 4 we analyze the two-type system with symmetric speeds on the star graph by proving the upper and lower bounds in Theorem 3. Section 5 houses the proofs of Theorems 4–6 for asymmetric speed two-type systems on the star graph.

2. One-type systems

One-type systems are fairly straightforward to precisely describe because at most one particle can occupy each site. Combining this feature with the simple geometry of the complete and star graphs makes it so the time to annihilate every particle decomposes as a sum of independent geometric random variables.

Proof of Proposition 1. We start with a general decomposition then explain how to prove (i) and (ii). Let τ_i be the first time there are $2i$ particles in the system for $0 \leq i \leq n$. Notice that

$\tau_0 = T^1(G)$ and $\tau_n = 0$ so that

$$T^1(G) = \tau_0 - \tau_n = \sum_{i=1}^{n} \tau_{i-1} - \tau_i. \tag{4}$$

To prove (i), notice that on K_{2n} we have $\tau_{i-1} - \tau_i \overset{d}{=} X((2i-1)/2n)$. This is true because, when there are $2i$ particles in the system, when a particle is selected there are $2i-1$ out of $2n$ sites with another particle that it could move to and cause a collision.

To prove (ii), for the star graph we claim that $\tau_{i-1} - \tau_i \overset{d}{=} 2X(q_i)$ with

$$q_i = \frac{(2i-1)(2n+1)}{4ni} \tag{5}$$

and the $X(q_i)$ independent. Assuming (5) holds and applying this to (4), we then have

$$\begin{aligned} ET^1(S_{2n}) = 2\sum_{i=1}^{n} \frac{1}{q_i} &= 2\sum_{i=1}^{n} \frac{4ni}{(2i-1)(2n+1)} \\ &= \frac{8n}{2n+1} \sum_{i=1}^{n} \frac{1}{2}\left(1 + \frac{1}{2i-1}\right) \\ &= 2n + \log n + \log 2 + o(1). \end{aligned}$$

It remains to justify that $\tau_{i-1} - \tau_i \overset{d}{=} 2X(q_i)$ with q_i as in (5). First notice that at time τ_i there is no particle at the core, and $2i$ particles at the leaves. This is because the only way an annihilation can occur on the star is by either (a) a particle moving from a leaf to the core, or (b) a particle moving from the core to a leaf. Since only one particle can occupy a given site, both (a) and (b) result in no particles at the core when annihilation occurs. Now, the next step from this configuration will necessarily be a particle from a leaf moving to the core. This particle is destroyed either if it is again selected and then moved to one of the $2i-1$ occupied leaves, or if a particle at a leaf is next selected; these occur with total probability

$$\frac{1}{2i}\frac{2i-1}{2n} + \frac{2i-1}{2i} = q_i.$$

To obtain a renewal, notice that if the particle at the core is not destroyed in this second step, then it must move back to an unoccupied leaf. On this event, we have no particle at the core, and $2i$ particles at the leaves, which was the configuration at time τ_i. The next two steps once again result in an annihilation with probability q_i, so we have $\tau_{i-1} - \tau_i \overset{d}{=} 2X(q_i)$. □

3. Two-type systems on the complete graph

Let A_t be the total number of particles remaining in the system after t steps. The fact that collisions occur in pairs ensures that $0 \le A_t \le 2n$ is even. It is also convenient to let R_t and B_t be the total number of sites with at least one red or blue particle, respectively. We start by giving the proof of the lower bound at (1), and then give the proof of the upper bound at (2).

Proof of Theorem 2 Eq. (1). We start by describing the transition probabilities for A_t conditional on the number of sites occupied by red and blue particles. Letting $0 \le r, b \le i \le n$, we have

$$P(A_{t+1} = A_t - 2 \mid R_t = r, B_t = b, A_t = 2i) = p\left(\frac{r}{2n}\right) + (1-p)\left(\frac{b}{2n}\right).$$

Otherwise $A_{t+1} = A_t$. Notice that the above equality does not depend on the value of A_t. When there are $2i$ particles remaining, the probability of an annihilation occurring on a given step is thus bounded by probability of an annihilation when $R_t = i = B_t$. This is equal to $p_i = i/2n$. Using p_i as a bound on the probability of an annihilation is equivalent to comparing to a process that always has red and blue particles occupying the maximal number of sites. For the case $p = 1$, the value p_i is the actual transition rate, because red particles do not move.

It follows that the number of steps to transition from $2i$ to $2(i-1)$ particles is stochastically larger than $X(p_i)$. Decomposing $T_p^2(K_{2n})$ into the time it takes to go from $2n$ to $2(n-1)$ to $2(n-2)$, and so on, we have

$$T_p^2(K_{2n}) \geq \sum_1^n X(p_i).$$

The well-known asymptotic behavior of the harmonic series ensures that

$$ET_p^2(K_{2n}) \geq E\sum_1^n X(p_i) = \sum_1^n \frac{2n}{i} = 2n\log n + 2\gamma n + O(n^{-1}). \quad \square$$

Proof of Theorem 2 Eq. (2). We first show that with high probability no site has more than $m := 6\log n/\log\log n$ particles through time n^3. Let Z_t be the number of blue particles at vertex 1 at time t. Note that Z_t cannot jump by more than 1 at any step, so we can dominate it by a birth and death chain. For $k \geq 0$,

$$P(Z_{t+1} = k+1 \mid Z_t = k, A_t = 2i) \leq \frac{p}{2n}, \tag{6}$$

since a blue particle must move to 1. To decrease the number of blue particles at 1, it suffices to choose a blue particle at 1 and move it somewhere else, so for $0 < k \leq i$ and $n \geq 2$,

$$P(Z_{t+1} = k-1 \mid Z_t = k, A_t = 2i) \geq p \cdot \frac{k}{i} \cdot \left(1 - \frac{1}{2n}\right) > \frac{pk}{2n}. \tag{7}$$

The ratio between the last two probabilities is at least

$$(pk/2n)/(p/2n) = k,$$

independent of i. Therefore, independent of (A_t), we have that Z_t is dominated by a birth and death chain that is k times as likely to move left as right when it is at k. Thus, whenever $Z_t = 1$, the probability that it hits m before hitting 0 is at most $1/(m-1)!$ (see, for instance, Example 5.3.9 in [16]). Using the bound $m! \geq (m/e)^m$, the probability that (Z_t) reaches m by time n^3 is at most

$$\frac{n^3}{(m-1)!} \leq n^3 \frac{e^m}{m^{m-1}} = n^3 \exp\left[-m\log m + m + \log m\right] \tag{8}$$

independent of (A_t). We obtain the same bound for the probability that the number of red particles at 1 hits m by time n^3 by repeating the same argument as above, replacing the p-factor with $1-p$ in (6) and (7).

Define the events

$$G_t = \{\text{every site has at most } m \text{ particles through time } t\},$$

so the union bound and (8) imply that for all $t \leq n^3$ and for all sufficiently large n, we have

$$\begin{aligned} P(G_t^c \mid A_t = 2i) &\leq 2n^4 \exp\left[-m\log m + m + \log m\right] \\ &\leq n^{-1}. \end{aligned}$$

On the event $\{A_t = 2i\} \cap G_t$, there are at least i/m sites that contain at least one blue particle, and at least i/m sites that contain at least one red particle. A collision occurs if a particle is selected and moves to one such site containing the opposite type. For $1 \le i \le n$ and $t \le n^3$, and assuming n is sufficiently large, we have

$$\begin{aligned} P(A_{t+1} = 2i-2 \mid A_t = 2i) &\ge P(A_{t+1} = 2i-2 \mid \{A_t = 2i\} \cap G_t) \cdot P(G_t \mid A_t = 2i) \\ &\ge \frac{i/m}{2n} \cdot (1 - 1/n) \\ &\ge \frac{i}{3nm} =: r_i. \end{aligned}$$

Therefore, for $t \le n^3$, we have

$$P(T_p^2(K_{2n}) \ge t) \le P\left(\sum_{i=1}^{n} X(r_i) \ge t\right).$$

For $t > n^3$, letting $S_n = X_1 + \cdots + X_n$ be the sum of n i.i.d. Geometric$(1/2n)$ random variables, we have the trivial upper bound,

$$P(T_p^2(K_{2n}) \ge t) \le P(S_n \ge t) \le n P(X_1 \ge \lfloor t/n \rfloor) = n(1 - 1/2n)^{\lfloor t/n \rfloor - 1} \le e^{-t/(4n^2)}$$

for large n. Summing over t and using the last two inequalities for $t \le n^3$ and $t > n^3$, respectively, we have

$$ET_p^2(K_{2n}) \le E\sum_{i=1}^{n} X(r_i) + \frac{e^{-n/4}}{1 - e^{-1/(4n^2)}}. \tag{9}$$

The first term is equal to

$$\sum_{i=1}^{n} \frac{1}{r_i} = 3\ nm \sum_{i=1}^{n} \frac{1}{i} \le 3nm(\log n + 1).$$

The second term in (9) is bounded by $8n^2 e^{-n/4}$ for large n, so tends to 0 as $n \to \infty$. We have proved that for large n,

$$ET_p^2(K_{2n}) \le 3nm(\log n + 2) \le \frac{20n(\log n)^2}{\log\log n}. \quad \square$$

4. The star graph with symmetric speeds

We start by fixing some notation. Again let A_t be the total number of particles in the system after t steps. Let C_t be the number of particles that are at the core after t steps. Additionally, let M_t be the number of times up to time t that a particle at the core is sampled to move, but is not annihilated after taking a step. Let $Z_t = 1$ if blue is sampled at time t, and -1 if red is sampled (note that $(Z_t)_{t \ge 1}$ is an i.i.d. sequence, defined even for $t \ge T_p^2(S_{2n})$ after there are no particles remaining). Define the quantities

$$W_t = \textstyle\sum_{s=1}^{t} Z_s \text{ and } D_t = |W_t|, \tag{10}$$

so that D_t has the same law as the displacement of a p-biased random walk. When we write D_{2n} it is implicit that this is the value of D_t at $t = 2n$ for the process on S_{2n}.

The quantities A_t, C_t and M_t are related by the following identity, which will be useful for proving both lower and upper bounds on $ET_p^2(S_{2n})$.

Lemma 7. *For all $p \in [1/2, 1]$ and $t \leq T_p^2(S_{2n})$ we have*

$$A_t = 2n - t + C_t + 2M_t. \tag{11}$$

Moreover, for $t \leq 2n$, we have

$$A_t \geq 2n - t,$$

and consequently $T_p^2(S_{2n}) \geq 2n$.

Proof. Clearly the formula holds for $t = 0$. We proceed inductively from here. Suppose that (11) holds through step t. At step $t + 1$, we sample a particle to move; call this particle x. If the step taken by x results in a collision, then by (11),

$$A_{t+1} = A_t - 2 = 2n - t - 2 + C_t + 2M_t.$$

Either the collision happens at the core, or x moves from the core to a leaf. In either scenario we have $C_{t+1} = C_t - 1$ and $M_{t+1} = M_t$, so

$$A_{t+1} = 2n - t - 2 + (C_{t+1} + 1) + 2M_{t+1} = 2n - (t + 1) + C_{t+1} + 2M_{t+1}.$$

This is the desired statement at time $t + 1$.

Now, suppose that the step taken by x does not result in a collision. If x moves from a leaf to the core, then $C_{t+1} = C_t + 1$, and if x moves from the core to a leaf, then $M_{t+1} = M_t + 1$ and $C_{t+1} = C_t - 1$. In the first case, we have

$$A_{t+1} = A_t = 2n - t + C_t + 2M_t = 2n - t + C_{t+1} - 1 + 2M_{t+1}.$$

In the second case we have

$$A_{t+1} = A_t = 2n - t + (C_{t+1} + 1) + 2(M_{t+1} - 1).$$

Simplifying either case gives $A_{t+1} = 2n - (t + 1) + C_{t+1} + 2M_{t+1}$, as desired.

The second and third statements follow from (11) by observing that $C_t \geq 0$ and $M_t \geq 0$ for $t = 0, 1, \ldots, 2n - 1$. □

4.1. A lower bound for the symmetric case

The starting point for our lower bound is a simple observation that relates $T_p^2(S_{2n})$ to the process stopped at a given time.

Lemma 8. *For all $p \in [1/2, 1]$ and $t \leq T_p^2(S_{2n})$ it holds that*

$$T_p^2(S_{2n}) \geq t + A_t/2.$$

Proof. At most two particles can be removed from the system at each step. Thus, if there are A_t particles at time t, then it deterministically takes at least $A_t/2$ more time steps to remove them all. □

We can further bound M_{2n} in terms of D_{2n}.

Lemma 9. *For any $p \in [1/2, 1]$, we have $EM_{2n} \geq (1/8)ED_{2n} - 1$.*

Proof. Without loss of generality, suppose that blue is sampled $n + D_{2n}/2$ times through time $2n$. Let α be the number of times through time $2n$ that a blue particle moves from the core to a leaf. By Lemma 7 we have $T_p^2(S_{2n}) \geq 2n$, so $(n + D_{2n}/2) - \alpha$ is the number of times that a blue particle moves from a leaf to the core. Since we have only n blue particles initially, we must have

$$[(n + D_{2n}/2) - \alpha] - \alpha \leq n,$$

so $\alpha \geq D_{2n}/4$. Since red can occupy at most half of the leaves, each core selection of blue has at least a $1/2$ chance of increasing the count of M_{2n}. In particular, M_{2n} stochastically dominates a binomial thinning of $D_{2n}/4$ with success probability $1/2$. The same holds when red is sampled D_{2n} times more than blue. Using the bound $\lfloor ED_{2n}/8 \rfloor \geq ED_{2n}/8 - 1$ gives the claimed inequality. □

It is a well-known estimate that ED_t grows like $\sqrt{t}$ when $p = 1/2$. We give a combinatorial proof of this fact below.

Lemma 10. *Suppose $p = 1/2$. It holds that $ED_t \leq \sqrt{t}$ for all $t \geq 0$. Moreover, as $n \to \infty$ it holds that $ED_{2n} \sim \sqrt{2n/\pi}$.*

Proof.

Recalling the definition at (10), it is a standard exercise to show that $W_t^2 - t$ is a martingale, and thus $EW_t^2 = t$. We then have

$$ED_t \leq \sqrt{EW_t^2} = \sqrt{t}.$$

Next we prove the asymptotic claim. Observe that for every integer $x > 0$, $E[D_{n+1} \mid D_n = x] = x$, while $E[D_{n+1} \mid D_n = 0] = 1$. We then have

$$ED_{n+1} = ED_n + P(D_n = 0).$$

Using the parity observation that $D_{2n+1} \neq 0$, and that $ED_1 = 1$, gives the equation

$$ED_{2n} = 1 + \sum_{k=1}^{n-1} P(D_{2k} = 0) = 1 + \sum_{k=1}^{n-1} 2^{-2k} \binom{2k}{k}.$$

Stirling's approximation then yields

$$2^{-2k} \binom{2k}{k} \sim \frac{1}{\sqrt{\pi k}}.$$

Integrating $1/\sqrt{\pi k}$ from 1 to n gives the claimed asymptotic formula for ED_{2n}. □

Proof of Theorem 3 (i). Evaluating the formula in Lemma 7 at $t = 2n$ and ignoring the C_{2n} term give $A_{2n} \geq 2M_{2n}$. Lemma 9 then tells us that

$$EM_{2n} \geq (1/8)ED_{2n} - 1,$$

which by Lemma 10 is $(1/8)\sqrt{2n/\pi} + o(\sqrt{n})$. Thus, for any $C' < (32\pi)^{-1/2}$ we have

$$EA_{2n} - C'\sqrt{n} = \Omega(1).$$

The result then follows by applying the above bound on EA_{2n} to the inequality $ET_{1/2}^2(S_{2n}) \geq 2n + EA_{2n}/2$ implied by Lemmas 7 and 8. □

4.2. An upper bound for the symmetric case

We start with a simple observation that provides a stochastic upper bound on $T_p^2(S_{2n})$ by stopping the process at a given time and then using a worst-case upper bound related to the number of particles still in the system at that time.

Lemma 11. *For all $p \in [1/2, 1)$ we have*

$$T_p^2(S_{2n}) \preceq t + 2\sum_{i=1}^{A_t/2} X_i(1-p),$$

where $(X_i(1-p) : i \geq 1)$ are i.i.d. Geometric$(1-p)$, and are independent of A_t; the sum on the right is 0 when $A_t = 0$.

Proof. If the core is occupied, then the probability of a collision in the next step is at least $(1-p)\wedge p = 1-p$. If the core is not occupied, then after one step it becomes occupied, and the probability of a collision on the next step is at least $1-p$. Therefore, from any configuration of particles, the probability of a collision occurring in the next two steps is always at least $1-p$, and each collision reduces A_t by 2. The formula follows. □

Keeping in mind the identity in Lemma 7, we will require a bound on C_{2n}. Because C_t always has a slight drift towards zero, it can be dominated by the displacement of a simple random walk.

Lemma 12. *Fix $p = 1/2$. Let D'_t be the displacement from the origin of a simple symmetric random walk on $\mathbb{Z}$ started at 0. There exists a coupling such that*

$$C_t \leq D'_t + 1$$

for all $t \geq 0$.

Proof. We explain how to construct D'_t from C_t. Notice that the probability C_t increases is equal to the probability of picking a particle at a leaf that is the same color as those currently occupying the core, or 1 if no particles are there. Thus, we define $D'_{t+1} = D'_t + 1$ if one of the following occurs:

(a) $D'_t = 0$,
(b) $C_t > 0$ and $C_{t+1} = C_t + 1$,
(c) $C_t > 0$ and $C_{t+1} = C_t - 1$ because of a particle moving away from the core,
(d) with probability $1/2$ if $C_t = 0$ and $D'_t > 0$.

Otherwise $D'_{t+1} = D'_t - 1$.

It is easy to check that D'_t is the displacement of a simple random walk, since, when it is nonzero, it transitions up or down with equal probability (the probabilities in (b) and (c) sum to $p = 1/2$, the probability of choosing the color at the core). Moreover, D'_t and C_t are coupled so that D'_t increases whenever C_t does with one exception. The only situation in which C_t can exceed D'_t is if $C_{t-1} = 0$ and $D'_{t-1} = 1$ and $D'_t = 0$. When this occurs we have $C_t = D'_t + 1$. However, the gap cannot become any larger than this, because while C_t is larger than D'_t, case (d) is prohibited, so D'_t increases whenever C_t does. □

We will soon require an estimate on a sum via comparison to an integral. We provide the antiderivative and asymptotic behavior of that integral now.

Lemma 13. *It holds that* $\int_1^{2n} x^{-1}\sqrt{2n-x}\,dx = O(\sqrt{n}\log n)$.

Proof. This follows immediately from setting $C = 2n$ in the equation

$$\int_1^C \frac{\sqrt{C-x}}{x}\,dx = \sqrt{C}\log C + 2\sqrt{C}\log\left(\sqrt{\frac{C-1}{C}}+1\right) - 2\sqrt{C-1}. \tag{12}$$

We obtain this formula by computing the indefinite integral $\int x^{-1}\sqrt{C-x}\,dx$. Start with the substitution $u = \sqrt{C-x}$ so that the integral becomes

$$-2\int \frac{u^2}{C-u^2}\,du = 2\int \frac{C}{u^2-C}+1\,du = -2\int \frac{1}{1-\frac{u^2}{C}}\,du + 2u.$$

Next, make the substitution $s = iu/\sqrt{C}$ with $i = \sqrt{-1}$ so that the above is equal to

$$2u - 2\int \frac{1}{1-\frac{u^2}{C}}\,du = 2u + 2i\sqrt{C}\int \frac{1}{s^2+1}\,ds = 2u + 2i\sqrt{C}\tan^{-1}(s).$$

Substituting back u and then x yields

$$\int x^{-1}\sqrt{C-x}\,dx = 2\sqrt{C-x} - 2\sqrt{C}\tanh^{-1}\left(\frac{\sqrt{C-x}}{\sqrt{C}}\right) + C_0.$$

We obtain the claimed formula at (12) by applying the identity

$$\tanh^{-1}(z) = \frac{1}{2}\left(\log(1+z) - \log(1-z)\right),$$

combining logarithmic terms, and then computing the definite integral. □

We now put this inequality to work in bounding EM_{2n}.

Lemma 14. *Fix* $p = 1/2$. *It holds that* $EM_{2n} = O(\sqrt{n}\log n)$.

Proof. Let $G_t = \{M_t = M_{t-1} + 1\}$ be the event that a particle at the core is sampled at time t and the particle is not annihilated after taking a step. This tracks when M_t increases; accordingly, at time $2n$ we have

$$M_{2n} = \sum_{t=1}^{2n} \mathbf{1}\{G_t\}.$$

Notice that $P(G_t)$ is at most the probability of sampling a particle at the core at time t. Given $A_t = a$ and $C_t = c$, this probability is equal to c/a, since $p = 1/2$. It follows that

$$P(G_t) = E[E[\mathbf{1}\{G_t\} \mid A_t, C_t]] \le E[C_t/A_t]. \tag{13}$$

Using this for $t < 2n$ and the bound $P(G_{2n}) \le 1$ as we bring the expectation inside the sum, we obtain

$$EM_{2n} \le 1 + \sum_{t=1}^{2n-1} E\left[\frac{C_t}{A_t}\right] \le 1 + \sum_{t=1}^{2n-1} \frac{EC_t}{2n-t}.$$

The second inequality uses the deterministic bound $A_t \geq 2n - t$ from Lemma 7. Bounding $EC_t \leq ED_t + 1$ via Lemma 12 and then bounding ED_t with Lemma 10, we obtain

$$EM_{2n} \leq 1 + \sum_{t=1}^{2n-1} \frac{ED_t + 1}{2n - t} \leq 1 + 2\sum_{t=1}^{2n-2} \frac{\sqrt{t}}{2n - t}.$$

Reindexing with $s = 2n - t$ gives

$$EM_{2n} \leq 1 + 2\sum_{s=1}^{2n} \frac{\sqrt{2n - s}}{s}$$

By comparison to the integral in Lemma 13, the summation above is $O(\sqrt{n}\log n)$. □

Remark 15. Note that in the previous argument at (13) we made the bound $P(G_t) \leq E[C_t/A_t]$. One might suspect that the logarithmic factor comes from this estimate. However, the exact formula is

$$P(G_t) = E\left[\frac{C_t}{A_t}\frac{(2n - U_t)}{2n}\right]$$

where U_t is the number of sites occupied by particles of the opposite color from the core at the leaves of S_{2n} at time t. Exactly describing the quantity U_t is subtle since it depends on clustering at the leaves and on the current particle type occupying the core. Regardless, we have $1/2 \leq (2n - U_t)/2n \leq 1$ for all t since $0 \leq U_t \leq n$. So, $P(G_t) \geq (1/2)E[C_t/A_t]$. Thus, the estimate we make on $P(G_t)$ is not the source of the logarithmic factor.

In any case, we now we have the necessary ingredients to prove our upper bound.

Proof of Theorem 3 (ii). By Lemma 7 we have

$$EA_{2n} = EC_{2n} + 2EM_{2n}.$$

It follows from Lemmas 10 and 12 that $EC_{2n} = O(\sqrt{n})$, and from Lemma 14 we have $EM_{2n} = O(\sqrt{n}\log n)$. Thus, $EA_{2n} = O(\sqrt{n}\log n)$. Applying Lemma 11 with $t = 2n$ gives

$$T^2_{1/2}(S_{2n}) \leq 2n + 2\sum_{i=1}^{A_{2n}/2} X_i(1/2).$$

It follows from Wald's lemma and our bound on EA_{2n} that

$$ET^2_{1/2}(S_{2n}) \leq 2n + 4EA_{2n} = 2n + O(\sqrt{n}\log n). \quad \square$$

5. Asymmetric speeds on the star graph

We break this section into three subsections. The first two subsections contain technical estimates for a modified coupon collector problem, and also describe how these connect back to the two-type system. The third subsection contains the proofs of Theorems 4–6.

5.1. Lemmas for the asymptotic lower bound

The idea behind the lower bound is that after $t \approx -4\log(1 - p)n$ steps approximately $r = (1 - p)t$ red particles will move from their starting location. To eliminate all of the red

particles by time t, blue particles must visit all $n-r$ of the sites with red particles that did not move. We identify the sites initially occupied by red particles as coupons, and view each jump from the core by a blue particle as an attempt to collect one of these coupons. So $T_p^2(S_{2n})$ is lower bounded by the number of steps needed for a coupon collector to collect $n-r$ coupons, which we prove has expected value on the order of t.

Lemma 16. *Fix any $\epsilon \in (0,1)$, let $t_p = t_p(n) = -n\log(1-p)$ and let $R \stackrel{d}{=} \mathrm{Bin}(4(1-\epsilon)t_p, 1-p)$. Consider a coupon collector process in which there are n coupons. At each step, with probability $1/2$ no coupon is selected, and otherwise one is picked uniformly at random. Let T'_p be the number of steps needed to sample $n-R$ distinct coupons. Then there exists $p'(\epsilon) < 1$ such that for all $p > p'(\epsilon)$ we have*

$$P\left(T'_p \le 2(1-\epsilon)t_p\right) \to 0$$

as $n \to \infty$.

Proof. The time between each new coupon discovery is a geometric random variable, so

$$T'_p \stackrel{d}{=} \sum_{i=0}^{n-R-1} X\left(\frac{n-i}{2n}\right)$$

with the convention that the sum is zero if $n - R \le 0$ and the X's are independent. Let $a = 5(1-\epsilon)(1-p)\log(1-p)^{-1}$, and observe that $a \to 0$ as $p \to 1$. Noting that $ER = 4an/5$, by a standard Chernoff bound for the binomial distribution, we have

$$P(R \ge an) \le e^{-cn} \tag{14}$$

for some $c = c(\epsilon, p)$. Letting

$$Y = \sum_{i=0}^{n-an-1} X\left(\frac{n-i}{2n}\right),$$

then using the bound $\log m \le \sum_{i=1}^{m} i^{-1} \le 1 + \log m$ we have

$$EY = 2n \sum_{j=an+1}^{n} \frac{1}{j} \ge 2n\left[\log(1/a) - 1\right] > (2-\epsilon)n\log\left(\frac{1}{1-p}\right),$$

and

$$\mathrm{Var}(Y) \le 4n^2 \sum_{j=an+1}^{\infty} \frac{1}{j^2} \le \frac{4n}{a}.$$

for all p sufficiently close to 1 and n large enough, depending on p. Therefore, by Chebychev's inequality,

$$\begin{aligned} P(Y < -2(1-\epsilon)\log(1-p)\cdot n) &\le P\left(|Y - EY| > \frac{\epsilon}{2-\epsilon}EY\right) \\ &\le \frac{4}{a\epsilon^2(\log((1-p)^{-1}))^2}n^{-1}. \end{aligned} \tag{15}$$

Combining (14) and (15), we arrive at

$$P\left(T'_p < -2(1-\epsilon)\log(1-p)\cdot n\right) \le P(R \ge an) + P(Y < -2(1-\epsilon)\log(1-p)\cdot n) \to 0$$

as $n \to \infty$. □

Lemma 17. *For any $\epsilon \in (0,1)$, there exists $p'(\epsilon) < 1$ such that for each $p > p'(\epsilon)$ and all sufficiently large n, we have*

$$ET_p^2(S_{2n}) \ge 4(1-\epsilon)^2 n \log\left(\frac{1}{1-p}\right).$$

Proof. Let t_p, R and T'_p be as in Lemma 16. Let `Red_Moved` be the number of red particles that have moved from their starting locations through time $4(1-\epsilon)t_p$, and let `Red_Sites_Visited_by_Blue` be the number of leaf vertices that were initially occupied by red particles and were visited by at least one blue particle through time $4(1-\epsilon)t_p$ (whether or not they are occupied by red at the time of blue's visit). The number of red particles extinguished through time $4(1-\epsilon)t_p$ cannot exceed

$$\texttt{Red\_Moved} + \texttt{Red\_Sites\_Visited\_by\_Blue},$$

so if all particles are to be removed by time $4(1-\epsilon)t_p$, we must have

$$\texttt{Red\_Sites\_Visited\_by\_Blue} \ge n - \texttt{Red\_Moved}.$$

Note that `Red_Moved` cannot exceed the number of times that red particles are chosen to move through time $4(1-\epsilon)t_p$, which has the same distribution as R, so we have $\texttt{Red\_Moved} \preceq R$. Also, through time $4(1-\epsilon)t_p$, a blue particle has moved to a uniformly sampled leaf on at most $2(1-\epsilon)t_p$ steps, since each visit to a leaf requires two moves by a blue particle. The random variable `Red_Sites_Visited_by_Blue` is therefore stochastically dominated by the number of distinct coupons collected after $2(1-\epsilon)t_p$ steps, where at each step, with probability $1/2$ no coupon is selected, and otherwise one of n coupons is selected uniformly at random. This number can be taken to be independent of R, as the number of steps taken by the coupon collector is deterministic, so we are in the setting of Lemma 16, and we have

$$P(T_p^2(S_{2n}) \le 4(1-\epsilon)t_p) \le P(T'_p \le 2(1-\epsilon)t_p) \to 0$$

as $n \to \infty$. Letting n be sufficiently large so that the probability above is smaller than ϵ gives the desired lower bound on the expectation. □

5.2. Lemmas for the asymptotic upper bound

The idea behind the upper bound is to run the process for $t = -8n\log(1-p)$ steps. At this point, we prove that blue has moved to nearly pn of the sites that were initially red, and at most $r = (1-p)t = (1-p)\log((1-p)^{-8})n$ red particles have moved to avoid a collision. This means that at most $n - pn + r$ red particles have avoided collision through time t. We then use the bound at Lemma 11 to show that the expected time to destroy the remaining particles is $O(n)$ with leading constant that does not depend on p.

Lemma 18. *Consider the following coupon collection process like that of Lemma 16. Let* $t_p = -n\log(1-p)$ *and for* $r > 4$, *let* $B \stackrel{d}{=} \mathrm{Bin}(rt_p, p)$. *Let* $N = (B-n)/2$, *and let* V *be the number of the* n *coupons that are* not *collected through* N *steps, where we set* $V = n$ *if* $N \le 0$. *For all fixed* p *sufficiently close to* 1 *and for all* n *sufficiently large,*

$$P(V \ge (1-p)n) \le 3/n.$$

Proof. Note that B is a binomial random variable with mean rpt_p, so for p sufficiently close to 1, standard large deviation estimates for the binomial distribution imply that

$$P\left(B \le rpt_p\left(1-(rpt_p)^{-1/4}\right)\right) \le e^{-(rpt_p)^{1/2}/16} \le e^{-n^{1/2}} \tag{16}$$

for all large enough n. Let $a = a(r,p) := \frac{1}{2}(rp\log\frac{1}{1-p} - 2)$, which is large for p close to 1, so for all large enough n, we have

$$\begin{aligned} P(N \le an) &\le P\left(N \le \frac{1}{2}[rpt_p(1-(rpt_p)^{-1/4}) - n]\right) \\ &= P\left(B \le rpt_p\left(1-(rpt_p)^{-1/4}\right)\right) \\ &\le e^{-n^{1/2}}. \end{aligned}$$

Letting V' be the number of coupons *not* collected through an steps, we have

$$\begin{aligned} EV' &= n(1-1/2n)^{an} \\ &\le ne^{-a/2} \\ &\le n\exp\left[-\frac{1}{4}rp\log\frac{1}{1-p} + 1\right] \\ &\le n\frac{1}{2}(1-p), \end{aligned}$$

where in the last line we use the assumption $r > 4$ and take p such that $rp/4 > 1$.

In anticipation of our variance bound, observe that for n sufficiently large, by Taylor's theorem, we have

$$\begin{aligned} (1-1/n)^{an} - (1-1/2n)^{2an} &\le e^{-a}(e^{a/2n} - e^{a/8n}) \\ &\le e^{-a}(a/n - a/8n) \\ &\le ae^{-a}/n. \end{aligned}$$

The probability that coupons labeled 1 and 2 (say) are not chosen through an steps is $(1-2/2n)^{an}$, so we have for p close to 1,

$$\begin{aligned} \mathrm{Var}(V') &\le EV' + n^2\left[(1-1/n)^{an} - (1-1/2n)^{2an}\right] \\ &\le n(e^{-a/2} + ae^{-a}) \\ &\le (1-p)n. \end{aligned}$$

Finally, noting that $V \le V'\mathbf{1}_{\{N>an\}} + n\mathbf{1}_{\{N\le an\}}$, we have

$$\begin{aligned} P(V \ge (1-p)n) &\le P(V' \ge (1-p)n) + P(N \le an) \\ &\le P(|V' - EV'| \ge (1-p)n/2) + e^{-n^{1/2}} \\ &\le 3/n. \quad \square \end{aligned}$$

Lemma 19. *For any $\epsilon > 0$ there exists $p'(\epsilon) < 1$ such that for each $p > p'(\epsilon)$ and all sufficiently large n, we have*

$$ET_p^2 \leq (12+\epsilon)n \log \frac{1}{1-p}.$$

Proof. Let $r > 4$ and t_p be as in Lemma 18. Let $B \stackrel{d}{=} \mathrm{Bin}(rt_p, p)$ be the number of times that blue is chosen to move through time $t = rt_p$ (based on the values of $(Z_s)_{s\leq t}$ defined at the start of Section 4). Like in the proof of Lemma 17, we let `Red_Moved` be the number of red particles that have moved from their starting locations through time t, and let `Red_Sites_Visited_by_Blue` be the number of leaf vertices that were initially occupied by red particles and were visited by at least one blue particle through time t.

The number of times that blue particles are chosen to move from the core to a leaf through time t must be at least $N = (B-n)/2$, provided particles persist through time t. To see this, let $C^{\rightarrow}$ be the number of jumps from the core to leaves up to time t, and $C^{\leftarrow}$ be the number of jumps from leaves to the core up to time t. Adding the equations

$$C^{\rightarrow} + C^{\leftarrow} = B$$
$$C^{\rightarrow} - C^{\leftarrow} \geq -n,$$

and solving for $C^{\rightarrow}$ gives the claimed inequality $C^{\rightarrow} \geq N$. We now satisfy the hypotheses of Lemma 18. Therefore,

$$P(\texttt{Red\_Sites\_Visited\_by\_Blue} \leq pn, A_t > 0) \leq P(V \geq (1-p)n) \leq 3/n.$$

Moreover, by (16), for p close to 1 and sufficiently large n we have

$$\begin{aligned} P\left(\texttt{Red\_Moved} \geq nr(1-p)\left(\log\frac{1}{1-p}+1\right)\right) &\leq P\left(B \leq rpt_p\left(1-(rpt_p)^{-1/4}\right)\right) \\ &\leq e^{-n^{1/2}}. \end{aligned}$$

Observe that $A_t \leq 2(n - \texttt{Red\_Sites\_Visited\_by\_Blue} + \texttt{Red\_Moved})$, since red particles that have not moved and are at sites that are visited by blue particles by time t must be eliminated. Combining this observation with the last two inequalities gives

$$\begin{aligned} P\left(A_t \geq 2n\left((1-p)+r(1-p)\left(\log\frac{1}{1-p}+1\right)\right)\right) &\leq P(\texttt{Red\_Sites\_Visited\_by\_Blue} \leq pn, A_t > 0) \\ &\quad + P\left(\texttt{Red\_Moved} \geq nr(1-p)\left(\log\frac{1}{1-p}+1\right)\right) \\ &\leq 3/n + e^{-n^{1/2}} \\ &\leq 4/n. \end{aligned}$$

Since we have $A_t \leq 2n$, we arrive at

$$EA_t \leq 2n\left((1-p)+r(1-p)\left(\log\frac{1}{1-p}+1\right)\right) + 2n(4/n).$$

Combining this bound with Lemma 11 applied at time $t = rt_p$ and Wald's equation gives

$$\begin{aligned} ET_2^p &\le nr \log \frac{1}{1-p} + 2E[A_t/2]\frac{1}{1-p} \\ &\le n\left[3r \log \frac{1}{1-p} + 2 + 2r\right] + \frac{1}{2(1-p)}. \end{aligned}$$

Taking r close to 4, then p close enough to 1, then n sufficiently large completes the proof. □

5.3. Proofs

Proof of Theorem 4. The upper bound follows from Lemma 11 by setting $t = 0$ and taking expectation. The lower bound follows from similar reasoning as the proof of Theorem 3 (i). Lemmas 7–9 together imply that

$$ET_p^2(S_{2n}) \ge 2n + (1/4)ED_t - 1. \tag{17}$$

Recalling the definition $D_t = |\sum_1^t Z_s|$ at the start of this section, we have

$$D_t \ge Z_1 + \cdots + Z_t.$$

Since $EZ_1 = 2p - 1$, we then have

$$ED_{2n} \ge (2p-1)2n.$$

Applying this inequality at (17), it follows that

$$ET_p^2(S_{2n}) \ge 2n + \frac{2p-1}{2}n - 1. \quad \square$$

Proof of Theorem 5. The lower bound follows from Lemma 17 and the upper bound from Lemma 19. □

Proof of Theorem 6. Let M denote the value of M_t at time $t = T_1^2(S_{2n})$. Evaluating Lemma 7 at $t = T_1^2(S_{2n})$ and rearranging give

$$T_1^2(S_{2n}) = 2n + 2M. \tag{18}$$

Let $M(i)$ be the number of times a particle is sampled at the core and moves without being annihilated when there are $2i$ particles in the system. Since red particles do not move, each time a particle at the core is sampled there is an $i/2n$ chance of annihilation, and annihilations cannot occur in any other way. It follows that $M(i)$ has distribution $X(i/2n) - 1$ and thus

$$M \overset{d}{=} \sum_1^n (X(i/2n) - 1).$$

In light of (18) and the above equality, we have

$$T_1^2(S_{2n}) \overset{d}{=} 2n + 2\sum_1^n (X(i/2n) - 1) = 2\sum_1^n X(i/2n),$$

which has expectation $4n \log n + 4\gamma n + o(n)$. □

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Thanks to Rick Durrett for helpful conversations, encouragement, and advice. Matthew Junge was partially supported by NSF-DMS grant 1855516. David Sivakoff was partially supported by NSF TRIPODS grant CCF1740761.

References

[1] Richard Arratia, Limiting point processes for rescalings of coalescing and annihilating random walks on $\mathbb{Z}^d$, Ann. Probab. (1981) 909–936.
[2] Richard Arratia, Site recurrence for annihilating random walks on $\mathbb{Z}^d$, Ann. Probab. 11 (3) (1983) 706–713.
[3] Riti Bahl, Philip Barnet, Tobias Johnson, Matthew Junge, Diffusion-limited annihilating systems and the increasing convex order, 2021, ArXiv ID: 2104.12797.
[4] Maury Bramson, David Griffeath, Asymptotics for interacting particle systems on $\mathbb{Z}^d$, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 53 (2) (1980) 183–196.
[5] Maury Bramson, Joel L. Lebowitz, Asymptotic behavior of densities in diffusion-dominated annihilation reactions, Phys. Rev. Lett. 61 (21) (1988) 2397.
[6] Maury Bramson, Joel L. Lebowitz, Asymptotic behavior of densities in diffusion dominated two-particle reactions, Physica A 168 (1) (1990) 88–94.
[7] Maury Bramson, Joel L. Lebowitz, Asymptotic behavior of densities for two-particle annihilating random walks, J. Stat. Phys. 62 (1–2) (1991) 297–372.
[8] Maury Bramson, Joel L. Lebowitz, Spatial structure in diffusion-limited two-particle reactions, Ann. Appl. Probab. 11 (1) (2001) 121–181.
[9] M. Cabezas, L.T. Rolla, V. Sidoravicius, Non-equilibrium phase transitions: Activated random walks at criticality, J. Stat. Phys. 155 (6) (2014) 1112–1125.
[10] M. Cabezas, L.T. Rolla, V. Sidoravicius, Recurrence and density decay for diffusion-limited annihilating systems, Probab. Theory Related Fields 170 (3) (2018) 587–615.
[11] Colin Cooper, Robert Elsasser, Hirotaka Ono, Tomasz Radzik, Coalescing random walks and voting on connected graphs, SIAM J. Discrete Math. 27 (4) (2013) 1748–1758.
[12] Colin Cooper, Alan Frieze, Tomasz Radzik, Multiple random walks and interacting particle systems, in: International Colloquium on Automata, Languages, and Programming, Springer, 2009, pp. 399–410.
[13] Colin Cooper, Alan Frieze, Tomasz Radzik, Multiple random walks in random regular graphs, SIAM J. Discrete Math. 23 (4) (2009) 1738–1761.
[14] J. Theodore Cox, Coalescing random walks and voter model consensus times on the torus in $\mathbb{Z}^d$, Ann. Probab. 17 (4) (1989) 1333–1366.
[15] Michael Damron, Janko Gravner, Matthew Junge, Hanbaek Lyu, David Sivakoff, et al., Parking on transitive unimodular graphs, Ann. Appl. Probab. 29 (4) (2019) 2089–2113.
[16] Rick Durrett, Probability: Theory and Examples, Vol. 49, Cambridge University Press, 2019.
[17] P. Erdos, P. Ney, Some problems on random intervals and annihilating particles, Ann. Probab. 2 (5) (1974) 828–839.
[18] Christina Goldschmidt, Michał Przykucki, Parking on a random tree, Combin. Probab. Comput. 28 (1) (2019) 23–45.
[19] Tobias Johnson, Matthew Junge, Hanbaek Lyu, David Sivakoff, Particle density in diffusion-limited annihilating systems, 2020, arXiv:2005.06018, arXiv e-prints.
[20] Varun Kanade, Frederik Mallmann-Trenn, Thomas Sauerwald, On coalescence time in graphs: When is coalescing as fast as meeting? in: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2019, pp. 956–965.
[21] K. Kang, S. Redner, Scaling approach for the kinetics of recombination processes, Phys. Rev. Lett. 52 (1984) 955–958.
[22] Alan G. Konheim, Benjamin Weiss, An occupancy discipline and applications, SIAM J. Appl. Math. 14 (6) (1966) 1266–1274.

[23] Marie-Louise Lackner, Alois Panholzer, Parking functions for mappings, J. Combin. Theory Ser. A 142 (2016) 1–28.
[24] Benjamin P. Lee, John Cardy, Renormalization group study of the $A + B \to 0$ diffusion-limited reaction, J. Stat. Phys. 80 (5) (1995) 971–1007.
[25] J.-C. Lootgieter, Problèmes de récurrence concernant des mouvements aléatoires de particules sur z avec destruction, in: Annales de l'IHP Probabilités et statistiques, Vol. 13, No. 2, 1977, pp. 127–139.
[26] A.A. Ovchinnikov, Ya B. Zeldovich, Role of density fluctuations in bimolecular reaction kinetics, Chem. Phys. 28 (1–2) (1978) 215–218.
[27] Michał Przykucki, Alexander Roberts, Alex Scott, Parking on the integers, 2019, ArXiv preprint arXiv:1907.09437.
[28] Diane Schwartz, On hitting probabilities for an annihilating particle model, Ann. Probab. (1978) 398–403.
[29] Doug Toussaint, Frank Wilczek, Particle–antiparticle annihilation in diffusive motion, J. Chem. Phys. 78 (5) (1983) 2642–2647.