

Single-cell directional sensing at ultra-low chemoattractant concentrations from extreme first-passage events

Vincent Fiorino¹, Sean D. Lawley², Alan E. Lindsay^{1*}

^{1*}Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, 46656, IN, USA.

²Department of Mathematics, University of Utah, Salt Lake City, 84112, UT, USA.

*Corresponding author(s). E-mail(s): a.lindsay@nd.edu;
Contributing authors: vfiorino@nd.edu; lawley@utah.edu;

Abstract

We investigate single-cell directional sensing from diffusing chemoattractant signals released by a localized source. We focus on the low-concentration regime in which receptor activity is discrete and cellular decisions are made on timescales far shorter than those required for steady-state concentration profiles or receptor saturation to emerge. We derive analytic expressions for the joint distribution of receptor binding times and binding locations, conditional on the position of the source. We show that early binding events carry disproportionately more information about source directionality than later arrivals. Motivated by this observation, we propose and analyze several source-localization estimates that exploit early receptor binding statistics. For a directional estimate formed from an equally weighted average of the earliest impacts, we demonstrate that this estimate is accurate even from a small number of impacts, and derive an expression for the number of binding events that minimizes the mean squared error. Our results demonstrate that cells possess sufficient information to rapidly and accurately infer the directionality of sparse signals from a localized source through a small number of discrete receptor engagements, well before a steady-state gradient has been established.

Keywords: Cell motility; Directional sensing; First passage times; Extreme statistics.

MSC Classification: 60G70 , 62G32 , 62P10 , 92C17.

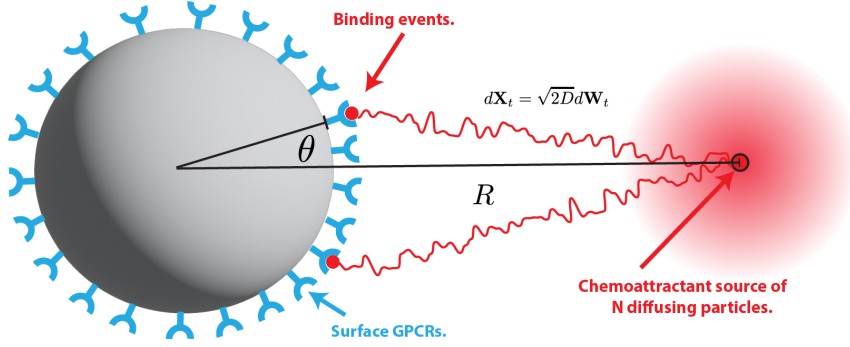


Fig. 1 Schematic of source detection from the discrete arrival of signaling molecules to a circular cell. At $t = 0$, a source located $R > 1$ cell lengths away emits N diffusing molecules of diffusivity $D > 0$. These signaling molecules diffuse freely before being absorbed at a surface receptor (GPCR). The extreme directional sensing problem is the following: given arrival times $\{T_{j,N}\}_{j=1}^k$ and corresponding polar locations $\{\theta_{j,N}\}_{j=1}^k$ of the first k impacts based on N initial walkers, what estimates can the cell form for the directionality of the source?

1 Introduction

The aim of this paper is to explore the theoretical limits by which cells can establish the directionality of chemical cues at ultra-low chemoattractant concentrations. A key cellular function is the ability to rapidly and accurately interpret the directionality of chemical or electric gradients to inform decisions on where to move, when to divide or whether to initiate a response to a threat [1–6]. The information regulating these decisions is encoded by the noisy binding activity of diffusing signals at membrane bound receptors [7–12].

The ability to orient towards a source has been demonstrated across many model systems, including yeast [2, 13, 14], Dictyostelium [3] and neutrophils [15–17]. In the case of chemotaxis in neutrophils, numerous experimental studies [15, 16, 18] have demonstrated that cells rapidly identify the directionality of an instantaneous chemoattractant source (administered by pipette). For example, neutrophils demonstrate asymmetric recruitment of membrane components within five seconds of exposure to a $10\mu\text{M}$ concentration of the chemoattractant FMLP presented by a micropipette at around two cell radii distance [18, 19]. The timescale of such a response is well before a steady-state gradient is established [20].

A commonality across these systems is the role of membrane bound G protein-coupled receptors (GPCRs) which bind to signaling molecules [21] and transmit information to downstream signaling cascades. These “sensors” therefore act as intermediaries between extracellular chemoattractants and downstream pathways which either inhibit or initiate necessary cellular responses.

Previous experiments utilized chemoattractant concentrations which saturated GPCRs. However, as the precision of experimental measurement has expanded, it is now clear that GPCRs can be activated at remarkably low levels of chemoattractant concentrations [22], including at the femtomolar (10^{-15} M) and attomolar (10^{-18} M) levels [23, 24]. At such low concentration levels, receptor engagement and consequently

cellular signaling decision making, are regulated by a small number of discrete binding events governed by extreme first passage time statistics. This work expands upon several recent studies that have employed extreme first passage time theory to understand decision making processes in biological systems [25–28].

The aim of this paper is to give a mathematical analysis of the directional information that is encoded in signals associated with such low concentrations of chemoattractant. To this end, we consider an idealized non-dimensional scenario akin to the experiments of [18, 19] where a circular cell (normalized to unit radius) is centered at the origin while a source of diffusing chemoattractant is placed at $\mathbf{x}_0 = R[\cos \theta_0, \sin \theta_0]$ for $R > 1$ (see Fig. 1). The quantity R represents the multiple of the cell radius at which the source is placed. Intuitively, we expect that detection at larger values of R is more challenging than source detection for values $R \approx 1$. A key element of our approach in the low chemoattractant limit is to consider the signaling molecules to be discrete. In this setting we assume at the initial time $t = 0$ that N molecules are released from the source and undergo unbiased random motion before eventually reaching a receptor on the cell surface. As an example of typical values, a release of 1mL of chemoattractant at femtomolar to attomolar concentrations would represent a release of approximately $N \approx 10^3 - 10^6$ molecules. In our analysis we will assume a dense covering of GPCRs so that a continuum approximation is valid.

Each of the N molecules released at the source eventually reaches the cell surface and binds a receptor, producing two observable quantities: the time of the binding event and its location on the cell surface. Directional sensing is then an inverse problem: from the earliest binding events, what can the cell infer about the source? We will show that the two observables carry complementary information. The arrival times reflect the source distance, while the arrival locations reflect the source direction. We introduce the formal notation for the sets of arrival times and locations in Section 2, where the extreme-value statistics of the earliest arrivals are developed.

The most pertinent biological information on the source is its direction relative to the cell. It was noticed in [20] that when recovering the location of a diffusive source from the time-dependent signal, the short time signal ($0 < t \ll 1$) gave a strong estimate on the directionality of the source but yielded poor information on the distance. Heuristically, this is because molecules which are quickly absorbed at the cell surface have likely taken very direct or near-ballistic paths. This idea was further explored in [29] which showed in a three dimensional setting that the directional information is contained primarily in a small fraction of the earliest signals received by the cell. Impacts that arrive after sufficiently long sojourns have largely lost information about their initial location and therefore dilute the directional information carried by earlier arrivals. This observation highlights a challenge in recovering source direction from equilibrium quantities (such as splitting probabilities), since such estimators necessarily average over signals of lower informational quality [30–32]. By using only the early fraction of the signal, it was shown in [29] that a simple average of just a few impact locations can provide an accurate directional estimate. Here we build on these studies by deriving the joint statistics of early receptor binding times and impact locations in 2D.

We find that the information a cell can extract about the source separates into two channels: the times of early binding events, which encode the source distance R , and the locations of those events, which encode the source direction θ_0 . Section 3 asks how accurately the source distance can be inferred from arrival times alone, using maximum-likelihood estimation and the associated Fisher information both to quantify the best achievable accuracy and to compare candidate estimators. Sections 4–5 turn to the directional problem, deriving the distribution of arrival locations and analyzing estimators of source direction. In both cases we quantify the achievable accuracy as a function of the number k of early binding events.

The main mathematical result of this work is a derivation of the limiting distribution of the first k arrivals in this extreme statistics regime. Specifically, in Sec. 4, we find that the impact location of the k^{th} arrival to a circular cell has the limiting behavior

$$\theta_{k,N} \sim \mathcal{N}(\theta_0, \sigma_{k,N}^2), \quad \sigma_{k,N}^2 = \frac{2D}{R} (b_N + a_N \log k), \quad (1)$$

as $N \rightarrow \infty$. Here a_N and b_N are quantities found in terms of the limiting properties of the survival probability for a single particle which we also obtain. For the specific case of an all absorbing circular cell of unit radius, the leading order limiting behavior is found to be

$$\theta_{k,N} \sim \mathcal{N}(\theta_0, \sigma_{k,N}^2), \quad \sigma_{k,N}^2 \sim \frac{(R-1)^2}{RW} \left(1 + \frac{2 \log k}{1+W}\right), \quad W = W_0\left(\frac{2N^2}{\pi R}\right), \quad (2)$$

as $N \rightarrow \infty$. This behavior is based on the limiting forms $a_N \sim (R-1)^2/(DW(1+W))$ and $b_N \sim (R-1)^2/(2DW)$ as $N \rightarrow \infty$. Here $W_0(z)$ is the principal branch of the LambertW function [33], which has behavior $W_0(z) = \log z - \log \log z + \dots$ as $z \rightarrow \infty$. Hence $W_0 \sim 2 \log N$ as $N \rightarrow \infty$. The result reveals the limits on the accuracy of directional inference based on the first few early arrivals together with the dependence on key parameters, including source distance (R), number of early arrivals (k) and the strength of the initial signal (N). To our knowledge, this is the first derivation and analysis of the joint asymptotic distribution of early hitting times and impact locations for diffusive capture in two dimensions.

A surprising consequence of (2) is that the breadth of angular space sampled by the early arrivals is independent of the single particle diffusivity D . We will demonstrate that a larger diffusivity delivers the k^{th} impact proportionally sooner by exactly the factor by which it broadens the angular distribution. Hence the parameter D sets only the overall timescale of capture, not the geometry of the early-impact distribution.

Using the detailed statistical properties of the early arrivals, we investigate the accuracy of several estimates that a cell could form to infer the source direction. We find that the impact times largely inform on the source distance while the impact locations encode the directionality. Our analysis shows that a cell can employ simple averaging over the first k arrivals to reduce the variance (error) in the accuracy.

Maximum likelihood estimation (MLE) can be used to form more accurate estimates of the source [34]. However, implementation of these estimators necessitates additional computational capabilities, such as the solution of nonlinear equations. Our analysis

is able to quantify the accuracy in a range of estimators, from simple averages to relatively sophisticated MLEs. As an example of our theory, we analyze the simple average of the first k impacts given by $\mathbf{n}^{\text{res}} = \frac{1}{k} \sum_{j=1}^k (\cos \theta_j, \sin \theta_j)$. We find that the error $\rho_k^{\text{res}} = 1 - \mathbf{n}^{\text{res}} \cdot (\cos \theta_0, \sin \theta_0)$ associated with this estimate has mean and variance

$$\mathbb{E}[\rho_k^{\text{res}}] \approx \frac{D}{R} \left(b_N + a_N (\log k - 1) \right), \quad (3a)$$

$$\text{Var}[\rho_k^{\text{res}}] \approx \frac{2D^2}{R^2 k} \left((a_N \log k + b_N - a_N)^2 + a_N^2 \right). \quad (3b)$$

The error (3a) grows with k since the first impact has the lowest error while subsequent individual arrivals have logarithmically growing error. However, while the error increases in k , the variance in the error (3b) decreases. These opposing trends are a bias–variance trade-off. By considering the mean squared error $\mathbb{E}[(\rho_k^{\text{res}})^2]$, we show that this trade off results in an optimal $k = k^*$ which satisfies a transcendental equation. As $N \rightarrow \infty$, this critical number has asymptotic behavior $k^* \sim \log N$. For $k > k^*$, additional impacts produce a larger error in the resultant $\mathbf{n}^{\text{res}}$. Hence an accurate estimate of the source direction can be formed from relatively few impacts k .

The remainder of the article is structured as follows. In Sec. 2 we describe preliminary results on the distributions of extreme arrival times, including those of individual arrivals and the joint distribution of k arrivals $\mathcal{T}_N$ and their differences. We emphasize that it is necessary to first describe the statistics of arrival times before arrival locations since the former depends on the latter. In Sec. 3 we introduce and analyze several estimates of the source distance using only the arrival times. These quantities primarily encode information on the distance of the source. In Sec. 4, we derive the distribution (1) of angular arrivals to a circular cell in the case of both Dirichlet and Robin boundary conditions. In Sec. 5, we use the result (1) to investigate the accuracy of several estimators of the source direction. Finally, in Sec. 6 we discuss future avenues of investigation that arise from this study.

2 Statistics of extreme first passage times.

In this section, we introduce key results on the distribution of the set of extreme hitting times which have been a topic of significant recent research in the literature [25, 35–37]. It is necessary to characterize the hitting times first before the hitting locations. This is due to the fact that particles arriving quickly to the cell have hitting distributions tightly focused on the source direction, while particles that undergo longer sojourns are distributed more evenly over the cell surface.

We denote the arrival times of the N molecules by $\{T_{j,N}\}_{j=1}^N$ and their arrival locations on the circular cell by $\{(\cos \theta_{j,N}, \sin \theta_{j,N})\}_{j=1}^N$, where the index $j = 1, \dots, N$ labels the molecules and N is the number released.

For an independent and identically distributed (i.i.d.) sequence of first passage times $\{\tau_1, \dots, \tau_N\}$ with survival probability $S(t) := \mathbb{P}(\tau_1 > t)$, the first arrival is defined as

$$T_{1,N} := \min\{\tau_1, \dots, \tau_N\}.$$

The i.i.d. property of the times means that

$$\mathbb{P}(T_{1,N} > \tau) = \mathbb{P}(\tau_1 > \tau) \times \cdots \times \mathbb{P}(\tau_N > \tau) = [S(\tau)]^N.$$

For $N \gg 1$ and $0 \leq S(\tau) \leq 1$, the mass becomes concentrated near $t = 0$. Hence, describing extreme statistics requires detailed knowledge of the short time asymptotic behavior of $S(t)$ as $t \rightarrow 0^+$. For a large class of diffusion problems in various dimensions, geometric settings and boundary conditions [35, 38] this behavior can be mapped onto the family of functions

$$1 - S(t) \sim At^p e^{-\frac{C}{t}}, \quad \text{as } t \rightarrow 0^+. \quad (4)$$

Here A , C , p are constants with A and C being strictly positive. A simple example we will make extensive use of is diffusion on the half line with release at $x = R$ and capture at $x = 0$. The survival probability is $S(t) = 1 - \operatorname{erfc}(\frac{R}{\sqrt{4Dt}})$ and the limiting behavior as $t \rightarrow 0^+$ is given by

$$1 - S(t) \sim \sqrt{\frac{4Dt}{\pi R^2}} e^{-\frac{R^2}{4Dt}}, \quad \text{as } t \rightarrow 0^+.$$

More generally, the k^{th} fastest arrival time at the cell is defined as

$$T_{k,N} := \min \{ \{\tau_1, \dots, \tau_N\} \setminus \cup_{j=1}^{k-1} \{T_{j,N}\} \}, \quad k \in \{1, \dots, N\}. \quad (5a)$$

Refined statistics on the distributions of the k^{th} arrival $T_{k,N}$ have recently been determined in the limit as $N \rightarrow \infty$. A key result [39, Thm. 4] utilized in the present work is the following convergence in distribution,

$$\left(\frac{T_{1,N} - b_N}{a_N}, \dots, \frac{T_{k,N} - b_N}{a_N} \right) \rightarrow_d \mathbf{X}^{(k)} := (X_1, \dots, X_k) \in \mathbb{R}^k, \quad \text{as } N \rightarrow \infty. \quad (5b)$$

Here the joint probability density function of $\mathbf{X}^{(k)} \in \mathbb{R}^k$ is

$$f_{\mathbf{X}^{(k)}}(x_1, \dots, x_k) = \begin{cases} \exp(-e^{x_k}) \prod_{r=1}^k e^{x_r}, & x_1 \leq \dots \leq x_k; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

The scale parameters a_N and b_N set the distribution of the early arrival times in the limiting law (5). The constants a_N and b_N are defined as

$$b_N = \frac{C}{\log(AN)}, \quad a_N = \frac{b_N}{\alpha}, \quad \alpha = \log(AN), \quad \text{if } p = 0; \quad (7a)$$

$$b_N = \frac{C}{pW}, \quad a_N = \frac{b_N}{\alpha}, \quad \alpha = p(1+W), \quad \text{if } p \neq 0. \quad (7b)$$

We will show these quantities are defined in terms of the source distance and diffusivity so that estimating a_N from the arrival times is therefore equivalent to estimating the source distance R for fixed D . The constant W is defined as

$$W = \begin{cases} W_0((C/p)(AN)^{1/p}), & p > 0, \\ W_{-1}((C/p)(AN)^{1/p}), & p < 0, \end{cases} \quad (8)$$

where $W_0(z)$ and $W_{-1}(z)$ denote the principal and lower branches of the LambertW function respectively. The principal branch of the LambertW function has large argument asymptotic representation [33]

$$W_0(z) = \log z - \log \log z + \mathcal{O}\left(\frac{\log \log z}{\log z}\right) + \dots \quad \text{as } z \rightarrow \infty. \quad (9)$$

In the case $p > 0$, we will often invoke the leading order behavior $W \sim \frac{1}{p} \log N$ as $N \rightarrow \infty$ to obtain expressions that are less accurate but offer more biological interpretability. To retain qualitative accuracy at moderate values of N , the full form described in (8) is retained.

For the individual arrival times, we have that the k^{th} fastest arrival has the limiting behavior

$$\frac{T_{k,N} - b_N}{a_N} \rightarrow_d X_k, \quad \text{as } N \rightarrow \infty, \quad (10)$$

where X_k has the probability density function

$$f_{X_k}(x) = \frac{\exp(kx - e^x)}{(k-1)!}. \quad (11)$$

In particular, we have that (Theorem 5 of [39])

$$\mathbb{E}[T_{k,N}] = b_N + \psi(k)a_N + o(a_N), \quad (12a)$$

$$\text{Var}(T_{k,N}) = \psi'(k)a_N^2 + o(a_N^2), \quad (12b)$$

where $\psi(k) = H_{k-1} - \gamma_e$ is the digamma function, $H_k = \sum_{r=1}^k \frac{1}{r}$ is the k^{th} harmonic number, and $\gamma_e \approx 0.577$ is the Euler-Mascheroni constant. In Fig. 2(a) we show Monte-Carlo simulations that demonstrate the convergence of $T_{1,N}$ to the limiting distribution (11) for $R = 5$ and $D = 1$ in the case of the one-dimensional survival probability $S(t) = 1 - \text{erfc}(\frac{R}{\sqrt{4Dt}})$. Over the biologically relevant range of N ($10^3 - 10^6$), we observe that the two distributions have broad agreement in shape. This is in agreement with related convergence studies performed in [39]. In addition, we plot in Fig. 2(b) the quantity $\sigma_{1,N} = \sqrt{2Db_N}/R$ which characterizes the angular extent of the first impact. This quantity was stated in (1) and will be derived in Sec. 4. The key observation is that this quantity exhibits modest variations in value while the number of source particles N varies over several decades. Hence the biologically relevant range of values for N is well captured in the asymptotic limit $N \rightarrow \infty$.

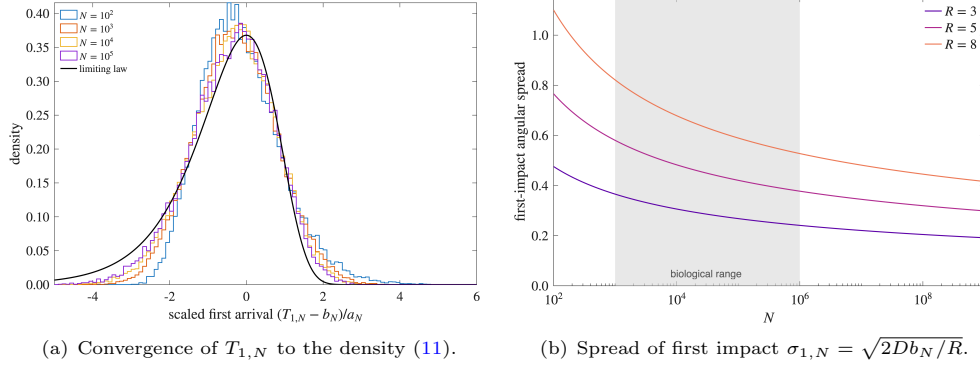


Fig. 2 Convergence to the limiting distribution as $N \rightarrow \infty$. In Panel (a), we show convergence of empirical distributions of $(T_{1,N} - b_N)/a_N$ to the limiting distribution $f_{X_1}(x) = \exp(x - \exp(x))$ for a range of N . In panel (b), we plot the angular spread (1) of the first arrival $\sigma_{1,N} = \sqrt{2Db_N/R}$ to the cell. Over several decades in N we observe that the limiting distribution accurately captures the low concentration limit and that the angular spread varies modestly, hence providing a robust measure of directional sensing capabilities.

The full record of arrival times $\{T_{j,N}\}_{j=1}^k$ is a set of dependent random variables satisfying the joint distribution (5) in the limit as $N \rightarrow \infty$. This implies that a cell would need to store a sequence of times in order to process them to reconstruct information on the source. As a consequence, it is natural to ask whether simpler measurements of time can be used for inference. To this end, we derive (see Appendix A) the distribution of the differences between arrival times. For $1 \leq k \ll N$, we have that the time differences are exponentially distributed, specifically,

$$T_{k+1,N} - T_{k,N} \sim_d \text{Exp} \left[\frac{k}{a_N} \right], \quad (13)$$

with mean a_N/k where a_N is given in (7). Hence, if we only know the interarrival times, then this value is related to the single parameter a_N which shapes the limiting distribution for $N \gg 1$.

3 Maximum likelihood estimates of source location from extreme time data

In this section, we describe a maximum likelihood estimation (MLE) process for determining the source parameters from the arrival statistics. The arrival time statistics depend on the source only through a_N which itself depends on R and D . Therefore inferring source distance reduces to estimating the single parameter a_N . The MLE and its Fisher information that we derive in this section are able to characterize the best achievable accuracy and allow for a systematic comparison of the information content by different summaries of the data: the full arrival times versus the differences.

To set the stage for our MLE approach, consider a single cellular observation of k impact times $\mathcal{T}^{(k)} = (T_1, T_2, \dots, T_k)$ based on N initial walkers. To simplify notation,

we omit in this section the dependence of both $\mathcal{T}^{(k)}$ and $a = a_N$ on N . Suppose we have n independent and identically distributed realizations $\{\mathcal{T}_1^{(k)}, \dots, \mathcal{T}_n^{(k)}\}$ of this observable, corresponding to repeated releases from a source at the same distance. Maximum likelihood estimation infers the unknown parameter a by maximizing the likelihood of the observed data. For the case of the full set of arrival times, the full likelihood is the product over the n independent observations,

$$\mathcal{L}(a|\{\mathcal{T}_j^{(k)}\}_{j=1}^n) = \prod_{j=1}^n f_{\mathcal{T}^{(k)}}(\mathcal{T}_j^{(k)}|a) \quad (14)$$

where $f_{\mathcal{T}^{(k)}}$ is the joint density of the first k arrivals for each observation and given explicitly in (6) after rescaling.

The maximum likelihood estimate $\hat{a}$ is the maximizer of (14). This method is consistent: given a sufficient number n of observations, it converges in the sense that [40]:

$$\sqrt{n}(\hat{a} - a) \rightarrow_d \mathcal{N}(0, I^{-1}), \quad I(a) = -\mathbb{E}\left[\frac{\partial^2}{\partial a^2} \log \mathcal{L}(a|\mathcal{T}^{(k)}) \middle| a\right] \quad (15)$$

as $n \rightarrow \infty$. The quantity $I(a)$ is the (per-observation) *Fisher information*, a simple quantity that determines the quality of the estimation with larger values corresponding to a more accurate estimate on the correct value. The total information from n independent observations is $nI(a)$.

In the following two sections we build estimators from two different summaries of this data. In Section 3.1 we use only the interarrival differences $T_{i+1} - T_i$, which are mutually independent (Appendix A) and yield a particularly simple estimator; in Section 3.2 we use the full arrival times through the joint likelihood (14). Comparing their Fisher information quantifies how much information about the source distance is discarded by retaining only the interarrival differences.

3.1 Assuming the cell accesses the time differences

In the scenario we consider the situation where the time interval between arrivals is exponentially distributed as given in (13). Let $\{\mathcal{T}_1^{(k)}, \mathcal{T}_2^{(k)}, \dots, \mathcal{T}_n^{(k)}\}$ be a set of n i.i.d. observations for first k arrival times after the release of N molecules at $t = 0$. Here $\mathcal{T}_j^{(k)} = (T_{1,j}, \dots, T_{k,j})$ where we have assumed the k arrival times are sorted ($T_{i,j} < T_{i+1,j}$). The dependence of N is not explicitly recorded. Under the assumption that the n observations are independent and that, within each observation, the successive interarrival times are mutually independent with $T_{i+1,j} - T_{i,j} \sim_d \text{Exp}(i/a)$ (rate $\lambda_i = i/a$; see Appendix A), the joint probability density (likelihood) of the complete set of interarrival times

$$\mathcal{L}(a|\{\mathcal{T}_j^{(k)}\}_{j=1}^n) = \prod_{j=1}^n \prod_{i=1}^{k-1} \lambda_i e^{-\lambda_i (T_{i+1,j} - T_{i,j})}, \quad \lambda_i = \frac{i}{a}.$$

For convenience, the subscript on the unknown parameter $a = a_N$ is suppressed. The negative log of the likelihood is more amenable to analysis and is given by

$$\ell(a|\{\mathcal{T}_j^{(k)}\}_{j=1}^n) = -\log \left[\mathcal{L}(a|\{\mathcal{T}_j^{(k)}\}_{j=1}^n) \right]$$

With this definition, we calculate that

$$\begin{aligned} \ell(a|\{\mathcal{T}_j^{(k)}\}_{j=1}^n) &= -\log \prod_{j=1}^n \prod_{i=1}^{k-1} \lambda_i e^{-\lambda_i (T_{i+1,j} - T_{i,j})} \\ &= \sum_{j=1}^n \sum_{i=1}^{k-1} \left[\log a + \frac{i}{a} (T_{i+1,j} - T_{i,j}) - \log i \right] \\ &= n(k-1) \log a + \frac{1}{a} \sum_{j=1}^n \sum_{i=1}^{k-1} i (T_{i+1,j} - T_{i,j}) - n \sum_{i=1}^{k-1} \log i. \end{aligned}$$

The first and second derivatives with respect to a yield the score function and the Fisher information respectively. These quantities are given by

$$\frac{\partial \ell}{\partial a} = \frac{n(k-1)}{a} - \frac{1}{a^2} \sum_{j=1}^n \sum_{i=1}^{k-1} i (T_{i+1,j} - T_{i,j}) \quad (16a)$$

$$\frac{\partial^2 \ell}{\partial a^2} = -\frac{n(k-1)}{a^2} + \frac{2}{a^3} \sum_{j=1}^n \sum_{i=1}^{k-1} i (T_{i+1,j} - T_{i,j}). \quad (16b)$$

The zeros of the score function yield the MLE estimate $\hat{a}$ given by

$$\hat{a}_n = \frac{1}{n(k-1)} \sum_{j=1}^n \sum_{i=1}^{k-1} i (T_{i+1,j} - T_{i,j}). \quad (17)$$

The Fisher information of the full observation set is given by

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial a^2} \right] = \frac{n(k-1)}{a^2}, \quad (18)$$

so that the per-observation information from a single independent observation is

$$I_{\text{diff}}(a) = \frac{k-1}{a^2}.$$

To obtain (18), we have calculated that

$$\mathbb{E} \left[\sum_{j=1}^n \sum_{i=1}^{k-1} i(T_{i+1,j} - T_{i,j}) \right] = na(k-1),$$

where from (12a), we have used that $\mathbb{E}[T_{k,j}] = \mathbb{E}[T_{1,j}] + H_{k-1}a$. Hence, applying the result (15), we conclude that the estimate (17) satisfies

$$\sqrt{n}(\hat{a}_n - a) \rightarrow_d \mathcal{N}\left(0, \frac{a^2}{k-1}\right), \quad \text{as } n \rightarrow \infty. \quad (19)$$

As a confirmation of the result (19), we simulated $n = 10^3$ arrival events to create an estimate for $\hat{a}$. This process was repeated over $M = 10^4$ trials to build a distribution for these estimates. In Fig. 3 we show excellent agreement between the theoretically predicted distribution (solid green) and that obtained by simulation (green histogram).

3.2 Estimation using the full times

In this section, we describe a scenario where the cell can access the entire set of arrival times. The likelihood function of n independent arrival times $\{T_{1,j}, T_{2,j}, \dots, T_{k,j}\}$ for $j = 1, \dots, n$, is given by

$$\mathcal{L}(a | \{\mathcal{T}_j^{(k)}\}_{j=1}^n) = \prod_{j=1}^n \frac{e^{-e^{x_{k,j}}}}{a^k} \prod_{i=1}^k e^{x_{i,j}}, \quad x_{i,j} = \frac{T_{i,j} - \alpha a}{a},$$

where $\alpha = p(1+W)$ is assumed constant. The negative log likelihood is given by

$$\ell = -\log \left(\prod_{j=1}^n \frac{e^{-e^{x_{k,j}}}}{a^k} \prod_{i=1}^k e^{x_{i,j}} \right) = nk \log a + \sum_{j=1}^n e^{\frac{T_{k,j} - \alpha a}{a}} - \sum_{j=1}^n \sum_{i=1}^k \frac{T_{i,j} - \alpha a}{a}. \quad (20)$$

The first and second derivatives of ℓ are calculated to be

$$\begin{aligned} \frac{\partial \ell}{\partial a} &= \frac{nk}{a} - \frac{1}{a^2} \sum_{j=1}^n T_{k,j} e^{\frac{T_{k,j} - \alpha a}{a}} + \frac{1}{a^2} \sum_{j=1}^n \sum_{i=1}^k T_{i,j}, \\ \frac{\partial^2 \ell}{\partial a^2} &= -\frac{nk}{a^2} + \frac{1}{a^4} \sum_{j=1}^n (T_{k,j})^2 e^{\frac{T_{k,j} - \alpha a}{a}} + \frac{2}{a^3} \sum_{j=1}^n T_{k,j} e^{\frac{T_{k,j} - \alpha a}{a}} - \frac{2}{a^3} \sum_{j=1}^n \sum_{i=1}^k T_{i,j}. \end{aligned}$$

The zeros of the score function yield the MLE estimate $\hat{a}$ which is given by

$$0 = \hat{a} - \frac{1}{nk} \sum_{j=1}^n T_{k,j} e^{\frac{T_{k,j} - \alpha \hat{a}}{\hat{a}}} + \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k T_{i,j}. \quad (22)$$

Introducing the change of variable $z = \frac{1}{a}$ reduces (22) to

$$e^\alpha - \frac{z}{nk} \sum_{j=1}^n T_{k,j} e^{zT_{k,j}} + \frac{z}{nk} e^\alpha \sum_{j=1}^n \sum_{i=1}^k T_{i,j} = 0. \quad (23)$$

In this case, the equation (23) is nonlinear and we utilize a bisection method to access its roots $\hat{a}_n$. The Fisher information is

$$\begin{aligned} I(a) &= \mathbb{E} \left[\frac{\partial^2}{\partial a^2} \ell(\mathcal{T}^{(k)}; a) \middle| a \right] \\ &= \mathbb{E} \left[\sum_{j=1}^n \left(\left(\frac{T_{k,j}^2}{a^4} + \frac{2T_{k,j}}{a^3} \right) e^{\frac{T_{k,j} - \alpha a}{a}} \right) \right] - \frac{2}{a^3} \mathbb{E} \left[\sum_{j=1}^n \sum_{i=1}^k T_{i,j} \right] - \frac{nk}{a^2} \\ &= n \underbrace{\mathbb{E} \left[\left(\left(\frac{T_k^2}{a^4} + \frac{2T_k}{a^3} \right) e^{\frac{T_k - \alpha a}{a}} \right) \right]}_{I_1} - \frac{2n}{a^3} \underbrace{\mathbb{E} \left[\sum_{i=1}^k T_i \right]}_{I_2} - \frac{nk}{a^2}. \end{aligned} \quad (24)$$

We now proceed to calculate the expectations I_1 and I_2 separately.

Calculation of I_1 :

The probability density function of T_k , given by (11), is

$$\mathbb{P}(T_k = t) = \frac{e^{k(\frac{t - \alpha a}{a}) - e^{\frac{t - \alpha a}{a}}}}{a(k-1)!}.$$

Hence, by defining the auxiliary functions $h(t)$ and $g(t)$ as

$$h(t) = \frac{1}{a(k-1)!} \left(\frac{t^2}{a^4} + \frac{2t}{a^3} \right), \quad g(t) = (k+1)(t - \alpha a) - ae^{(a^{-1}(t - \alpha a))},$$

we can express the integral I_1 as

$$I_1 = \mathbb{E} \left[\left(\left(\frac{T_k^2}{a^4} + \frac{2T_k}{a^3} \right) e^{\frac{T_k - \alpha a}{a}} \right) \right] = \int_{t=0}^{\infty} h(t) e^{a^{-1}g(t)} dt.$$

Applying Laplace's method for $a \ll 1$, we approximate this integral as

$$\begin{aligned} I_1 &\approx \sqrt{\frac{2\pi a}{|g''(t_c)|}} h(t_c) \exp[a^{-1}g(t_c)] \\ &= a \sqrt{\frac{2\pi}{k+1}} h(t_c) \exp \left[(k+1)(\log(k+1) - 1) \right] \quad \text{for} \quad a \ll 1. \end{aligned}$$

Here we have calculated

$$t_c = a[\alpha + \log(k+1)], \quad g(t_c) = a(k+1)[\log(k+1) - 1], \quad |g''(t_c)| = \frac{k+1}{a}.$$

After some rearrangement and simplification, we obtain that

$$I_1 \approx \sqrt{\frac{2\pi}{k+1}} \left(\frac{t_c^2}{a^4} + \frac{2t_c}{a^3} \right) \frac{1}{(k-1)!} \left(\frac{k+1}{e} \right)^{k+1}. \quad (25)$$

This expression can be further simplified using Stirling's approximation

$$(k+1)! \approx \sqrt{2\pi(k+1)} \left(\frac{k+1}{e} \right)^{k+1},$$

which yields

$$I_1 = \left((\log(k+1) + 1 + \alpha)^2 - 1 \right) \frac{k}{a^2}. \quad (26)$$

Calculation of I_2 :

Recalling from (7) that $b = \alpha a$ and from (12a) that $\mathbb{E}[T_k] \approx a(\alpha + \psi(k))$, we calculate

$$\begin{aligned} I_2 &= \mathbb{E} \left[\sum_{i=1}^k T_i \right] = ka\alpha + a[\psi(1) + \psi(2) + \dots + \psi(k)] \\ &= ka(\alpha - \gamma_e) + a \left[\sum_{i=1}^{k-1} H_i \right] = ka(\alpha - \gamma_e) + a(kH_{k-1} - k + 1), \end{aligned}$$

where we have used $\sum_{i=1}^{k-1} H_i = kH_{k-1} - (k-1)$. Combining I_1 and I_2 in (24) together with simplification, we obtain the Fisher information for the total observation of full times

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial a^2} \right] = \frac{nk}{a^2} \left((\log(k+1) + 1 + \alpha)^2 - 2(H_{k-1} - \gamma_e + \alpha) - \frac{2}{k} \right). \quad (27a)$$

If we consider an approximation for large k , we can apply $H_k \approx \log k + \gamma_e + \mathcal{O}(k^{-1})$ for $k \gg 1$ and obtain the following compact representation

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial a^2} \right] \approx \frac{nk}{a^2} \left((\log k + \alpha)^2 + 1 \right). \quad (27b)$$

The per-observation Fisher information is then given by

$$I_{\text{full}} \approx \frac{k}{a^2} \left((\log k + \alpha)^2 + 1 \right). \quad (28)$$

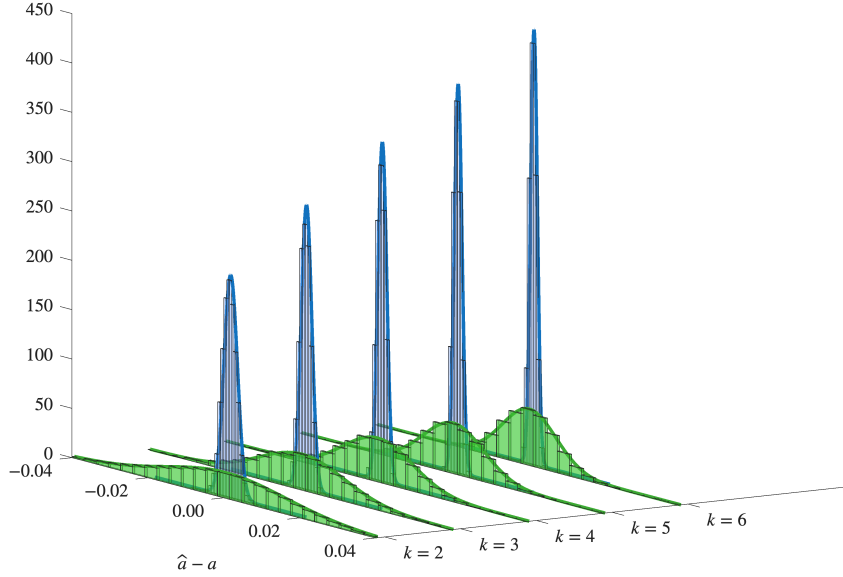


Fig. 3 Recovery of the extreme distribution parameter for test case $a = 0.5$ for $n = 10^3$. Distribution in errors $(\hat{a} - a)$ using the full times (blue) and the time differences (green). Each histogram is formed by $M = 10^4$ repetitions and for each we overlay the distribution $\mathcal{N}(0, 1/\sqrt{nI(a)})$ where the Fisher Information $I(a)$ is given in (29).

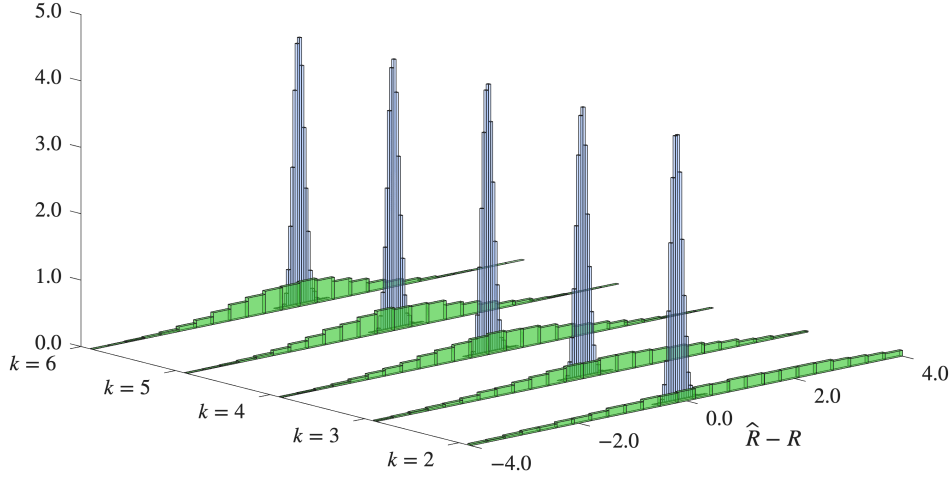
Hence we have derived two consistent estimators $\hat{a}$ based on either the difference in times (17) or the full arrival times (23). The approximate convergence behavior for these estimates as $n \rightarrow \infty$ is

$$\sqrt{n}(\hat{a}_n - a) \rightarrow_d \mathcal{N}(0, I^{-1}), \quad I(a) = \frac{1}{a^2} \begin{cases} k - 1, & \text{time differences;} \\ k(\log k + \alpha)^2 + k, & \text{full times,} \end{cases} \quad (29)$$

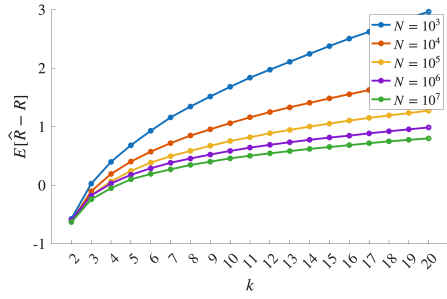
where $I(a)$ is the per-observation Fisher information.

Before proceeding to numerical validation of the main result (29), we make some comments on its biological significance. Clearly $I_{\text{full}} > I_{\text{diff}}$ for any $k \geq 1$ and hence knowledge of the full times yields more information on the source position. However, exploiting this fact requires storing these times and solving the nonlinear system (23). Conversely, a cell can utilize a simpler procedure of a weighted average of the inter-arrival times (17). While it exhibits higher variance (lower accuracy), this estimate requires substantially less computational work to obtain.

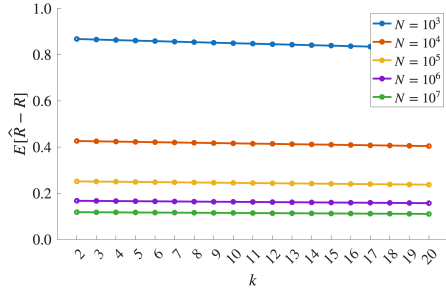
As a confirmation of the estimates (29), we take the fixed value of $a = 1/2$ and generate $n = 10^3$ samples (from (6)) of the first k arrival times for $k = 2, 3, 4, 5, 6$. From these observations, we create estimates $\hat{a}$ using both the MLE estimate based on the full times and the time differences. To explore the entire distribution predicted by (29) we then repeat this process $M = 10^4$ times and plot the results in Fig. 3 and observe that our result (29) accurately predicts the distribution of errors. In addition, these results confirm that the full time method generates estimates with much lower error and that the error is further reduced as k increases.



(a) The distribution of errors $\hat{R} - R$ from full times (blue) and time differences (green) data for $N = 10^6$ initial walkers of diffusivity $D = 1$ and source $R = 5$.



(b) Average error versus k using the time differences.



(c) Average error versus k using the full set of times.

Fig. 4 Accuracy in the prediction (30) of source distance using the full arrival times and the differences in arrival times in one-dimension. Panel (a): The distribution of errors for source detection when $R = 5$ and $N = 10^6$. The estimate is vastly more accurate using the full times (blue) than the time differences (green). Distributions formed from $M = 10^4$ independent estimates. In panels (b-c) we plot the average error against k for initial walker numbers ranging from $N = 10^3 - 10^7$.

3.3 Finite N

The previous analysis and its demonstration were conducted purely on the limiting distribution as the number of walkers $N \rightarrow \infty$. To consider the problem for a finite number of walkers, we consider an idealized one-dimensional problem, where N walkers with diffusivity D originate at the point $x = R$ and are absorbed at $x = 0$. The survival probability for this situation is given by $S(t) = 1 - \text{erfc}(\frac{R}{\sqrt{4Dt}})$. The limiting behavior as $t \rightarrow 0^+$ is given by

$$1 - S(t) \sim \sqrt{\frac{4Dt}{\pi R^2}} e^{-\frac{R^2}{4Dt}}, \quad \text{as } t \rightarrow 0^+.$$

By comparing with (4), we identify the constants

$$A = \sqrt{\frac{4D}{\pi R^2}}, \quad p = \frac{1}{2}, \quad C = \frac{R^2}{4D}.$$

Taking these values and applying them to the result (7), we obtain that

$$a_N = \frac{R^2}{DW(1+W)}, \quad W = W_0\left(\frac{2N^2}{\pi}\right),$$

where $W_0(z)$ is the principal branch of the LambertW function. We additionally notice that $\alpha = \frac{1}{2}(1+W)$ is independent of R and that $b_N = \alpha a_N$. Hence, given the first k arrival times, we can form an estimate $\hat{a}$, either using the full times or their differences, to arrive at an estimate

$$\hat{R} = [DW(1+W)\hat{a}]^{\frac{1}{2}}. \quad (30)$$

In Fig. 4 we display results on the accuracy of the estimate $\hat{R}$ for a range of k and initial walker numbers N with the true value $R = 5$. Across all examples, we see that exploiting the full arrival times yields a far more accurate estimate of the source distance R . The observation that the full-time distribution is far narrower than the difference-based one is clear from the ratio $I_{\text{full}}/I_{\text{diff}} \approx (\log k + \alpha)^2 \approx 10^2$ for the values in Fig. 4. In Fig. 4(c) we observe a very slight dependence on k as the error is dominated by N when the MLE estimator is so much more efficient. The algorithms we have introduced and analyzed in the previous section, are based on the limiting distribution of the arrival times in the limit as $N \rightarrow \infty$. In Figs. 4(b-c) we observe the convergence of the error for a sequence of initial walker number N and again notice that the full arrival time estimate performs markedly better than using only the differences.

4 Estimation of source direction from extreme time data in 2D

In this section, we move on to the problem of estimating the source direction from the information contained in the first k arrival impacts. The arrival locations are time dependent since quick arrivals are more focused around the source direction whereas longer sojourns arrive more uniformly on the cell surface.

4.1 Statistics of arrival locations for a circular cell.

In order to determine the impact location distribution, we consider a circular cell of unit radius and a source placed at $\mathbf{x}_0 = (R, 0)$ for $R > 1$. The placement of the source at an angle θ_0 relative to the horizontal is a simple rotation in the counter-clockwise direction. In polar coordinates $\mathbf{x} = r(\cos \theta, \sin \theta)$, the quantity $u(r, \theta, t)rdrd\theta$ is the probability of a particle being in $(r, r+dr) \times (\theta, \theta+d\theta)$ at time t . Then $u(r, \theta, t)$ solves

the diffusion equation

$$\frac{\partial u}{\partial t} = D \left[\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} \right], \quad r \in (1, \infty), \quad \theta \in (-\pi, \pi) \quad t > 0; \quad (31a)$$

$$u = \frac{1}{r} \delta(r - R) \delta(\theta), \quad r \in (1, \infty), \quad \theta \in (-\pi, \pi) \quad t = 0; \quad (31b)$$

$$u(r, \theta + 2\pi, t) = u(r, \theta, t), \quad r \in (1, \infty), \quad \theta \in (-\pi, \pi), \quad t \geq 0; \quad (31c)$$

$$\partial_\theta u(r, \theta + 2\pi, t) = \partial_\theta u(r, \theta, t), \quad r \in (1, \infty), \quad \theta \in (-\pi, \pi), \quad t \geq 0. \quad (31d)$$

The choice of boundary conditions on the surface of the cell can be influenced by many factors including the distribution and affinities of the receptors. For the purposes of the present study, we consider a continuum of receptors evenly distributed on the cellular surface. Under these conditions, and assuming N_{rec} receptors of non-dimensional common radius ε_{rec} , the following Robin boundary condition has recently [20] been derived

$$D \frac{\partial u}{\partial r} = \kappa u, \quad r = 1; \quad D\kappa^{-1} = -\frac{2}{N_{\text{rec}}} \log \left(\frac{N_{\text{rec}} \varepsilon_{\text{rec}}}{4} \right). \quad (31e)$$

When the number of receptors is large, ($N_{\text{rec}} \approx 10^5$), each with radius $10^{-8} - 10^{-10}m$ [41], the value of κ is large, resulting in an effectively fully absorbing boundary. In the limiting $\kappa \rightarrow \infty$ case, the Dirichlet boundary condition $u = 0$ can be applied on $r = 1$. Further analysis on the homogenization of porous membranes is described in [42, 43]. The treatment of receptors as a uniform distribution rather than considering their discrete positions (e.g. [20, 30]) affords many simplifications. It is frequently observed that receptors can have non-uniform surface distributions that are modulated by the curvature of the membrane [44]. Hence, the analysis of the uniform case presented here serves as a baseline for future studies on the role of geometry in chemotaxis.

With the assumption of the uniform boundary condition (31e), we are able to determine the short time asymptotics of the solution to (31). The two key quantities that describe the hitting times and hitting locations, are the survival probability $S(t)$ and surface flux $\mathcal{J}(\theta, t)$, respectively. These quantities are defined from the solution of (31) by

$$S(t) = \int_{r=1}^{\infty} \int_{\theta=-\pi}^{\pi} u(r, \theta, t) r \, dr \, d\theta \quad (32a)$$

$$\mathcal{J}(\theta, t) = D \partial_r u(r = 1, \theta, t). \quad (32b)$$

To derive the short time asymptotic forms of (32) as $t \rightarrow 0$, we apply the method of moments [20]. The first step is to apply the Laplace transform $\hat{u}(r, \theta; s) = \int_{t=0}^{\infty} u(r, \theta, t) e^{-st} dt$ to equation (31). We follow a standard process of writing the solution in terms of the cosine-Bessel series [45] where we enforce continuity and a jump condition at $r = R$, require the solution to be finite as $r \rightarrow \infty$ and also enforce the

boundary condition (31e). The Laplace transform of the flux is found to be

$$\hat{\mathcal{J}}(\theta; s) = \frac{1}{2\pi} \chi_0(q) + \frac{1}{\pi} \sum_{n=1}^{\infty} \chi_n(q) \cos n\theta, \quad \chi_n(q) = \frac{\kappa(K_n(qR)/K'_n(q))}{\kappa(K_n(q)/K'_n(q)) - Dq}. \quad (33)$$

where $q = \sqrt{s/D}$. Our goal is to approximate the solution of (33) for $q \rightarrow \infty$ which corresponds to the limit $t \rightarrow 0^+$ in the time domain. We define the moments $\widehat{M}_n(s) = \int_{-\pi}^{\pi} \hat{\mathcal{J}}(\theta; s) \theta^n d\theta$ and apply $\cos n\theta \approx 1 - \frac{1}{2}n^2\theta^2 + \frac{1}{24}n^4\theta^4$ in (33) to obtain

$$\chi_n(q) = \int_{-\pi}^{\pi} \hat{\mathcal{J}}(\theta; s) \cos n\theta d\theta \approx \widehat{M}_0(s) - \frac{n^2}{2} \widehat{M}_2(s) + \frac{n^4}{24} \widehat{M}_4(s). \quad (34)$$

The large argument asymptotics [46, §9.7.2] of $K_n(z)$ for $z \rightarrow \infty$ yields the behavior

$$\frac{K_n(qR)}{K'_n(q)} \sim \frac{e^{-q(R-1)}}{R^{\frac{1}{2}}} \left[-1 + \frac{(4n^2 + 3)(R-1) + 4}{8Rq} + \frac{(-16n^4 + 8n^2 - 33)(R-1)^2 + (32n^2 - 72)(R-1) + (64n^2 - 48)}{128R^2q^2} \right].$$

Notice that the term $K_n(q)/K'_n(q)$ arises from substituting $R = 1$ into this expression. We insert this approximation into the expression for χ_n given in (33) and then apply to (34). Gathering terms at the relevant orders of n , we obtain the following expressions for the moments

$$\widehat{M}_0 \sim \frac{\kappa e^{-q(R-1)}}{R^{\frac{1}{2}}(\kappa + Dq)}, \quad \widehat{M}_2 \sim \frac{\kappa(R-1)e^{-q(R-1)}}{R^{\frac{3}{2}}q(\kappa + Dq)}, \quad \widehat{M}_4 \sim \frac{3\kappa(R-1)^2 e^{-q(R-1)}}{R^{\frac{5}{2}}q^2(\kappa + Dq)},$$

in the limit as $q \rightarrow \infty$. Inverting these Laplace transforms explicitly as $q \rightarrow \infty$ ($t \rightarrow 0^+$), we identify the following limits for the variance and kurtosis of the flux

$$\text{Var}[\mathcal{J}(\theta, t)] := \frac{M_2}{M_0} \sim \frac{2Dt}{R}, \quad \text{Kur}[\mathcal{J}(\theta, t)] := \frac{M_4 M_0}{[M_2]^2} \sim 3, \quad t \rightarrow 0^+.$$

The kurtosis is the normalized fourth moment $\text{Kur}[\mathcal{J}] = M_4 M_0 / M_2^2$, equal to 3 for a Gaussian; its convergence to 3 as $t \rightarrow 0^+$ confirms the limiting flux is Gaussian. Hence, in the limit as $t \rightarrow 0^+$, we have derived from (31) the short time behavior

$$\mathcal{J}(\theta, t) \sim \frac{M_0(t)}{\sigma \sqrt{2\pi}} e^{-\frac{\theta^2}{2\sigma^2}}, \quad \sigma^2 \sim \frac{2Dt}{R}, \quad (35a)$$

as $t \rightarrow 0^+$. Here the total arrival rate $M_0(t)$ at the cell surface is given by

$$M_0(t) \sim \begin{cases} \frac{\kappa(R-1)}{\sqrt{\pi R D t}} e^{-\frac{(R-1)^2}{4Dt}} \left[\frac{1}{2\kappa t + (R-1)} \right], & \kappa \in (0, \infty); \\ \frac{(R-1)}{\sqrt{4\pi R D t^3}} e^{-\frac{(R-1)^2}{4Dt}} & \kappa = \infty, \end{cases} \quad (35b)$$

as $t \rightarrow 0$. The survival probability satisfies $S'(t) = -\int_{\theta=-\pi}^{\pi} \mathcal{J}(\theta, t) d\theta \sim -M_0(t)$. From this relationship, we obtain that

$$S(t) \sim 1 - \begin{cases} \frac{4\kappa}{\sqrt{\pi R}} \frac{\sqrt{D}}{(R-1)^2} t^{\frac{3}{2}} e^{-\frac{(R-1)^2}{4Dt}}, & \kappa \in (0, \infty); \\ \frac{\sqrt{4D}}{(R-1)\sqrt{\pi R}} t^{\frac{1}{2}} e^{-\frac{(R-1)^2}{4Dt}}, & \kappa = \infty. \end{cases} \quad (35c)$$

Hence, comparing between (35c) and (4), the constants from the short time survival probability $S(t) \sim 1 - At^p e^{-\frac{C}{t}}$ that characterize the distribution of arrival times are

$$\begin{aligned} p = \frac{3}{2}, \quad A = \frac{4\kappa}{\sqrt{\pi R}} \frac{\sqrt{D}}{(R-1)^2}, \quad C = \frac{(R-1)^2}{4D}, \quad \kappa \in (0, \infty); \\ p = \frac{1}{2}, \quad A = \frac{\sqrt{4D}}{(R-1)\sqrt{\pi R}}, \quad C = \frac{(R-1)^2}{4D}, \quad \kappa = \infty. \end{aligned} \quad (36)$$

These variables now define the scalars a_N and b_N that characterize the distribution of arrival times through the definitions (7).

4.2 The angular distribution of impacts

Here we analyze the k^{th} arrival location characterized by the impact time $T_{k,N}$ and determine the distribution of the corresponding angular location $\theta_{k,N}$. As discussed previously, the distribution of the angular location depends on the time of impact. Hence, we have that

$$\mathbb{P}[\theta_{k,N} = \Theta] = \int_{t=0}^{\infty} \mathbb{P}[\theta_{k,N} = \Theta | T_{k,N} = t] \mathbb{P}[T_{k,N} = t] dt.$$

The limiting distribution for $\mathbb{P}[T_{k,N} = t]$ is given in (11) and takes the form

$$\mathbb{P}[T_{k,N} = t] = \frac{1}{a_N(k-1)!} \exp[\varphi(t)], \quad \varphi(t) = \frac{k(t - b_N)}{a_N} - \exp\left[\frac{t - b_N}{a_N}\right].$$

The distribution of angular arrivals now satisfies

$$\mathbb{P}[\theta_{k,N} = \Theta] = \frac{1}{a_N(k-1)!} \int_{t=0}^{\infty} \mathbb{P}[\theta_{k,N} = \Theta | T_{k,N} = t] \exp[\varphi(t)] dt.$$

In the limit as $N \rightarrow \infty$, which corresponds to $a_N \rightarrow 0$, we apply Laplace's method to simplify this integral. The value of the integral is dominated by the sharp peak of the function $\varphi(t)$ around the peak (mode) $t_c = b_N + a_N \log k$. In the vicinity of t_c , the function has a local quadratic shape with $\varphi''(t_c) = -k/a_N^2$. This allows for a standard application of Laplace's method. From the distribution of fluxes along the surface of the cell derived in (35a), we conclude that

$$\mathbb{P}[\theta_{k,N} = \Theta | T_{k,N} = t_c] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\theta^2}{2\sigma^2}}, \quad \sigma^2 \sim \frac{2Dt_c}{R}. \quad (37)$$

Finally, we conclude that the impact locations have a Gaussian distribution given by

$$\theta_{k,N} \sim \mathcal{N}(\theta_0, \sigma_{k,N}^2), \quad \sigma_{k,N}^2 = \frac{2D}{R} (b_N + a_N \log k), \quad (38)$$

in the limit as $N \rightarrow \infty$. From equation (38) we observe that since $a_N > 0$, particles arriving earlier to the cell have impact locations more tightly distributed (lower variance) around the source direction. Consequently, the first particle to arrive at the cell has the most directional information, followed by the second, and then followed by the third, etc. As a confirmation of this, we plot in Fig. 5(a) a comparison between particle simulations of the arrival distribution and the result (38). In Fig. 5(b) we plot the limiting distribution of the k^{th} impact (38) for fixed R and a range of k values. The variance in the impact location (38) is independent of the particle diffusivity D . This reflects a cancellation of two competing effects: the angular spread of the flux conditioned on arrival at time t grows as $2Dt_c/R$ while the characteristic arrival time of the k^{th} molecule, $t_c = b_N + a_N \log k$, scales as $1/D$. A larger diffusivity therefore delivers the k^{th} impact proportionally sooner by exactly the factor by which it broadens the angular distribution. The result is that the k^{th} arrival always samples the same range of angular space hence D sets the timescale of capture but not the geometry of the early-impact distribution.

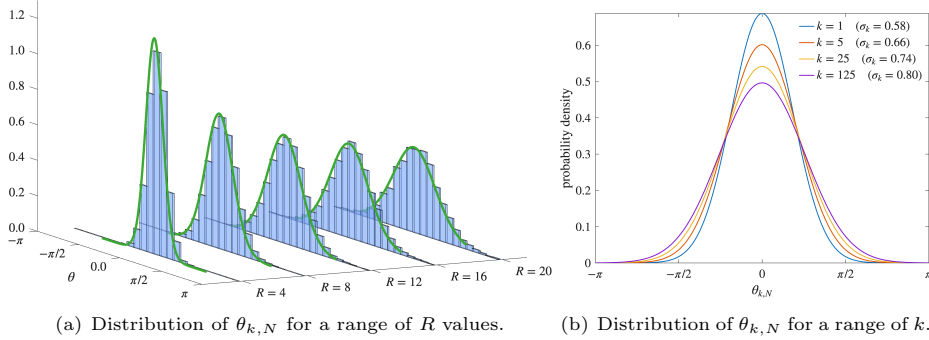


Fig. 5 The distribution of impact angles $\theta_{k,N}$. Panel (a) shows results for $k = 10$, $D = 1$, $\theta_0 = 0$ and $N = 10^6$ for various R . The histograms are based on $M = 10^4$ impacts sampled directly from the case of Dirichlet boundary condition ($\kappa = \infty$). The solid curve is the predicted limiting distribution (38) with good agreement observed. In panel (b) we show the limiting impact distribution (38) with fixed parameters $R = 5$, $D = 1$, $N = 10^3$ and over a range of k values. The distribution of arrival angles broadens as k increases.

5 Recovery of source direction from extreme arrivals

In this section, we use the main result (38) to explore the accuracy of several estimators on the direction of the source.

5.1 Assuming the cell accesses and averages the full angle values

Let us consider a situation where the cell uses the first k angles $\{\theta_{j,N}\}_{j=1}^k$ from N initial walkers to form an estimate of the source direction. Without loss of generality, we set the source to have direction $\theta_0 = 0$. A first quantity to consider is the average angle Θ_k^{avg} inferred by these impacts. It has the following definition:

$$\Theta_k^{\text{avg}} = \frac{1}{k} \sum_{j=1}^k \theta_{j,N}, \quad [\text{Average angle}]. \quad (39)$$

Applying the distribution of each $\theta_{j,N}$ for $j = 1, \dots, k$ given in (38), we obtain that

$$\Theta_k^{\text{avg}} \sim \mathcal{N}\left(0, \frac{2D}{Rk^2} (b_N k + a_N \log k!)\right) \approx \mathcal{N}\left(0, \frac{2D}{Rk} (b_N - a_N + a_N \log k)\right), \quad (40)$$

where on the last step we have applied the approximation $\log k! \approx k \log k - k$. We notice that by increasing the number k of points averaged, the variance decreases like $k^{-1} \log k$. This indicates that although the variance associated with individual impacts increases with k , the averaged variance is overall decreasing.

5.2 Estimates based on the difference in angles

Here we consider an estimate based on the average of the differences in the angle between impacts. Specifically, we derive the distribution of

$$\Theta_k^{\text{diff}} = \frac{1}{k} \sum_{j=1}^{k-1} [\theta_{j+1,N} - \theta_{j,N}] \quad (41)$$

From (38), we have that $\theta_{k,N} \sim \mathcal{N}(0, \sigma_{k,N}^2)$, and hence we observe that

$$\theta_{j+1,N} - \theta_{j,N} \sim \mathcal{N}\left(0, \frac{2D}{R} (b_N + a_N \log(j+1)) + \frac{2D}{R} (b_N + a_N \log(j))\right). \quad (42)$$

By summing the first $k-1$ differences in angle we get the following simplification

$$\begin{aligned} \sum_{j=1}^{k-1} [\theta_{j+1,N} - \theta_{j,N}] &= (\theta_{2,N} - \theta_{1,N}) + (\theta_{3,N} - \theta_{2,N}) + \dots + (\theta_{k,N} - \theta_{k-1,N}) \\ &= \theta_{k,N} - \theta_{1,N}. \end{aligned}$$

The average of the differences then has the distribution

$$\Theta_k^{\text{diff}} = \frac{1}{k} \sum_{j=1}^{k-1} [\theta_{j+1,N} - \theta_{j,N}] = \frac{1}{k} (\theta_{k,N} - \theta_{1,N}) \sim \mathcal{N}\left(0, \frac{2D}{Rk^2} (2b_N + a_N \log k)\right). \quad (43)$$

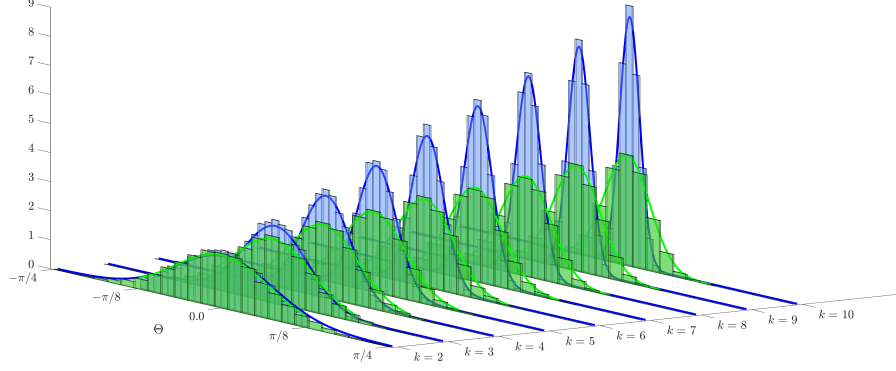


Fig. 6 Distribution of angular estimates in predictions using the average angles Θ_k^{avg} (green histograms) defined in (40) and the differences estimate Θ_k^{diff} (blue histograms) defined in (43).

In Fig. 6 we show confirmation of the expressions (40) for the average Θ_k^{avg} and (43) for the average difference Θ_k^{diff} . We generated $M = 10^4$ samples of the k^{th} arrival angle $\theta_{k,N}$ to a circular cell based on $N = 10^6$ initial walkers. These were sampled exactly from the closed form flux expression (33) by applying a numerical inverse Laplace transform (see [47]). We observe that for $k = 1$, both estimates provide similar accuracy on the source direction. However, as k increases the difference estimate (43) improves because fewer values are averaged.

5.3 The statistics of the resultant vector

In this example, we analyze the resultant formed by averaging over a vector of each arrival. Specifically, let

$$\mathbf{n}^{\text{res}} \equiv \frac{1}{k} \sum_{j=1}^k \mathbf{n}_j, \quad \mathbf{n}_j = (\cos \theta_j, \sin \theta_j), \quad j = 1, \dots, k,$$

be the equally weighted average of the first k impacts at the cell. When utilizing a polar coordinate system where the source has location $\mathbf{x}_0 = R(1, 0)$, the true source is in the direction $\mathbf{e}_1 = (1, 0)$. Hence, the angular difference between the vector $\mathbf{n}^{\text{res}}$ and the source is estimated by

$$\Theta^{\text{res}} = \langle \mathbf{n}^{\text{res}}, \mathbf{e}_1 \rangle = \frac{1}{k} \sum_{j=1}^k \cos \theta_j. \quad (44)$$

In the limit as $N \rightarrow \infty$, we assume the angular distribution is tightly distributed around the region $|\theta| \ll 1$. Expanding for $|\theta_j| \ll 1$, we have that

$$\Theta^{\text{res}} = \frac{1}{k} \sum_{j=1}^k \cos \theta_j \approx 1 - \frac{1}{2k} \sum_{j=1}^k \theta_j^2. \quad (45)$$

The expression (45), which represents a combination of squared Gaussian random variables, is known to follow the *generalized chi-squared* distribution. In the simpler scenario in which θ_j are i.i.d., the random variable follows the chi-squared distribution. Hence the error ρ_k^{res} in the resultant vector formed by k arrivals is described by the generalized chi-squared variable [48]

$$\rho_k^{\text{res}} = \frac{1}{2k} \sum_{j=1}^k \theta_j^2 = \frac{D}{Rk} \sum_{j=1}^k (b_N + a_N \log(j)) Z_j^2, \quad (46)$$

where $Z_j \sim \mathcal{N}(0, 1)$. The expected value and variance of this random variable satisfy

$$\mathbb{E}[\rho_k^{\text{res}}] = \frac{D}{Rk} \sum_{j=1}^k (b_N + a_N \log j) \approx \frac{D}{R} (b_N + a_N (\log k - 1)), \quad (47a)$$

$$\text{Var}[\rho_k^{\text{res}}] = \frac{2D^2}{R^2 k^2} \sum_{j=1}^k (b_N + a_N \log j)^2 \approx \frac{2D^2}{R^2 k} \left((a_N \log k + b_N - a_N)^2 + a_N^2 \right). \quad (47b)$$

In arriving at (47), we have applied the approximations

$$\sum_{j=1}^k \log j \approx k \log k - k, \quad \sum_{j=1}^k [\log j]^2 \approx k (\log^2 k - 2 \log k + 2), \quad k \gg 1.$$

In Fig. 7, we plot results for the estimate Θ^{res} for $R = 4$ and $k = 5$. We see that the distribution of the resultant is closely approximated by the generalized chi-squared distribution (46). This good agreement validates the predicted mean and variance of the estimator.

Bias-variance trade-off for finite k

The observation that the error (47a) increases gradually with k , while the variance (47b) decreases suggests a classic bias-variance trade-off where a small number of arrivals can combine in a simple vector average to give an accurate estimate of the source direction. To investigate this further, we calculate the mean squared error (MSE), to be

$$\begin{aligned} \mathbb{E}[(\rho_k^{\text{res}})^2] &= \mathbb{E}[\rho_k^{\text{res}}]^2 + \text{Var}[\rho_k^{\text{res}}] \\ &= \frac{D^2}{R^2} \left[(b_N + a_N (\log k - 1))^2 + \frac{2}{k} \left((a_N (\log k - 1) + b_N)^2 + a_N^2 \right) \right] \\ &= \frac{D^2 a_N^2}{R^2} \mu(k), \quad \left[\mu := \left(\alpha + \log k - 1 \right)^2 \left(1 + \frac{2}{k} \right) + \frac{2}{k} \right], \end{aligned} \quad (48)$$

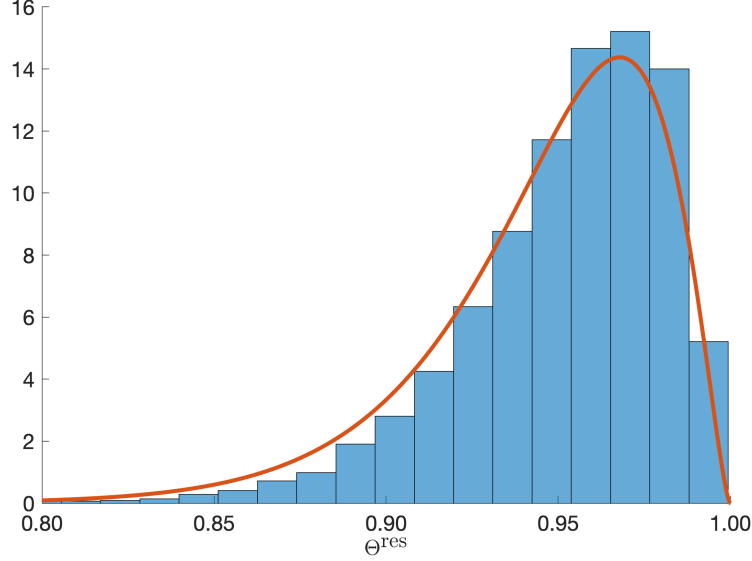


Fig. 7 The distribution of the resultant vector based on the average of the first k arrivals for $k = 5$ and $N = 10^6$ initial walkers at a distance of $R = 4$. The histogram shows estimates $\Theta^{\text{res}} = \mathbf{n}^{\text{res}} \cdot \mathbf{e}_1$, defined in (44) for $M = 10^4$ realizations. The generalized chi-squared distribution (solid) defined in (45) gives a close approximation to the estimate.

where we have used $b_N/a_N = \alpha = p(1 + W)$ from (7). The critical points k^* of the mean squared error satisfy $\mu'(k^*) = 0$. After algebra, this condition reduces to

$$k^* = \alpha + \log k^* - 3 + \frac{1}{\alpha + \log k^* - 1}. \quad (49)$$

For k^* fixed and $N \gg 1$, we have $W \gg 1$ and $\log k^* \ll W$. In Appendix B we use these properties to find that the solution of (49) to have behavior

$$k^* = \log N + (1 - p) \log \log N + [p + \log(AC^p) - 3] + o(1) \quad \text{as } N \rightarrow \infty. \quad (50)$$

Remarkably, we find that in the regime $N \gg 1$, the optimal number of receptor events is $k^* \approx \log N$, independent of geometric factors which provide $\mathcal{O}(1)$ corrections. In Appendix B we explore the dependence of k^* on these lower order corrections for some particular choices of boundary conditions.

In Fig. 8 we display results for the case $\alpha \approx \log N$. In Fig. 8(a) we show the minimization of the MSE (48) with k for several values of $N \gg 1$. The critical values $k^* \approx \log N$ accurately capture the threshold beyond which further impacts increase the MSE. In Fig. 8(b), we plot the relative bias and relative variance for fixed $N = 10^6$. These quantities are defined as

$$\text{Rel. bias : } \frac{\mathbb{E}[\rho_k^{\text{res}}]^2}{\mathbb{E}[\rho_k^{\text{res}}]^2 + \text{Var}[\rho_k^{\text{res}}]}, \quad \text{Rel. variance: } \frac{\text{Var}[\rho_k^{\text{res}}]}{\mathbb{E}[\rho_k^{\text{res}}]^2 + \text{Var}[\rho_k^{\text{res}}]}. \quad (51)$$

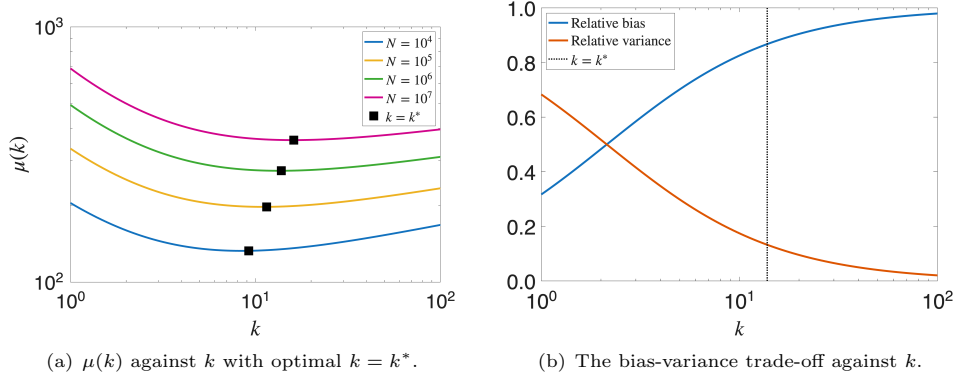


Fig. 8 Bias-variance trade-off in the mean squared error (MSE) (48) as a function of impacts k in the case $\alpha = \log N$. In panel (a) we plot $\mu(k)$ against k for several values of N . We observe the MSE is minimized for $k^* \approx \log N$. In panel (b), we fix $N = 10^6$ and show the relative importance of bias and variance (51) in the MSE as k increases. For $k < k^*$, the MSE is decreasing as k increases. Conversely, for $k > k^*$, the MSE increases with the number of impacts.

As seen in Fig. 8(b), the bias in the estimate increases with k while the variance decreases with k . The vertical line $k = k^* \approx \log N$ is the critical value beyond which more impacts deteriorate the estimate. In Fig. 8(a) the optimization landscape is seen to be relatively shallow indicating that for even $k \approx 10$, the MSE is very close to that of the optimal choice k^* over several decades in N .

6 Discussion

In this work, we investigated the limits of single-cell directional sensing in the ultra-low chemoattractant regime, where receptor binding is discrete and cellular decisions occur long before steady state. A central mathematical contribution of this study is the derivation and validation of explicit asymptotic distributions for both the hitting times and hitting locations of diffusing signaling molecules at the cell surface. By coupling extreme first-passage time statistics with short-time angular localization, we obtained a joint statistical description of early receptor binding events that is valid in the biologically relevant pre-equilibrium regime.

The biological significance of this mathematical framework is that it provides a quantitative framework for comparing the directional information available to the cell under different hypothesized sensing mechanisms. Rather than assuming equilibrium gradients or time-averaged receptor occupancy, our results make it possible to directly analyze how much information is encoded in specific features of early binding data, such as first-arrival times, interarrival intervals, binding locations, or combinations thereof. This enables a systematic comparison of inference strategies, including statistically optimal estimators and simpler biologically plausible estimators, on equal footing. As an example, we found that the interarrival times conveyed much less information to the cell than the full arrival times, but exploiting the full arrival record requires nonlinear processing of the arrival times.

Our analysis demonstrates that directional information is strongly front-loaded in early arrivals: molecules that bind soon after release are preferentially localized around the true source direction, while later arrivals lose directional bias. Additionally, source distance is primarily encoded in the arrival times while the source direction is encoded in arrival locations. As a result, accurate directional inference can be achieved from a small number of early binding events without the need for long-time averaging or high receptor occupancy. This provides a plausible mechanism that explains the experimentally observed rapid symmetry breaking and polarization responses that occur on timescales much shorter than those required to establish steady-state concentration profiles.

A central simplification of our analysis is the treatment of the cell surface as a continuum of receptors, encoded in the homogenized Robin condition (31e) with effective permeability κ . This description is justified when receptors are numerous and densely and uniformly distributed, for which $\kappa \gg 1$ and the surface is effectively absorbing [20]. In this regime the fine-grained arrangement of individual receptors is averaged out, and the arrival statistics depend only on the source geometry and the bulk diffusion. Receptor polarization is characterized by a non-uniform surface density, possibly modulated by membrane curvature [44] or as an active element of the downstream chemotactic response [5]. Capturing these effects requires moving beyond the homogenized boundary condition, for instance through discrete-trap or narrow-escape treatments [30] or a matched-asymptotic analysis that resolves individual receptors while retaining the short-time structure in the survival probability. This is an important direction for extending the present results to more biologically realistic receptor configurations.

Several other extensions of this framework remain open for further analysis. In particular, source detection in environments containing multiple chemoattractant sources presents new challenges, as early arrivals may reflect competing extreme events whose statistics overlap in time and space [7]. In congested environments, it is known that a community of cells can collaborate to better decipher environmental cues [1, 9, 12, 49]. Likewise, crowded or heterogeneous environments may modify first-passage statistics through obstruction, intermittency, or anomalous transport. Extending the joint hitting-time and hitting-location analysis to these settings would enable a quantitative assessment of how robust early-time directional sensing remains under realistic environmental complexity. Finally, an interesting avenue for future research will be to extend these results to more general cell shapes to further inform on the observed roles that aberrations to radial symmetry can play in cellular signaling processes [1, 5, 6, 50, 51].

Acknowledgements

AEL acknowledges support from the NSF under grant DMS-2052636. SDL acknowledges support from the NSF under grant DMS-2325258.

Data availability statement

This manuscript has no associated experimental data. The code used to generate the figures and the numerical results are available from the corresponding author upon request.

A Extreme interarrival times are exponential

As stated in equation (5) (see also [39, Thm. 4]), the times $\{T_{1,N}, \dots, T_{k,N}\}$ obey the following convergence in distribution,

$$\left(\frac{T_{1,N} - b_N}{a_N}, \dots, \frac{T_{k,N} - b_N}{a_N} \right) \rightarrow_d (X_1, \dots, X_k) \quad \text{as } N \rightarrow \infty, \quad (52)$$

where the probability density function $f_{\mathbf{X}^{(k)}}$ of $\mathbf{X}^{(k)} = (X_1, \dots, X_k)$ is given in (6). The continuous mapping theorem ensures that continuous functions preserve convergence in distribution [52]. Hence, applying the following continuous function $g : \mathbb{R}^k \rightarrow \mathbb{R}^{k-1}$,

$$g(x_1, \dots, x_k) = (x_2 - x_1, \dots, x_k - x_{k-1}),$$

to (52) yields the following convergence in distribution,

$$\left(\frac{T_{2,N} - T_{1,N}}{a_N}, \dots, \frac{T_{k,N} - T_{k-1,N}}{a_N} \right) \rightarrow_d (X_2 - X_1, \dots, X_k - X_{k-1}) \quad \text{as } N \rightarrow \infty.$$

It thus remains to determine the joint distribution of $(X_2 - X_1, \dots, X_k - X_{k-1})$. For $k = 2$, we have that for $z \geq 0$,

$$\begin{aligned} \mathbb{P}(X_2 - X_1 > z) &= \mathbb{P}(X_2 > X_1 + z) = \int_{-\infty}^{\infty} dx_2 e^{x_2} \exp(-e^{x_2}) \int_{-\infty}^{x_2 - z} dx_1 e^{x_1} \\ &= e^{-z} \int_{-\infty}^{\infty} dx_2 e^{2x_2} \exp(-e^{x_2}) = e^{-z}. \end{aligned}$$

Hence, $X_2 - X_1$ is an exponential random variable with unit mean. For $k = 3$, we have that for $z_1 \geq 0, z_2 \geq 0$,

$$\begin{aligned} \mathbb{P}(X_3 - X_2 > z_2, X_2 - X_1 > z_1) &= \mathbb{P}(X_3 > X_2 + z_2, X_2 > X_1 + z_1) \\ &= \int_{-\infty}^{\infty} dx_3 e^{x_3} \exp(-e^{x_3}) \int_{-\infty}^{x_3 - z_2} dx_2 e^{x_2} \int_{-\infty}^{x_2 - z_1} dx_1 e^{x_1} \\ &= e^{-z_1} \int_{-\infty}^{\infty} dx_3 e^{x_3} \exp(-e^{x_3}) \int_{-\infty}^{x_3 - z_2} dx_2 e^{2x_2} \\ &= e^{-z_1 - 2z_2} \int_{-\infty}^{\infty} dx_3 e^{3x_3} \exp(-e^{x_3}) / 2 = e^{-z_1 - 2z_2}. \end{aligned}$$

Hence, $2(X_3 - X_2)$ and $(X_2 - X_1)$ are i.i.d. exponential random variables with unit means. For a general $k \geq 2$, we have that for $z_1 \geq 0, \dots, z_{k-1} \geq 0$,

$$\begin{aligned}
\mathbb{P}(X_k - X_{k-1} > z_{k-1}, \dots, X_2 - X_1 > z_1) &= \mathbb{P}(X_k > X_{k-1} + z_{k-1}, \dots, X_2 > X_1 + z_1) \\
&= \int_{-\infty}^{\infty} dx_k e^{x_k} \exp(-e^{x_k}) \int_{-\infty}^{x_k - z_{k-1}} dx_{k-1} e^{x_{k-1}} \int_{-\infty}^{x_{k-1} - z_{k-2}} dx_{k-2} e^{x_{k-2}} \dots \int_{-\infty}^{x_2 - z_1} dx_1 e^{x_1} \\
&= e^{-z_1 - 2z_2 - 3z_3 - \dots - (k-1)z_{k-1}} \int_{-\infty}^{\infty} dx_k e^{kx_k} \exp(-e^{x_k}) / ((k-1)!) \\
&= e^{-z_1} e^{-2z_2} e^{-3z_3} \dots e^{-(k-1)z_{k-1}}.
\end{aligned}$$

Here, we have used that

$$\begin{aligned}
&\int_{-\infty}^{x_k - z_{k-1}} dx_{k-1} e^{x_{k-1}} \int_{-\infty}^{x_{k-1} - z_{k-2}} dx_{k-2} e^{x_{k-2}} \dots \int_{-\infty}^{x_2 - z_1} dx_1 e^{x_1} \\
&= e^{-z_1 - 2z_2 - 3z_3 - \dots - (k-1)z_{k-1}} e^{(k-1)x_k} / ((k-1)!),
\end{aligned}$$

which can be established by induction. We have also used that the change of variables $u = e^x$ yields

$$\int_{-\infty}^{\infty} dx_k e^{kx_k} \exp(-e^{x_k}) = \int_0^{\infty} du u^{k-1} e^{-u} = \Gamma(k) = (k-1)!.$$

We therefore conclude that

$$\{i(X_{i+1} - X_i)\}_{i=1}^{k-1} = \{X_2 - X_1, 2(X_3 - X_2), 3(X_4 - X_3), \dots, (k-1)(X_k - X_{k-1})\}$$

are i.i.d. exponential random variables with unit means.

A.1 X_1 and $X_2 - X_1$ are dependent

Suppose we want to sample $(X_1, \dots, X_k)$ from the pdf in (6). As a first attempt, one might first sample $X_1 \sim \text{Gumbel}(0, 1)$ and then let $X_2 = X_1 + E$ where E is an exponential random variable independent of X_1 . However, this sampling method assumes that X_1 and $X_2 - X_1$ are independent. However, this is false. In particular, while all the *interarrival* times are independent, the first arrival time and the first interarrival time are *not* independent. This is intuitive; if the first particle takes a long time to hit the target (i.e. X_1 is large), then it seems likely that the first interarrival time will be small (i.e. $X_2 - X_1$ is small). To see that X_1 and $X_2 - X_1$ are dependent, we compute

$$\begin{aligned}
\mathbb{E}[X_1(X_2 - X_1)] &= \int \int x_1(x_2 - x_1) f(x_1, x_2) dx_1 dx_2 = -1 - \gamma_e \\
&\neq \mathbb{E}[X_1] \mathbb{E}[X_2 - X_1] = (-\gamma_e) \times 1 = -\gamma_e.
\end{aligned}$$

B Next order asymptotics of optimal k^*

We now derive the next order asymptotics of the optimal k^* derived in section 5. Recall from (50) that $k^* \sim \alpha = p(1 + W) \sim \log N$ as $N \rightarrow \infty$ where $W = W_0(z)$ and $z = (C/p)(AN)^{1/p}$ for $p > 0$. Setting $L = \log z$, the LambertW function has the following large argument expansion [33],

$$W_0(z) = L - \log L + O\left(\frac{\log L}{L}\right) \quad \text{as } L \rightarrow \infty. \quad (53)$$

Since

$$L = \log z = \log((C/p)(AN)^{1/p}) = \frac{1}{p} \log N + \frac{1}{p} \log A + \log(C/p) \sim \frac{1}{p} \log N \quad \text{as } N \rightarrow \infty,$$

we have that

$$\log L = \log \log N - \log p + O\left(\frac{1}{\log N}\right) \quad \text{as } N \rightarrow \infty.$$

Combining this with (53) yields

$$\alpha = p(1 + W) = \log N - p \log \log N + p + \log A + p \log C + o(1) \quad \text{as } N \rightarrow \infty, \quad (54)$$

which, after combining with the relation $k^* = \alpha + \log k^* - 3 + 1/(\alpha + \log k^* - 1)$ in (49), yields

$$k^* = \log N + (1 - p) \log \log N + [p + \log A + p \log C - 3] + o(1) \quad \text{as } N \rightarrow \infty.$$

It is straightforward to check that this estimate is self-consistent with (49). Hence, the geometry-dependent constants A and C affect the optimal k^* through an additive constant. For the case of a circular cell with Dirichlet boundary conditions, the constant in (36) ($\kappa = \infty$) yields

$$k_{\text{Dir}}^* = \log N + \frac{1}{2} \log \log N - \frac{1}{2} \log(\pi R) - 5/2 + o(1) \quad \text{as } N \rightarrow \infty.$$

Notice that all of the factors of the diffusivity D have canceled. Hence, in accord with the rest of our results, the diffusivity controls the timescale of when particles arrive, but it does not affect the optimal number to use for source inference. Notice also that farther sources (larger R) decreases the number of arrivals worth retaining, which has the interpretation that directional information degrades faster with respect to the ordered arrivals for more distant sources.

References

- [1] Copos, C., Sun, Y.-H., Zhu, K., Zhang, Y., Reid, B., Draper, B., Lin, F., Yue, H., Bernadskaya, Y., Zhao, M., Mogilner, A.: Galvanotactic directionality of cell

- groups depends on group size. *Proceedings of the National Academy of Sciences* **122**(21), 2416440122 (2025)
- [2] Henderson, N.T., Pablo, M., Ghose, D., Clark-Cotton, M.R., Zyla, T.R., Nolen, J., Elston, T.C., Lew, D.J.: Ratiometric gpcr signaling enables directional sensing in yeast. *PLOS Biology* **17**(10), 1–35 (2019)
 - [3] Parent, C.A., Devreotes, P.N.: A cell’s sense of direction. *Science* **284**(5415), 765–770 (1999)
 - [4] Sun, Y., Reid, B., Zhang, Y., Zhu, K., Ferreira, F., Estrada, A., Sun, Y., Draper, B.W., Yue, H., Copos, C., Lin, F., Bernadskaya, Y., Zhao, M., Mogilner, A.: Electric field-guided collective motility initiation of large epidermal cell groups. *Molecular Biology of the Cell* **34**(5), 48 (2023) <https://doi.org/10.1091/mbc.E22-09-0391> . PMID: 36989037
 - [5] Nwogbaga, I., Camley, B.A.: Cell shape and orientation control galvanotactic accuracy. *Soft Matter* **20**, 8866–8887 (2024) <https://doi.org/10.1039/D4SM00952E>
 - [6] Kaiyrbekov, K., Camley, B.A.: Does nematic order allow groups of elongated cells to sense electric fields better? *PLoS One* **20**(6), 0325800 (2025)
 - [7] Kashyap, A., Wang, W., Camley, B.A.: Trade-offs in concentration sensing in dynamic environments. *Biophysical Journal* **123**(10), 1184–1194 (2024)
 - [8] Mugler, A., Levchenko, A., Nemenman, I.: Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proc Natl Acad Sci* **113**(6), 689–695 (2016)
 - [9] Camley, B.A.: Collective gradient sensing and chemotaxis: modeling and recent developments. *J. Phys. Condens. Matter* **30**(22), 223001 (2018)
 - [10] Berg, H.C., Purcell, E.M.: Physics of chemoreception. *Biophys J* **20**(2), 193–219 (1977)
 - [11] Aquino, G., Wingreen, N.S., Endres, R.G.: Know the single-receptor sensing limit? think again. *J Stat Phys* **162**(5), 1353–1364 (2016)
 - [12] Fancher, S., Mugler, A.: Fundamental limits to collective concentration sensing in cell populations. *Phys Rev Lett* **118**(7), 078101 (2017)
 - [13] Lakhani, V., Elston, T.C.: Testing the limits of gradient sensing. *PLoS Comput. Biol.* **13**(2), 1–30 (2017)
 - [14] Ismael, A., Tian, W., Waszczak, N., Wang, X., Cao, Y., Suchkov, D., Bar, E., Metodiev, M.V., Liang, J., Arkowitz, R.A., Stone, D.E.: $G\beta$ promotes pheromone receptor polarization and yeast chemotropism by inhibiting receptor

- phosphorylation. *Sci. Signal.* **9**(423), 1–17 (2016)
- [15] Millius, A., Weiner, O.D.: In: Jin, T., Hereld, D. (eds.) *Chemotaxis in Neutrophil-Like HL-60 Cells*, pp. 167–177. Humana Press, Totowa, NJ (2009)
 - [16] Servant, G., Weiner, O.D., Herzmark, P., Balla, T., Sedat, J.W., Bourne, H.R.: Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science* **287**(5455), 1037–1040 (2000)
 - [17] Levchenko, A., Iglesias, P.A.: Models of eukaryotic gradient sensing: Application to chemotaxis of amoebae and neutrophils. *Biophys. J.* **82**(1), 50–63 (2002) [https://doi.org/10.1016/S0006-3495\(02\)75373-3](https://doi.org/10.1016/S0006-3495(02)75373-3)
 - [18] Weiner, O.D., Servant, G., Welch, M.D., Mitchison, T.J., Sedat, J.W., Bourne, H.R.: Spatial control of actin polymerization during neutrophil chemotaxis. *Nature Cell Biology* **1**(2), 75–81 (1999)
 - [19] Weiner, O.D.: Regulation of cell polarity during eukaryotic chemotaxis: the chemotactic compass. *Curr Opin Cell Biol* **14**(2), 196–202 (2002)
 - [20] Lindsay, A.E., Bernoff, A.J., Navarro Hernández, A.: Short-time diffusive fluxes over membrane receptors yields the direction of a signalling source. *Royal Society Open Science* **10**(4), 221619 (2023)
 - [21] Boltz, H.-H., Sirbu, A., Stelzer, N., Lanerolle, P., Winkelmann, S., Annibale, P.: The impact of membrane protein diffusion on gpcr signaling. *Cells* **11**(10) (2022)
 - [22] Civciristov, S., Ellisdon, A.M., Suderman, R., Pon, C.K., Evans, B.A., Kleinfeld, O., Charlton, S.J., Hlavacek, W.S., Canals, M., Halls, M.L.: Preassembled gpcr signaling complexes mediate distinct cellular responses to ultralow ligand concentrations. *Science signaling* **11**(551), 1188 (2018)
 - [23] Civciristov, S., Halls, M.L.: Signalling in response to sub-picomolar concentrations of active compounds: Pushing the boundaries of gpcr sensitivity. *Br J Pharmacol* **176**(14), 2382–2401 (2019)
 - [24] Calabrese, E.J., Giordano, J.: Ultra low doses and biological amplification: Approaching avogadro’s number. *Pharmacological Research* **170**, 105738 (2021) <https://doi.org/10.1016/j.phrs.2021.105738>
 - [25] Linn, S., Lawley, S.D., Karamched, B.R., Kilpatrick, Z.P., Josić, K.: Fast decisions reflect biases; slow decisions do not. *Phys. Rev. E* **110**, 024305 (2024) <https://doi.org/10.1103/PhysRevE.110.024305>
 - [26] Morgan, J., Lindsay, A.E.: Modulation of antigen discrimination by duration of immune contacts in a kinetic proofreading model of t cell activation with extreme statistics. *PLOS Computational Biology* **19**(8), 1–17 (2023)

- [27] Lawley, S.D., Johnson, J.: Slowest first passage times, redundancy, and menopause timing. *Journal of Mathematical Biology* **86**(6), 90 (2023)
- [28] Newby, J., Allard, J.: First-passage time to clear the way for receptor-ligand binding in a crowded environment. *Phys Rev Lett* **116**(12), 128101 (2016)
- [29] Bernoff, A.J., Jilkine, A., Navarro Hernández, A., Lindsay, A.E.: Single-cell directional sensing from just a few receptor binding events. *Biophysical Journal* **122**(15), 3108–3116 (2023)
- [30] Lawley, S.D., Lindsay, A.E., Miles, C.E.: Receptor organization determines the limits of single-cell source location detection. *Phys. Rev. Lett.* **125**, 018102 (2020) <https://doi.org/10.1103/PhysRevLett.125.018102>
- [31] Dobramysl, U., Holcman, D.: Reconstructing the gradient source position from steady-state fluxes to small receptors. *Scientific Reports* **8**(1), 941 (2018)
- [32] Dobramysl, U., Holcman, D.: Mixed analytical-stochastic simulation method for the recovery of a brownian gradient source from probability fluxes to small windows. *Journal of Computational Physics* **355**, 22–36 (2018)
- [33] Corless, R., Gonnet, G., Hare, D., Jeffrey, D., Knuth, D.: On the LambertW function. *Advances in Computational mathematics* **5**(1), 329–359 (1996)
- [34] Endres, R.G., Wingreen, N.S.: Maximum likelihood and the single receptor. *Phys. Rev. Lett.* **103**, 158101 (2009) <https://doi.org/10.1103/PhysRevLett.103.158101>
- [35] Linn, S., Lawley, S.D.: Extreme hitting probabilities for diffusion. *Journal of Physics A: Mathematical and Theoretical* **55**(34), 345002 (2022)
- [36] MacLaurin, J., Newby, J.: Extreme first passage times for populations of identical rare events. *SIAM Journal on Applied Mathematics* **85**(1), 109–142 (2025)
- [37] Tung, H.-R., Lawley, S.D.: First passage times with fast immigration. *SIAM Journal on Applied Mathematics* **85**(5), 2145–2166 (2025)
- [38] Lawley, S.D.: Extreme first-passage times for random walks on networks. *Physical Review E* **102**(6), 062118 (2020)
- [39] Lawley, S.D.: Distribution of extreme first passage times of diffusion. *Journal of Mathematical Biology* (2020) <https://doi.org/10.1007/s00285-020-01496-9>
- [40] Vaart, A.W.: *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3. Cambridge University Press, Cambridge (1998)
- [41] Odoemelam, C.S., Percival, B., Wallis, H., Chang, M.-W., Ahmad, Z., Scholey, D., Burton, E., Williams, I.H., Kamerlin, C.L., Wilson, P.B.: G-protein coupled receptors: structure and function in drug discovery. *RSC Adv.* **10**, 36337–36348

- (2020) <https://doi.org/10.1039/D0RA08003A>
- [42] Brennan, M., Yeo, E.F., Pearce, P., Dalwadi, M.P.: Effective permeability conditions for diffusive transport through impermeable membranes with gaps. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **482**(2331), 20250703 (2026)
 - [43] Lindsay, A.E., Kolokolnikov, T., Tzou, J.C.: Narrow escape problem with a mixed trap and the effect of orientation. *Phys Rev E* **91**(3) (2015) <https://doi.org/10.1103/PhysRevE.91.032111> . Accessed 2015-08-10
 - [44] Alonso, A., Endres, R.G., Kirkegaard, J.B.: Local clustering and global spreading of receptors for optimal spatial gradient sensing. *Phys. Rev. Lett.* **134**, 158401 (2025) <https://doi.org/10.1103/PhysRevLett.134.158401>
 - [45] Stepien, T.L., Zmurchok, C., Hengeniuss, J.B., Caja Rivera, R.M., D’Orsogna, M.R., Lindsay, A.E.: Moth mating: Modeling female pheromone calling and male navigational strategies to optimize reproductive success. *Applied Sciences* **10**(18) (2020) <https://doi.org/10.3390/app10186543>
 - [46] Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth dover printing, tenth gpo printing edn. Dover, New York City (1964)
 - [47] Cherry, J., Lindsay, A.E., Quaife, B.D.: Boundary integral methods for particle diffusion in complex geometries: Shielding, confinement, and escape. *Journal of Computational Physics* **534**, 114032 (2025) <https://doi.org/10.1016/j.jcp.2025.114032>
 - [48] Das, A.: New methods to compute the generalized chi-square distribution. *Journal of Statistical Computation and Simulation* **95**(12), 2608–2642 (2025) <https://doi.org/10.1080/00949655.2025.2501401>
 - [49] Camley, B.A., Zimmermann, J., Levine, H., Rappel, W.J.: Collective Signal Processing in Cluster Chemotaxis: Roles of Adaptation, Amplification, and Co-attraction in Collective Guidance. *PLoS Comput. Biol.* **12**(7), 1–28 (2016)
 - [50] Chakraborty, S., Kolokolnikov, T., Lindsay, A.E.: The effect of target orientation on the mean first passage time of a brownian particle to a small elliptical absorber. *European Journal of Applied Mathematics*, 1–25 (2025) <https://doi.org/10.1017/S095679252510020X>
 - [51] Bernoff, A.J., Lindsay, A.E.: Kinetic monte carlo methods for three-dimensional diffusive capture problems in exterior domains. *Royal Society Open Science* **12**(2), 241033 (2025)
 - [52] Billingsley, P.: *Convergence of Probability Measures*, 2nd edn. Wiley Series in

Probability and Statistics. John Wiley & Sons, ??? (2013)