

## Searching circular DNA strands

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2007 J. Phys.: Condens. Matter 19 065140

(<http://iopscience.iop.org/0953-8984/19/6/065140>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

### Download details:

IP Address: 129.252.86.83

The article was downloaded on 28/05/2010 at 16:04

Please note that [terms and conditions apply](#).

## Searching circular DNA strands

Iddo Eliazar<sup>1,2</sup>, Tal Koren<sup>2</sup> and Joseph Klafter<sup>2</sup>

<sup>1</sup> Department of Technology Management, Holon Institute of Technology, Holon 58102, Israel

<sup>2</sup> School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Received 8 September 2006, in final form 6 October 2006

Published 22 January 2007

Online at [stacks.iop.org/JPhysCM/19/065140](http://stacks.iop.org/JPhysCM/19/065140)

### Abstract

We introduce and explore a model of an ensemble of enzymes searching, in parallel, a circular DNA strand for a target site. The enzymes performing the search combine local scanning—conducted by a 1D motion along the strand—and random relocations on the strand—conducted via a confined motion in the medium containing the strand. Both the local scan mechanism and the relocation mechanism are considered general.

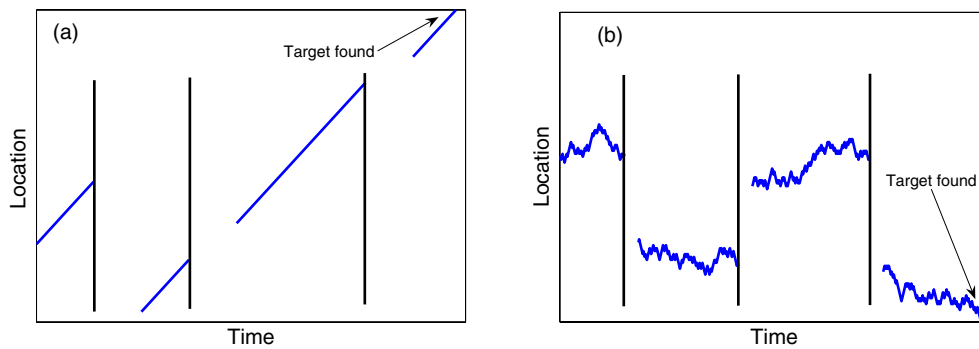
The search durations are analysed, and their limiting probability distributions—for long DNA strands—are obtained in closed form. The results obtained (i) encompass the cases of single, parallel and massively parallel searches, taking place in the presence of either finite-mean or heavy-tailed relocation times, (ii) are applicable to a wide spectrum of local scan mechanisms including linear, Brownian, selfsimilar, and sub-diffusive motions, (iii) provide a quantitative theoretical justification for the necessity of the relocation mechanism, and (iv) facilitate the derivation of optimal relocation strategies.

(Some figures in this article are in colour only in the electronic version)

### 1. Introduction

The biological mechanism of DNA-searching has attracted major interest in recent years from both experimentalists—investigating the mechanism *in vivo* and *in vitro* [1–10]—and theoreticians—constructing mathematical models of the mechanism [11–16].

DNA-searching is usually performed in parallel by an ensemble of enzymes operating simultaneously. In a process referred to as '*facilitated diffusion*', the enzymes alternate between a 1D motion along the DNA strand and a confined motion within the cellular medium containing the DNA strand [2, 6–11]. An enzyme conducting a 1D motion along the DNA strand occasionally disengages it, and commences the confined motion. The confined motion—after a period of time henceforth called the *relocation time*—'lands' the enzyme on a new location on the DNA strand, from which the enzyme resumes the 1D motion (the new location may or may not be correlated with the enzyme's last 'disengagement location' [6, 16]).



**Figure 1.** An illustration of the facilitated diffusion mechanism. The  $x$ -axis denotes time; the  $y$ -axis denotes the enzyme's location along the strand. (a) Local scanning conducted by a linear motion. (b) Local scanning conducted by a random motion. The vertical bars represent the disengagement epochs; the horizontal gaps following the vertical bars represent the relocation durations.

The goal of the search is to find a target site on the DNA strand. In principle, the length of the target site is far smaller than the length of the embedding DNA strand. The variable of interest is the enzyme-ensemble *search duration*: the time elapsing from search initiation till the 'first discovery' of the target site by one of the searching enzymes.

The facilitated diffusion mechanism combines two key features: (i) *local scanning*—conducted by the 1D motion along the DNA strand—and (ii) *relocation*—caused by the confined motion in the surrounding medium. See figure 1 for a schematic illustration of the facilitated diffusion mechanism. This mechanism turns out to be highly effective and efficient [2, 6–14]. Reference [15] provides a careful study of DNA searching in which the parameters of the local scanning and the relocation are externally controlled.

In this paper we introduce and explore a stochastic model of *parallel* and *massively parallel* searching of *circular DNA strands* (including, in particular, DNA plasmids [17]). The parallel search is performed by an ensemble of enzymes operating simultaneously and combining *local scanning* and *relocation*. The search is rendered massively parallel when the size of the agent ensemble is large.

The model proposed is fairly general and robust, yet tractable and amenable to mathematical and statistical analysis. It goes beyond existing theoretical models (i) by considering *arbitrary* local-scanning and relocation mechanisms—including *anomalous* mechanisms—and (ii) by yielding closed-form results for the *limiting distributions* of the search durations—whereas the existing models focus mainly on the *mean* search durations. From a more abstract perspective, the model proposed falls into the category of general search models [18–21].

The paper is organized as follows. We begin, in section 2, with the formulation of the model and with the statistical analysis of the search duration in the case of a single searching enzyme. In section 3 long DNA strands are considered, and probabilistic limit laws for the search durations are obtained in the case of single, parallel, and massively parallel searches. In section 4 the role of the relocation mechanism is investigated, and the question 'is relocation beneficial?' is formulated and answered quantitatively. In section 5 three specific classes of local scanning mechanisms are examined: linear motions, Brownian motions, and selfsimilar motions (this last class includes fractional Brownian motions and fractional Lévy motions). Section 6 modifies the model proposed by incorporating a general 'halting mechanism' affecting the enzymes' local scanning. This modification accommodates, in particular, the class

of sub-diffusive local scanning mechanisms. We conclude, in section 7, with an investigation of massively parallel searches of stationary DNA–enzyme systems.

#### *A note regarding notation and nomenclature*

Throughout the following,  $\mathbf{P}(\cdot)$  = Probability;  $\mathbf{E}[\cdot]$  = Expectation; and  $\tilde{X}(\theta) := \mathbf{E}[\exp\{-\theta X\}]$  ( $\theta \geq 0$ ) denotes the Laplace transform of a non-negative valued random variable  $X$ .

A non-negative random variable  $X$  is said to be *heavy tailed* of order  $\alpha$  ( $0 < \alpha < 1$ ) if it is asymptotically Pareto with exponent  $\alpha$ :  $\mathbf{P}(X > x) \sim a/x^\alpha$ , as  $x \rightarrow \infty$  (where  $a$  is an arbitrary positive constant).

## 2. Modelling and analysis

Consider a circular DNA strand composed of  $l + n$  base-pairs. The strand consists of a target site  $l$  base-pairs long, and its remainder is  $n$  base-pairs long. A *single enzyme* searches the strand for the target site, as follows.

The enzyme begins its search from a random initial position along the strand. If, by chance, the initial position is within the target site then the search is concluded. Otherwise, (i) the enzyme initiates a local scan and (ii) an exponential timer is set. If the local scan traces the target site before the timer expires then the search is concluded. Otherwise, upon the timer's expiration, the enzyme relocates to a new random position along the strand, and the search begins anew.

We introduce the following random variables.

- $S$  := the duration of a local scan, initiated from a random position along the  $n$  ‘non-target’ base-pairs of the strand.
- $R$  := the relocation time.
- $T$  := the overall search duration.

Moreover, we set  $\lambda$  to denote the relocation rate (i.e. the rate of the exponential timer), and set  $p := n/(l + n)$  to denote the fraction of the ‘non-target’ base-pairs.

Both the local-scan mechanism and the relocation mechanism are considered general—implying, in turn, that  $S$  and  $R$  are arbitrarily distributed non-negative valued random variables.

The following proposition gives the precise mathematical connection, in the Laplace domain, between the distributions of the ‘input’ random variables  $S$  and  $R$  and the distribution of the ‘output’ random variable  $T$ .

**Proposition 1.** *The Laplace transform of the overall search duration  $T$  is given by*

$$\tilde{T}(\theta) = \frac{1 - p(1 - \tilde{S}(\lambda + \theta))}{1 - p(1 - \tilde{S}(\lambda + \theta)) \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)} \quad (1)$$

( $\theta \geq 0$ ).

As we shall demonstrate in the following, there is a deep and marked difference between the case of *finite-mean* relocation times ( $\mathbf{E}[R] < \infty$ ) and the case of *infinite-mean* relocation times ( $\mathbf{E}[R] = \infty$ ). In the latter case we shall henceforth assume that the relocation time  $R$  is *heavy tailed*. Proposition 1 implies the following corollaries.

**Finite-mean relocation times.** *If the relocation time  $R$  possesses a finite mean then so does the overall search duration  $T$ , and*

$$E[T] = \frac{p(1 - \tilde{S}(\lambda))}{1 - p(1 - \tilde{S}(\lambda))} \cdot \left( E[R] + \frac{1}{\lambda} \right). \quad (2)$$

Hence, the mean overall search duration  $E[T]$  is an affine function of the mean relocation time  $E[R]$ . We emphasize that the scanning duration  $S$  is not required to possess a finite mean in order to assure that the overall search duration  $T$  possesses a finite mean.

**Infinite-mean relocation times.** *If the relocation time  $R$  is heavy tailed of order  $\alpha$  ( $0 < \alpha < 1$ ) then so is the overall search duration  $T$ , and*

$$P(T > t) \underset{t \rightarrow \infty}{\sim} \frac{p(1 - \tilde{S}(\lambda))}{1 - p(1 - \tilde{S}(\lambda))} \cdot P(R > t). \quad (3)$$

Hence, the overall search duration's survival probability  $P(T > t)$  is, asymptotically, a linear function of the relocation time's survival probability  $P(R > t)$ . Heavy-tailed relocation durations can be considered to be induced by anomalous confined motions [22].

The proofs of proposition 1 and its corollaries are given in appendix A.

### 3. Searching long DNA strands

Often, the length of the DNA strand is considerably larger than the length of its target site: namely,  $n \gg l$ . In such cases it is of interest to study the asymptotic behaviour, as  $n \rightarrow \infty$ , of the overall search duration. To this end we set  $(T_n, S_n, p_n)$  to be the random variables/parameter corresponding to a DNA strand with  $n$  non-target base-pairs (the relocation time  $R$  and the relocation rate  $\lambda$  are independent of  $n$ ).

The following technical condition is required to hold:

$$\lim_{n \rightarrow \infty} n \cdot \tilde{S}_n(\lambda + \delta_n) = \psi(\lambda) \quad (4)$$

for all  $\lambda > 0$  and all non-negative valued sequences  $\{\delta_n\}_{n=1}^{\infty}$  decaying to zero. Namely, it is required that the limit appearing in the left-hand side of equation (4) exists. The limit  $\psi(\lambda)$  shall henceforth be referred to as the 'scan function'.

The asymptotic analysis of the overall search durations  $T_n$ , in the limit  $n \rightarrow \infty$ , is presented in section 3.1. The results obtained are then generalized to the case of parallel searches (section 3.2) and massively parallel searches (section 3.3)—conducted by a multitude of enzymes searching the DNA strand in parallel. Examples and analysis of scanning mechanisms satisfying the technical condition of equation (4) are presented in sections 5 and 6 below.

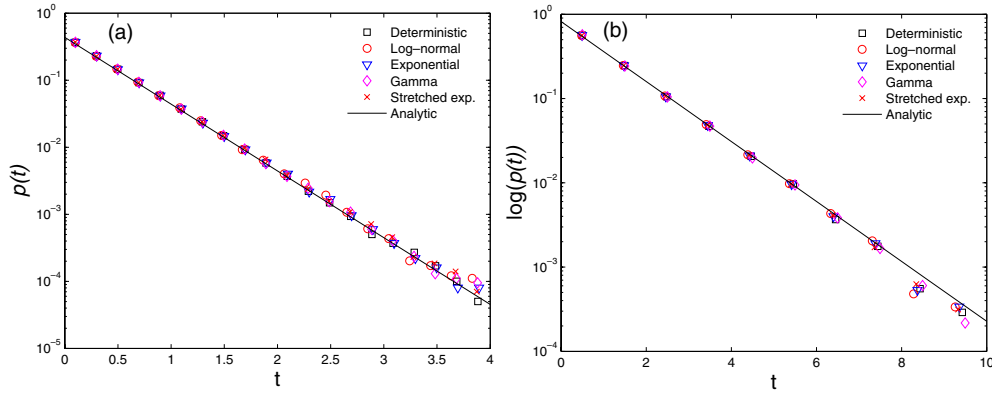
#### 3.1. Asymptotic analysis

The asymptotic behaviour, as  $n \rightarrow \infty$ , of the overall search durations  $T_n$  is highly contingent on whether or not the relocation time possesses a finite mean.

**Proposition 2 (Finite-mean relocation times).** *If the relocation time  $R$  possesses a finite mean then the scaled random variables  $T_n/n$  converge, in law, as  $n \rightarrow \infty$ , to a limiting random variable  $T_\infty$  which is exponentially distributed:*

$$P(T_\infty > t) = \exp \left\{ - \frac{l + \psi(\lambda)}{E[R] + 1/\lambda} \cdot t \right\} \quad (5)$$

( $t \geq 0$ ).



**Figure 2.** Simulations of a single enzyme searching a circular DNA strand of length  $n + l = 2000$ . (a) Local scanning conducted by a linear motion with velocity  $v = 5$ ; the length of the target site is  $l = 5$ ; number of simulations  $2 \times 10^5$ . (b) Local scanning conducted by a Brownian motion with diffusion parameter  $D = 1$ ; the length of the target site is  $l = 10$ ; number of simulations  $5 \times 10^4$ . The empirical probability density function  $p(t)$  of the scaled overall search durations is depicted on a logarithmic plot:  $\log(p(t))$  versus  $t$ . The simulations were carried out for finite-mean relocation times, with mean  $\mathbf{E}[R] = 10$ , drawn from various probability laws: deterministic; log-normal; exponential; gamma; and stretched exponential. The solid line depicts the theoretical exponential limit predicted by proposition 2—to which the simulations closely fit. The simulations also demonstrate the phenomena of *universality* with respect to the relocation mechanism.

**Proposition 3 (Infinite-mean relocation times).** *If the relocation time  $R$  is heavy tailed of order  $\alpha$  ( $0 < \alpha < 1$ ) then the scaled random variables  $T_n/n^{1/\alpha}$  converge, in law, as  $n \rightarrow \infty$ , to a limiting random variable  $T_\infty$  which is heavy tailed of order  $\alpha$ :*

$$P(T_\infty > t) \underset{t \rightarrow \infty}{\sim} \frac{1}{l + \psi(\lambda)} \cdot P(R > t). \tag{6}$$

In fact, in the infinite-mean case the limiting random variable  $T_\infty$  admits the multiplicative probabilistic representation  $T_\infty = \mathcal{L} \cdot \mathcal{E}^{1/\alpha}$ , where  $\mathcal{L}$  and  $\mathcal{E}$  are independent random variables and (i)  $\mathcal{L}$  is Lévy distributed with exponent  $\alpha$  and (ii)  $\mathcal{E}$  is exponentially distributed with rate  $l + \psi(\lambda)$ .

Two remarks are in place.

- The finite-mean and infinite-mean cases are markedly different—requiring different scalings ( $n$  in the finite-mean case, and  $n^{1/\alpha}$  in the infinite-mean case), and yielding different limiting distributions (exponential in the finite-mean case, and heavy tailed in the infinite-mean case).
- The limiting distribution, in the finite-mean case, is *universal* with respect to the relocation mechanism—being contingent on the mean relocation time  $\mathbf{E}[R]$  alone (rather than on the *distribution* of the relocation time). Such universality fails to hold in the infinite-mean case—where the limiting distribution is contingent on the survival probability of the relocation time.

See figure 2 for computer simulations illustrating the universality phenomena taking place in the case of finite-mean relocation times. The proofs of propositions 2 and 3 are given in appendix A.

### 3.2. Parallel searching

We now generalize the asymptotic results of section 3.1 to the case of an *ensemble of enzymes* searching a long circular DNA strand, *in parallel*, for the target site.

Consider a DNA strand, with  $n$  non-target base-pairs, being searched by an ensemble of  $m$  enzymes operating *simultaneously* and *independently*. Assume that once a ‘search command’ is given it takes each enzyme (independently of all other enzymes) a random time of length  $R_0$  to initialize its search procedure. Thus, the time it takes an enzyme to trace the target site is  $R_0 + T_n$  ( $T_n$  being the overall search duration defined and analysed above).

Let  $T_n^m$  denote the enzyme-ensemble search duration—i.e. the time elapsing till the ‘first discovery’ of the target site. Since the enzymes operate independently, the random variable  $T_n^m$  is the *minimum* of  $m$  independent copies of the random variable  $R_0 + T_n$ . This observation implies that

$$\mathbf{P}(T_n^m > t) = (\mathbf{P}(R_0 + T_n > t))^m. \quad (7)$$

For long strands ( $n \gg l$ )—assuming the technical condition of equation (4) holds—equation (7) implies, respectively, the following corollaries of propositions 2 and 3.

**Finite-mean relocation times.** *If the relocation time  $R$  possesses a finite mean then the scaled random variables  $T_n^m/n$  converge, in law, as  $n \rightarrow \infty$ , to a limiting random variable  $T_\infty^m$  which is asymptotically exponential:*

$$\mathbf{P}(T_\infty^m > t) \underset{t \rightarrow \infty}{\sim} \exp \left\{ -m \frac{l + \psi(\lambda)}{E[R] + 1/\lambda} \cdot t \right\}. \quad (8)$$

**Infinite-mean relocation times.** *If the relocation time  $R$  is heavy tailed of order  $\alpha$  ( $0 < \alpha < 1$ ) then the scaled random variables  $T_n^m/n^{1/\alpha}$  converge, in law, as  $n \rightarrow \infty$ , to a limiting random variable  $T_\infty^m$  which is asymptotically Pareto with exponent  $m\alpha$ :*

$$\mathbf{P}(T_\infty^m > t) \underset{t \rightarrow \infty}{\sim} \left( \frac{\mathbf{P}(R > t)}{l + \psi(\lambda)} \right)^m. \quad (9)$$

### 3.3. Massively parallel searching

We now turn to explore the case of a *large enzyme ensemble* ( $m \gg 1$ ) searching a long circular DNA strand ( $n \gg l$ ). The setting is that of the previous subsection, and we study the asymptotic behaviour of the enzyme-ensemble search duration  $T_n^m$  in the double limit  $n, m \rightarrow \infty$ .

The proper scaling of the parameters  $n$  and  $m$ , in the double limit  $n, m \rightarrow \infty$ , turns out to be  $\lim_{n, m \rightarrow \infty} (m/n) = \kappa$ , where  $\kappa$  is a positive constant. The meaning of the limiting ratio  $\kappa$  is ‘enzyme concentration per DNA base-pairs’. Assuming the technical condition of equation (4) holds, we have the following.

**Proposition 4.** *The random variables  $T_n^m$  converge, in law, as  $n, m \rightarrow \infty$ , to a limiting random variable  $T_*$  with survival probability*

$$\mathbf{P}(T_* > t) = \exp\{-\Phi(t)\} \quad (10)$$

( $t \geq 0$ ). The function  $\Phi(t)$ , in turn, is characterized by the Laplace transform

$$\int_0^\infty \exp\{-\theta t\} \Phi(t) dt = \kappa \frac{\tilde{R}_0(\theta)}{\theta} \cdot \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)} \quad (11)$$

( $\theta \geq 0$ ).

The distribution of the limiting random variable  $T_*$  is highly contingent on whether or not the relocation time possesses a finite mean, as the following corollaries of proposition 4 assert.

**Finite-mean relocation times.** *If the relocation time  $R$  possesses a finite mean then the function  $\Phi(t)$  follows, asymptotically, a linear growth:*

$$\Phi(t) \underset{t \rightarrow \infty}{\sim} \kappa \frac{l + \psi(\lambda)}{\mathbf{E}[R] + 1/\lambda} \cdot t. \quad (12)$$

Hence, the distribution of the limiting random variable  $T_*$  is asymptotically exponential.

**Infinite-mean relocation times.** *If the relocation time  $R$  is heavy tailed of order  $\alpha$  ( $0 < \alpha < 1$ ) then the function  $\Phi(t)$  follows, asymptotically, a power-law growth with exponent  $\alpha$ :*

$$\Phi(t) \underset{t \rightarrow \infty}{\sim} \kappa \frac{\sin(\pi\alpha)}{\pi\alpha} \cdot \frac{l + \psi(\lambda)}{P(R > t)}. \quad (13)$$

Hence, the distribution of the limiting random variable  $T_*$  is asymptotically stretched exponential with exponent  $\alpha$ .

Several remarks are in place.

- Transcending from parallel to massively parallel searches has a dramatic effect on the *timescales* of the overall search durations  $T_n^m$ —reducing them from the orders  $O(n)$  and  $O(n^{1/\alpha})$  (in the case of parallel searches) to the order  $O(1)$  (in the case of massively parallel searches).
- The cases of finite-mean and infinite-mean relocation times yield different limiting distributions—*asymptotically exponential* in the former case, and *symptomatically stretched exponential* in the latter case.
- The limiting distribution, in the finite-mean case, is *asymptotically universal* with respect to the relocation mechanism—being contingent on the mean relocation time  $\mathbf{E}[R]$  alone (rather than on the distribution of the relocation time). Such asymptotic universality fails to hold in the infinite-mean case (where the limiting distribution is contingent on the survival probability of the relocation time).
- Heuristically, equation (12) can be derived from equation (8) via the naive substitution  $t' = nt$ . The rigorous derivation however—properly taking into account the double limit  $n, m \rightarrow \infty$ —requires the analysis undertaken in the proof of proposition 4.
- The asymptotically stretched-exponential limiting distribution of equation (13) resembles the survival probability obtained in the ‘target problem’ [23, 24].

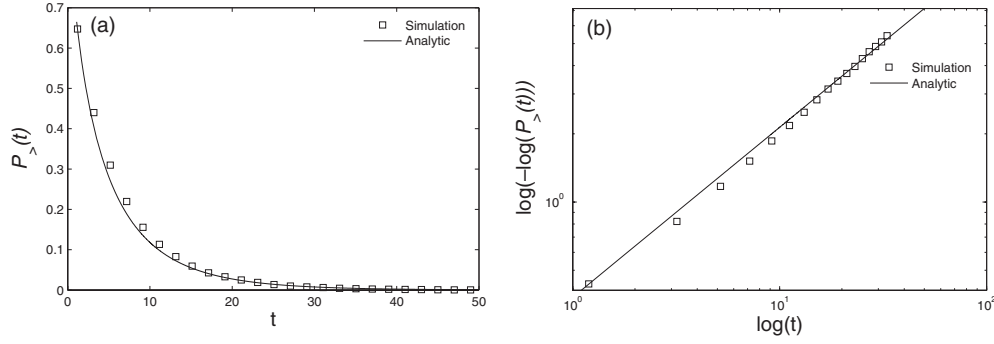
See figure 3 for computer simulations of massively parallel searches taking place in the presence of heavy-tailed relocation times. The proofs of proposition 4 and its corollaries are given in appendix A.

The asymptotic results obtained in this section are summarized in table 1.

#### 4. The relocation mechanism

Let us examine now the relocation mechanism in the case of finite-mean relocation times.





**Figure 3.** Simulations of a massively parallel search of a circular DNA strand of length  $n + l = 2000$ , conducted by an enzyme ensemble of size  $m = 100$  performing linear scanning with velocity  $v = 5$ . The length of the target site is  $l = 5$ , and the ratio of ‘enzymes per base-pairs’ is  $\kappa \approx 0.05$ . The relocation times are heavy tailed—drawn from a one-sided Lévy distribution with exponent  $\alpha = 0.75$ . Number of simulations:  $4 \times 10^4$ . The empirical survival probability  $P_{>}(t)$  of the overall search durations is depicted on (a) a standard plot,  $P_{>}(t)$  versus  $t$ ; (b) a logarithmic plot,  $\log(-\log(P_{>}(t)))$  versus  $\log(t)$ . The solid line depicts the theoretical stretched-exponential limit predicted by proposition 4 and equation (13)—to which the simulations closely fit.

**Table 1.** Searching long DNA strands ( $n \gg 1$ ): classification of the asymptotic probabilistic limit laws of the overall search durations. The rows indicate the search type: single, parallel ( $m$  denoting the number of searching enzymes), or massively parallel. The columns indicate the type of the relocation time: finite mean or heavy tailed (in the ‘heavy-tailed column’ the numbers in parentheses are the values of the corresponding exponents). The table summarizes the asymptotic results obtained in section 3—the appropriate scaling of the overall search durations, and the emerging probabilistic limit laws.

	Finite-mean relocation time	Heavy-tailed relocation time ( $\alpha$ )
Single/parallel search ( $m$ )	Scaling: $O(n)$ Limit: exponential	Scaling: $O(n^{1/\alpha})$ Limit: Pareto ( $m\alpha$ )
Massively parallel search	Scaling: $O(1)$ Limit: exponential	Scaling: $O(1)$ Limit: stretched exp. ( $\alpha$ )

#### 4.1. Mean performance

Is relocation beneficial? Does the presence of the relocation mechanism improve the search performance? One way of answering these questions is by comparing the mean search duration with relocation to the mean search duration *without* relocation.

Consider a single searching enzyme. With the relocation mechanism ‘shut down’—corresponding to the degenerate case  $\lambda = 0$ —the search is conducted via scanning alone, and the mean overall search duration is given by

$$(1 - p_n) \cdot 0 + p_n \cdot \mathbf{E}[S_n] \underset{n \rightarrow \infty}{\sim} \mathbf{E}[S_n]. \quad (14)$$

On the other hand, when the relocation mechanism is ‘active’ then the mean overall search duration is  $\mathbf{E}[T_n]$ , and proposition 2 implies that

$$\mathbf{E}[T_n] \underset{n \rightarrow \infty}{\sim} \frac{\mathbf{E}[R] + 1/\lambda}{l + \psi(\lambda)} \cdot n. \quad (15)$$

Thus, in order to evaluate the ‘mean impact’ of the relocation mechanism, one has to compare the asymptotic behaviour (as  $n \rightarrow \infty$ ) of the sequences  $\mathbf{E}[S_n]$  and  $\mathbf{E}[T_n]$ . When

the sequence  $\mathbf{E}[T_n]$  grows *slower* than the sequence  $\mathbf{E}[S_n]$  then the presence of the relocation mechanism—from a mean-performance perspective—is indeed beneficial.

See sections 5 and 6 for an examination of specific classes of local scan mechanisms.

#### 4.2. Search optimization

Set  $\phi(\lambda) = \lambda \cdot \psi(\lambda)$ . The ‘search function’

$$F(\lambda) = \frac{l + \psi(\lambda)}{\mathbf{E}[R] + 1/\lambda} = \frac{\phi(\lambda) + l\lambda}{1 + \mathbf{E}[R]\lambda} \quad (16)$$

governs the asymptotically exponential behaviour of the overall search duration’s limiting distribution in the case of a single searching enzyme (equation (5)), in the case of parallel searches (equation (8)), and in the case of massively parallel searches (equation (12)).

It is evident from these equations that the *asymptotically optimal relocation rate* is that at which the search function  $F(\lambda)$  is *maximized*. However, does the search function  $F(\lambda)$  attain a global maximum? The following proposition gives necessary conditions for the search function  $F(\lambda)$  to be unimodal.

**Proposition 5.** *Assume that  $\lim_{\lambda \rightarrow \infty} \psi(\lambda) = 0$  and that the function  $\phi(\lambda)$  initiates at the origin ( $\phi(0) = 0$ ), is concave ( $\phi''(\lambda) < 0$ ), and satisfies  $\lim_{\lambda \rightarrow \infty} \phi(\lambda) = \infty$ . Then we have the following.*

*The search function  $F(\lambda)$  is unimodal: it initiates at the origin ( $F(0) = 0$ ), increases monotonically to a global maximum, and thereafter decreases monotonically to the asymptotic level  $\lim_{\lambda \rightarrow \infty} F(\lambda) = l/\mathbf{E}[R]$ .*

Examples of classes of local scanning mechanisms yielding unimodal search functions  $F(\lambda)$  are given in section 5. The proof of proposition 5 is given in appendix A.

### 5. The local scanning mechanism

So far we have drawn general asymptotic results and conclusions for systems with general local scanning and relocation mechanisms, based on the assumption that the technical condition of equation (4) holds. In this section we explore three specific classes of local scanning mechanisms: linear motions, Brownian motions, and selfsimilar motions.

#### 5.1. Linear scanning

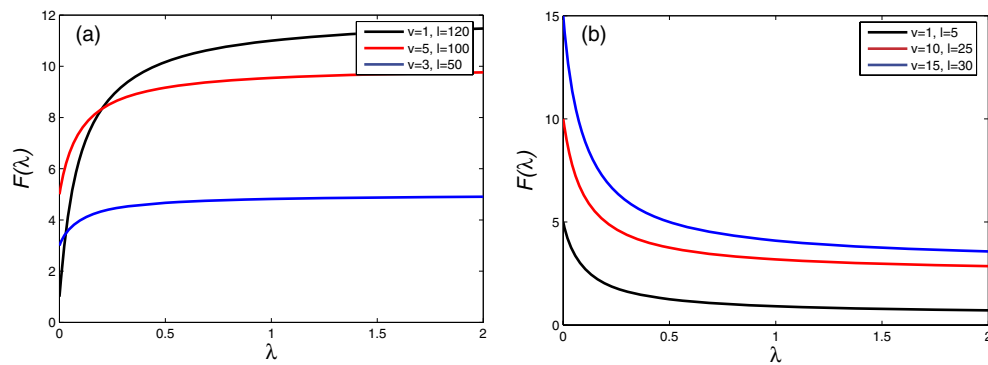
Consider a deterministic local scan conducted by a linear directional motion with velocity  $v$ .

Since the enzyme begins scanning from a random initial position along the strand, the resulting duration  $S_n$  of the local scan is uniformly distributed on the interval  $[0, n/v]$ . This, in turn, implies that

$$\tilde{S}_n(\theta) = \frac{1 - \exp\{-\frac{n}{v}\theta\}}{\frac{n}{v}\theta}. \quad (17)$$

The technical condition of equation (4) is satisfied—yielding the scan function  $\psi(\lambda) = v/\lambda$ , and, in turn, the search function

$$F(\lambda) = \frac{v + l\lambda}{1 + \mathbf{E}[R]\lambda}. \quad (18)$$



**Figure 4.** Search performance as a function of the relocation rate  $\lambda$ , in the presence of finite-mean relocation times with mean  $\mathbf{E}[R] = 10$ , using a linear scanning mechanism (equation (18)). (a) The monotonically increasing scenario  $\mathbf{E}[R] < l/v$ . (b) The monotonically decreasing scenario  $\mathbf{E}[R] > l/v$ .

### Is relocation beneficial?

Since the scanning duration  $S_n$  is uniformly distributed on the interval  $[0, n/v]$  its mean is given by  $\mathbf{E}[S_n] = n/2v$ . Hence, the sequence  $\mathbf{E}[S_n]$  has the same order of growth— $O(n)$ —as the sequence  $\mathbf{E}[T_n]$ . Thus, from a mean-performance perspective, the presence of a relocation mechanism does not make much of a difference. However, the mean performance provides a rather narrow perspective. Indeed, as we shall demonstrate shortly, the asymptotically optimal search strategies exhibit the following *dichotomy*: either the relocation mechanism should be used alone (abolishing scanning altogether), or the local scan mechanism should be used alone (never relocating).

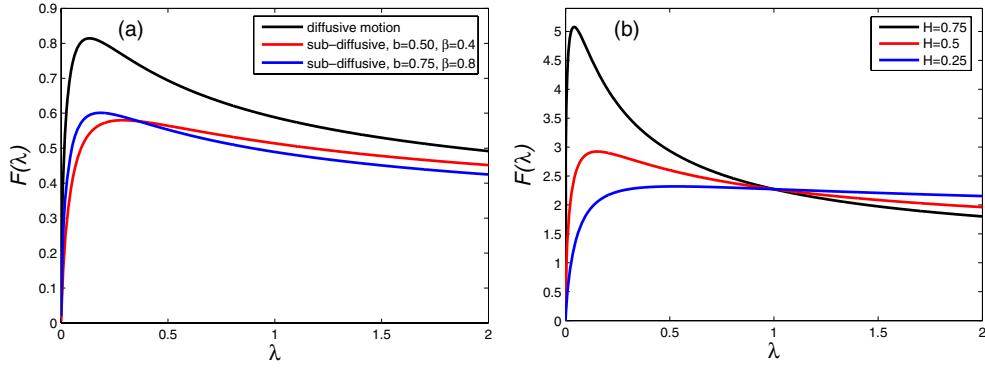
### Search optimization

The search function  $F(\lambda)$  admits three qualitatively different forms—the corresponding graphs are depicted in figure 4. The optimal search strategies, in these three qualitatively different ‘scenarios’, are as follows.

- $\mathbf{E}[R] < l/v$ . In this scenario the search function  $F(\lambda)$  increases monotonically from the level  $F(0) = v$  to the level  $\lim_{\lambda \rightarrow \infty} F(\lambda) = l/\mathbf{E}[R]$ . Thus, the optimum of the search function  $F(\lambda)$  is attained at the limit  $\lambda \rightarrow \infty$ . This, in turn, implies that the asymptotically optimal search strategy is to *repeatedly relocate*—spending no time on local scanning—till ‘landing’ on the target site.
- $\mathbf{E}[R] = l/v$ . In this ‘borderline’ scenario  $F(\lambda) \equiv v$ , and hence the asymptotic search performance is independent of the relocation rate  $\lambda$ .
- $\mathbf{E}[R] > l/v$ . In this scenario the search function  $F(\lambda)$  decreases monotonically from the level  $F(0) = v$  to the level  $\lim_{\lambda \rightarrow \infty} F(\lambda) = l/\mathbf{E}[R]$ . Thus, the optimum of the search function  $F(\lambda)$  is attained at the point  $\lambda = 0$ . This, in turn, implies that the asymptotically optimal search strategy is to continue the local scan—*never relocating*—till ‘hitting’ the target site.

### 5.2. Brownian scanning

Consider a random local scan conducted by a Brownian motion with diffusion parameter  $D$  (i.e., the motion’s mean square displacement, after ‘running’ for  $t$  units of time, is  $D \cdot t$ ).



**Figure 5.** Search performance as a function of the relocation rate  $\lambda$ , in the presence of finite-mean relocation times with mean  $\mathbf{E}[R] = 10$ . (a) Brownian scanning (equation (20)) and sub-diffusive scanning (equation (28)) with parameters  $l = 2$  and  $D = 10$ . (b) Selfsimilar scanning (equation (23)) with parameters  $l = 10$  and  $c_H = 15$ .

A scaling argument implies that the scanning duration  $S_n$  is equal, in law, to  $(n^2/D) \cdot \tau$ —where  $\tau$  is the first passage time of a standard Brownian motion from the unit interval, when initiated from a random location on the unit interval. This, in turn, implies that

$$\tilde{S}_n(\theta) = \frac{\tanh(n\sqrt{\frac{\theta}{2D}})}{n\sqrt{\frac{\theta}{2D}}}. \quad (19)$$

The scaling argument used, as well as the derivation of equation (19), are explained in appendix B.

The technical condition of equation (4) is satisfied—yielding the scan function  $\psi(\lambda) = \sqrt{2D/\lambda}$ , and, in turn, the search function

$$F(\lambda) = \frac{\sqrt{2D\lambda} + l\lambda}{1 + \mathbf{E}[R]\lambda}. \quad (20)$$

#### Is relocation beneficial?

Since the scanning duration  $S_n$  is equal in law to  $(n^2/D) \cdot \tau$ , and since  $\mathbf{E}[\tau] = 1/6$  (see appendix B for the details), we have  $\mathbf{E}[S_n] = n^2/6D$ . The growth of the sequence  $\mathbf{E}[S_n]$  is thus of order  $O(n^2)$ , whereas the growth of the sequence  $\mathbf{E}[T_n]$  is of order  $O(n)$ . Hence, the presence of the relocation mechanism induces an *order-of-magnitude reduction* in the mean search time. Therefore, from a mean-performance perspective, relocation turns out to be very beneficial indeed.

#### Search optimization

The scan function  $\psi(\lambda) = \sqrt{2D/\lambda}$  satisfies the conditions of proposition 5, and hence the search function  $F(\lambda)$  is unimodal and attains a global maximum—see figure 5. The optimal relocation rate  $\lambda_*$  and its corresponding global maximum level  $F(\lambda_*)$  are given, respectively, by

$$\lambda_* = \frac{1}{\mathbf{E}[R]} + \frac{l^2}{D\mathbf{E}[R]^2} \left( 1 + \sqrt{1 + 2\frac{D\mathbf{E}[R]}{l^2}} \right) \quad (21)$$

and

$$F(\lambda_*) = \frac{l}{\mathbf{E}[R]} + \frac{D}{l} \left( 1 + \sqrt{1 + 2 \frac{D\mathbf{E}[R]}{l^2}} \right)^{-1}. \quad (22)$$

### 5.3. Selfsimilar scanning

A stochastic process  $(Z(t))_{t \geq 0}$  is said to be  $\mathbb{H}$ -selfsimilar if for any positive constant  $c$  the process  $(Z(ct))_{t \geq 0}$  is equal, in law, to the process  $(c^{\mathbb{H}}Z(t))_{t \geq 0}$ . The parameter  $\mathbb{H}$  ( $\mathbb{H} > 0$ ) is called the process' *Hurst exponent* [25]. The best known examples of selfsimilar processes are stable Lévy processes—including, in particular, Brownian motion. Examples of selfsimilar processes with *continuous* sample-path trajectories are Brownian motion, fractional Brownian motions and fractional Lévy motions.

In the case of fractional Brownian motions the Hurst parameter takes values in the range  $0 < \mathbb{H} < 1$ . In the sub-range  $0 < \mathbb{H} < 1/2$  the motion is negatively correlated and highly intermittent; in the sub-range  $1/2 < \mathbb{H} < 1$  the motion is positively correlated and long-range dependent; the 'border case'  $\mathbb{H} = 1/2$  corresponds to standard Brownian motion.

In the case of fractional Lévy motions the Hurst parameter takes values in the range  $1/\nu < \mathbb{H} < 1$ , where  $\nu$  is the process's Lévy exponent (taking values in the range  $1 < \nu < 2$ ). The motion is positively correlated and is long-range dependent.

For a detailed exposition of selfsimilar processes readers are referred to [25].

Consider a random local scan conducted by an  $\mathbb{H}$ -selfsimilar motion with continuous sample-path trajectories.

A scaling argument implies that the scanning duration  $S_n$  is equal, in law, to  $n^{1/\mathbb{H}} \cdot \tau_{\mathbb{H}}$ —where  $\tau_{\mathbb{H}}$  is the first passage time of the  $\mathbb{H}$ -selfsimilar motion from the unit interval, when initiated from a random location on the unit interval.

If the limit  $\lim_{k \rightarrow \infty} k^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k) := c_{\mathbb{H}}$  exists then the technical condition of equation (4) is satisfied—yielding the scan function  $\psi(\lambda) = c_{\mathbb{H}}/\lambda^{\mathbb{H}}$ , and, in turn, the search function

$$F(\lambda) = \frac{c_{\mathbb{H}}\lambda^{1-\mathbb{H}} + l\lambda}{1 + \mathbf{E}[R]\lambda}. \quad (23)$$

The scaling argument used, as well as the derivation of the scan function  $\psi(\lambda)$ , are explained in appendix B. Proving that the limit  $\lim_{k \rightarrow \infty} k^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k) := c_{\mathbb{H}}$  indeed exists is beyond the scope of this research, and is only conjectured here.

#### *Is relocation beneficial?*

Since the scanning duration  $S_n$  is equal in law to  $n^{1/\mathbb{H}} \cdot \tau_{\mathbb{H}}$  we have  $\mathbf{E}[S_n] = n^{1/\mathbb{H}} \cdot \mathbf{E}[\tau_{\mathbb{H}}]$ . The growth of the sequence  $\mathbf{E}[S_n]$  is thus of order  $O(n^{1/\mathbb{H}})$ , whereas the growth of the sequence  $\mathbf{E}[T_n]$  is of order  $O(n)$ . Hence, when the Hurst exponent is in the range  $0 < \mathbb{H} < 1$ , the presence of the relocation mechanism *reduces the order* of the expected search time. Therefore, from a mean-performance perspective, relocation turns out to be beneficial.

#### *Search optimization*

When the Hurst exponent is in the range  $0 < \mathbb{H} < 1$  the scan function  $\psi(\lambda) = c_{\mathbb{H}}/\lambda^{\mathbb{H}}$  satisfies the conditions of proposition 5, and hence the search function  $F(\lambda)$  is unimodal and attains a global maximum—see figure 5. The optimal relocation rate  $\lambda_*$  is the unique positive root of the algebraic equation

$$l \cdot x^{\mathbb{H}} - c_{\mathbb{H}}\mathbb{H}\mathbf{E}[R] \cdot x + c_{\mathbb{H}}(1 - \mathbb{H}) = 0. \quad (24)$$

## 6. Local scanning with random halting

The searching enzymes, while scanning the DNA strand, may incur random halts. In this section we modify the searching model to accommodate such a phenomenon.

### 6.1. The halting mechanism

We introduce a ‘*halting mechanism*’ which acts on the searching enzymes while scanning the DNA strand. The halting mechanism is analogous, in model, to the relocation mechanism, and operates according to the following pair of rules:

- enzyme-halting—during the periods of local scanning—occurs at an exponential rate  $\eta$  and
- once halted, the enzyme ‘freezes’ for a random duration of time  $H$ , after which the local scanning is resumed.

A comprehensive analysis of such halting mechanisms is conducted in [26]. This analysis implies that if the Laplace transform of the *un-halted* scanning duration is  $S_n(\theta)$ , then the Laplace transform of the *halted* scanning duration is  $\tilde{S}_n(\theta + \mathcal{H}(\theta))$ , where

$$\mathcal{H}(\theta) := \eta (1 - \tilde{H}(\theta)) \quad (25)$$

( $\theta \geq 0$ ). Moreover, if the mean *un-halted* scanning duration is  $\mathbf{E}[S_n]$ , then the mean *halted* scanning duration is  $(1 + \eta \mathbf{E}[H]) \cdot \mathbf{E}[S_n]$ .

Hence, the incorporation of the halting mechanism corresponds (i) in Laplace space to the nonlinear transformation  $\theta \mapsto \theta + \mathcal{H}(\theta)$  and (ii) in mean to the multiplication by the factor  $(1 + \eta \mathbf{E}[H])$ .

It is straightforward to deduce that if the *un-halted* local-scanning mechanism satisfies the technical condition of equation (4) with the scan function  $\psi(\lambda)$ , then so does the *halted* local scanning mechanism—but with the ‘*halted scan function*’

$$\psi_{\text{halt}}(\lambda) := \psi(\lambda + \mathcal{H}(\lambda)). \quad (26)$$

Thus, from an asymptotic point of view, the incorporation of halting into the search model boils down to the replacement of the *un-halted* scan function  $\psi(\lambda)$  by the *halted* scan function  $\psi_{\text{halt}}(\lambda)$ .

### *Is relocation beneficial?*

If the halting time  $H$  possesses a finite mean (i.e.  $\mathbf{E}[H] < \infty$ ) then the mean scanning durations—in the *un-halted* and *halted* cases—differ only by the multiplicative factor  $(1 + \eta \mathbf{E}[H])$ . Hence, the comparison of the mean overall search durations with and without relocation is essentially the same—from a mean-performance perspective—for the *un-halted* and *halted* cases.

If the halting time  $H$  possesses an infinite mean (i.e.  $\mathbf{E}[H] = \infty$ ) then so does the *halted* scanning duration. Nonetheless, the *halted* overall search duration has a *finite mean*, and  $\mathbf{E}[T_n]$  is of order  $O(n)$ . Thus, in the case of infinite-mean halting times, the presence of the relocation mechanism is far more than beneficial—it is *essential!*

In the case of infinite-mean halting times the relocation mechanism has a most vital role: it ‘bails out’ the scanning enzyme when ‘trapped’ in long halts—thus reducing the mean search duration from infinite (with no relocation) to finite (with relocation).

### 6.2. Sub-diffusive scanning

Consider now the case of a halted Brownian scan, in which the halting times are heavy tailed with exponent  $\beta$  ( $0 < \beta < 1$ ): namely,  $\mathbf{P}(H > t) \sim b/t^\beta$  as  $t \rightarrow \infty$ .

In this case the resulting halted Brownian scan is a *sub-diffusive* motion [26]: its asymptotic mean square displacement follows the sub-linear temporal growth  $D_\beta \cdot t^\beta$ —in sharp contrast to the linear temporal growth  $D \cdot t$  of the underlying Brownian motion<sup>3</sup>. Moreover, the halted scanning duration is heavy tailed—its probability tails admitting the asymptotic form  $(\eta n^2/6D) \cdot \mathbf{P}(H > t)$  (as  $t \rightarrow \infty$ ).

The sub-diffusive scan function  $\psi_{\text{halt}}(\lambda)$  satisfies

$$\psi_{\text{halt}}(\lambda) \underset{\lambda \rightarrow 0}{\sim} \frac{\sqrt{\frac{2D}{\eta b \Gamma(1-\beta)}}}{\sqrt{\lambda^\beta}} \quad \text{and} \quad \psi_{\text{halt}}(\lambda) \underset{\lambda \rightarrow \infty}{\sim} \frac{\sqrt{2D}}{\sqrt{\lambda}}. \quad (27)$$

Note that for small timer rates ( $\lambda \ll 1$ ) the halted scan function admits a power-law form corresponding to an  $\mathbb{H}$ -selfsimilar scan with Hurst exponent  $\mathbb{H} = \beta/2$  (taking values in the range  $0 < \mathbb{H} < 1/2$ ), whereas for large timer rates ( $\lambda \gg 1$ ) the halted scan function admits a power-law form corresponding to a Brownian scan.

‘Selfsimilar’ halting mechanisms attain the power-law form  $\mathcal{H}(\theta) = b\theta^\beta$  with exponent  $0 < \beta < 1$  [26]. The resulting halted Brownian scan is sub-diffusive, its halted scan function is given by  $\psi_{\text{halt}}(\lambda) = \sqrt{2D/(\lambda + b\lambda^\beta)}$ , and its halted search function is given by

$$F_{\text{halt}}(\lambda) = \frac{\lambda}{(1 + \mathbf{E}[R]\lambda)} \left( l + \frac{\sqrt{2D}}{\sqrt{\lambda + b\lambda^\beta}} \right). \quad (28)$$

The halted search function  $F_{\text{halt}}(\lambda)$  is unimodal: it initiates at the origin, increases monotonically to a global maximum, and thereafter decreases monotonically to the asymptotic level  $l/\mathbf{E}[R]$ —see figure 5.

## 7. Massively parallel searching of stationary systems

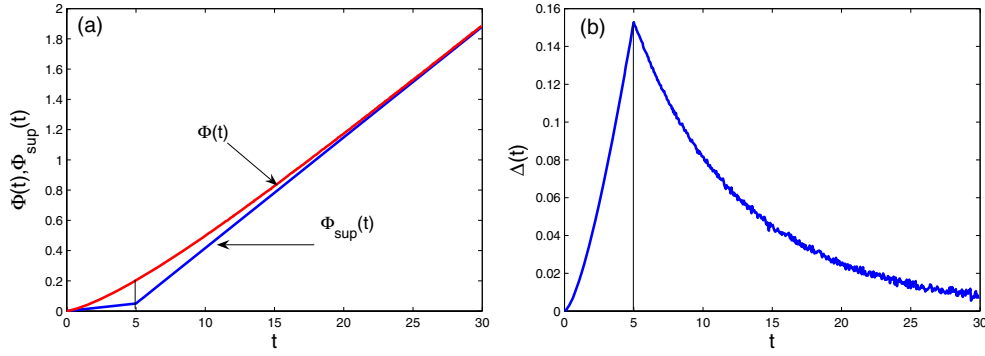
The trajectory of a searching enzyme alternates between a 1D motion along the DNA strand, and a confined motion in the medium containing the strand. The duration of the one-dimensional motion,  $\mathcal{E}$ , is exponentially distributed with rate  $\lambda$  (mean  $1/\lambda$ ), and the duration of the confined motion is the relocation time  $R$ . The enzyme-trajectory is thus the concatenation of independent and identically distributed random cycles of length  $\mathcal{E} + R$ .

Consider, in the case of finite-mean relocation times, the ‘stationary setting’ in which

- the system is initiated at time  $t = -\infty$ ,
- the ‘search command’ is given at an arbitrary time epoch (say  $t = 0$ ), and
- the search command can be communicated to the searching enzymes only while in the confined-motion phase.

Thus, each enzyme is informed about the search command upon its first confined-motion phase taking place after time  $t = 0$ . Recall that we denoted by  $R_0$  the enzymes’ initialization time. Namely,  $R_0$  is the time it takes an enzyme, once the search command is given, to commence its search procedure. Standard argumentation from the theory of renewal processes implies that—in the ‘stationary setting’ just defined—the initialization time  $R_0$  is the *residual lifetime* of the random variable  $\mathcal{E} + R$  [27].

<sup>3</sup> The precise value of the sub-diffusive coefficient is  $D_\beta = \frac{D}{\eta b} \frac{\sin(\pi\beta)}{\pi\beta}$ .



**Figure 6.** Massively parallel search of stationary systems: the logarithm of the survival probability of the overall search duration. (a) The function  $\Phi(t)$  (equation (29)) and its piecewise-linear support function  $\Phi_{\text{sup}}(t)$  (equation (30)). (b) The difference  $\Delta(t) = \Phi(t) - \Phi_{\text{sup}}(t)$ . The maximal vertical distance is attained at the bending point  $t = 5$  of the support function  $\Phi_{\text{sup}}(t)$ . The parameters taken are  $\kappa = 0.2$ ,  $\lambda = 0.1$ ,  $\gamma = 0.5$ ,  $l = 1$ , and  $\mathbf{E}[R] = 10$ .

Proposition 4 asserted that, in the case of massively parallel searching, the survival probability of the limiting overall search duration is  $\exp\{-\Phi(t)\}$ , where the function  $\Phi(t)$  is characterized by the Laplace transform of equation (11). In general, it is not possible to explicitly invert the Laplace transform of the function  $\Phi(t)$ . However, if the scan function admits the power-law form  $\psi(\lambda) = c/\lambda^\gamma$ , with exponent  $0 < \gamma \leq 1$ , and if the ‘stationary setting’ defined above holds, then the function  $\Phi(t)$  admits the probabilistic representation

$$\Phi(t) = \frac{\kappa\lambda}{1 + \mathbf{E}[R]\lambda} \left\{ l \cdot t + \frac{c}{\lambda^\gamma} \cdot \mathbf{E} \left[ (t - \mathcal{G}_{\lambda,\gamma})_+ \right] \right\}, \quad (29)$$

where  $\mathcal{G}_{\lambda,\gamma}$  is a *Gamma*( $\lambda, \gamma$ )-distributed random variable<sup>4</sup>, and where  $(x)_+ := \max\{0, x\}$  denotes the positive part of the real number  $x$ . The probabilistic representation of equation (29) is highly amenable to Monte Carlo simulation.

The power-law form  $\psi(\lambda) = c/\lambda^\gamma$  (with exponent  $0 < \gamma \leq 1$ ) is attained by the following classes of local scan mechanisms (described in section 5 above): *linear scanning* ( $c = v$  and  $\gamma = 1$ ); *Brownian scanning* ( $c = \sqrt{2D}$  and  $\gamma = 1/2$ ); and *selfsimilar scanning* ( $c = c_{\mathbb{H}}$  and  $\gamma = \mathbb{H}$ ).

To readers familiar with risk-neutral valuation of financial derivatives [28], we note that the term  $\mathbf{E}[(t - \mathcal{G}_{\lambda,\gamma})_+]$  corresponds to the price of a put option, with strike  $t$ , contingent on an underlying risky asset governed by a *Gamma*( $\lambda, \gamma$ ) risk-neutral probability law.

#### Properties of the function $\Phi(t)$

The function  $\Phi(t)$  initiates at the origin ( $\Phi(0) = 0$ ), is monotonically increasing ( $\Phi'(t) > 0$ ), is convex ( $\Phi''(t) > 0$ ), and is supported from below by the piecewise-linear function

$$\Phi_{\text{sup}}(t) = \frac{\kappa\lambda}{1 + \mathbf{E}[R]\lambda} \left\{ l \cdot t + \frac{c}{\lambda^\gamma} \cdot \left( t - \frac{\gamma}{\lambda} \right)_+ \right\} \quad (30)$$

(that is,  $\Phi(t) \geq \Phi_{\text{sup}}(t)$  for all  $t \geq 0$ )—see figure 6.

The ‘support’ function  $\Phi_{\text{sup}}(t)$  is tangent to the function  $\Phi(t)$  at the origin ( $t = 0$ ) and at infinity ( $t \rightarrow \infty$ ). The maximal distance between the function  $\Phi(t)$  and its support  $\Phi_{\text{sup}}(t)$  is

<sup>4</sup> Namely, the random variable  $\mathcal{G}_{\lambda,\gamma}$  is governed by the probability density function  $(\lambda^\gamma / \Gamma(\gamma)) \exp\{-\lambda x\} x^{\gamma-1}$  ( $x > 0$ ).



attained at the ‘bending point’  $t = \gamma/\lambda$  of the piecewise-linear support function, and is given by

$$\max_{t \geq 0} \{ \Phi(t) - \Phi_{\text{sup}}(t) \} = \frac{\kappa c \lambda^{1-\gamma}}{1 + \mathbf{E}[R]\lambda} \cdot \mathbf{E} \left[ \left( \frac{\gamma}{\lambda} - \mathcal{G}_{\lambda, \gamma} \right)_+ \right]. \quad (31)$$

The support function  $\Phi_{\text{sup}}(t)$  also provides us with the following upper bound to the mean of the limiting search duration  $T_*$ :

$$\mathbf{E}[T_*] \leq \frac{1}{\kappa l} \left( \mathbf{E}[R] + \frac{1}{\lambda} \right) \left( 1 - \frac{c}{c + l\lambda^\gamma} \exp \left\{ -\frac{\kappa l \gamma}{1 + \mathbf{E}[R]\lambda} \right\} \right). \quad (32)$$

The proofs of the assertions made in this section are given in appendix B.

## 8. Conclusions

We have introduced and explored a stochastic model of an ensemble of enzymes searching a circular DNA strand for a target site. The enzymes operate in parallel and independently, following a search algorithm combining *local scanning* and *random relocation*.

Focusing on the case of long DNA strands, an asymptotic analysis of the search durations was conducted—yielding closed-form formulae for the corresponding limiting probability distributions. These formulae, in turn, enabled a quantitative formulation and answer to the question ‘is relocation beneficial?’, and facilitated the derivation of the asymptotically optimal relocation rates.

The theory developed encompasses the cases of both parallel and massively parallel ensemble-searches, taking place in the presence of either finite-mean or heavy-tailed relocation times. Moreover, the theory is applicable to a wide spectrum of local-scanning mechanisms, including linear motions, Brownian motions, selfsimilar motions, and sub-diffusive motions.

## Appendix A

### A.1. Proposition 1 and its corollaries

Let  $\mathcal{E}$  denote the timer’s duration (i.e.,  $\mathcal{E}$  is exponentially distributed with rate  $\lambda$ ). Three ‘scenarios’, regarding the overall search duration  $T$ , are possible.

- (1) The random initial position from which the search begins is within the target site. In this scenario  $T = 0$ . The probability of this scenario is  $1 - p = l/(l + n)$ .
- (2) The random initial position from which the search begins is *not* within the target site *and* the local scan traced the target site *before* the timer expired. In this scenario  $T = S$ .
- (3) The random initial position from which the search begins is *not* within the target site *and* the timer expired *before* the local scan traced the target site. In this scenario  $\mathcal{E}$  units of time were spent on an unsuccessful local scan,  $R$  units of time were spent on relocation, and then a new search began anew. Hence  $T = \mathcal{E} + R + T'$  ( $T'$  being a copy of  $T$  which is independent of  $\mathcal{E}$ ,  $S$ , and  $R$ ).

Thus, we have

$$\mathbf{E}[f(T)] = (1 - p) \cdot \mathbf{E}[f(0)] + p \cdot \mathbf{E}[f(S)\mathbf{I}\{S \leq \mathcal{E}\}] + p \cdot \mathbf{E}[f(\mathcal{E} + R + T')\mathbf{I}\{S > \mathcal{E}\}], \quad (33)$$

where  $f(\cdot)$  is an arbitrary function, and where  $\mathbf{I}\{E\}$  denotes the indicator function of the event  $E$ .

With equation (33) at hand, we are now in a position to compute the Laplace transform of the overall search duration  $T$ .

*Laplace transform*

Take  $f(t) = \exp\{-\theta t\}$ , where  $\theta$  is a non-negative parameter.

Using conditioning, and the fact that  $\mathcal{E}$  is exponentially distributed with rate  $\lambda$ , we have

$$\begin{aligned} \mathbf{E}[\exp\{-\theta S\}\mathbf{I}\{S \leq \mathcal{E}\}] &= \mathbf{E}[\exp\{-\theta S\}\mathbf{P}(S \leq \mathcal{E}|S)] \\ &= \mathbf{E}[\exp\{-\theta S\} \exp\{-\lambda S\}] = \mathbf{E}[\exp\{-(\lambda + \theta)S\}] = \tilde{S}(\lambda + \theta); \end{aligned} \tag{34}$$

and

$$\begin{aligned} \mathbf{E}[\exp\{-\theta \mathcal{E}\}\mathbf{I}\{S > \mathcal{E}\}] &= \mathbf{E}\left[\int_0^S \exp\{-\theta t\} \cdot \lambda \exp\{-\lambda t\} dt\right] \\ &= \mathbf{E}\left[\frac{\lambda}{\lambda + \theta} (1 - \exp\{-(\lambda + \theta)S\})\right] = \frac{\lambda}{\lambda + \theta} (1 - \tilde{S}(\lambda + \theta)). \end{aligned} \tag{35}$$

Moreover, since  $T'$  is a copy of  $T$  which is independent of  $\mathcal{E}$ ,  $S$  and  $R$ , we further have

$$\begin{aligned} \mathbf{E}[\exp\{-\theta(\mathcal{E} + R + T')\}\mathbf{I}\{S > \mathcal{E}\}] &= \mathbf{E}[\exp\{-\theta \mathcal{E}\}\mathbf{I}\{S > \mathcal{E}\}] \cdot \mathbf{E}[\exp\{-\theta R\}] \cdot \mathbf{E}[\exp\{-\theta T\}] \\ &= \frac{\lambda}{\lambda + \theta} (1 - \tilde{S}(\lambda + \theta)) \cdot \tilde{R}(\theta) \cdot \tilde{T}(\theta). \end{aligned} \tag{36}$$

Now, substituting equations (34) and (36) into equation (33) yields

$$\tilde{T}(\theta) = (1 - p) + p\tilde{S}(\lambda + \theta) + p\frac{\lambda}{\lambda + \theta} (1 - \tilde{S}(\lambda + \theta)) \tilde{R}(\theta)\tilde{T}(\theta). \tag{37}$$

Equation (37), in turn, implies that

$$\tilde{T}(\theta) = \frac{1 - p (1 - \tilde{S}(\lambda + \theta))}{1 - p (1 - \tilde{S}(\lambda + \theta)) \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)}.$$

*Mean*

The mean of the overall search duration  $T$  can be obtained either (i) by differentiation of the Laplace transform  $\tilde{T}(\theta)$  at the origin or (ii) by taking  $f(t) = t$  in equation (33). We follow the second alternative.

Using conditioning, and the fact that  $\mathcal{E}$  is exponentially distributed with rate  $\lambda$ , we have

$$\mathbf{E}[S\mathbf{I}\{S \leq \mathcal{E}\}] = \mathbf{E}[S\mathbf{P}(S \leq \mathcal{E}|S)] = \mathbf{E}[S \exp\{-\lambda S\}] = -\tilde{S}'(\lambda); \tag{38}$$

$$\begin{aligned} \mathbf{E}[\mathcal{E}\mathbf{I}\{S > \mathcal{E}\}] &= \mathbf{E}\left[\int_0^S t \cdot \lambda \exp\{-\lambda t\} dt\right] = \mathbf{E}\left[-S \exp\{-\lambda S\} + \frac{1}{\lambda}(1 - \exp\{-\lambda S\})\right] \\ &= \tilde{S}'(\lambda) + \frac{1}{\lambda} (1 - \tilde{S}(\lambda)); \end{aligned} \tag{39}$$

and

$$\mathbf{P}(S > \mathcal{E}) = \mathbf{E}[\mathbf{P}(S > \mathcal{E}|S)] = \mathbf{E}[1 - \exp\{-\lambda S\}] = 1 - \tilde{S}(\lambda). \tag{40}$$

Moreover, since  $T'$  is a copy of  $T$  which is independent of  $\mathcal{E}$ ,  $S$  and  $R$ , we further have

$$\begin{aligned} \mathbf{E}[(\mathcal{E} + R + T')\mathbf{I}\{S > \mathcal{E}\}] &= \mathbf{E}[\mathcal{E}\mathbf{I}\{S > \mathcal{E}\}] + \mathbf{E}[R] \cdot \mathbf{P}(S > \mathcal{E}) + \mathbf{E}[T] \cdot \mathbf{P}(S > \mathcal{E}) \\ &= \tilde{S}'(\lambda) + \frac{1}{\lambda}(1 - \tilde{S}(\lambda)) + \mathbf{E}[R](1 - \tilde{S}(\lambda)) + \mathbf{E}[T](1 - \tilde{S}(\lambda)). \end{aligned} \tag{41}$$

Now, substituting equations (38) and (41) into equation (33) yields

$$\mathbf{E}[T] = p \left( \frac{1}{\lambda} (1 - \tilde{S}(\lambda)) + \mathbf{E}[R](1 - \tilde{S}(\lambda)) + \mathbf{E}[T](1 - \tilde{S}(\lambda)) \right). \quad (42)$$

Equation (42), in turn, implies that

$$\mathbf{E}[T] = \frac{p(1 - \tilde{S}(\lambda))}{1 - p(1 - \tilde{S}(\lambda))} \cdot \left( \mathbf{E}[R] + \frac{1}{\lambda} \right).$$

*Heavy tails*

Set  $q := p(1 - \tilde{S}(\lambda))$ , and note that

$$\tilde{T}(\theta) \underset{\theta \rightarrow 0}{\sim} \frac{1 - q}{1 - q\tilde{R}(\theta)} = \frac{1}{1 + \frac{q}{1-q}(1 - \tilde{R}(\theta))} \underset{\theta \rightarrow 0}{\sim} 1 - \frac{q}{1 - q} (1 - \tilde{R}(\theta)). \quad (43)$$

Equation (43), in turn, implies that

$$1 - \tilde{T}(\theta) \underset{\theta \rightarrow 0}{\sim} \frac{q}{1 - q} (1 - \tilde{R}(\theta)). \quad (44)$$

Since the relocation time  $R$  is heavy tailed of order  $\alpha$  ( $0 < \alpha < 1$ ), Karamata's Tauberian theorem for random variables (see corollary 8.1.7 in [29]) implies that

$$1 - \tilde{R}(\theta) \underset{\theta \rightarrow 0}{\sim} \Gamma(1 - \alpha) \cdot \mathbf{P}(R > 1/\theta), \quad (45)$$

and hence

$$1 - \tilde{T}(\theta) \underset{\theta \rightarrow 0}{\sim} \Gamma(1 - \alpha) \cdot \frac{q}{1 - q} \mathbf{P}(R > 1/\theta). \quad (46)$$

Applying Karamata's Tauberian theorem for random variables to equation (46) yields

$$\mathbf{P}(T > t) \underset{t \rightarrow \infty}{\sim} \frac{q}{1 - q} \mathbf{P}(R > t).$$

*A.2. Propositions 2 and 3*

We shall make use of the following result.

Let  $\{\delta_n\}_{n=1}^\infty$  be a non-negative valued sequence satisfying

$$n(1 - \tilde{R}(\delta_n)) \xrightarrow{n \rightarrow \infty} A \quad \text{and} \quad n\delta_n \xrightarrow{n \rightarrow \infty} B \quad (47)$$

(where  $A$  and  $B$  are non-negative valued limits). Then

$$\tilde{T}_n(\delta_n) \xrightarrow{n \rightarrow \infty} \left( 1 + \frac{A + B/\lambda}{l + \psi(\lambda)} \right)^{-1}. \quad (48)$$

The proof of equation (48) is given below.

*Finite-mean relocation times*

Fix  $\theta \geq 0$  and set  $\delta_n = \theta/n$ . Since the relocation time possesses a finite mean we have

$$n(1 - \tilde{R}(\delta_n)) \xrightarrow{n \rightarrow \infty} \mathbf{E}[R]\theta. \quad (49)$$

On the other hand we have  $\lim_{n \rightarrow \infty} n\delta_n = \theta$ . Hence, equation (48) implies that

$$\mathbf{E} \left[ \exp \left\{ -\theta \frac{T_n}{n} \right\} \right] = \tilde{T}_n(\delta_n) \xrightarrow{n \rightarrow \infty} \left( 1 + \frac{\mathbf{E}[R] + 1/\lambda}{l + \psi(\lambda)} \cdot \theta \right)^{-1}. \quad (50)$$

Noting that the right-hand side of equation (50) is the Laplace transform of an exponential distribution with mean  $(l + \psi(\lambda))/(\mathbf{E}[R] + 1/\lambda)$  concludes the proof.

*Infinite-mean relocation times*

Assume that the relocation rate is heavy tailed with exponent  $\alpha$  ( $0 < \alpha < 1$ ): namely,  $\mathbf{P}(R > t) \sim a/t^\alpha$  as  $t \rightarrow \infty$ . Fix  $\theta \geq 0$  and set  $\delta_n = \theta/n^{1/\alpha}$ . Karamata's Tauberian theorem for random variables (see corollary 8.1.7 in [29]) implies that

$$1 - \tilde{R}(\delta_n) \underset{n \rightarrow \infty}{\sim} \Gamma(1 - \alpha) \cdot \mathbf{P}(R > 1/\delta_n) \underset{n \rightarrow \infty}{\sim} \Gamma(1 - \alpha)a\theta^\alpha \cdot \frac{1}{n}, \tag{51}$$

and hence

$$n(1 - \tilde{R}(\delta_n)) \xrightarrow{n \rightarrow \infty} \Gamma(1 - \alpha)a\theta^\alpha. \tag{52}$$

On the other hand we have  $\lim_{n \rightarrow \infty} n\delta_n = 0$ . Hence, equation (48) implies that

$$\mathbf{E} \left[ \exp \left\{ -\theta \frac{T_n}{n^{1/\alpha}} \right\} \right] = \tilde{T}_n(\delta_n) \xrightarrow{n \rightarrow \infty} \left( 1 + \frac{\Gamma(1 - \alpha)a}{l + \psi(\lambda)} \cdot \theta^\alpha \right)^{-1}. \tag{53}$$

The right-hand side of equation (53) is the Laplace transform of a limiting random variable  $T_\infty$ . Moreover,

$$1 - \tilde{T}_\infty(\theta) \underset{\theta \rightarrow 0}{\sim} \frac{\Gamma(1 - \alpha)a}{l + \psi(\lambda)} \cdot \theta^\alpha = \Gamma(1 - \alpha) \cdot \frac{\mathbf{P}(R > 1/\theta)}{l + \psi(\lambda)}, \tag{54}$$

and hence Karamata's Tauberian theorem for random variables implies that

$$\mathbf{P}(T_\infty > t) \underset{t \rightarrow \infty}{\sim} \frac{\mathbf{P}(R > 1/\theta)}{l + \psi(\lambda)}.$$

The limiting random variable  $T_\infty$  admits the multiplicative probabilistic representation  $T_\infty = \mathcal{L} \cdot \mathcal{E}^{1/\alpha}$  where (i) the random variable  $\mathcal{L}$  is Lévy distributed with Laplace transform  $\tilde{\mathcal{L}}(\theta) = \exp\{-\Gamma(1 - \alpha)a\theta^\alpha\}$ , (ii) the random variable  $\mathcal{E}$  is exponentially distributed with mean  $1/(l + \psi(\lambda))$ , and (iii)  $\mathcal{L}$  and  $\mathcal{E}$  are independent random variables. Indeed, using conditioning we obtain that

$$\begin{aligned} \mathbf{E} [\exp \{-\theta \mathcal{L} \cdot \mathcal{E}^{1/\alpha}\}] &= \mathbf{E} [\mathbf{E} [\exp \{-\theta \mathcal{L} \cdot \mathcal{E}^{1/\alpha}\} | \mathcal{E}]] = \mathbf{E} [\exp \{-\Gamma(1 - \alpha)a\theta^\alpha \mathcal{E}\}] \\ &= \left( 1 + \frac{\Gamma(1 - \alpha)a}{l + \psi(\lambda)} \cdot \theta^\alpha \right)^{-1}. \end{aligned}$$

*Proof of equation (48)*

A straightforward calculation gives

$$\begin{aligned} \tilde{T}_n(\delta_n) &= \frac{1 - p_n(1 - \tilde{S}_n(\lambda + \delta_n))}{1 - p_n(1 - \tilde{S}_n(\lambda + \delta_n)) \frac{\lambda \tilde{R}(\delta_n)}{\lambda + \delta_n}} \\ &= (\lambda + \delta_n) \cdot \frac{[1 - p_n(1 - \tilde{S}_n(\lambda + \delta_n))]}{\lambda \tilde{R}(\delta_n) [1 - p_n(1 - \tilde{S}_n(\lambda + \delta_n))] + [\delta_n + \lambda(1 - \tilde{R}(\delta_n))]} \\ &= (\lambda + \delta_n) \cdot \left( \lambda \tilde{R}(\delta_n) + \frac{\delta_n + \lambda(1 - \tilde{R}(\delta_n))}{1 - p_n(1 - \tilde{S}_n(\lambda + \delta_n))} \right)^{-1} \\ &= \left( 1 + \frac{1}{\lambda} \delta_n \right) \cdot \left( \tilde{R}(\delta_n) + \frac{n(1 - \tilde{R}(\delta_n)) + \frac{1}{\lambda} \cdot n\delta_n}{p_n \cdot (l + n\tilde{S}_n(\lambda + \delta_n))} \right)^{-1}. \end{aligned} \tag{55}$$

The limit  $\lim_{n \rightarrow \infty} n\delta_n = B$  implies that  $\lim_{n \rightarrow \infty} \delta_n = 0$ , and hence (i)  $\lim_{n \rightarrow \infty} \tilde{R}(\delta_n) = 1$  and (ii)  $\lim_{n \rightarrow \infty} n\tilde{S}_n(\lambda + \delta_n) = \psi(\lambda)$ . These limits, in turn, imply the limit of equation (48):

$$\left( 1 + \frac{1}{\lambda} \delta_n \right) \cdot \left( \tilde{R}(\delta_n) + \frac{n(1 - \tilde{R}(\delta_n)) + \frac{1}{\lambda} \cdot n\delta_n}{p_n \cdot (l + n\tilde{S}_n(\lambda + \delta_n))} \right)^{-1} \xrightarrow{n \rightarrow \infty} \left( 1 + \frac{A + B/\lambda}{l + \psi(\lambda)} \right)^{-1}.$$

A.3. Proposition 4 and its corollaries

Set  $\Phi_n(t) := \mathbf{P}(R_0 + T_n \leq t)$  ( $t \geq 0$ ) to be the cumulative distribution function of the random sum  $R_0 + T_n$ .

Laplace transform

The Laplace transform of the function  $m\Phi_n(t)$  is given by

$$\begin{aligned} \int_0^\infty \exp\{-\theta t\} (m\Phi_n(t)) dt &= \frac{m}{\theta} \int_0^\infty \exp\{-\theta t\} \Phi_n'(t) dt = \frac{m}{\theta} (\widetilde{R_0 + T_n})(\theta) \\ &= \frac{m}{\theta} \widetilde{R_0}(\theta) \widetilde{T_n}(\theta) = \frac{m}{n} \cdot \frac{\widetilde{R_0}(\theta)}{\theta} \cdot (n\widetilde{T_n}(\theta)). \end{aligned} \tag{56}$$

Using proposition 1, together with the technical condition of equation (4), we obtain the limit

$$\begin{aligned} n\widetilde{T_n}(\theta) &= n \cdot \frac{1 - p_n (1 - \widetilde{S}_n(\lambda + \theta))}{1 - p_n (1 - \widetilde{S}_n(\lambda + \theta)) \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta)} \\ &= \frac{l + n\widetilde{S}_n(\lambda + \theta)}{(1 - \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta)) + \frac{1}{n} (l + n\widetilde{S}_n(\lambda + \theta))} \xrightarrow{n \rightarrow \infty} \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta)}, \end{aligned} \tag{57}$$

which, in turn, implies that

$$\int_0^\infty \exp\{-\theta t\} (m\Phi_n(t)) dt \xrightarrow{n, m \rightarrow \infty} \kappa \cdot \frac{\widetilde{R_0}(\theta)}{\theta} \cdot \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta)}. \tag{58}$$

Thus, we have obtained that  $\lim_{n, m \rightarrow \infty} m\Phi_n(t) = \Phi(t)$ , where the Laplace transform of the limit function  $\Phi(t)$  is given by the right-hand side of equation (58).

Using equation (7), and the limit  $\lim_{n, m \rightarrow \infty} m\Phi_n(t) = \Phi(t)$ , we conclude that

$$\begin{aligned} \mathbf{P}(T_n^m > t) &= (\mathbf{P}(R_0 + T_n > t))^m = (1 - \Phi_n(t))^m \\ &= \left(1 - \frac{m\Phi_n(t)}{m}\right)^m \xrightarrow{n, m \rightarrow \infty} \exp\{-\Phi(t)\}. \end{aligned} \tag{59}$$

Finite-mean relocation times

If the relocation time  $R$  possesses a finite mean then

$$1 - \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta) \underset{\theta \rightarrow 0}{\sim} \left(\mathbf{E}[R] + \frac{1}{\lambda}\right) \cdot \theta, \tag{60}$$

and hence

$$\int_0^\infty \exp\{-\theta t\} \Phi(t) dt = \kappa \frac{\widetilde{R_0}(\theta)}{\theta} \cdot \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \widetilde{R}(\theta)} \underset{\theta \rightarrow 0}{\sim} \kappa \frac{l + \psi(\lambda)}{\mathbf{E}[R] + 1/\lambda} \cdot \frac{1}{\theta^2}. \tag{61}$$

Karamata's Tauberian theorem for functions (see theorem 1.7.1 in [29]), in turn, implies that

$$\Phi(t) \underset{t \rightarrow \infty}{\sim} \kappa \frac{l + \psi(\lambda)}{\mathbf{E}[R] + 1/\lambda} \cdot t.$$

*Infinite-mean relocation times*

Assume that the relocation rate is heavy tailed with exponent  $\alpha$  ( $0 < \alpha < 1$ ): namely,  $\mathbf{P}(R > t) \sim a/t^\alpha$  as  $t \rightarrow \infty$ . Karamata's Tauberian theorem for random variables (see theorem 1.7.1 in [29]) implies that

$$1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta) \underset{\theta \rightarrow 0}{\sim} \Gamma(1 - \alpha)a\theta^\alpha, \tag{62}$$

and hence

$$\int_0^\infty \exp\{-\theta t\} \Phi(t) dt = \kappa \frac{\tilde{R}_0(\theta)}{\theta} \cdot \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)} \underset{\theta \rightarrow 0}{\sim} \kappa \frac{l + \psi(\lambda)}{\Gamma(1 - \alpha)a} \cdot \frac{1}{\theta^{1+\alpha}}. \tag{63}$$

Karamata's Tauberian theorem for functions (see theorem 1.7.1 in [29]), in turn, implies that

$$\Phi(t) \underset{t \rightarrow \infty}{\sim} \kappa \frac{l + \psi(\lambda)}{\Gamma(1 - \alpha)a} \cdot \frac{t^\alpha}{\Gamma(1 + \alpha)}. \tag{64}$$

Using the identity  $\Gamma(1 - \alpha)\Gamma(1 + \alpha) = \pi\alpha / \sin(\pi\alpha)$  (which holds in the range  $0 < \alpha < 1$ ), we conclude that

$$\Phi(t) \underset{t \rightarrow \infty}{\sim} \kappa \frac{\sin(\pi\alpha)}{\pi\alpha} \cdot \frac{l + \psi(\lambda)}{\mathbf{P}(R > t)}.$$

*A.4. Proposition 5*

We shall make use of the following result.

Let  $a(\lambda)x + b(\lambda)$  ( $x > 0$ ) be the tangent line to the function  $\phi(x)$ , passing through the point  $\lambda$ . Then

$$a(\lambda) = \phi'(\lambda) \underset{\lambda \rightarrow \infty}{\longrightarrow} 0 \quad \text{and} \quad b(\lambda) = \phi(\lambda) - \lambda\phi'(\lambda) \underset{\lambda \rightarrow \infty}{\longrightarrow} \infty. \tag{65}$$

The proof of equation (65) is given below.

Consider the search function

$$F(\lambda) = \frac{l + \psi(\lambda)}{\mathbf{E}[R] + 1/\lambda} = \frac{\phi(\lambda) + l\lambda}{1 + \mathbf{E}[R]\lambda}. \tag{66}$$

The fact that  $\phi(0) = 0$  implies that  $F(0) = 0$ , and the fact that  $\lim_{\lambda \rightarrow \infty} \psi(\lambda) = 0$  implies that  $\lim_{\lambda \rightarrow \infty} F(\lambda) = l/\mathbf{E}[R]$ .

The derivative of the function  $F(\lambda)$  is given by

$$F'(\lambda) = \frac{l - G(\lambda)}{(1 + \mathbf{E}[R]\lambda)^2}, \tag{67}$$

where

$$G(\lambda) := \mathbf{E}[R]\phi(\lambda) - (1 + \mathbf{E}[R]\lambda)\phi'(\lambda) = \mathbf{E}[R]b(\lambda) - a(\lambda). \tag{68}$$

Now we have the following.

- (i) The fact that the function  $\phi(x)$  is non-negative valued, concave, and satisfies  $\phi(0) = 0$  implies that  $\phi'(0) \geq 0$ . This, in turn, implies that  $G(0) \leq 0$ .
- (ii) Equation (65) implies that  $\lim_{\lambda \rightarrow \infty} G(\lambda) = \infty$ .
- (iii) Since  $G'(\lambda) = -(1 + \mathbf{E}[R]\lambda)\phi''(\lambda)$  the fact that the function  $\phi(\lambda)$  is concave implies that  $G'(\lambda) > 0$ .

Thus, the function  $G(\lambda)$  initiates from a non-positive level and increases monotonically to infinity. This, in turn, implies that the function  $F(\lambda)$  is monotonically increasing in the range  $\lambda < G^{-1}(l)$ , attains its global maximum at the point  $\lambda = G^{-1}(l)$  and is monotonically decreasing in the range  $\lambda > G^{-1}(l)$ .

*Proof of equation (65)*

The fact that the function  $\phi(x)$  is concave and satisfies  $\phi(0) = 0$ , combined with the mean-value theorem, implies that

$$\psi(\lambda) = \frac{\phi(\lambda) - \phi(0)}{\lambda - 0} = \phi'(\theta_\lambda) \geq \phi'(\lambda) = a(\lambda) \tag{69}$$

(for some  $0 < \theta_\lambda < \lambda$ ). On the other hand, the function  $\phi(x)$ —since it is positive valued and concave—is monotonically increasing:  $a(x) = \phi'(x) \geq 0$ . Hence, the limit  $\lim_{\lambda \rightarrow \infty} \psi(\lambda) = 0$  implies that  $\lim_{\lambda \rightarrow \infty} a(\lambda) = 0$ .

We turn now to prove, by contradiction, that  $\lim_{\lambda \rightarrow \infty} b(\lambda) = \infty$ . Assume that the function  $b(\lambda)$  is bounded: namely, that  $b(\lambda) \leq M$  for all  $\lambda > 0$  (where  $M$  is some positive constant). By concavity, the function  $\phi(x)$  lies below the tangent line  $a(\lambda)x + b(\lambda)$ , and hence

$$\phi(x) \leq a(\lambda)x + b(\lambda) \leq a(\lambda)x + M. \tag{70}$$

Now, taking  $\lambda \rightarrow \infty$ , while using the fact that  $\lim_{\lambda \rightarrow \infty} a(\lambda) = 0$ , implies that  $\phi(x) \leq M$ . Thus, we obtained that the function  $\phi(\lambda)$  is bounded—in contradiction to the fact that  $\lim_{\lambda \rightarrow \infty} \phi(\lambda) = \infty$ . Hence  $\lim_{\lambda \rightarrow \infty} b(\lambda) = \infty$ .

**Appendix B**

*B.1. Brownian motion on the unit interval*

Consider a standard Brownian motion on the unit interval, initiated from the starting point  $x \in [0, 1]$ . Let  $\tau_x$  denote its first passage time (FPT) to the boundary  $\{0, 1\}$  of the unit interval.

Fix  $\theta \geq 0$  and set  $U(x) := \tilde{\tau}_x(\theta)$ . The celebrated Feynman–Kac formula (see, for example, [30]) asserts that the function  $U(x)$  satisfies the differential equation  $\frac{1}{2}U''(x) = \theta U(x)$ , with the boundary conditions  $U(0) = 1 = U(1)$ . Solving this equation yields

$$\tilde{\tau}_x(\theta) = \frac{(\exp\{l\sqrt{2\theta}\} - 1) \cdot \exp\{-x\sqrt{2\theta}\} + (1 - \exp\{-l\sqrt{2\theta}\}) \cdot \exp\{x\sqrt{2\theta}\}}{\exp\{l\sqrt{2\theta}\} - \exp\{-l\sqrt{2\theta}\}}. \tag{71}$$

Assume now that the starting point is randomly located on the unit interval, and let  $\tau$  denote the corresponding FPT. Averaging over equation (71) we obtain that

$$\tilde{\tau}(\theta) = \int_0^1 \tilde{\tau}_x(\theta) dx = \frac{\tanh(\sqrt{\theta/2})}{\sqrt{\theta/2}}. \tag{72}$$

Moreover, expanding the right-hand side of equation (72) into a power series yields

$$\tilde{\tau}(\theta) = 1 - \frac{1}{6} \cdot \theta + \frac{1}{15} \cdot \frac{\theta^2}{2} - \dots, \tag{73}$$

which, in turn, implies that

$$\mathbf{E}[\tau] = \frac{1}{6} \quad \text{and} \quad \mathbf{Var}[\tau] = \frac{7}{180}. \tag{74}$$

*B.2. Brownian local scanning*

Consider a local Brownian scan initiated from a random starting point on the ‘non-target’ part of a DNA strand. We parametrize the strand’s ‘non-target’ part by the interval  $[0, n]$ . Clearly, the duration of a local scan  $S_n$  is the FPT of the interval  $[0, n]$ —by a Brownian motion with diffusion parameter  $D$ , initiated from a random position on the interval.

Let  $(B(t))_{t \geq 0}$  be a standard Brownian motion, and let  $X$  be a uniformly distributed random variable on the unit interval. Then, using the definition of the FPTs  $S_n$  and  $\tau$ , and using the scaling properties of the Brownian motion, we have (the equalities below being in law)

$$\begin{aligned} S_n &= \inf \left\{ t \geq 0 \mid nX + \sqrt{D} \cdot B(t) \notin [0, n] \right\} = \inf \left\{ t \geq 0 \mid X + \frac{\sqrt{D}}{n} \cdot B(t) \notin [0, 1] \right\} \\ &= \inf \left\{ t \geq 0 \mid X + B\left(\frac{D}{n^2}t\right) \notin [0, 1] \right\} \\ &= \frac{n^2}{D} \cdot \inf \left\{ t' \geq 0 \mid X + B(t') \notin [0, 1] \right\} = \frac{n^2}{D} \cdot \tau \end{aligned} \tag{75}$$

(using the change of variables  $t' := Dn^{-2} \cdot t$ ).

Hence, combining equations (72) and (75) together we conclude that

$$\tilde{S}_n(\theta) = \frac{\tanh(n\sqrt{\frac{\theta}{2D}})}{n\sqrt{\frac{\theta}{2D}}}.$$

### B.3. Selfsimilar local scan

Consider a local scan conducted by an  $\mathbb{H}$ -selfsimilar stochastic process  $Z = (Z(t))_{t \geq 0}$  with continuous sample-path trajectories. Let  $\tau_{\mathbb{H}}$  denote the first passage time of the  $\mathbb{H}$ -selfsimilar motion from the unit interval, when initiated from a random location on the unit interval.

In full analogy with equation (75) we have (the equalities below being in law)

$$\begin{aligned} S_n &= \inf \left\{ t \geq 0 \mid nX + Z(t) \notin [0, n] \right\} = \inf \left\{ t \geq 0 \mid X + \frac{1}{n} \cdot Z(t) \notin [0, 1] \right\} \\ &= \inf \left\{ t \geq 0 \mid X + Z(n^{-1/\mathbb{H}}t) \notin [0, 1] \right\} \\ &= n^{1/\mathbb{H}} \cdot \inf \left\{ t' \geq 0 \mid X + Z(t') \notin [0, 1] \right\} := n^{1/\mathbb{H}} \cdot \tau_{\mathbb{H}}, \end{aligned} \tag{76}$$

(using the change of variables  $t' := n^{-1/\mathbb{H}} \cdot t$ ).

Let  $\{\delta_n\}_{n=1}^\infty$  be an arbitrary non-negative valued sequence decaying to zero. The scaling  $S_n = n^{1/\mathbb{H}} \cdot \tau_{\mathbb{H}}$  implies that

$$\begin{aligned} n \cdot \tilde{S}_n(\lambda + \delta_n) &= n \cdot \tilde{\tau}_{\mathbb{H}}((\lambda + \delta_n)n^{1/\mathbb{H}}) = \left(\frac{k_n}{\lambda + \delta_n}\right)^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k_n) \\ &= \left(\frac{1}{\lambda + \delta_n}\right)^{\mathbb{H}} \cdot (k_n^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k_n)), \end{aligned} \tag{77}$$

where  $k_n := (\lambda + \delta_n)n^{1/\mathbb{H}}$ . Hence, if the limit  $\lim_{k \rightarrow \infty} k^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k) := c_{\mathbb{H}}$  exists then equation (77) yields

$$\lim_{n \rightarrow \infty} n \cdot \tilde{S}_n(\lambda + \delta_n) = \frac{1}{\lambda^{\mathbb{H}}} \cdot \lim_{k \rightarrow \infty} k^{\mathbb{H}} \cdot \tilde{\tau}_{\mathbb{H}}(k) = \frac{c_{\mathbb{H}}}{\lambda^{\mathbb{H}}}.$$

### B.4. Massively parallel searching of stationary systems

In the ‘stationary setting’ the initialization time  $R_0$  is the residual lifetime of the random variable  $\mathcal{E} + R$ . Hence the Laplace transform of the initialization time  $R_0$  is given by [27]

$$\tilde{R}_0(\theta) = \frac{1 - \widetilde{(\mathcal{E} + R)}(\theta)}{\mathbf{E}[\mathcal{E} + R] \cdot \theta} = \frac{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)}{(\frac{1}{\lambda} + \mathbf{E}[R]) \cdot \theta} \tag{78}$$



( $\theta \geq 0$ ). Substituting equation (78) and  $\psi(\lambda) = c/\lambda^\gamma$  into equation (11) yields

$$\begin{aligned} \int_0^\infty \exp\{-\theta t\} \Phi(t) dt &= \kappa \frac{\tilde{R}_0(\theta)}{\theta} \cdot \frac{l + \psi(\lambda + \theta)}{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)} = \kappa \frac{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)}{\left(\frac{1}{\lambda} + \mathbf{E}[R]\right) \theta^2} \cdot \frac{l + \frac{c}{(\lambda + \theta)^\gamma}}{1 - \frac{\lambda}{\lambda + \theta} \tilde{R}(\theta)} \\ &= \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ \frac{l}{\theta^2} + \frac{c}{\lambda^\gamma} \frac{1}{\theta^2} \left(\frac{\lambda}{\lambda + \theta}\right)^\gamma \right\}. \end{aligned} \tag{79}$$

Now,

(1)

$$\frac{1}{\theta^2} = \int_0^\infty \exp\{-\theta t\} f_1(t) dt \tag{80}$$

where  $f_1(t) = t$ ,

(2)

$$\left(\frac{\lambda}{\lambda + \theta}\right)^\gamma = \int_0^\infty \exp\{-\theta t\} f_2(t) dt \tag{81}$$

where  $f_2(t) = (\lambda^\gamma / \Gamma(\gamma)) \exp\{-\lambda t\} t^{1-\gamma}$  is the probability density function of the Gamma( $\lambda, \gamma$ ) distribution, and

(3)

$$\frac{1}{\theta^2} \left(\frac{\lambda}{\lambda + \theta}\right)^\gamma = \int_0^\infty \exp\{-\theta t\} f_3(t) dt \tag{82}$$

where  $f_3(t)$  is the convolution of the functions  $f_1(t)$  and  $f_2(t)$ . Hence

$$f_3(t) = \int_0^t f_1(t-s) f_2(s) ds = \int_0^\infty (t-s)_+ f_2(s) ds = \mathbf{E} \left[ (t - \mathcal{G}_{\lambda, \gamma})_+ \right], \tag{83}$$

where  $\mathcal{G}_{\lambda, \gamma}$  is a Gamma( $\lambda, \gamma$ )-distributed random variable (and where  $(x)_+ := \max\{0, x\}$  denotes the positive part of a real number  $x$ ).

Combining the above together yields

$$\begin{aligned} \int_0^\infty \exp\{-\theta t\} \Phi(t) dt &= \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ \frac{l}{\theta^2} + \frac{c}{\lambda^\gamma} \frac{1}{\theta^2} \left(\frac{\lambda}{\lambda + \theta}\right)^\gamma \right\} \\ &= \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ l \cdot \int_0^\infty \exp\{-\theta t\} f_1(t) dt + \frac{c}{\lambda^\gamma} \cdot \int_0^\infty \exp\{-\theta t\} f_3(t) dt \right\} \\ &= \int_0^\infty \exp\{-\theta t\} \left( \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ l \cdot f_1(t) + \frac{c}{\lambda^\gamma} \cdot f_3(t) \right\} \right) dt, \end{aligned} \tag{84}$$

which, in turn, implies that

$$\Phi(t) = \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ l \cdot f_1(t) + \frac{c}{\lambda^\gamma} \cdot f_3(t) \right\} = \frac{\kappa \lambda}{1 + \mathbf{E}[R] \lambda} \left\{ l \cdot t + \frac{c}{\lambda^\gamma} \cdot \mathbf{E} \left[ (t - \mathcal{G}_{\lambda, \gamma})_+ \right] \right\}. \tag{85}$$

We now turn to prove the assertions made in section 7 regarding the function  $\Phi(t)$  and its piecewise-linear support  $\Phi_{\text{sup}}(t)$ .

*Monotonicity and convexity*

Differentiating the function  $\Phi(t)$  yields

$$\Phi'(t) = \frac{\kappa\lambda}{1 + \mathbf{E}[R]\lambda} \left\{ l + \frac{c}{\lambda^\gamma} \cdot \mathbf{P}(\mathcal{G}_{\lambda,\gamma} \leq t) \right\} > 0 \tag{86}$$

- implying that the function  $\Phi(t)$  is monotonically increasing, and

$$\Phi''(t) = \frac{c\kappa\lambda}{\Gamma(\gamma)(1 + \mathbf{E}[R]\lambda)} \exp\{-\lambda t\} t^{1-\gamma} > 0 \tag{87}$$

- implying that the function  $\Phi(t)$  is convex.

*Support from below*

Jensen's inequality (see, for example, [31]) gives

$$\mathbf{E} \left[ (t - \mathcal{G}_{\lambda,\gamma})_+ \right] \geq (t - \mathbf{E}[\mathcal{G}_{\lambda,\gamma}])_+ = \left( t - \frac{\gamma}{\lambda} \right)_+, \tag{88}$$

which, in turn, implies that the function  $\Phi(t)$  is supported from below as follows:

$$\begin{aligned} \Phi(t) &= \frac{\kappa\lambda}{1 + \mathbf{E}[R]\lambda} \left\{ l \cdot t + \frac{c}{\lambda^\gamma} \cdot \mathbf{E} \left[ (t - \mathcal{G}_{\lambda,\gamma})_+ \right] \right\} \\ &\geq \frac{\kappa\lambda}{1 + \mathbf{E}[R]\lambda} \left\{ l \cdot t + \frac{c}{\lambda^\gamma} \cdot \left( t - \frac{\gamma}{\lambda} \right)_+ \right\} := \Phi_{\text{sup}}(t). \end{aligned} \tag{89}$$

*The support function  $\Phi_{\text{sup}}(t)$*

Set  $\Delta(t) := \Phi(t) - \Phi_{\text{sup}}(t)$ . By the definition of the functions  $\Phi(t)$  and  $\Phi_{\text{sup}}(t)$  we have

$$\Delta(t) = c \left( \mathbf{E} \left[ (t - \mathcal{G}_{\lambda,\gamma})_+ \right] - (t - \mu)_+ \right) \tag{90}$$

and

$$\Delta'(t) = c \left( \mathbf{P}(\mathcal{G}_{\lambda,\gamma} \leq t) - \mathbf{I}\{t - \mu\} \right), \tag{91}$$

where  $c_1 := \kappa c \lambda^{1-\gamma} / (1 + \mathbf{E}[R]\lambda)$  and  $\mu := \mathbf{E}[\mathcal{G}_{\lambda,\gamma}] = \gamma/\lambda$  (and where  $\mathbf{I}\{E\}$  denotes the indicator function of the event  $E$ ).

Now we have the following.

- (1)  $\Delta(0) = 0$  and  $\Delta'(0) = 0$ . Hence, the support function  $\Phi_{\text{sup}}(t)$  is tangent to the function  $\Phi(t)$  at the origin.
- (2)

$$\begin{aligned} \lim_{t \rightarrow \infty} \Delta(t) &= c_1 \lim_{t \rightarrow \infty} \left( \mathbf{E} \left[ (t - \mathcal{G}_{\lambda,\gamma})_+ \right] - (t - \mu)_+ \right) \\ &= c_1 \lim_{t \rightarrow \infty} \left( \left( t - \int_0^t \mathbf{P}(\mathcal{G}_{\lambda,\gamma} > s) \, ds \right) - (t - \mu) \right) \\ &= c_1 \lim_{t \rightarrow \infty} \left( \mu - \int_0^t \mathbf{P}(\mathcal{G}_{\lambda,\gamma} > s) \, ds \right) \\ &= c_1 \left( \mu - \int_0^\infty \mathbf{P}(\mathcal{G}_{\lambda,\gamma} > s) \, ds \right) = c_1 (\mu - \mathbf{E}[\mathcal{G}_{\lambda,\gamma}]) = 0 \end{aligned} \tag{92}$$

and

$$\lim_{t \rightarrow \infty} \Delta'(t) = c_1 \lim_{t \rightarrow \infty} \left( \mathbf{P}(\mathcal{G}_{\lambda,\gamma} \leq t) - \mathbf{I}\{t - \mu\} \right) = 0. \tag{93}$$

Hence, the support function  $\Phi_{\text{sup}}(t)$  is tangent to the function  $\Phi(t)$  at infinity.

(3)

$$\Delta'(t) = c_1 \cdot \begin{cases} \mathbf{P}(\mathcal{G}_{\lambda,\gamma} \leq t) & t < \mu \\ -\mathbf{P}(\mathcal{G}_{\lambda,\gamma} > t) & t > \mu \end{cases} \quad (94)$$

and thus the function  $\Delta(t)$  is monotonically increasing in the interval  $(0, \mu)$ , and monotonically decreasing in the interval  $(\mu, \infty)$ . Hence, the global maximum of the function  $\Delta(t)$  is attained at the point  $t = \mu$  and its value is

$$\Delta(\mu) = \frac{\kappa c \lambda^{1-\gamma}}{1 + \mathbf{E}[R]\lambda} \cdot \mathbf{E} \left[ \left( \frac{\gamma}{\lambda} - \mathcal{G}_{\lambda,\gamma} \right)_+ \right]. \quad (95)$$

*Upper bound for  $\mathbf{E}[T_*]$*

Using the limiting distribution of proposition 4, and the inequality  $\Phi(t) \geq \Phi_{\text{sup}}(t)$ , we obtain that

$$\mathbf{E}[T_*] = \int_0^\infty \mathbf{P}(T_* > t) dt = \int_0^\infty \exp\{-\Phi(t)\} dt \leq \int_0^\infty \exp\{-\Phi_{\text{sup}}(t)\} dt. \quad (96)$$

Since the support function  $\Phi_{\text{sup}}(t)$  is piecewise linear we further have (using the shorthand notation  $c_1 := \kappa c \lambda^{1-\gamma} / (1 + \mathbf{E}[R]\lambda)$  and  $c_2 := \kappa l \lambda / (1 + \mathbf{E}[R]\lambda)$ )

$$\begin{aligned} & \int_0^\infty \exp\{-\Phi_{\text{sup}}(t)\} dt \\ &= \int_0^\mu \exp\{-c_2 t\} dt + \int_\mu^\infty \exp\{-(c_2 t + c_1(t - \mu))\} dt \\ &= \frac{1}{c_2} (1 - \exp\{-c_2 \mu\}) + \frac{1}{c_2 + c_1} \exp\{-c_2 \mu\} = \frac{1}{c_2} \left( 1 - \frac{c_1}{c_2 + c_1} \exp\{-c_2 \mu\} \right) \\ &= \frac{1}{\kappa l} \left( \mathbf{E}[R] + \frac{1}{\lambda} \right) \left( 1 - \frac{c}{c + l \lambda^\gamma} \exp\left\{ -\frac{\kappa l \gamma}{1 + \mathbf{E}[R]\lambda} \right\} \right). \end{aligned} \quad (97)$$

## References

- [1] Riggs A D, Bourgeois S and Cohn M 1970 *J. Mol. Biol.* **53** 401
- [2] von Hippel P H and Berg O G 1989 *J. Biol. Chem.* **264** 675
- [3] Shimamoto N 1999 *J. Biol. Chem.* **274** 15293
- [4] Bustamante C *et al* 1999 *J. Biol. Chem.* **274** 16665
- [5] Misteli T 2001 *Science* **291** 843
- [6] Halford S E and Szczelkun M D 2002 *Eur. Biophys. J.* **31** 257
- [7] Halford S E and Marko J F 2004 *Nucleic Acids Res.* **32** 3040
- [8] Gowers D M, Wilson G G and Halford S E 2005 *Proc. Natl Acad. Sci.* **102** 15883
- [9] Widom J 2005 *Proc. Natl Acad. Sci.* **102** 16909
- [10] Blainey P C *et al* 2006 *Proc. Natl Acad. Sci.* **103** 5752
- [11] Richter P H and Eigen M 1974 *Biophys. Chem.* **2** 255
- [12] Coppey M *et al* 2004 *Biophys. J.* **87** 1640
- [13] Slutsky M and Mirny L A 2004 *Biophys. J.* **87** 4021
- [14] Lomholt M A, Ambjörnsson T and Metzler R 2005 *Phys. Rev. Lett.* **95** 260603
- [15] Sokolov I M *et al* 2005 *Biophys. J.* **89** 895
- [16] Hu T, Grosberg A Y and Shklovskii B I 2006 *Biophys. J.* **90** 2731
- [17] Vologodskii A V 2004 *Circular DNA* Biophysics textbook online: <http://www.biophysics.org/education/vologodskii.pdf>
- [18] Levandowsky M, Klafter J and White B S 1988 *Bull. Mater. Sci.* **43** 758
- [19] Viswanathan G M *et al* 1999 *Nature* **401** 911
- [20] Bénichou O *et al* 2005 *Phys. Rev. Lett.* **94** 198101
- [21] Sokolov I M *et al* 2005 *Phys. Rev. E* **72** 041102

- 
- [22] Golding I and Cox E C 2006 *Phys. Rev. Lett.* **96** 098102
  - [23] Shlesinger M F and Montroll E W 1984 *Proc. Natl Acad. Sci.* **81** 1280
  - [24] Blumen A, Zumofen G and Klafter J 1984 *Phys. Rev. B* **80** 5379
  - [25] Embrechts P and Maejima M 2002 *Selfsimilar Processes* (Princeton, NJ: Princeton University Press)
  - [26] Eliazar I and Klafter J 2004 *Physica D* **187** 30
  - [27] Ross S M 1970 *Applied Probability Models with Optimization Applications* (San Francisco, CA: Holden-Day)
  - [28] Bingham N H and Kiesel R 2004 *Risk-neutral Valuation: Pricing and Hedging of Financial Derivatives* 2nd edn (Berlin: Springer)
  - [29] Bingham N H, Goldie C M and Teugels J L 1987 *Regular Variation* (Cambridge: Cambridge University Press)
  - [30] Oksendal B 1995 *Stochastic Differential Equations* (Berlin: Springer)
  - [31] Feller W 1971 *An Introduction to Probability Theory and its Applications* 2nd edn, vol 2 (New York: Wiley)