



## A SPATIAL MUTATION MODEL WITH INCREASING MUTATION RATES

BRIAN CHAO ,\* *Cornell University*

JASON SCHWEINSBERG ,\*\* *University of California San Diego*

### Abstract

We consider a spatial model of cancer in which cells are points on the  $d$ -dimensional torus  $\mathcal{T} = [0, L]^d$ , and each cell with  $k - 1$  mutations acquires a  $k$ th mutation at rate  $\mu_k$ . We assume that the mutation rates  $\mu_k$  are increasing, and we find the asymptotic waiting time for the first cell to acquire  $k$  mutations as the torus volume tends to infinity. This paper generalizes results on waiting for  $k \geq 3$  mutations in Foo *et al.* (2020), which considered the case in which all of the mutation rates  $\mu_k$  are the same. In addition, we find the limiting distribution of the spatial distances between mutations for certain values of the mutation rates.

*Keywords:* Mutation; cancer; spatial population model

2020 Mathematics Subject Classification: Primary 60J99

Secondary 60G55; 92D15; 92D25

### 1. Introduction

Cancer is often caused by genetic mutations which disrupt regular cell division and apoptosis, in which case cancerous cells divide much more rapidly compared to healthy cells. This can happen, for example, as soon as several distinct mutations occur and dramatically disrupt cell function. Thus, it is sometimes reasonable to model cancer as occurring after  $k$  distinct mutations appear in sequence within a large body.

Mathematical models in which cancer occurs once some cell acquires  $k$  mutations date back to the famous 1954 paper [1], which proposed a multi-stage model of carcinogenesis in which, once a cell has acquired  $k - 1$  mutations, it acquires a  $k$ th mutation at rate  $\mu_k$ . In this model, the probability of acquiring the  $k$ th mutation during a small time interval  $(t, t + dt)$  is

$$\frac{\mu_1 \mu_2 \cdots \mu_k t^{k-1}}{(k-1)!} dt.$$

That is, the incidence rate of the  $k$ th mutation (at which point the individual becomes cancerous) is proportional to  $\mu_1 \mu_2 \cdots \mu_k t^{k-1}$ . This means that cancer risk is proportional to both the mutation rates and the  $(k - 1)$ th power of age. More sophisticated models, taking into account the possibilities of cell division and cell death, were later analyzed in [6, 7, 9, 12, 13, 15, 17, 18, 24].

To model some types of cancer, it is important to also include spatial structure in the model. In 1972, [25] introduced a spatial model of skin cancer now known as the biased voter model.

---

Received 21 August 2021; revision received 18 October 2022.

\* Postal address: 310 Malott Hall, Ithaca, NY 14853. Email: [bc492@cornell.edu](mailto:bc492@cornell.edu)

\*\* Postal address: Department of Mathematics, 0112; University of California, San Diego; 9500 Gilman Drive; La Jolla, CA 92093-0112. Email: [jschweinsberg@ucsd.edu](mailto:jschweinsberg@ucsd.edu)

At each site on a lattice, there is an associated binary state indicating whether the site is cancerous or healthy. Each cell divides at a certain rate, and when cell division occurs, the daughter cell replaces one of the neighboring cells chosen at random. The model is biased in that a cancerous cell spreads  $\kappa > 1$  times as fast as a healthy cell. Computer simulations for this model were presented in [25], and the model was later analyzed mathematically [3, 4].

More recently, [5], building on earlier work in [8, 14], studied a spatial Moran model which is a generalization of the biased voter model. Cells are modeled as points of the discrete torus  $(\mathbb{Z} \bmod L)^d$ , and each cell is of type  $i \in \mathbb{N} \cup \{0\}$ . A cell of type  $i - 1$  mutates to type  $i$  at rate  $\mu_i$ . Type  $i$  cells have fitness level  $(1 + s)^i$ , where  $s > 0$  measures the selective advantage of one cell over its predecessors. Each cell divides at a rate proportional to its fitness, and then, as in the biased voter model, the daughter cell replaces a randomly chosen neighboring cell. The authors considered the question of how long it takes for some type 2 cell to appear. To simplify the analysis, they introduced a continuous model where cells live inside the torus  $[0, L]^d$ . This continuous stochastic model approximates the biased voter model because of the Bramson–Griffeath shape theorem [3, 4], which implies that, conditioned on the survival of the mutations, the cluster of cells in  $\mathbb{Z}^d$  with a particular mutation has an asymptotic shape that is a convex subset of  $\mathbb{R}^d$ . In [5, Section 4], the authors used the continuous model to compute the distribution of the time that the first type 2 cell appears, under certain assumptions on the mutation rates.

We describe here in more detail this continuous approximation to the biased voter model. The spread of cancer is modeled on the  $d$ -dimensional torus  $\mathcal{T} := [0, L]^d$ , where the points 0 and  $L$  are identified. Note that this is the continuous analog of the space  $(\mathbb{Z} \bmod L)^d$  considered in [5]. We write  $N := L^d$  to denote the volume of  $\mathcal{T}$ . Each point in  $\mathcal{T}$  is assigned a type, indicating the number of mutations the cell has acquired. At the initial time  $t = 0$ , all points in  $\mathcal{T}$  are type 0, meaning they have no mutations. A so-called type 1 mutation then occurs at rate  $\mu_1$  per unit volume. Once each type 1 mutation appears, it spreads out in a ball at rate  $\alpha$  per unit time. This means that  $t$  time units after a mutation appears, all points within a distance  $\alpha t$  of the site where the mutation occurred will have acquired the mutation. Type 1 points then acquire a type 2 mutation at rate  $\mu_2$  per unit volume, and this process continues indefinitely. In general, type  $k$  mutations overtake type  $k - 1$  mutations at rate  $\mu_k$  per unit volume, and each type  $k$  mutation then spreads outward in a ball at rate  $\alpha$  per unit time. A full mathematical construction of this process, starting from Poisson point processes which govern the mutations, is given at the beginning of Section 3.

Let  $\sigma_k$  denote the first time that some cell becomes type  $k$ ; [11] obtained the asymptotic distribution of  $\sigma_2$  under a wide range of values for the parameters  $\alpha$ ,  $\mu_1$ , and  $\mu_2$ , extending the results in [5], and also found the asymptotic distribution of  $\sigma_k$  for  $k \geq 3$  assuming equal mutation rates  $\mu_i = \mu$  for all  $i$ . In this paper, we will further generalize the results in [11] for  $k \geq 3$  by considering the case where the mutation rates are increasing. We will see that several qualitatively different types of behavior are possible, depending on how fast the mutation rates increase.

We mention two biological justifications for assuming increasing mutation rates. A general phenomenon in carcinogenesis was suggested in [16] where there is favorable selection for certain mutations in genes responsible for repairing DNA damage. The increasing genetic instability disrupting DNA repair, in the context of the present paper, would correspond to increasing mutation rates. Also, our model would be of interest in the situation described in [22], which hypothesized that cancer cells express a mutator phenotype, which causes cells to mutate at a much higher rate, and proposed targeting the mutator phenotype as part of cancer

therapy, possibly with the goal of further increasing the mutation rate to the point where the mutations incapacitate or kill malignant cells.

As in [11], we assume that the rate of mutation spread  $\alpha$  is constant across mutation types, so that successive mutations have equal selective advantage. One possible generalization of our model would be to allow each type  $i$  mutation to have a different rate of spread  $\alpha_i$ . However, this more general model is non-trivial even to formulate unless  $(\alpha_i)_{i=1}^\infty$  is decreasing, because if  $\alpha_{i+1} > \alpha_i$ , then regions of type  $i+1$  could completely swallow the surrounding type  $i$  region. Consequently, it would be necessary to model what happens not only when mutations of types  $i+1$  and  $i$  compete, but also how mutations of types  $i+1$  and  $j \in \{1, \dots, i-1\}$  compete. We do not pursue this generalization here.

After computing the limiting distribution of  $\sigma_k$ , we also find the limiting distribution of the distances between the first mutation of type  $i$  and the first mutation of type  $j$ , where  $i < j$ . The distribution of distances between mutations is relevant in studying a phenomenon known as the “cancer field effect”, which refers to the increased risk for certain regions to acquire primary tumors. These regions are called premalignant fields, and they have a high risk of becoming malignant despite appearing to be normal [10]. The size of the premalignant field is clinically relevant when a patient is diagnosed with cancer, because it will determine the area of tissue to be surgically removed in order to avoid cancer recurrence. Surgical removal of premalignant fields, put in the context of this paper, is akin to removing the region with at least  $i$  mutations once the first type  $j > i$  mutation appears. The case in which  $i = 1$  and  $j = 2$  was considered in [10], which characterized the sizes of premalignant fields conditioned on  $\{\sigma_2 = t\}$  in  $d \in \{1, 2, 3\}$  spatial dimensions. These ideas were applied to head and neck cancer in [23].

We note that the model that we are studying in this paper independently appeared in the statistical physics literature, where it is known as the polynuclear growth model. It has been studied most extensively in  $d = 1$  when all of the  $\mu_k$  are the same [2, 19, 20], but the model was also formulated in higher dimensions in [21]. Most of this work in the statistical physics literature focuses on the long-run growth properties of the surface, and detailed information about the fluctuations has been established when  $d = 1$ . This is quite different from our goal of understanding the time to acquire a fixed number of mutations.

In Section 2 we introduce some basic notation and state our main results, as well as some heuristics explaining why these results are true. In Section 3 we prove the limit theorems regarding the time to wait for  $k$  mutations, and in Section 4 we prove the limit theorems for the distances between mutations.

## 2. Main results and heuristics

We first introduce some notation that we will need before stating the results. Given two sequences of non-negative real numbers  $(a_N)_{N=1}^\infty$  and  $(b_N)_{N=1}^\infty$ , we write:

$$a_N \sim b_N \text{ if } \lim_{N \rightarrow \infty} a_N/b_N = 1;$$

$$a_N \ll b_N \text{ if } \lim_{N \rightarrow \infty} a_N/b_N = 0 \text{ and } a_N \gg b_N \text{ if } \lim_{N \rightarrow \infty} a_N/b_N = \infty;$$

$$a_N \asymp b_N \text{ if } 0 < \liminf_{N \rightarrow \infty} a_N/b_N \leq \limsup_{N \rightarrow \infty} a_N/b_N < \infty;$$

$$a_N \lesssim b_N \text{ if } \limsup_{N \rightarrow \infty} a_N/b_N < \infty.$$

We also define the following notation:

- If  $X_N$  converges to  $X$  in distribution, we write  $X_N \Rightarrow X$ .
- If  $X_N$  converges to  $X$  in probability, we write  $X_N \rightarrow_p X$ .
- $\gamma_d$  denotes the volume of the unit ball in  $\mathbb{R}^d$ .
- For each  $k \geq 1$  and  $j \geq 1$ , we define

$$\beta_k := \left( N\alpha^{(k-1)d} \prod_{i=1}^k \mu_i \right)^{-1/((k-1)d+k)}, \quad \kappa_j := (\mu_j \alpha^d)^{-1/(d+1)}. \quad (1)$$

We explain how  $\beta_k$  and  $\kappa_j$  arise in Sections 2.3 and 2.5, respectively.

- $\sigma_k$  denotes the first time a mutation of type  $k$  appears, and  $\sigma_k^{(2)}$  denotes the second time a mutation of type  $k$  appears. More rigorous definitions of  $\sigma_k$  and  $\sigma_k^{(2)}$  are given in Sections 3 and 4, respectively.

All limits in this paper will be taken as  $N \rightarrow \infty$ . The mutation rates  $(\mu_i)_{i=1}^\infty$  and the rate of mutation spread  $\alpha$  will depend on  $N$ , even though this dependence is not recorded in the notation. Throughout the paper we will assume that the mutation rates  $(\mu_i)_{i=1}^\infty$  are asymptotically increasing, i.e.

$$\mu_1 \lesssim \mu_2 \lesssim \mu_3 \lesssim \dots \quad (2)$$

## 2.1. Theorem 1: Low mutation rates

Assume

$$\mu_1 \ll \frac{\alpha}{N^{(d+1)/d}} \text{ and } \frac{\mu_i}{\mu_1} \rightarrow c_i \in (0, \infty] \text{ for all } i \in \{1, \dots, k\}.$$

The first time a mutation of type 1 appears is exponentially distributed with rate  $N\mu_1$ . The maximal distance between any two points on the torus  $\mathcal{T} = [0, L]^d$  is  $\sqrt{d}L/2$ . Also note that  $L = N^{1/d}$ , where  $N$  is the volume of  $\mathcal{T}$ . Consequently, once the first type 1 mutation appears, it will spread to the entire torus in time  $\sqrt{d}L/(2\alpha) = \sqrt{d}N^{1/d}/(2\alpha)$ . Hence, as noted in [11], the time required for a type 1 mutation to fixate once it has first appeared is much shorter than  $\sigma_1$  precisely when  $N^{1/d}/\alpha \ll 1/(N\mu_1)$ , which is equivalent to  $\mu_1 \ll \alpha/N^{(d+1)/d}$ .

Now, because of the second assumption  $\mu_i/\mu_1 \rightarrow c_i \in (0, \infty]$ , mutations of types  $i \in \{2, \dots, k\}$  appear at least as fast as the first mutation. If  $c_i < \infty$ , then the waiting times  $\sigma_1$  and  $\sigma_i - \sigma_{i-1}$  are on the same order of magnitude. Because we have  $\sigma_1 \sim \text{Exponential}(N\mu_1 c_1)$ , it follows that  $\sigma_i - \sigma_{i-1}$  is also exponentially distributed and that  $\sigma_i - \sigma_{i-1} \sim \text{Exponential}(N\mu_1 c_i)$ . Otherwise, if  $c_i = \infty$ , then the first type  $i$  mutation appears so quickly that its waiting time  $\sigma_i - \sigma_{i-1}$  is negligible as  $N \rightarrow \infty$ . Putting everything together gives us the following theorem. This result is a very slight generalization of [11, Theorem 1], and is proved by the same method.

**Theorem 1.** *Suppose (2) holds, and  $\mu_1 \ll \alpha/N^{(d+1)/d}$ . Suppose that, for all  $i \in \{1, \dots, k\}$ , we have  $\mu_i/\mu_1 \rightarrow c_i \in (0, \infty]$ . Let  $W_1, \dots, W_k$  be independent random variables with  $W_i \sim \text{Exponential}(c_i)$  if  $c_i < \infty$  and  $W_i = 0$  if  $c_i = \infty$ . Then  $N\mu_1 \sigma_k \Rightarrow W_1 + \dots + W_k$ .*

Figure 1 illustrates that once a type  $i$  mutation appears, it quickly fills up the whole torus, and then a type  $i + 1$  mutation occurs.

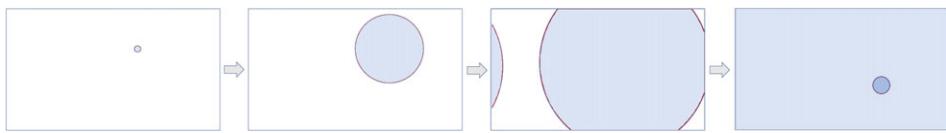


FIGURE 1. Mutations transition from type  $i$  to type  $i + 1$ . Higher mutation types are colored darker than lower mutation types.

## 2.2. Theorem 2: Type $j \geq 2$ mutations occur rapidly after $\sigma_1$

Assume

$$\mu_1 \gg \frac{\alpha}{N^{(d+1)/d}}, \quad \mu_2 \gg \frac{(N\mu_1)^{d+1}}{\alpha^d}. \quad (3)$$

In contrast to Theorem 1, the assumption  $\mu_1 \gg \alpha/N^{(d+1)/d}$  means that the time it takes for type 1 mutations to spread to the entire torus is much longer than  $\sigma_1$ . As a result, there will be many growing balls of type 1 mutations before any of these balls can fill the entire torus. However, if mutations of types 2, 3,  $\dots$ ,  $k$  appear quickly after the first type 1 mutation appears, then the time to wait for the first type  $k$  mutation will be close to the time to wait for the first type 1 mutation. We consider here the conditions under which this will be the case.

First, consider the ball of type 1 cells resulting from the initial type 1 mutation at time  $\sigma_1$ . Assuming  $t$  is small enough that, by time  $\sigma_1 + t$ , the ball has not started overlapping itself by wrapping around the torus, the ball will have volume  $\gamma_d(\alpha t)^d$  at time  $t$ . Then the probability that the first type 2 mutation appears in that ball before time  $t$  is

$$1 - \exp\left(-\int_0^t \mu_2 \gamma_d(\alpha r)^d dr\right) = 1 - \exp\left(-\frac{\gamma_d}{d+1} \mu_2 \alpha^d t^{d+1}\right). \quad (4)$$

It follows that the first time a type 2 mutation occurs in this ball is on the order of  $(\mu_2 \alpha^d)^{-1/(d+1)}$ . Hence, whenever  $(\mu_2 \alpha^d)^{-1/(d+1)} \ll 1/(N\mu_1)$ , which is equivalent to the second assumption in (3), it follows that  $\sigma_2 - \sigma_1$  is much quicker than  $\sigma_1$ . From this heuristic, we see that  $N\mu_1(\sigma_2 - \sigma_1) \rightarrow_p 0$ . Repeating this reasoning with types  $j - 1$  and  $j$  in place of types 1 and 2, we see that  $\sigma_j - \sigma_{j-1}$  is much quicker than  $\sigma_1$  when  $(\mu_j \alpha^d)^{-1/(d+1)} \ll 1/(N\mu_1)$ , or, equivalently,  $\mu_j \gg (N\mu_1)^{d+1}/\alpha^d$ . However, this follows from the second assumption in (3) because of (2). Hence, we also have  $N\mu_1(\sigma_j - \sigma_{j-1}) \rightarrow_p 0$ . Putting everything together, when  $N$  is large,

$$N\mu_1\sigma_k = N\mu_1\sigma_1 + N\mu_1(\sigma_2 - \sigma_1) + \dots + N\mu_1(\sigma_k - \sigma_{k-1}) \approx N\mu_1\sigma_1.$$

This gives us the following theorem. We note that the  $k = 2$  case was proved in [5, Theorem 3] using essentially the same reasoning as above.

**Theorem 2.** Suppose (2) holds. Suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$ . For all  $k \geq 2$ ,  $N\mu_1\sigma_k \Rightarrow W$ , where  $W \sim \text{Exponential}(1)$ .

A pictorial representation is given in Fig. 2, where the nested circles correspond to mutations of types 1,  $\dots$ ,  $k$  for  $k = 4$ .

## 2.3. Theorem 3: Type $j \in \{1, \dots, k - 1\}$ mutations appear many times

Assume

$$\mu_1 \gg \frac{\alpha}{N^{(d+1)/d}}, \quad \mu_k \ll \frac{1}{\alpha^d \beta_{k-1}^{d+1}}. \quad (5)$$

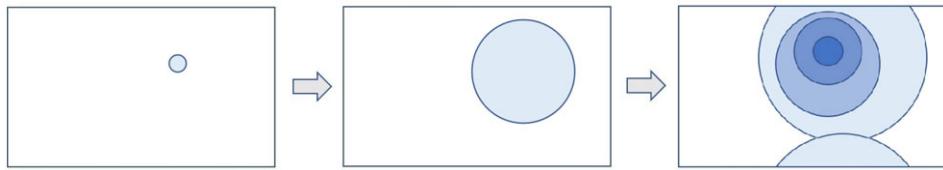


FIGURE 2. Once the first type 1 mutation appears, the type 2, 3, and 4 mutations all happen quickly. Higher mutation types are colored darker than lower mutation types.

As in Theorem 2, the first assumption ensures that  $\sigma_1$  is shorter than the time it takes for type 1 mutations to fixate once they appear. The second assumption ensures that all mutations of types up to  $k$  do not appear too quickly, so that we are not in the setting of Theorem 2. In particular, note that when  $k = 2$ , we have  $\beta_{k-1} = (N\mu_1)^{-1}$ , and the second assumption reduces to  $\mu_2 \ll (N\mu_1)^{d+1}/\alpha^d$ . When (5) holds, for  $j \in \{2, \dots, k\}$  there will be many small balls of type  $j - 1$  before any type  $j$  mutation appears. In this case, we will be able to use a ‘law of large numbers’ established in [11] to approximate the total volume of type  $j - 1$  regions with its expectation.

To explain what happens in this case, we review a derivation from [11]. We want to define an approximation  $v_j(t)$  to the total volume of regions with at least  $j$  mutations at time  $t$ . We set  $v_0(t) \equiv N$ . Next, let  $t > 0$ . For times  $r \in [0, t]$ , type  $j$  mutations occur at rate  $\mu_j v_{j-1}(r)$ , and these type  $j$  mutations each grow into a ball of size  $\gamma_d(\alpha(t-r))^d$  by time  $t$ . Therefore, we define

$$v_j(t) = \int_0^t \mu_j v_{j-1}(r) \gamma_d(\alpha(t-r))^d dr.$$

Note that this gives a good approximation to the volume of the type  $j$  region because we have many mostly non-overlapping balls of type  $j$ . In [11] it is shown using induction that

$$v_j(t) = \frac{\gamma_d^j (d!)^j}{(j(d+1))!} \left( \prod_{i=1}^j \mu_i \right) N \alpha^{jd} t^{j(d+1)},$$

which gives us the approximation

$$\begin{aligned} \mathbb{P}(\sigma_k > t) &\approx \exp \left( - \int_0^t \mu_k v_{k-1}(r) dr \right) \\ &= \exp \left( - \frac{\gamma_d^{k-1} (d!)^{k-1}}{((k-1)d+k)!} \left( \prod_{i=1}^k \mu_i \right) N \alpha^{(k-1)d} t^{(k-1)d+k} \right). \end{aligned}$$

It will follow that if we define  $\beta_k$  as in (1), then we have the following result.

**Theorem 3.** Suppose (2) holds. Let  $k \geq 2$ , and suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and  $\mu_k \ll 1/(\alpha^d \beta_{k-1}^{d+1})$ . Then, for  $t > 0$ ,

$$\mathbb{P}(\sigma_k > \beta_k t) \rightarrow \exp \left( - \frac{\gamma_d^{k-1} (d!)^{k-1}}{((k-1)d+k)!} t^{(k-1)d+k} \right).$$

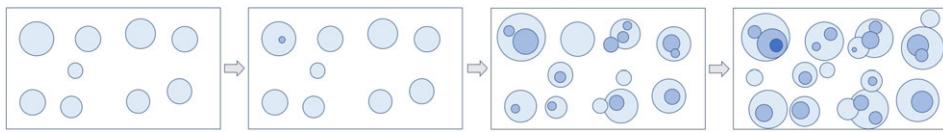


FIGURE 3. Mutations of types 1, 2, and 3 appear in succession. Higher mutation types are colored darker than lower mutation types.

When we have equal mutation rates (i.e.  $\mu_i = \mu$  for all  $i$ ), the result above is covered by [11, Theorem 10, part 3]. The form of the result and the strategy of the proof are exactly the same in the more general case when the mutation rates can differ. Theorem 3 is illustrated in Fig. 3 for  $k = 3$ .

#### 2.4. Theorem 4: An intermediate case between Theorems 2 and 3

Assume  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . We first define

$$l := \max \left\{ j \geq 2 : \mu_j \ll \frac{1}{\alpha^d \beta_{j-1}^{d+1}} \right\}. \quad (6)$$

It follows from (2) that if  $\mu_j \ll 1/(\alpha^d \beta_{j-1}^{d+1})$ , then  $\mu_{j-1} \ll 1/(\alpha^d \beta_{j-1}^{d+1})$ , which by Lemma 2 below implies that  $\mu_{j-1} \ll 1/(\alpha^d \beta_{j-2}^{d+1})$ . It follows that

$$l = \max \left\{ j \geq 2 : \mu_2 \ll \frac{1}{\alpha^d \beta_1^{d+1}}, \mu_3 \ll \frac{1}{\alpha^d \beta_2^{d+1}}, \dots, \mu_j \ll \frac{1}{\alpha^d \beta_{j-1}^{d+1}} \right\}. \quad (7)$$

Intuitively,  $l$  is the largest index for which mutations of types 1, 2,  $\dots$ ,  $l$  behave exactly as in Theorem 3. The definition of  $l$  in (6) omits the possibility  $l = 1$ , since  $\beta_0$  is undefined. However, if we define  $l = 1$  when the set over which we take the maximum in (6) is empty, then Theorem 4 below when  $l = 1$  is the same as Theorem 2. On the other hand, if  $l \in \{k, k+1, \dots\} \cup \{\infty\}$ , then by (7) we have  $\mu_k \ll 1/(\alpha^d \beta_{k-1}^{d+1})$ , in which case Theorem 3 applies. Hence, we assume  $l \in \{2, \dots, k-1\}$  and

$$\mu_{l+1} \gg \frac{1}{\alpha^d \beta_l^{d+1}}. \quad (8)$$

The situation in Theorem 4 is a hybrid of Theorems 2 and 3. A mutation of type  $j \in \{1, \dots, l-1\}$  takes a longer time to fixate in the torus than the interarrival time  $\sigma_j - \sigma_{j-1}$ . As a result, if  $j \in \{2, \dots, l\}$ , there will be many mostly non-overlapping balls of type  $j-1$  before time  $\sigma_j$ . Using this fact, we proceed as in Theorem 3 and find  $\lim_{N \rightarrow \infty} \mathbb{P}(\sigma_l > \beta_l t)$ . Next, our assumption in (8) places us in the regime of Theorem 2; all mutations of types  $l+1, \dots, k$  happen so quickly that for all  $\varepsilon > 0$  we have  $\mathbb{P}(\sigma_k - \sigma_l > \beta_l \varepsilon) \rightarrow 0$ . Then, combining these two results yields the following theorem.

**Theorem 4.** Suppose (2) holds, and suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . Suppose also that  $l \in \{2, \dots, k-1\}$  and that  $\mu_{l+1} \gg 1/(\alpha^d \beta_l^{d+1})$ . Then, for  $t > 0$ ,

$$\mathbb{P}(\sigma_k > \beta_l t) \rightarrow \exp \left( -\frac{\gamma_d^{l-1} (d!)^{l-1}}{((l-1)d + l)!} t^{(l-1)d+l} \right).$$

In pictures, Theorem 4 looks like Fig. 3 for mutations up to type  $l$ . Then, once the first type  $l$  mutation appears and spreads in a circle, all the subsequent mutations become nested within that circle, similar to Fig. 2.

**Remark 1.** Theorems 1–4 cover most of the possible cases in which (2) holds. However, we assume that either  $\mu_1 \ll \alpha/N^{(d+1)/d}$  or  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . In the case  $\mu_1 \asymp \alpha/N^{(d+1)/d}$ , we expect that at the time a type 2 mutation appears, there could be several overlapping type 1 balls whose size is comparable to the size of the torus, and we do not expect the limiting distribution of  $\sigma_k$  to have a simple expression. Consequently, we do not pursue this case here. We note that if  $\mu_1 \asymp \alpha/N^{(d+1)/d}$  and all mutation rates are equal (i.e.  $\mu_i = \mu$  for all  $i$ ), then it is proven, as a special case of [11, Theorem 12], that  $N\mu\sigma_k$  converges in distribution to a non-degenerate random variable for every  $k \geq 1$ . Likewise, we do not consider the case in which, instead of (8), we have  $\mu_{l+1} \asymp 1/(\alpha^d \beta_l^{d+1})$ . In this case we believe there could be several overlapping type  $l$  balls at the time the first type  $l+1$  mutation occurs, again preventing there from being a simple expression for the limit distribution.

## 2.5. Distances between mutations

For  $1 \leq i < j$ , define  $D_{i,j}$  to be the distance in the torus between the location of the first mutation of type  $j$  and the location of the first mutation of type  $i$ . Also define  $D_{i+1} := D_{i,i+1}$ .

Consider the setting of Theorem 2. We will assume a stronger version of (2):

$$\mu_2 \ll \mu_3 \ll \mu_4 \ll \dots \quad (9)$$

Recall that the mutations appear in nested balls as in Fig. 2. Because the first type  $j+1$  mutation will therefore appear before the second type  $j$  mutation with high probability, we can calculate, as in (4), that

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > t) \approx \exp\left(-\frac{\mu_{j+1} \gamma_d \alpha^d}{d+1} t^{d+1}\right).$$

It follows that if we define  $\kappa_{j+1}$  as in (1), then

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > \kappa_{j+1} t) \approx \exp\left(-\frac{\gamma_d}{d+1} t^{d+1}\right).$$

With this, we can calculate the approximate density  $f(t)$  of  $(\sigma_{j+1} - \sigma_j)/\kappa_{j+1}$ . This allows us to calculate

$$\mathbb{P}\left(\frac{D_{j+1}}{\alpha \kappa_{j+1}} \leq s\right) \approx \int_0^\infty \mathbb{P}\left(\frac{D_{j+1}}{\alpha \kappa_{j+1}} \leq s \mid \frac{\sigma_{j+1} - \sigma_j}{\kappa_{j+1}} = t\right) f(t) dt.$$

The location of the first type  $j+1$  mutation conditioned on  $\sigma_{j+1} - \sigma_j = \kappa_{j+1} t$  is a uniformly random point on a  $d$ -dimensional ball of radius  $\alpha \kappa_{j+1} t$ . This allows us to calculate  $\lim_{N \rightarrow \infty} \mathbb{P}(D_{j+1} \leq \alpha \kappa_{j+1} s)$ . Next, because of (9), mutations of types  $j+2, j+3, j+4, \dots$  appear rapidly once the first type  $j+1$  appears. This means that  $D_{j+2} + \dots + D_{j+k}$  is small relative to  $D_{j+1}$ , and therefore that  $D_{j,k}$  has the same limiting distribution as  $D_{j+1}$ . These heuristics lead to the following theorem.

**Theorem 5.** Suppose (9) holds. Suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$ . Suppose  $1 \leq j < k$ . Then, for all  $s > 0$ ,

$$\mathbb{P}\left(\frac{D_{j,k}}{\alpha \kappa_{j+1}} \leq s\right) \rightarrow \int_0^\infty \gamma_d (t \wedge s)^d \exp\left(-\frac{\gamma_d t^{d+1}}{d+1}\right) dt.$$

Recall the definition of  $l$  in (6). Theorem 4 is similar to Theorem 2 in that once the first type  $l$  mutation appears, all the subsequent type  $l+1, l+2, \dots$  mutations happen quickly. Therefore, it is reasonable to expect that the type  $l, l+1, l+2, \dots$  mutations behave similarly to the type  $1, 2, 3, \dots$  mutations in Theorem 2. Hence, analogous to (9), assume that

$$\mu_{l+1} \ll \mu_{l+2} \ll \mu_{l+3} \ll \dots \quad (10)$$

We then obtain the following result.

**Theorem 6.** *Suppose (10) holds, and suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . Define  $l$  as in (6), and suppose also that  $l \geq 2$  and that  $\mu_{l+1} \gg 1/(\alpha^d \beta_l^{d+1})$ . Suppose  $l \leq j < k$ . Then, for all  $s > 0$ ,*

$$\mathbb{P}\left(\frac{D_{j,k}}{\alpha \kappa_{j+1}} \leq s\right) \rightarrow \int_0^\infty \gamma_d(t \wedge s)^d \exp\left(-\frac{\gamma_d t^{d+1}}{d+1}\right) dt.$$

**Remark 2.** Theorems 5 and 6 hold in the settings of Theorems 2 and 4 respectively. In the setting of Theorem 1, each type  $i \geq 1$  mutation fills up the entire torus before a type  $i+1$  mutation occurs, and so the first type  $i+1$  mutation appears at a uniformly distributed point on the torus, independently of where all previous mutations originated. Therefore, the problem of finding the distribution of the distances between mutations becomes trivial in this case. On the other hand, in the setting of Theorem 3, type  $i$  mutations appear in small and mostly non-overlapping circles before the first type  $i+1$  mutation appears. Thus, calculating the distribution of  $D_{i+1}$  requires understanding not only the total volume of the type  $i$  region, but also the sizes and locations of many small type  $i$  regions. We do not pursue this case here, but we conjecture that because the first type  $i+1$  mutation is likely not to appear within the type  $i$  region generated by the first type  $i$  mutation, the locations of the first type  $i$  and the first type  $i+1$  mutations should be nearly independent of each other, as in the setting of Theorem 1.

### 3. Proofs of limit theorems for $\sigma_k$

In this section we prove Theorems 1–4. We begin by introducing the structure of the torus  $\mathcal{T} = [0, L]^d$ , following the notation of [11]. We define a pseudometric on the closed interval  $[0, L]$  by  $d_L(x, y) := \min\{|x - y|, L - |x - y|\}$ . The  $d$ -dimensional torus of side length  $L$  will be denoted by  $\mathcal{T} = [0, L]^d$ . For  $x = (x^1, \dots, x^d) \in \mathcal{T}$  and  $y = (y^1, \dots, y^d) \in \mathcal{T}$ , we define a pseudometric by

$$|x - y| := \sqrt{\sum_{i=1}^d d_L(x^i, y^i)^2}.$$

The torus should be viewed as  $\mathcal{T}$  modulo the equivalence relation  $x \sim y$  if and only if  $|x - y| = 0$ , or more simply  $\mathcal{T} = (\mathbb{R} \text{ mod } L)^d$ . However, we will continue to write  $\mathcal{T} = [0, L]^d$ , keeping in mind that certain points are considered to be the same via the equivalence relation defined above. It will be useful to observe the following:

- We have  $d_L(x, y) \leq L/2$  for all  $x, y \in [0, L]$ . As a result, the distance between any two points  $x, y \in \mathcal{T}$  is at most

$$\sup\{|x - y| : x, y \in \mathcal{T}\} = \sqrt{\sum_{i=1}^d \left(\frac{L}{2}\right)^2} = \frac{\sqrt{d}L}{2}.$$

- Therefore, once a mutation of type  $j$  appears, the entire torus will become type  $j$  in time

$$\frac{\text{maximal distance between any } x, y \in \mathcal{T}}{\text{rate of mutation spread per unit time}} = \frac{\sqrt{d}L}{2\alpha}. \quad (11)$$

We use  $|A|$  to denote the Lebesgue measure of some subset  $A$  of  $\mathcal{T}$  or  $\mathcal{T} \times [0, \infty)$ , so that  $N = L^d = |\mathcal{T}|$  is the torus volume. Each  $x \in \mathcal{T}$  at time  $t$  has a type  $k \in \{0, 1, 2, \dots\}$ , which we denote by  $T(x, t)$ , corresponding to the number of mutations the site has acquired. The set of type  $i$  sites is defined by  $\chi_i(t) := \{x \in \mathcal{T} : T(x, t) = i\}$ . The set of points whose type is at least  $i$  is defined by

$$\psi_i(t) := \{x \in \mathcal{T} : T(x, t) \geq i\} = \bigcup_{j=i}^{\infty} \chi_j(t).$$

At time  $t$ , we denote the total volume of type  $i$  sites by  $X_i(t) := |\chi_i(t)|$ , and the total volume of sites with type at least  $i$  by  $Y_i(t) := |\psi_i(t)|$ . The first time a type  $k$  mutation appears in the torus can be expressed as  $\sigma_k = \inf\{t > 0 : Y_k(t) > 0\}$ .

Still following [11], we now explicitly describe the construction of the process which gives rise to mutations in the torus. We model mutations as random space-time points  $(x, t) \in \mathcal{T} \times [0, \infty)$ . Let  $(\Pi_k)_{k=1}^{\infty}$  be a sequence of independent Poisson point processes on  $\mathcal{T} \times [0, \infty)$ , where  $\Pi_k$  has intensity  $\mu_k$ . That is, for any space-time region  $A \subseteq \mathcal{T} \times [0, \infty)$ , the probability that  $A$  contains  $j$  points of type  $k$  is  $e^{-\mu_k|A|}(\mu_k|A|)^j/j!$ . Each  $(x, t) \in \Pi_k$  is a space-time point at which  $x \in \mathcal{T}$  can acquire a  $k$ th mutation at time  $t$ . We say that  $x$  mutates to type  $k$  at time  $t$  precisely when  $x \in \chi_{k-1}(t)$  and  $(x, t) \in \Pi_k$ . Once an individual obtains a type  $k$  mutation, it spreads the type  $k$  mutations outward in a ball at rate  $\alpha$  per unit time.

### 3.1. Proof of Theorem 1

In the setting of Theorem 1, once the first mutation appears, with high probability it spreads to the entire torus before another mutation appears. The proof of Theorem 1 uses [11, Theorem 1], which we restate below as Theorem 7. Theorem 1 is very similar to Theorem 7 when  $j = 1$ . However, Theorem 7 requires  $\mu_j \ll \alpha/N^{(d+1)/d}$  for all  $j \in \{1, \dots, k\}$ , whereas Theorem 1 requires this condition only for  $j = 1$ . This is why Theorem 1 cannot be deduced directly from Theorem 7, even though the proofs of the results are essentially the same.

**Theorem 7.** ([11, Theorem 1].) *Suppose  $\mu_i \ll \alpha/N^{(d+1)/d}$  for  $i \in \{1, \dots, k-1\}$ . Suppose there exists  $j \in \{1, \dots, k\}$  such that  $\mu_j \ll \alpha/N^{(d+1)/d}$  and  $\mu_i/\mu_j \rightarrow c_i \in (0, \infty]$  for all  $i \in \{1, \dots, k\}$ . Let  $W_1, \dots, W_k$  be independent random variables such that  $W_i$  has an exponential distribution with rate parameter  $c_i$  if  $c_i < \infty$  and  $W_i = 0$  if  $c_i = \infty$ . Then  $N\mu_j\sigma_k \Rightarrow W_1 + \dots + W_k$ .*

*Proof of Theorem 1.* Let  $r := \max\{j \in \{1, \dots, k\} : \mu_j \lesssim \mu_1\}$ . For all  $j \in \{1, \dots, r\}$ , we have  $\mu_j \ll \alpha/N^{(d+1)/d}$ . By Theorem 7,  $N\mu_1\sigma_r \Rightarrow W_1 + \dots + W_r$ . If  $r = k$ , then the conclusion follows. Otherwise,  $r \leq k-1$ , and by the maximality of  $r$  and (2), we have  $\mu_l/\mu_1 \rightarrow \infty$  for all  $l \in \{r+1, \dots, k\}$ . Then the result follows if we show that  $N\mu_1(\sigma_k - \sigma_r) \rightarrow_p 0$ . We have

$$0 \leq N\mu_1(\sigma_k - \sigma_r) = N\mu_1 \sum_{j=r}^{k-1} (\sigma_{j+1} - \sigma_j). \quad (12)$$

We will find an upper bound for the right-hand side of (12). For  $i \geq 1$ , let  $t_i = \inf\{t > 0 : Y_i(t) = N\}$  be the first time which every point in  $\mathcal{T}$  is of at least type  $i$ . Define  $\hat{t}_i := t_i - \sigma_i$ , which is

the time elapsed between  $\sigma_i$  and when mutations of type  $i$  fixate in the torus. Also define  $\hat{\sigma}_i = \inf\{t > 0: \Pi_i \cap (\mathcal{T} \times [t_{i-1}, t]) \neq \emptyset\}$ , which is the first time there is a potential type  $i$  mutation after  $t_{i-1}$ . Observe that, because we always have  $\sigma_i \leq \hat{\sigma}_i$ ,

$$\sigma_{j+1} - \sigma_j \leq \hat{\sigma}_{j+1} - \sigma_j = \hat{\sigma}_{j+1} - \sigma_j + t_j - t_j = \hat{t}_j + (\hat{\sigma}_{j+1} - t_j).$$

Also observe that, by (11), we have  $\hat{t}_j \leq \sqrt{dN^{1/d}}/(2\alpha)$ . Consequently, the right-hand side of (12) has the upper bound

$$N\mu_1 \left( \sum_{j=r}^{k-1} \hat{t}_j + \sum_{j=r}^{k-1} (\hat{\sigma}_{j+1} - t_j) \right) \leq N\mu_1(k-r) \frac{\sqrt{dN^{1/d}}}{2\alpha} + N\mu_1 \sum_{j=r}^{k-1} (\hat{\sigma}_{j+1} - t_j).$$

The result follows if the right-hand side of the above expression converges to 0 in probability. The first term tends to zero because  $\mu_1 \ll \alpha/N^{(d+1)/d}$ . The second term tends to zero because  $\hat{\sigma}_{j+1} - t_j \sim \text{Exponential}(N\mu_{j+1})$ , so  $N\mu_1(\hat{\sigma}_{j+1} - t_j) \sim \text{Exponential}(\mu_{j+1}/\mu_1) \rightarrow_p 0$ .  $\square$

### 3.2. Proof of Theorem 2

**Lemma 1.** *Let  $t_N$  be a random time that is  $\sigma(\Pi_1, \dots, \Pi_{j-1})$ -measurable and satisfies  $t_N \geq \sigma_{j-1}$ . Then*

$$\mathbb{P}(\sigma_j > t_N) = \mathbb{E} \left[ \exp \left( - \int_{\sigma_{j-1}}^{t_N} \mu_j Y_{j-1}(s) \, ds \right) \right].$$

*Proof.* Write  $\mathcal{G} := \sigma(\Pi_1, \dots, \Pi_{j-1})$ . Define the set  $A := \{(x, r) \in \psi_{j-1}(r) \times [\sigma_{j-1}, t_N]\}$ , and note that the Lebesgue measure of this set, which we denote by  $|A|$ , is a  $\mathcal{G}$ -measurable random variable. The event  $\{\sigma_j > t_N\}$  occurs precisely when  $\Pi_j \cap A = \emptyset$ . Let  $X$  be the number of points of  $\Pi_j$  in the set  $A$ . Because  $\Pi_j$  is independent of  $\Pi_1, \dots, \Pi_{j-1}$ , the conditional distribution of  $X$  given  $\mathcal{G}$  is Poisson( $\mu_j |A|$ ). Therefore,

$$\mathbb{P}(\sigma_j > t_N \mid \mathcal{G}) = \mathbb{P}(X = 0 \mid \mathcal{G}) = \exp(-\mu_j |A|) = \exp \left( - \int_{\sigma_{j-1}}^{t_N} \mu_j Y_{j-1}(s) \, ds \right).$$

Taking expectations of both sides completes the proof.  $\square$

*Proof of Theorem 2.* Write  $N\mu_1\sigma_k$  as a telescoping sum,

$$N\mu_1\sigma_k = N\mu_1\sigma_1 + \sum_{j=2}^k N\mu_1(\sigma_j - \sigma_{j-1}).$$

We have  $N\mu_1\sigma_1 \sim \text{Exponential}(1)$ . Hence, it suffices to show that, for each  $j \geq 2$ , the random variable  $N\mu_1(\sigma_j - \sigma_{j-1})$  converges in probability to zero. Let  $t > 0$ . Then, by Lemma 1,

$$\mathbb{P}(N\mu_1(\sigma_j - \sigma_{j-1}) > t) = \mathbb{P} \left( \sigma_j > \frac{t}{N\mu_1} + \sigma_{j-1} \right) = \mathbb{E} \left[ \exp \left( - \int_{\sigma_{j-1}}^{t/(N\mu_1) + \sigma_{j-1}} \mu_j Y_{j-1}(s) \, ds \right) \right].$$

We want to show that the term on the right-hand side tends to zero. By the dominated convergence theorem, it suffices to show that as  $N \rightarrow \infty$ ,

$$\int_{\sigma_{j-1}}^{t/(N\mu_1) + \sigma_{j-1}} \mu_j Y_{j-1}(s) \, ds \rightarrow \infty.$$

Notice that because  $\mu_1 \gg \alpha/N^{(d+1)/d}$ , for all sufficiently large  $N$  we have  $t/(N\mu_1) \leq N^{1/d}/(2\alpha)$ . Therefore, at time  $\sigma_{j-1} + t/(N\mu_1)$ , there is a ball of type  $j-1$  mutations of radius  $\alpha(t - \sigma_{j-1})$  which has not yet begun to wrap around the torus and overlap itself. Hence, we have  $Y_{j-1}(s) \geq \gamma_d \alpha^d (s - \sigma_{j-1})^d$  for  $s \in [\sigma_{j-1}, \sigma_{j-1} + t/(N\mu_1)]$ , and therefore

$$\begin{aligned} \int_{\sigma_{j-1}}^{t/(N\mu_1) + \sigma_{j-1}} \mu_j Y_{j-1}(s) \, ds &\geq \int_{\sigma_{j-1}}^{t/(N\mu_1) + \sigma_{j-1}} \mu_j \gamma_d \alpha^d (s - \sigma_{j-1})^d \, ds \\ &= \int_0^{t/(N\mu_1)} \mu_j \gamma_d \alpha^d u^d \, du = \frac{\mu_j \gamma_d \alpha^d}{d+1} \left( \frac{t}{N\mu_1} \right)^{d+1}. \end{aligned}$$

It remains to show that

$$\frac{\mu_j \gamma_d \alpha^d}{d+1} \left( \frac{t}{N\mu_1} \right)^{d+1} \rightarrow \infty \text{ as } N \rightarrow \infty.$$

For the above to hold, it suffices to have  $\mu_j \gg (N\mu_1)^{d+1}/\alpha^d$ , which holds due to the second assumption in the theorem and (2). This completes the proof.  $\square$

### 3.3. Proof of Theorem 3

We recall the definition of  $\beta_k$  as in (1) of Section 2. In the setting of Theorem 3,  $\beta_k$  is the order of magnitude of the time it takes for the  $k$ th mutation to appear.

Much of the proof of Theorem 3 will rely on [11, Lemma 9], which approximates a monotone stochastic process by a deterministic function under a certain time scaling. In order to apply this lemma, it is important to ensure that  $Y_k(t)$  is well approximated by its expectation, which is [11, Lemma 8].

Before proving Theorem 3, we state several lemmas, some of which are from [11]. First, we need to ensure that the last assumption,  $\mu_k \alpha^d \beta_{k-1}^{d+1} \rightarrow 0$ , in Theorem 3 implies  $\mu_k \alpha^d \beta_k^{d+1} \rightarrow 0$ , so that we are able to use part (ii) of Lemma 5 to approximate  $Y_{k-1}(\beta_k t)$  by its expectation.

**Lemma 2.** *For  $k \geq 2$ , we have  $\mu_k \ll 1/(\alpha^d \beta_k^{d+1})$  if and only if  $\mu_k \ll 1/(\alpha^d \beta_{k-1}^{d+1})$ .*

*Proof.* By using the definition of  $\beta_k$  from (1), we get

$$\begin{aligned} \mu_k \ll \frac{1}{\alpha^d \beta_k^{d+1}} &\iff \mu_k^{(k-1)d+k} \ll \frac{1}{\alpha^{d[(k-1)d+k]}} \left( N \alpha^{(k-1)d} \prod_{i=1}^k \mu_i \right)^{d+1} \\ &\iff \mu_k^{(k-2)d+(k-1)} \ll \frac{1}{\alpha^d} N^{d+1} \left( \prod_{i=1}^{k-1} \mu_i \right)^{d+1} \\ &\iff \mu_k^{(k-2)d+(k-1)} \ll \frac{\alpha^{d(d+1)(k-2)}}{\alpha^{d[(k-2)d+(k-1)]}} N^{d+1} \left( \prod_{i=1}^{k-1} \mu_i \right)^{d+1} \\ &\iff \mu_k^{(k-2)d+(k-1)} \ll \frac{1}{\alpha^{d[(k-2)d+(k-1)]}} \left( N \alpha^{(k-2)d} \prod_{i=1}^{k-1} \mu_i \right)^{d+1} \\ &\iff \mu_k \ll \frac{1}{\alpha^d \beta_{k-1}^{d+1}}, \end{aligned}$$

as claimed.  $\square$

We also need [11, Lemma 9], which is restated as Lemma 3. This lemma gives necessary conditions under which a monotone stochastic process is well approximated by a deterministic function.

**Lemma 3.** *Suppose, for all positive integers  $N$ ,  $(Y_N(t), t \geq 0)$  is a non-decreasing stochastic process such that  $\mathbb{E}[Y_N(t)] < \infty$  for each  $t > 0$ . Assume there exist sequences of positive numbers  $(v_N)_{N=1}^{\infty}$  and  $(s_N)_{N=1}^{\infty}$  and a continuous non-decreasing function  $g > 0$  such that, for each fixed  $t > 0$  and  $\varepsilon > 0$ , we have*

$$\lim_{N \rightarrow \infty} \mathbb{P}(|Y_N(s_N t) - \mathbb{E}[Y_N(s_N t)]| > \varepsilon \mathbb{E}[Y_N(s_N t)]) = 0, \quad (13)$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{v_N} \mathbb{E}[Y_N(s_N t)] = g(t). \quad (14)$$

Then, for all  $\varepsilon > 0$  and  $\delta > 0$ , we have

$$\lim_{N \rightarrow \infty} \mathbb{P}(v_N g(t)(1 - \varepsilon) \leq Y_N(s_N t) \leq v_N g(t)(1 + \varepsilon) \text{ for all } t \in [\delta, \delta^{-1}]) = 1.$$

Next, we state a criterion which guarantees that, for fixed  $t > 0$ , the probability  $\mathbb{P}(\sigma_k > \beta_k t)$  converges to a deterministic function as  $N \rightarrow \infty$ .

**Lemma 4.** *For a continuous non-negative function  $g$ , a sequence  $(v_N)_{N=1}^{\infty}$  of positive real numbers, and  $\delta, \varepsilon > 0$ , define the event*

$$B_N^{k-1}(\delta, \varepsilon, g, v_N) = \{g(u)(1 - \varepsilon) v_N \leq Y_{k-1}(\beta_k u) \leq g(u)(1 + \varepsilon) v_N, \text{ for all } u \in [\delta, \delta^{-1}]\}.$$

If  $(v_N)_{N=1}^{\infty}$  and  $g$  are chosen such that  $\lim_{N \rightarrow \infty} \mathbb{P}(B_N^{k-1}(\delta, \varepsilon, g, v_N)) = 1$  and  $\lim_{N \rightarrow \infty} v_N \beta_k \mu_k$  exists, then

$$\lim_{N \rightarrow \infty} \mathbb{P}(\sigma_k > \beta_k t) = \lim_{N \rightarrow \infty} \exp\left(-v_N \beta_k \mu_k \int_0^t g(u) du\right).$$

*Proof.* Suppose  $\delta \leq t \leq \delta^{-1}$ . We reason as in the proof of [11, Theorem 10]. The upper and lower bounds from [11, (26) and (27)] are

$$\begin{aligned} \mathbb{P}(\sigma_k > \beta_k t) &\leq \exp\left(-\mu_k \beta_k v_N (1 - \varepsilon) \int_{\delta}^t g(u) du\right) + \mathbb{P}(B_N^{k-1}(\delta, \varepsilon, g, v_N)^c), \\ \mathbb{P}(\sigma_k > \beta_k t) &\geq \mathbb{P}(B_N^{k-1}(\delta, \varepsilon, g, v_N)) \exp\left(-v_N (1 + \varepsilon) \beta_k \mu_k \int_{\delta}^t g(u) du\right) \\ &\quad - \frac{\gamma_d^{k-1} (d!)^{k-1}}{(d(k-1) + k)!} \delta^{d(k-1)+k}. \end{aligned}$$

Taking  $N \rightarrow \infty$  and then  $\varepsilon, \delta \rightarrow 0$ , we get the desired result.  $\square$

We also need to approximate the expected volume of type  $k$  or higher regions,  $\mathbb{E}[Y_k(t)]$ , with a deterministic function, as well as making sure that  $Y_k(t)$  is well approximated by its expectation. Lemma 5 is a restatement of [11, Lemmas 5 and 8]. It is important to note that for this result, the time  $t$  may depend on  $N$ .

**Lemma 5.** *Fix a positive integer  $k$ . Suppose  $\mu_j \alpha^d t^{d+1} \rightarrow 0$  for all  $j \in \{1, \dots, k\}$ . Also suppose  $t \leq N^{1/d}/(2\alpha)$ . Then*

(i) Setting  $v_k(t) := \frac{\gamma_d^k (d!)^k}{(k(d+1))!} \left( \prod_{i=1}^k \mu_i \right) N \alpha^{kd} t^{k(d+1)}$ , we have  $\mathbb{E}[Y_k(t)] \sim v_k(t)$ .

(ii) If, in addition, we assume  $\left( \prod_{i=1}^k \mu_i \right) N \alpha^{(k-1)d} t^{(k-1)d+k} \rightarrow \infty$ , then, for all  $\varepsilon > 0$ ,

$$\lim_{N \rightarrow \infty} \mathbb{P}((1 - \varepsilon) \mathbb{E}[Y_k(t)] \leq Y_k(t) \leq (1 + \varepsilon) \mathbb{E}[Y_k(t)]) = 1.$$

**Remark 3.** Lemma 5 in [11] omits the necessary hypothesis  $t \leq N^{1/d}/(2\alpha)$ . This hypothesis ensures that a growing ball of mutations cannot begin to wrap around the torus and overlap itself before time  $t$ , which is needed for the formula for  $\mathbb{E}[\Lambda_{k-1}(t)]$  in [11, (15)] to be exact. This equation is used in the proof of [11, Lemma 5]. Note that the hypothesis  $t \leq N^{1/d}/(2\alpha)$  is also needed for [11, Lemma 8], because its proof uses [11, Lemma 5]. However, because it is easily verified that this hypothesis is satisfied in [11] whenever these lemmas are used, all of the main results in [11] are correct without additional hypotheses.

The next lemma states that if  $\mu_1 \gg \alpha/N^{(d+1)/d}$ , then  $\beta_l$  is much smaller than the time it takes for a mutation to spread to the entire torus.

**Lemma 6.** Suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and (2) holds. Then  $\beta_l \ll N^{1/d}/\alpha$  for any  $l \in \mathbb{N}$ .

*Proof.* By (2), we have  $\mu_1, \dots, \mu_l \gg \alpha/N^{(d+1)/d}$ . Thus

$$\prod_{i=1}^l \mu_i \gg \frac{\alpha^l}{N^{l(1+1/d)}}.$$

On the other hand by simplifying,

$$\beta_l \ll \frac{N^{1/d}}{\alpha} \iff N \alpha^{(l-1)d} \prod_{i=1}^l \mu_i \gg \left( \frac{\alpha}{N^{1/d}} \right)^{(l-1)d+l} \iff \prod_{i=1}^l \mu_i \gg \frac{\alpha^l}{N^{l(1+1/d)}}.$$

This proves the lemma.  $\square$

*Proof of Theorem 3.* In view of Lemma 4, we will choose  $(v_N)_{N=1}^\infty$  and a continuous non-negative function  $g_k$  such that  $\lim_{N \rightarrow \infty} v_N \beta_k \mu_k$  exists and  $\mathbb{P}(B_N^{k-1}(\delta, \varepsilon, g_k, v_N)) \rightarrow 1$  as  $N \rightarrow \infty$ . We set  $v_N = 1/(\beta_k \mu_k)$ , and, as in the proof of [11, Theorem 10], set

$$g_k(t) := \frac{\gamma_d^{k-1} (d!)^{k-1} t^{(k-1)(d+1)}}{((k-1)(d+1))!}.$$

A lengthy calculation shows that  $\beta_k \mu_k v_{k-1}(\beta_k t) = g_k(t)$ . On the other hand, by the last assumption in the theorem, we have  $\mu_k \alpha^d \beta_{k-1}^{d+1} \rightarrow 0$ . By Lemma 2, this is equivalent to  $\mu_k \alpha^d \beta_k^{d+1} \rightarrow 0$ . Because of (2), this implies that  $\mu_j \alpha^d (\beta_k t)^{d+1} \rightarrow 0$  for all  $j \in \{1, \dots, k\}$ . Also, because of Lemma 6, we have  $\beta_k \ll N^{1/d}/(2\alpha)$ . Hence the hypotheses of Lemma 5 are satisfied, and by the first result in Lemma 5 applied to  $k-1$ , it follows that  $v_{k-1}(\beta_k t) \sim \mathbb{E}[Y_{k-1}(\beta_k t)]$ , which implies

$$\lim_{N \rightarrow \infty} \beta_k \mu_k \mathbb{E}[Y_{k-1}(\beta_k t)] = \lim_{N \rightarrow \infty} \beta_k \mu_k v_{k-1}(\beta_k t) = g_k(t).$$

Hence, (14) of Lemma 3 is satisfied. A direct calculation gives

$$\left( \prod_{i=1}^{k-1} \mu_i \right) N \alpha^{(k-2)d} \beta_k^{(k-2)d+k-1} = \frac{1}{\mu_k \alpha^d \beta_k^{d+1}} \rightarrow \infty,$$

which by the second result of Lemma 5 is sufficient to give (13). Therefore, Lemma 3 guarantees that  $\mathbb{P}(B_N^{k-1}(\delta, \varepsilon, g_k, v_N)) \rightarrow 1$  as  $N \rightarrow \infty$ . Then, Lemma 4 gives us

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{P}(\sigma_k > \beta_k t) &= \lim_{N \rightarrow \infty} \exp \left( -\nu_N \beta_k \mu_k \int_0^t g_k(u) \, du \right) \\ &= \exp \left( - \int_0^t \frac{\gamma_d^{k-1} (d!)^{k-1} u^{(k-1)(d+1)}}{((k-1)(d+1))!} \, du \right) \\ &= \exp \left( - \frac{\gamma_d^{k-1} (d!)^{k-1}}{(d(k-1)+k)!} t^{d(k-1)+k} \right), \end{aligned}$$

completing the proof.  $\square$

### 3.4. Proof of Theorem 4

Now we turn to proving Theorem 4, which is a hybrid of Theorems 2 and 3. In particular, we assume that there is some  $l \in \mathbb{N}$  such that the mutation rates  $\mu_1, \mu_2, \dots, \mu_l$  fall under the regime of Theorem 3, and all subsequent mutation rates  $\mu_{l+1}, \dots, \mu_k$  are large enough that all mutations after the first type  $l$  mutation occur quickly, as in Theorem 2.

*Proof of Theorem 4.* For ease of notation, set, for  $j \in \mathbb{N}$  and  $t \geq 0$ ,

$$f_j(t) := \exp \left( - \frac{\gamma_d^{j-1} (d!)^{j-1} t^{d(j-1)+j}}{(d(j-1)+j)!} \right).$$

For  $\varepsilon > 0$ , we have the inequalities

$$\mathbb{P}(\sigma_l > \beta_l t) \leq \mathbb{P}(\sigma_k > \beta_l t) \leq \mathbb{P}(\sigma_l > \beta_l(t - \varepsilon)) + \mathbb{P}(\sigma_k - \sigma_l > \beta_l \varepsilon).$$

Taking  $N \rightarrow \infty$  and using Theorem 3 (noting that  $l \geq 2$ ), we have

$$f_l(t) \leq \lim_{N \rightarrow \infty} \mathbb{P}(\sigma_k > \beta_l t) \leq f_l(t - \varepsilon) + \lim_{N \rightarrow \infty} \mathbb{P}(\sigma_k - \sigma_l > \beta_l \varepsilon).$$

Since  $f_l$  is continuous, the result follows (by taking  $\varepsilon \rightarrow 0$ ) once we show that, for each fixed  $\varepsilon > 0$ ,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\sigma_k - \sigma_l > \beta_l \varepsilon) = 0. \tag{15}$$

Notice that because

$$\{\sigma_k - \sigma_l > \beta_l \varepsilon\} \subseteq \bigcup_{j=l}^{k-1} \left\{ \sigma_{j+1} - \sigma_j > \frac{\beta_l \varepsilon}{k-l} \right\},$$

it suffices to show that, for all  $j \in \{l, \dots, k-1\}$ ,  $\mathbb{P}(\sigma_{j+1} - \sigma_j > \beta_l \varepsilon) \rightarrow 0$ . By Lemma 1, we have

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > \beta_l \varepsilon) = \mathbb{E} \left[ \exp \left( - \int_{\sigma_j}^{\beta_l \varepsilon + \sigma_j} \mu_{j+1} Y_j(s) \, ds \right) \right].$$

Hence, by the dominated convergence theorem, to show that  $\mathbb{P}(\sigma_{j+1} - \sigma_j > \beta_l \varepsilon) \rightarrow 0$ , it suffices to show that  $\int_{\sigma_j}^{\beta_l \varepsilon + \sigma_j} \mu_{j+1} Y_j(s) \, ds \rightarrow \infty$  almost surely. By Lemma 6 we have  $\beta_l \ll N^{1/d}/\alpha$ , so  $\beta_l \varepsilon \leq N^{1/d}/(2\alpha)$  for large enough  $N$ . That is,  $\beta_l \varepsilon$  does not exceed the time it takes for a mutation to wrap around the torus. Hence, we have the lower bound  $Y_j(s) \geq \gamma_d \alpha^d (s - \sigma_j)^d$  for  $s \in [\sigma_j, \sigma_j + \beta_l \varepsilon]$ , and

$$\int_{\sigma_j}^{\beta_l \varepsilon + \sigma_j} \mu_{j+1} Y_j(s) \, ds \geq \int_{\sigma_j}^{\beta_l \varepsilon + \sigma_j} \mu_{j+1} \gamma_d \alpha^d (s - \sigma_j)^d \, ds = \frac{\mu_{j+1} \gamma_d \alpha^d}{d+1} (\beta_l \varepsilon)^{d+1}. \quad (16)$$

By the second assumption in the theorem, we have  $\mu_{l+1} \gg 1/(\alpha^d \beta_l^{d+1})$ . Because of (2), we have  $\mu_{j+1} \gg 1/(\alpha^d \beta_l^{d+1})$ . It follows that the right-hand side of (16) tends to infinity as  $N \rightarrow \infty$ , which completes the proof.  $\square$

#### 4. Proofs of limit theorems for distances between mutations

For each  $j \geq 1$ , we define  $\sigma_j^{(2)}$  to be the time at which the second type  $j$  mutation occurs, i.e.

$$\sigma_j^{(2)} := \inf\{t > \sigma_j : (x, t) \in \Pi_j \text{ and } x \in \psi_{j-1}(t) \text{ for some } x \in \mathcal{T}\}.$$

Note that  $\sigma_j^{(2)}$  is defined to be the first time, after time  $\sigma_j$ , that a point of  $\Pi_j$  lands in a region of type  $j-1$  or higher. If this point lands in a region of type  $j-1$ , then a new type  $j$  ball will begin to grow. If this point lands in a region of type  $j$  or higher, then the evolution of the process is unaffected. Also recall from (1) that  $\kappa_j := (\mu_j \alpha^d)^{-1/(d+1)}$ .

##### 4.1. Proof of Theorem 5

We begin with an upper bound for the time between first mutations of consecutive types.

**Lemma 7.** *Assume  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . Suppose  $i$  and  $j$  are positive integers. Then, for each fixed  $t > 0$ , we have, for all sufficiently large  $N$ ,*

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > \kappa_i t) \leq \exp\left(-\frac{\gamma_d}{d+1} \cdot \frac{\mu_{j+1}}{\mu_i} t^{d+1}\right).$$

*Proof.* Using Lemma 1, we have

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > \kappa_i t) = \mathbb{E}\left[\exp\left(-\int_{\sigma_j}^{\sigma_j + \kappa_i t} \mu_{j+1} Y_j(s) \, ds\right)\right].$$

Because of  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and (2), for all sufficiently large  $N$  we have  $\kappa_i t < L/(2\alpha)$ . Thus,  $Y_j(s) \geq \gamma_d \alpha^d (s - \sigma_j)^d$  for  $s \in [\sigma_j, \sigma_j + \kappa_i t]$ . Then,

$$\begin{aligned} \mathbb{P}(\sigma_{j+1} - \sigma_j > \kappa_i t) &\leq \exp\left(-\int_{\sigma_j}^{\sigma_j + \kappa_i t} \mu_{j+1} \gamma_d \alpha^d (s - \sigma_j)^d \, ds\right) \\ &= \exp\left(-\int_0^{\kappa_i t} \mu_{j+1} \gamma_d (\alpha s)^d \, ds\right) \end{aligned}$$

$$\begin{aligned}
&= \exp\left(-\frac{\mu_{j+1}\gamma_d\alpha^d}{d+1}(\kappa_i t)^{d+1}\right) \\
&= \exp\left(-\frac{\gamma_d}{d+1} \cdot \frac{\mu_{j+1}}{\mu_i} t^{d+1}\right).
\end{aligned}$$

This completes the proof.  $\square$

By Lemma 7, when  $\mu_1 \gg \alpha/N^{(d+1)/d}$  the interarrival time  $\sigma_j - \sigma_{j-1}$  is at most the same order of magnitude as  $\kappa_j$ . Lemma 8 further shows that if, in addition,  $\mu_j \ll \mu_{j+1} \ll \mu_{j+2} \ll \dots$ , then mutations of type  $m > j$  appear on an even faster time scale.

**Lemma 8.** *Assume  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . Suppose  $j$  is a positive integer, and  $\mu_j \ll \mu_{j+1} \ll \mu_{j+2} \ll \dots$ . Then  $(\sigma_m - \sigma_j)/\kappa_j \rightarrow_p 0$  for every  $m > j$ .*

*Proof.* Using (2), we get

$$\frac{\sigma_m - \sigma_j}{\kappa_j} = \sum_{i=j}^{m-1} \frac{\sigma_{i+1} - \sigma_i}{\kappa_j} \lesssim \sum_{i=j}^{m-1} \frac{\sigma_{i+1} - \sigma_i}{\kappa_i},$$

so it suffices to show that  $(\sigma_{i+1} - \sigma_i)/\kappa_i \rightarrow_p 0$  for all  $i \in \{j, j+1, \dots, m-1\}$ . Let  $\varepsilon > 0$ . Using Lemma 7, we have, for all sufficiently large  $N$ ,

$$\mathbb{P}(\sigma_{i+1} - \sigma_i > \kappa_i \varepsilon) \leq \exp\left(-\frac{\gamma_d}{d+1} \cdot \frac{\mu_{i+1}}{\mu_i} \varepsilon^{d+1}\right).$$

Then,  $\mathbb{P}(\sigma_{i+1} - \sigma_i > \kappa_i \varepsilon) \rightarrow 0$  because  $\mu_{i+1}/\mu_i \rightarrow \infty$ , as desired.  $\square$

Next, we want to show that the balls from different mutation types become nested, as in Fig. 2. That is, for any  $i \geq 1$  and  $j > i$ , we have  $\mathbb{P}(\sigma_j < \sigma_i^{(2)}) \rightarrow 1$ , meaning that a type  $j$  mutation appears before a second type  $i$  mutation can appear. We first prove the case when  $i = 1$  in Lemma 9, assuming the same hypotheses as in Theorem 2.

**Lemma 9.** *Suppose (2) holds, and suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$ . Then:*

- (i) *For all  $t > 0$ ,  $\mathbb{P}(\sigma_2 + \kappa_2 t < \sigma_1^{(2)}) \rightarrow 1$ .*
- (ii) *For every  $j \geq 2$ ,  $\mathbb{P}(\sigma_j < \sigma_1^{(2)}) \rightarrow 1$ .*

*Proof.* To prove (i), let  $\varepsilon > 0$ . It was shown in the proof of Theorem 2 that  $N\mu_1(\sigma_2 - \sigma_1) \rightarrow_p 0$ . Also, the assumption  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$  implies  $N\mu_1\kappa_2 t \rightarrow 0$ . Thus,  $N\mu_1(\sigma_2 - \sigma_1) + N\mu_1\kappa_2 t \rightarrow_p 0$ . Therefore, for all sufficiently large  $N$ ,

$$\mathbb{P}(N\mu_1(\sigma_2 - \sigma_1) + N\mu_1\kappa_2 t < \varepsilon) > 1 - \varepsilon.$$

On the other hand, because  $N\mu_1(\sigma_1^{(2)} - \sigma_1)$  has an Exponential(1) distribution,

$$\mathbb{P}(N\mu_1(\sigma_1^{(2)} - \sigma_1) > \varepsilon) = e^{-\varepsilon} > 1 - \varepsilon.$$

Combining the above, we find

$$\mathbb{P}(\sigma_2 + \kappa_2 t < \sigma_1^{(2)}) \geq \mathbb{P}(N\mu_1(\sigma_2 - \sigma_1) + N\mu_1\kappa_2 t < \varepsilon < N\mu_1(\sigma_1^{(2)} - \sigma_1)) > 1 - 2\varepsilon.$$

This proves (i). For (ii), by the proof of Theorem 2 again, we have  $N\mu_1(\sigma_i - \sigma_{i-1}) \rightarrow_p 0$  for every  $i \geq 2$ . Thus,

$$N\mu_1(\sigma_j - \sigma_1) = \sum_{i=2}^j N\mu_1(\sigma_i - \sigma_{i-1}) \rightarrow_p 0,$$

and the rest of the proof is essentially the same as that of (i); we just replace  $N\mu_1(\sigma_2 - \sigma_1) + N\mu_1\kappa_2 t$  with  $N\mu_1(\sigma_j - \sigma_1)$ .  $\square$

In Lemma 10, we establish that  $\sigma_j^{(2)} - \sigma_j > \kappa_j \delta$  with high probability. Then, for  $k > j$ , since  $(\sigma_k - \sigma_j)/\kappa_j \rightarrow_p 0$  by Lemma 8, it will follow that  $\sigma_k - \sigma_j < \kappa_j \delta$  with high probability. It will then follow that  $\mathbb{P}(\sigma_k < \sigma_j^{(2)}) \rightarrow 1$ , which we show in Lemma 11.

**Lemma 10.** *Suppose (2) holds, and suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$ . Let  $i \geq 2$ . Define the events  $A := \{\sigma_i - \sigma_{i-1} < \kappa_i t\}$  and  $B := \{\sigma_i + \kappa_i \delta < \sigma_{i-1}^{(2)}\}$ . Then, for any fixed  $t > 0$  and  $\delta > 0$ , for sufficiently large  $N$  we have*

$$\mathbb{P}(\sigma_i^{(2)} - \sigma_i > \kappa_i \delta) \geq \exp\left(-\frac{\gamma_d}{d+1}[(t+\delta)^{d+1} - t^{d+1}]\right) - \mathbb{P}(A^c) - \mathbb{P}(B^c).$$

*Proof.* Reasoning as in the proof of Lemma 1, we have

$$\mathbb{P}(\sigma_i^{(2)} - \sigma_i > \kappa_i \delta) = \mathbb{E}\left[\exp\left(-\int_0^{\kappa_i \delta} \mu_i Y_{i-1}(s + \sigma_i) ds\right)\right]. \quad (17)$$

Because of  $\mu_1 \gg \alpha/N^{(d+1)/d}$  and (2), for sufficiently large  $N$  we have  $\kappa_i(t + \delta) < L/(2\alpha)$ . Thus, on the event  $A$ , we have  $\kappa_i \delta + (\sigma_i - \sigma_{i-1}) < L/(2\alpha)$ . Therefore, on the event  $A \cap B$ , for all  $s \in [0, \kappa_i \delta]$  we have, for sufficiently large  $N$ ,

$$Y_{i-1}(s + \sigma_i) = \gamma_d \alpha^d (s + \sigma_i - \sigma_{i-1})^d \leq \gamma_d \alpha^d (s + \kappa_i t)^d.$$

Thus, for sufficiently large  $N$ ,

$$\begin{aligned} \exp\left(-\int_0^{\kappa_i \delta} \mu_i Y_{i-1}(s + \sigma_i) ds\right) &\geq \exp\left(-\int_0^{\kappa_i \delta} \mu_i \gamma_d \alpha^d (s + \kappa_i t)^d ds\right) \mathbf{1}_{A \cap B} \\ &\geq \exp\left(-\int_0^{\kappa_i \delta} \mu_i \gamma_d \alpha^d (s + \kappa_i t)^d ds\right) - \mathbf{1}_{A^c} - \mathbf{1}_{B^c}. \end{aligned} \quad (18)$$

It follows from (17) and (18) that, for sufficiently large  $N$ ,

$$\begin{aligned} \mathbb{P}(\sigma_i^{(2)} - \sigma_i > \kappa_i \delta) &\geq \exp\left(-\int_0^{\kappa_i \delta} \mu_i \gamma_d \alpha^d (s + \kappa_i t)^d ds\right) - \mathbb{P}(A^c) - \mathbb{P}(B^c) \\ &= \exp\left(-\frac{\mu_i \gamma_d \alpha^d}{d+1}[(\kappa_i \delta + \kappa_i t)^{d+1} - (\kappa_i t)^{d+1}]\right) - \mathbb{P}(A^c) - \mathbb{P}(B^c) \\ &= \exp\left(-\frac{\gamma_d}{d+1}[(t+\delta)^{d+1} - t^{d+1}]\right) - \mathbb{P}(A^c) - \mathbb{P}(B^c), \end{aligned}$$

as claimed.  $\square$

In Lemma 11 we give sufficient conditions for  $\mathbb{P}(\sigma_j < \sigma_i^{(2)}) \rightarrow 1$  for  $j > i \geq 1$ , which implies that we obtain nested balls, as in Fig. 2.

**Lemma 11.** *Assume  $\mu_1 \gg \alpha/N^{(d+1)/d}$ ,  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$ , and  $\mu_j \ll \mu_{j+1}$  for  $j \geq 2$ . Then, for every  $i \geq 1$  and  $j > i$ ,*

$$\mathbb{P}(\sigma_j < \sigma_i^{(2)}) \rightarrow 1. \quad (19)$$

*Proof.* The result in (19) when  $i = 1$  was proved in part (ii) of Lemma 9. To establish the result when  $i \geq 2$ , we will show that, for all  $i \geq 2$  and all  $\varepsilon > 0$ , there exists  $\delta > 0$  such that, for sufficiently large  $N$ ,

$$\mathbb{P}(\sigma_{i-1}^{(2)} > \sigma_i + \kappa_i \delta) > 1 - \varepsilon, \quad (20)$$

$$\mathbb{P}(\sigma_i^{(2)} > \sigma_i + \kappa_i \delta) > 1 - \varepsilon. \quad (21)$$

Assume for now that  $i \geq 2$ , and that (20) and (21) hold. Let  $\varepsilon > 0$ , and choose  $\delta > 0$  to satisfy (21). Lemma 8 implies that if  $j > i$  then  $\mathbb{P}(\sigma_j - \sigma_i < \kappa_i \delta) > 1 - \varepsilon$  for sufficiently large  $N$ . It follows that

$$\mathbb{P}(\sigma_j < \sigma_i^{(2)}) \geq \mathbb{P}(\sigma_i^{(2)} > \sigma_i + \kappa_i \delta > \sigma_j) > 1 - 2\varepsilon$$

for sufficiently large  $N$ , which implies (19).

It remains to prove (20) and (21); we proceed by induction. The result (20) when  $i = 2$  is part (i) of Lemma 9. Therefore, it suffices to show that (20) implies (21), and that if (21) holds for some  $i \geq 2$ , then (20) holds with  $i + 1$  in place of  $i$ .

To deduce (21) from (20), we first let  $\varepsilon > 0$  and use Lemma 7 to choose  $t > 0$  large enough that

$$\mathbb{P}(\sigma_i - \sigma_{i-1} > \kappa_i t) \leq \exp\left(-\frac{\gamma_d}{d+1} t^{d+1}\right) < \frac{\varepsilon}{3}.$$

Then choose  $\delta > 0$  small enough that (20) holds with  $\varepsilon/3$  in place of  $\varepsilon$  for sufficiently large  $N$ , and

$$\exp\left(-\frac{\gamma_d}{d+1} [(t + \delta)^{d+1} - t^{d+1}]\right) > 1 - \frac{\varepsilon}{3}.$$

It now follows from Lemma 10, that for sufficiently large  $N$ ,

$$\mathbb{P}(\sigma_i^{(2)} > \sigma_i + \kappa_i \delta) > 1 - \frac{\varepsilon}{3} - \frac{\varepsilon}{3} - \frac{\varepsilon}{3} = 1 - \varepsilon,$$

so (21) holds.

Next, suppose (21) holds for some  $i \geq 2$ . Let  $\varepsilon > 0$ . By (21), there exists  $\delta > 0$  such that, for sufficiently large  $N$ ,

$$\mathbb{P}(\sigma_i^{(2)} > \sigma_i + \kappa_i \delta) > 1 - \frac{\varepsilon}{2}.$$

By Lemma 8 and the fact that  $\mu_i \ll \mu_{i+1}$ , we have  $(\sigma_{i+1} - \sigma_i)/\kappa_i + \delta(\kappa_{i+1}/\kappa_i) \rightarrow_p 0$ . Thus, for sufficiently large  $N$ ,

$$\mathbb{P}\left(\frac{\sigma_{i+1} - \sigma_i}{\kappa_i} + \frac{\kappa_{i+1} \delta}{\kappa_i} < \delta\right) > 1 - \frac{\varepsilon}{2},$$

and therefore

$$\mathbb{P}(\sigma_i^{(2)} > \sigma_{i+1} + \kappa_{i+1}\delta) \geq \mathbb{P}(\sigma_i^{(2)} - \sigma_i > \kappa_i\delta > \sigma_{i+1} + \kappa_{i+1}\delta - \sigma_i) > 1 - \varepsilon,$$

which is (20) with  $i + 1$  in place of  $i$ .  $\square$

We now find the limiting distribution of distances between mutations of consecutive types.

**Lemma 12.** *Suppose  $\mu_1 \gg \alpha/N^{(d+1)/d}$ ,  $\mu_2 \gg (N\mu_1)^{d+1}/\alpha^d$ , and  $\mu_j \ll \mu_{j+1}$  for  $j \geq 2$ . Then, for all  $s > 0$ ,*

$$\mathbb{P}\left(\frac{D_{j+1}}{\alpha\kappa_{j+1}} \leq s\right) \rightarrow \int_0^\infty \gamma_d(t \wedge s)^d \exp\left(-\frac{\gamma_d t^{d+1}}{d+1}\right) dt. \quad (22)$$

*Proof.* Define the event  $A := \bigcap_{i=1}^j \{\sigma_{j+1} < \sigma_i^{(2)}\}$ . On the event  $A$ , the first type  $j + 1$  mutation appears before the second mutation of any type  $i \in \{1, \dots, j\}$ . By Lemma 11, we have  $\mathbb{P}(A) \rightarrow 1$ . As a result, it will be sufficient for us to consider a modified version of our process in which, for  $i \in \{1, \dots, j\}$ , only the first type  $i$  mutation is permitted to occur. Note that this modified process can be constructed from the same sequence of independent Poisson processes  $(\Pi_i)_{i=1}^\infty$  as the original process. However, in the modified process, all points of  $\Pi_i$  after time  $\sigma_i$  are disregarded. On the event  $A$ , the first  $j + 1$  mutations will occur at exactly the same times and locations in the original process as in the modified process. Therefore, because  $\mathbb{P}(A) \rightarrow 1$ , it suffices to prove (22) for this modified process. For the rest of the proof we will work with this modified process, which makes exact calculations possible.

Let  $K \in (s, \infty)$  be a constant which does not depend on  $N$ . Our assumptions imply that  $\mu_{j+1} \gg \mu_1 \gg \alpha/N^{(d+1)/d}$ . Thus, there is an  $N_K$  such that for  $N \geq N_K$  we have  $\kappa_{j+1}t < L/(2\alpha)$  for all  $t \in [0, K]$ . It follows that  $Y_j(s) = \gamma_d \alpha^d (s - \sigma_j)^d$  for  $s \in [\sigma_j, \sigma_j + \kappa_{j+1}K]$ . Therefore, reasoning as in the proof of Lemma 7, we get

$$\mathbb{P}(\sigma_{j+1} - \sigma_j > \kappa_{j+1}t) = \exp\left(-\frac{\gamma_d}{d+1} t^{d+1}\right). \quad (23)$$

It follows that, for  $N \geq N_K$ , the probability density of  $(\sigma_{j+1} - \sigma_j)/\kappa_{j+1}$  restricted to  $[0, K]$  is

$$f(t) := \gamma_d t^d \exp\left(-\frac{\gamma_d}{d+1} t^{d+1}\right).$$

For  $N \geq N_K$  and  $t \in [0, K]$ , conditional on the event  $\{\sigma_{j+1} - \sigma_j = \kappa_{j+1}t\}$ , the location of the first type  $j + 1$  mutation is a uniformly random point on a  $d$ -dimensional ball of radius  $\alpha\kappa_{j+1}t$ , which means

$$\mathbb{P}\left(\frac{D_{j+1}}{\alpha\kappa_{j+1}} \leq s \mid \frac{\sigma_{j+1} - \sigma_j}{\kappa_{j+1}} = t\right) = \mathbf{1}_{\{s > t\}} + \frac{\gamma_d (\alpha\kappa_{j+1}s)^d}{\gamma_d (\alpha\kappa_{j+1}t)^d} \mathbf{1}_{\{s \leq t\}} = \mathbf{1}_{\{s > t\}} + \frac{s^d}{t^d} \mathbf{1}_{\{s \leq t\}}.$$

It follows that, for  $N \geq N_K$ ,

$$\begin{aligned} \mathbb{P}\left(\frac{D_{j+1}}{\alpha\kappa_{j+1}} \leq s\right) &= \int_0^K f(t) \left( \mathbf{1}_{\{s > t\}} + \frac{s^d}{t^d} \mathbf{1}_{\{s \leq t\}} \right) dt + \mathbb{P}\left(\frac{D_{j+1}}{\alpha\kappa_{j+1}} \leq s, \frac{\sigma_{j+1} - \sigma_j}{\kappa_{j+1}} > K\right) \\ &= \int_0^K \gamma_d (t \wedge s)^d \exp\left(-\frac{\gamma_d t^{d+1}}{d+1}\right) dt + \mathbb{P}\left(\frac{D_{j+1}}{\alpha\kappa_{j+1}} \leq s, \frac{\sigma_{j+1} - \sigma_j}{\kappa_{j+1}} > K\right). \end{aligned}$$

Because (23) implies that

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P} \left( \frac{D_{j+1}}{\alpha \kappa_{j+1}} \leq s, \frac{\sigma_{j+1} - \sigma_j}{\kappa_{j+1}} > K \right) = 0,$$

the result (22) follows by letting  $N \rightarrow \infty$  and then  $K \rightarrow \infty$ .  $\square$

*Proof of Theorem 5.* Lemma 12 proves the case when  $k = j + 1$ , so assume that  $k \geq j + 2$ . The triangle inequality implies that

$$D_{j+1} - (D_{j+2} + \cdots + D_k) \leq D_{j,k} \leq D_{j+1} + (D_{j+2} + \cdots + D_k).$$

Suppose  $j + 2 \leq i \leq k$ . We know from Lemma 12 that  $D_i/(\alpha \kappa_i)$  converges in distribution to a non-degenerate random variable as  $N \rightarrow \infty$ . Because  $\kappa_{j+1}/\kappa_i \rightarrow \infty$  by the assumption in (9), it follows that  $D_i/(\alpha \kappa_{j+1}) \rightarrow_p 0$ . Therefore,

$$(D_{j+2} + \cdots + D_k)/(\alpha \kappa_{j+1}) \rightarrow_p 0.$$

Thus, Theorem 5 follows from Lemma 12 and Slutsky's theorem.  $\square$

#### 4.2. Proof of Theorem 6

Having found a limiting distribution for distances between mutations in the setting of Theorem 2, we now prove a similar result in the setting of Theorem 4, where once the first type  $l$  mutation appears, all subsequent mutations appear in nested balls.

We begin with a result that bounds  $\sigma_l^{(2)} - \sigma_l$  away from zero with high probability, on the time scale  $\beta_l$ .

**Lemma 13.** *Assume the same hypotheses as Theorem 4. Then, for all  $\varepsilon > 0$ , there is  $r > 0$  such that  $\liminf_{N \rightarrow \infty} \mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) > 1 - \varepsilon$ .*

*Proof.* Let  $\varepsilon > 0$ . Using Theorem 3, choose a large  $t > 0$  so that

$$\lim_{N \rightarrow \infty} \mathbb{P}(\sigma_l \leq \beta_l t) > 1 - \frac{\varepsilon}{2}. \quad (24)$$

Now set, as in the proof of Theorem 3,

$$g(s) := \frac{\gamma_d^{l-1} (d!)^{l-1} s^{(l-1)(d+1)}}{((l-1)(d+1))!}.$$

It is clear that we can choose a small  $r > 0$  so that

$$\exp \left( - \int_t^{t+r} g(s) \, ds \right) > 1 - \frac{\varepsilon}{2}. \quad (25)$$

Having chosen  $t > 0$  and  $r > 0$ , choose  $\delta > 0$  so that  $[t, t+r] \subseteq [\delta, \delta^{-1}]$ . Then, for any  $\lambda > 0$ , define, as in Lemma 4, the event

$$B := B_N^{l-1} \left( \delta, \lambda, g, \frac{1}{\mu_l \beta_l} \right) = \left\{ \frac{g(u)(1-\lambda)}{\beta_l \mu_l} \leq Y_{l-1}(\beta_l u) \leq \frac{g(u)(1+\lambda)}{\beta_l \mu_l}, \text{ for all } u \in [\delta, \delta^{-1}] \right\}.$$

Now we calculate

$$\begin{aligned}\mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) &= \mathbb{E} \left[ \exp \left( - \int_{\sigma_l}^{\sigma_l + \beta_l r} \mu_l Y_{l-1}(s) \, ds \right) \right] \\ &\geq \mathbb{E} \left[ \exp \left( - \int_{\sigma_l}^{\sigma_l + \beta_l r} \mu_l Y_{l-1}(s) \, ds \right) \mathbf{1}_{\{\sigma_l \leq \beta_l t\}} \mathbf{1}_B \right].\end{aligned}\quad (26)$$

Because  $Y_{l-1}(s)$  is monotone increasing in  $s$ , on the event  $\{\sigma_l \leq \beta_l t\} \cap B$  we have

$$\int_{\sigma_l}^{\sigma_l + \beta_l r} \mu_l Y_{l-1}(s) \, ds \leq \int_{\beta_l t}^{\beta_l(t+r)} \mu_l Y_{l-1}(s) \, ds \leq (1 + \lambda) \int_t^{t+r} g(s) \, ds.$$

Using the above and (26), we have

$$\mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) \geq \exp \left( -(1 + \lambda) \int_t^{t+r} g(s) \, ds \right) \mathbb{P}(\{\sigma_l \leq \beta_l t\} \cap B).$$

Now take  $N \rightarrow \infty$ . Using that  $\mathbb{P}(B) \rightarrow 1$  as shown in the proof of Theorem 3, and using (24), we have

$$\begin{aligned}\liminf_{N \rightarrow \infty} \mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) &\geq \exp \left( -(1 + \lambda) \int_t^{t+r} g(s) \, ds \right) \cdot \liminf_{N \rightarrow \infty} \mathbb{P}(\sigma_l \leq \beta_l t) \\ &> \exp \left( -(1 + \lambda) \int_t^{t+r} g(s) \, ds \right) \left( 1 - \frac{\varepsilon}{2} \right).\end{aligned}$$

Since  $\lambda > 0$  is arbitrary, (25) implies  $\liminf_{N \rightarrow \infty} \mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) > (1 - \varepsilon/2)^2 > 1 - \varepsilon$ , completing the proof.  $\square$

Using Lemma 13, we prove an analog of Lemma 9 in the setting of Theorem 4.

**Lemma 14.** *Assume the same hypotheses as Theorem 4. Then:*

- (i) *For all  $t > 0$ ,  $\mathbb{P}(\sigma_{l+1} + \kappa_{l+1} t < \sigma_l^{(2)}) \rightarrow 1$ .*
- (ii) *For every  $k \geq l + 1$ ,  $\mathbb{P}(\sigma_k < \sigma_l^{(2)}) \rightarrow 1$ .*

*Proof.* Let  $\varepsilon > 0$ . Lemma 13 implies that there is  $r > 0$  such that, for sufficiently large  $N$ ,

$$\mathbb{P}(\sigma_l^{(2)} - \sigma_l > \beta_l r) > 1 - \varepsilon. \quad (27)$$

Now note that  $(\sigma_{l+1} - \sigma_l)/\beta_l \rightarrow_p 0$  by (15) in the proof of Theorem 4. Also, our assumption that  $\mu_{l+1} \gg 1/(\alpha^d \beta_l^{d+1})$  is equivalent to  $\kappa_{l+1} \ll \beta_l$ . It follows that, for sufficiently large  $N$ ,  $\mathbb{P}((\sigma_{l+1} - \sigma_l + \kappa_{l+1} t)/\beta_l < r) > 1 - \varepsilon$ . This estimate along with (27) imply that  $\mathbb{P}(\sigma_{l+1} + \kappa_{l+1} t < \sigma_l^{(2)}) > 1 - 2\varepsilon$  for sufficiently large  $N$ . This proves the first statement. The second statement is proved similarly, using instead that  $(\sigma_k - \sigma_l)/\beta_l \rightarrow_p 0$  by (15).  $\square$

At this point, we have proved that, for  $k > l$ , the first type  $k$  mutation occurs before  $\sigma_l^{(2)}$  with probability tending to 1 as  $N \rightarrow \infty$ . This implies that in the setting of Theorem 4, we can

disregard the type  $1, \dots, l-1$  mutations and regard the first type  $l$  mutation as the first type 1 mutation, and then prove Theorem 6 by following the same argument used to prove Theorem 5.

*Proof of Theorem 6.* Relabel the type  $l, l+1, l+2, \dots$  mutations as type 1, 2, 3, … mutations, and repeat the arguments in Lemmas 7–12 and in the proof of Theorem 5. The only difference is that we have to apply Lemma 14 instead of Lemma 9. Note that type  $l$  mutations do not appear at the same rate as type 1 mutations, so we needed a different technique to establish  $\mathbb{P}(\sigma_j < \sigma_l^{(2)}) \rightarrow 1$  for  $j > l$ .  $\square$

### Acknowledgements

The authors thank Jasmine Foo for suggesting the problem of looking at the case of increasing mutation rates, and bringing to their attention the references [16, 22].

### Funding information

JS was supported in part by NSF Grant DMS-1707953.

### Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

### References

- [1] ARMITAGE, P. AND DOLL, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Brit. J. Cancer* **8**, 1–12.
- [2] BORODIN, A., FERRARI, P. L. AND SASAMOTO, T. (2008). Large time asymptotics of growth models on space-like paths II: PNG and parallel TASEP. *Commun. Math. Phys.* **283**, 417–449.
- [3] BRAMSON, M. AND GRIFFEATH, D. (1980). On the Williams–Bjerknes tumour growth model II. *Math. Proc. Camb. Phil. Soc.* **88**, 339–357.
- [4] BRAMSON, M. AND GRIFFEATH, D. (1981). On the Williams–Bjerknes tumour growth model I. *Ann. Prob.* **9**, 173–185.
- [5] DURRETT, R., FOO, J. AND LEDER, K. (2016). Spatial Moran models II: Cancer initiation in spatially structured tissue. *J. Math. Biol.* **72**, 1369–1400.
- [6] DURRETT, R. AND MAYBERRY, J. (2011). Traveling waves of selective sweeps. *Ann. Appl. Prob.* **21**, 699–744.
- [7] DURRETT, R. AND MOSELEY, S. (2010). Evolution of resistance and progression to disease during clonal expansion of cancer. *Theoret. Pop. Biol.* **77**, 42–48.
- [8] DURRETT, R. AND MOSELEY, S. (2015). Spatial Moran models I: Stochastic tunneling in the neutral case. *Ann. Appl. Prob.* **25**, 104–115.
- [9] DURRETT, R., SCHMIDT, D. AND SCHWEINSBERG, J. (2009). A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Prob.* **19**, 676–718.
- [10] FOO, J., LEDER, K. AND RYSER, M. D. (2014). Multifocality and recurrence risk: A quantitative model of field cancerization. *J. Theoret. Biol.* **355**, 170–184.
- [11] FOO, J., LEDER, K. AND SCHWEINSBERG, J. (2020). Mutation timing in a spatial model of evolution. *Stoch. Process. Appl.* **130**, 6388–6413.
- [12] IWASA, Y., MICHOR, F., KOMAROVA, N. L. AND NOWAK, M. A. (2005). Population genetics of tumor suppressor genes. *J. Theoret. Biol.* **233**, 15–23.
- [13] IWASA, Y., MICHOR, F. AND NOWAK, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579.
- [14] KOMAROVA, N. L. (2006). Spatial stochastic models for cancer initiation and progression. *Bull. Math. Biol.* **68**, 1573–1599.
- [15] KOMAROVA, N. L., SENGUPTA, A. AND NOWAK, M. A. (2003). Mutation-selection networks of cancer initiation: Tumor suppressor genes and chromosomal instability. *J. Theoret. Biol.* **223**, 433–450.
- [16] LOEB, K. R. AND LOEB, L. A. (2000). Significance of multiple mutations in cancer. *Carcinogenesis* **21**, 379–385.

- [17] MOOLGAVKAR, S. H. AND LUEBECK, G. (1990). Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. *Risk Analysis* **10**, 323–341.
- [18] MOOLGAVKAR, S. H. AND LUEBECK, E. G. (1992). Multistage carcinogenesis: Population-based model for colon cancer. *J. Nat. Cancer Inst.* **18**, 610–618.
- [19] PRÄHOFER, M. AND SPOHN, H. (2000). Statistical self-similarity of one-dimensional growth processes. *Physica A* **279**, 342–352.
- [20] PRÄHOFER, M. AND SPOHN, H. (2000). Universal distributions for growth processes in  $1 + 1$  dimensions and random matrices. *Phys. Rev. Lett.* **84**, 4882–4885.
- [21] PRÄHOFER, M. AND SPOHN, H. (2002). Scale invariance of the PNG droplet and the Airy process. *J. Stat. Phys.* **108**, 1071–1106.
- [22] PRINDLE, M. J., FOX, E. J. AND LOEB, L. A. (2010). The mutator phenotype in cancer: Molecular mechanisms and targeting strategies. *Curr. Drug Targets* **11**, 1296–1303.
- [23] RYSER, M. D., LEE, W. T., READY, N. E., LEDER, K. Z. AND FOO, J. (2016). Quantifying the dynamics of field cancerization in tobacco-related head and neck cancer: A multiscale modeling approach. *Cancer Res.* **76**, 7078–7088.
- [24] SCHWEINSBERG, J. (2008). Waiting for  $m$  mutations. *Electron. J. Prob.* **13**, 1442–1478.
- [25] WILLIAMS, T. AND BJERKNES, R. (1972). Stochastic model for abnormal clone spread through epithelial basal layer. *Nature* **236**, 19–21.