Receptor organization determines the limits of single-cell source location detection

Sean D. Lawley,[1] Alan E. Lindsay,[2] and Christopher E. Miles[3,*]

[1] *Department of Mathematics, University of Utah, Salt Lake City, UT 84112 USA*
[2] *Department of Applied & Computational Mathematics & Statistics, University of Notre Dame, South Bend, IN 46556 USA*
[3] *Courant Institute of Mathematical Sciences, New York University, New York, NY 10005 USA*

(Dated: June 12, 2020)

Many types of cells require the ability to pinpoint the location of an external stimulus from the arrival of diffusing signaling molecules at cell-surface receptors. *How does the organization (number and spatial configuration) of these receptors shape the limit of a cell's ability to infer the source location?* In the idealized scenario of a spherical cell, we apply asymptotic analysis to compute splitting probabilities between individual receptors and formulate an information-theoretic framework to quantify the role of receptor organization. Clustered configurations of receptors provide an advantage in detecting sources aligned with the clusters, suggesting a possible multiscale mechanism for single-cell source inference.

The ability to pinpoint the location of an external stimulus is critical for a variety of cell types. Canonical examples include eukaryotic gradient-directed cell migration (chemotaxis) [1], directional growth (chemotropism) in growing neurons [2] and yeast [3]. A unifying feature of these systems is that they must infer the spatial location of the external source through the noisy arrivals of diffusing particles to membrane receptors.

The spatial organization of receptors varies between these examples. GABA receptors in nerve cone growth begin relatively uniform on the membrane and dynamically reorganize by clustering receptors toward the source [4]. In budding yeast (*S. cerevisiae*), receptors are known to dynamically cluster towards the direction of a received signal in mate identification [3, 5]. In contrast, the receptors in *Dictyostelium* remain uniform throughout the process of identifying a source location. These differences raise the question: *what role does receptor clustering play in locating external stimuli?*

There has been considerable progress in answering this question. Clustered receptors can provide robustness against noise through rebinding cooperativity [6–10], or by reducing correlation from downstream signals [11]. These observations fit into the broader pursuit of understanding how complex downstream machinery, activated by noisy receptor input, robustly filters shallow gradients [12–20]. Here we study the limits of the most upstream stage: the diffusive arrival of signaling molecules to a fixed configuration of membrane receptors. We find that receptor configuration alone contributes significantly to the quality of signal acquired by the cell.

In this letter, we establish how receptor organization (number and spatial distribution) shapes the limits of a cell's ability to detect the source location of diffusing particles. Our approach draws from the conceptual model of Berg and Purcell [21] (and later [22]) consisting of a spherical cell with circular absorbing surface sites representing membrane receptors. We employ a matched asymptotic approach to compute the probability a signaling molecule hits a particular receptor [23, 24]. Within an information-theoretic framework [25, 26], we establish the informational limit of the fully absorbing cell and assess efficiency relative to this limit as a function of the surface fraction and number of receptors. We identify fundamental differences in the information content of clustered receptor configurations, suggesting higher information content in front of clustered receptors. This observation is verified by performing a maximum likelihood inference, showing that a source can be located with smaller average error in front of a cluster of receptors. This suggests a multiscale mechanism for source localization: if a cell can align toward an initial spatial cue (e.g. as observed in budding yeast [3]) with accuracy limited by the spacing between clusters, then it can exploit receptor nonuniformity to pinpoint the location with an accuracy limited by receptor spacing within a cluster.

Model. Let Ω be the unit sphere with N circular surface receptors of common radius ε. For a diffusing particle originating at $\mathbf{x}$, the splitting probability $p_n(\mathbf{x})$ gives the likelihood of its arrival at the n^{th} receptor, without reaching other receptors or escaping to spatial infinity. The splitting probabilities encode the cell's interaction

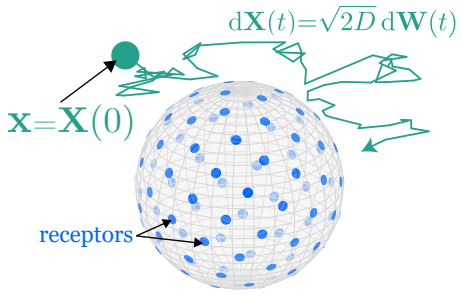


FIG. 1. *Model.* Diffusing particles are released from a source location $\mathbf{x}$ and either escape to spatial infinity or hit a perfectly absorbing cell-surface receptor.

with the extracellular environment. The probabilities $\{p_n(\mathbf{x})\}_{n=1}^N$ satisfy the mixed boundary value problem

$$\begin{aligned} \Delta p_n &= 0, \quad && \mathbf{x} \in \mathbb{R}^3 \setminus \Omega \\ p_n &= 1, && \text{on } n\text{th receptor} \\ p_n &= 0, && \text{all other receptors} \\ \partial_\nu p_n &= 0, && \text{elsewhere on cell surface} \end{aligned} \tag{1}$$

where $\partial_\nu \equiv \hat{\mathbf{n}} \cdot \nabla$ is the normal derivative. The receptor locations are fixed on the surface (cf. Fig. 1) with a general nonoverlapping configuration and centers $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. Dynamic rearrangement of receptors [3, 4] is not explicitly captured in this modeling paradigm, however, how such reorganizations may affect source detection and can be inferred by comparing different static configurations. We have derived and validated numerically [27] that as $\varepsilon \to 0$,

$$\begin{aligned} p_n(\mathbf{x};\mathcal{Y}) \sim 4\varepsilon G(\mathbf{x},\mathbf{y}_n) + \frac{4\varepsilon^2}{\pi}\Big[\big(\tfrac{3}{2} - \ln(2\varepsilon)\big)G(\mathbf{x},\mathbf{y}_n) \\ - 4\pi \sum_{\substack{k=1 \\ k\neq n}}^{N} G(\mathbf{y}_n,\mathbf{y}_k)G(\mathbf{x},\mathbf{y}_k)\Big] + \mathcal{O}(\varepsilon^3). \end{aligned} \tag{2}$$

Here $G(\mathbf{x},\boldsymbol{\xi})$ is the surface Green's function of the Laplacian, exterior to the unit sphere. For $|\boldsymbol{\xi}| = 1$, it is [28]

$$G(\mathbf{x},\boldsymbol{\xi}) = \frac{1}{2\pi}\left[\frac{1}{|\mathbf{x}-\boldsymbol{\xi}|} - \frac{1}{2}\ln\left[\frac{1 - \mathbf{x}\cdot\boldsymbol{\xi} + |\mathbf{x}-\boldsymbol{\xi}|}{|\mathbf{x}| - \mathbf{x}\cdot\boldsymbol{\xi}}\right]\right].$$

We distinguish between the unconditioned probabilities $p_n(\mathbf{x})$ and the conditioned probabilities

$$q_n(\mathbf{x};\mathcal{Y}) = \frac{p_n(\mathbf{x};\mathcal{Y})}{\sum_{k=1}^N p_k(\mathbf{x};\mathcal{Y})}. \tag{3}$$

The former incorporates the possibility for escape to infinity while the latter only reflects particles which have reached a receptor. By working with the conditioned probabilities, we adopt the biological assumption that the cell has no knowledge about particles that did not arrive at a receptor. The conditional probabilities do not vanish as the receptor radius tends to zero ($\varepsilon \to 0$) and $q_n(\mathbf{x};\mathcal{Y}) \to G(\mathbf{x},\mathbf{y}_n)/\sum_{k=1}^N G(\mathbf{x},\mathbf{y}_k)$.

Consider a fixed, unknown source location $\mathbf{x}$ from which particles are released and denote by c_n the count at the n^{th} receptor. When the number of receptors N and arriving particles $M = \sum_{n=1}^N c_n$ are finite, there is uncertainty in the acquired signal. To quantify this, we first take these quantities to be infinitely large and then consider the case where each is finite.

Case $M = \infty$, $N = \infty$: To establish the information content in this limit, we consider the arrival distribution to the sphere for a point source at distance $R > 1$ from the cell center. We adopt a coordinate system where the source is located at the north pole, and let $\theta \in [0,\pi)$ and $\phi \in [0, 2\pi)$ denote the arrival location on the sphere. The density describing (θ,ϕ) is equivalent to the classical charge distribution on a conducting sphere induced by a point charge [29–31]

$$f_R(\theta,\phi) = f_R(\theta) = \frac{1 - R^{-2}}{4\pi(1 - 2R^{-1}\cos\theta + R^{-2})^{\frac{3}{2}}}\sin\theta.$$

In the context of cellular decision making [32], we assume that the cell has a prior distribution of each receptor being equally likely to have an arrival of particles, i.e. the cell is initially uninformed about the source location. For the fully absorbing sphere, this yields

$$f_{\text{unif}}(\theta,\phi) = f_{\text{unif}}(\theta) = \frac{1}{4\pi}\sin\theta.$$

The directional information encoded by the arrivals of particles to the surface is therefore a measure of the *deviation* between the measured and prior distributions. The Kullback-Leibler (KL) divergence, or *relative entropy*, of q from p is defined by

$$d_{\text{re}}(p\,\|\,q) := \int p(x)\ln\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x.$$

The relative entropy $d_{\text{re}}(p\,\|\,q)$ interpreted in a Bayesian sense computes the amount of information gained revising the belief distribution from q to p. Consequently, the relative entropy from the uniform distribution of arrivals encodes the amount of directional information the cell has. This quantity is found explicitly as a function of R:

$$\begin{aligned} E(R) := d_{\text{re}}(f_R\,\|\,f_{\text{unif}}) &= \int_0^{2\pi}\int_0^{\pi} f_R \ln\left(\frac{f_R}{f_{\text{unif}}}\right)\mathrm{d}\theta\,\mathrm{d}\phi \\ &= \ln(R) + 3R\coth^{-1}(R) - \frac{1}{2}\ln\left(R^2-1\right) - 3. \end{aligned}$$

We note that $E(R)$ is positive and monotonically decreasing with intuitive limiting values. As the source approaches the absorbing sphere, $\lim_{R\to 1+} E(R) = \infty$. That is, the noise encoded from diffusion vanishes close to the cell and the arrivals encode the exact direction of the source. However, $E(R) \sim \frac{3}{2}R^{-2}$ as $R \to \infty$, so that for distant sources, diffusion induces more noise in the arrival locations and directional information is reduced.

Case $M = \infty$, $N < \infty$: We now consider a finite number of receptors. From the conditioned probabilities $q_n(\mathbf{x};\mathcal{Y})$ in (3), we define the information gained revising the prior belief from uniform to be the deviation

$$E(R,N) = \sum_{n=1}^{N}\int_{|\mathbf{x}|=R} q_n(\mathbf{x};\mathcal{Y})\ln\left(\frac{q_n(\mathbf{x};\mathcal{Y})}{1/N}\right)\mathrm{d}\mathbf{x}.$$

$E(R,N)$ averages over the angular position of the source. In the calculation for the fully absorbing sphere, the location of the source was chosen arbitrarily due to rotational symmetry. To compare directly to that quantity

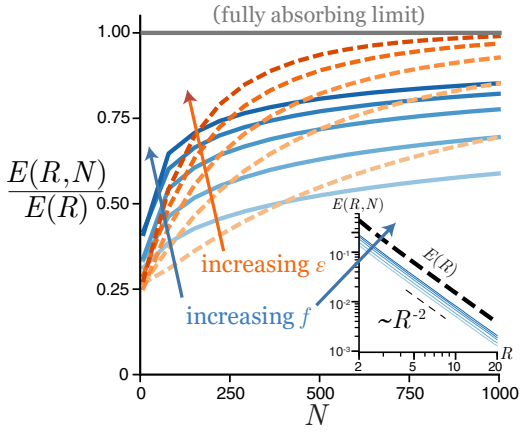


FIG. 2. *Directional information with finite number of receptors.* Relative entropy from uniform $E(R,N)$ for uniformly spaced receptors as a function of the number of receptors N normalized by the entropy in the fully absorbing sphere limit, $E(R)$. In the red, dashed: receptor radius $\varepsilon = \{1,2,3,4,5\} \times 10^{-2}$, blue, solid: surface fraction $f = \{1, 2.5, 5, 7.5, 10\}\%$ with $R = 5$ fixed. Inset: $E(R,N)$ as a function of R, with $N = 25$ receptors fixed and the same varied surface fractions.

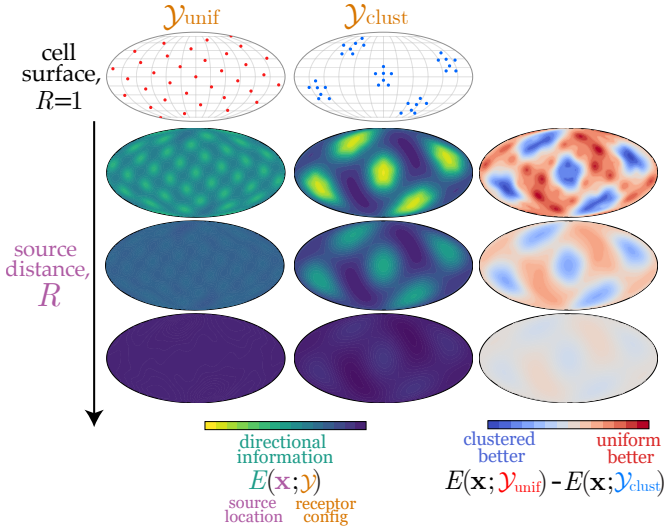


FIG. 3. *Relative entropy for two receptor configurations.* Left and center columns: The directional entropy (4) as a function of the source location for each configuration and $R = \{1.25, 1.75, 2.5\}$. Right column: The difference between the two entropies. At smaller R values, clustered configurations can receive more directional information. This difference diminishes at large R values suggesting receptor organization plays a negligible role when locating distant sources.

with explicit receptors, we average over angular positions of the source. Later, we explore the role of angular position relative to receptors in an unaveraged quantity. The probabilities $q_n(\mathbf{x};\mathcal{Y})$ are computed with a numerical solution to (1) for varying source locations and uniformly spaced receptor configurations $\mathcal{Y}$ centered at Fibonacci spiral points [33]. The results of computing $E(R,N)$ for varying number of receptors can be seen in Fig. 2. We first vary the receptor radius ε and see that the resulting behavior is intuitive: as the characteristic distance between receptors decreases, the resolution increases and the fully absorbing sphere serves as a limiting object for finite number of receptors: $E(R,N) \to E(R)$ as $N \to \infty$. Estimates of receptor numbers range from $N \approx 10^4 - 10^5$ for lymphotcytes [34], $N \approx 10^2$, for GABA receptors in neural cone growth [35] and $N \approx 10^4$ in budding yeast [3]. For $N = 1000$ and $\varepsilon = 0.05$, the largest values in the figure, the surface fraction coverage $f = \varepsilon^2 N/4$ is approximately $f \approx 70\%$ and the information content is effectively at the limit of the fully absorbing sphere.

Is this effect due to having more receptors or just a byproduct of increasing the absorbing surface area? We instead vary f (setting $\varepsilon = \sqrt{4f/N}$) and observe that the information content still increases as a function of N. In the case of 1% surface fraction coverage by $N = 1000$ receptors, the directional information content is over 50% of the fully absorbing limit [36]. This surprising result is analogous to the Berg and Purcell flux dependence on the absorbing surface *perimeter*. The probabilities q_n are influenced by arrivals to other receptors, the rate at which is controlled by the flux, meaning again the perimeter is the factor that influences the rate at which information is gained. In the inset of Fig. 2, we see $E(R,N) \sim R^{-2}$ as $R \to \infty$ alongside the perfectly absorbing limit.

Clustered receptor configurations. We have so far examined how a finite number of *uniformly distributed* receptors approaches the fully absorbing sphere limit. Receptor clustering reduces the total flux to the receptors [37], but it remains to determine the effect on directional sensing. The relative entropy in this case is the nonaveraged quantity

$$E(\mathbf{x};\mathcal{Y}) = \sum_{n=1}^{N} q_n(\mathbf{x};\mathcal{Y}) \ln\left(\frac{q_n(\mathbf{x};\mathcal{Y})}{1/N}\right), \tag{4}$$

For a given source location $\mathbf{x}$ and receptor configuration $\mathcal{Y}$, this measure should be interpreted as a prior distribution of uniform probabilities across receptors, which is not necessarily equivalent to any particular distribution of $\mathbf{x}$, the quantity being estimated. See [38] for a discussion of priors in direction sensing.

Let $\mathcal{Y}_{\text{clust}}$ and $\mathcal{Y}_{\text{unif}}$ denote clustered and uniform receptor configurations (Fig. 3). Clustered configurations are formed by placing receptors in a spherical cap and copying across the sphere at Fibonacci spiral points. For these configurations, the relative entropy (with respect to uniform probabilities) is computed using the asymptotic result (2) and shown in Fig. 3. The asymptotic result allows for rapid evaluation of these probabilities at a large number of source locations.

In Fig. 3, the directional relative entropy appears to be heterogeneous in space for $\mathcal{Y}_{\text{clust}}$ but directionally uni-

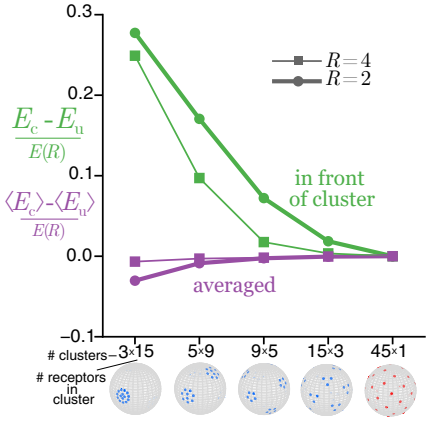


FIG. 4. *Comparison of entropies for clustered and uniform receptor configurations.* Difference in entropy is computed in front of the cluster (green) and averaged over source locations (purple). Clustered configurations have higher relative entropy in front of clusters but lower on average, with this effect most magnified close to the cell.

form for $\mathcal{Y}_{\text{unif}}$. The difference $E(\mathbf{x};\mathcal{Y}_{\text{clust}}) - E(\mathbf{x};\mathcal{Y}_{\text{unif}})$ (Fig. 3, right column) indicates that directional entropy in front of a cluster is higher than that of the uniform configuration, and this difference diminishes as source distance increases. This implies that informational content is richer for the clustered configuration when particles are arriving from sources in front of the cluster.

To explore further, we compute in Fig. 4 the difference in entropy both in front of a cluster and averaged over possible source locations (i.e. $E(R\mathbf{e}_x;\mathcal{Y})$ and $\int_{|\mathbf{x}|=R} E(\mathbf{x};\mathcal{Y})\mathrm{d}\mathbf{x}$). The results are shown for 45 total receptors and source distances $R = 2, 4$. In front of a cluster, nonuniform configurations have higher entropy levels, but on average, perform worse. The benefits of clustering become diminished as the source location becomes farther away or the configuration becomes less clustered. The resolution is therefore determined by the spacing between receptors. Thus, smaller receptor spacing within a cluster can resolve finer detail. Expectedly, as the source location moves away, the noise from diffusion makes both distributions converge to uniform probabilities and the difference vanishes.

Case $M < \infty$, $N < \infty$: For infinitely many incident particles ($M = \infty$), the q_n are discerned exactly. *How does source inference operate given a noisy sample formed by finite arrivals?* The probabilities of arrival at each receptor are multinomial (dependent on $\mathbf{x}$) with likelihood

$$L(\mathbf{x};\mathcal{Y}) = \sum_{n=1}^{N} \frac{c_n}{M} \ln(q_n(\mathbf{x},\mathcal{Y})),$$

The maximum likelihood estimate (MLE) of the source is

$$\hat{\mathbf{x}}_{\text{MLE}} = \arg\max_{\mathbf{x}} L(\mathbf{x};\mathcal{Y}). \tag{5}$$

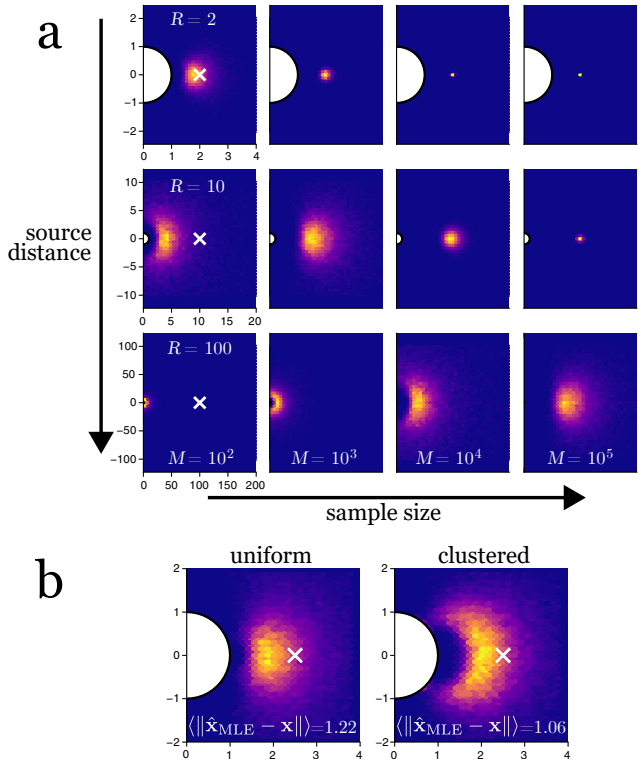


FIG. 5. *Frequency of maximum likelihood estimated locations.* a) For varying source locations $\mathbf{x}$ at a distance R (from the center of the cell) and sample sizes M, frequencies of MLE estimated locations (5) for uniform receptor covering. b) estimated locations for the configurations of receptors shown in Fig. 3. Colors correspond to relative frequency of estimated location.

We use this inference scheme only as a statistical abstraction to quantify the limits of uncertainty in the system. Cellular mechanisms for MLE-based [39] or Bayesian [40] inference have been proposed but are fundamentally downstream of diffusive arrivals and beyond the scope of this Letter. Relative entropy and Fisher information (the standard error of MLE) are related [41] so we expect the previous results about relative entropy to inform the error in the MLE estimate. In Fig. 5a, we vary the source location $\mathbf{x} = (R, 0, 0)$ for $R = 2, 10, 100$ and $M = 10^2, 10^3, 10^4, 10^5$. The receptor configuration remains the uniform configuration in Fig. 3. For each trial, we compute the MLE estimate numerically with from (5) with $z = 0$ fixed and plot the frequency of results. The error $\langle\|\hat{\mathbf{x}}_{\text{MLE}} - \mathbf{x}\|\rangle$ scales $\sim M^{-\frac{1}{2}}$, as predicted by the central limit theorem but also as $\sim R^{-2}$ [27], in accordance with the relative entropy results in the $M = \infty$ case. To verify the claim in Fig. 3 that certain source locations may be better detected by a clustered configuration, we fixed the source at $\mathbf{x} = (2.5, 0, 0)$ and took $M = 50$ particles. The frequency of predicted locations, shown in Fig. 5b, yields a lower mean error for the clustered configuration than that of the uniform.

Discussion. We have examined the role of receptor organization on detection of external stimuli. We demonstrate that a cell can operate near theoretical limits with a finite number of receptors and noisy arrival data. When receptors are not uniformly spaced, the information content is larger in front of clusters suggesting that resolution is limited by receptor spacing. A cell with clustered receptors can potentially benefit by forming a crude estimate and aligning itself in that direction.

Altogether, our results reinforce the notion that cells must balance trade-offs between directional signal coverage and robustness as seen in other work [8–11]. However, we emphasize that the only mechanism by which receptors are interacting in our model is through binding competition, as no downstream signaling or rebinding are included. Understanding the interplay between receptor organization and downstream signaling mechanisms is a natural direction for future investigations. Finally, it would be interesting to study the relative entropy of physiological or dynamical cluster configurations (e.g. [8]) compared to the synthetic ones utilized here. Our work suggests the spatial organization of membrane bound receptors plays a crucial role in cellular scale directional sensing and decision making.

Acknowledgements. Two authors are supported by the NSF (SDL by DMS-1814832 and DMS-1148230, AEL by DMS-1815216). Support and resources from The Convergence Accelerator Team program at the NSF-Simons Center for Multiscale Cell Research, funded by NSF grant DMS-1763272 and a grant from the Simons Foundation (594598, QN), and the AMS Mathematical Research Communities program are gratefully acknowledged.

* christopher.miles@cims.nyu.edu

[1] A. Levchenko and P. A. Iglesias, Biophys. J. **82**, 50 (2002).
[2] G. J. Goodhill, Trends Neurosci. **39**, 202 (2016).
[3] A. Ismael, W. Tian, N. Waszczak, X. Wang, Y. Cao, D. Suchkov, E. Bar, M. V. Metodiev, J. Liang, R. A. Arkowitz, and D. E. Stone, Sci. Signal. **9**, 1 (2016).
[4] C. Bouzigues, M. Morel, A. Triller, and M. Dahan, Proc. Natl. Acad. Sci. **104**, 11251 (2007).
[5] R. Yonashiro, A. Sugiura, M. Miyachi, T. Fukuda, N. Matsushita, R. Inatome, Y. Ogata, T. Suzuki, N. Dohmae, and S. Yanagi, Mol. Biol. Cell **20**, 4524 (2009).
[6] D. Bray, M. D. Levin, and C. J. Morton-Firth, Nature **393**, 85 (1998).
[7] B. R. Caré and H. A. Soula, BMC Syst. Biol. **5** (2011), 10.1186/1752-0509-5-48.
[8] G. Iyengar and M. Rao, Proc. Natl. Acad. Sci. U. S. A. **111**, 12402 (2014).
[9] H. Nguyen, P. Dayan, and G. J. Goodhill, J. Theor. Biol. **360**, 95 (2014).
[10] P. Recouvreux and P. F. Lenne, Curr. Opin. Cell Biol. **38**, 18 (2016).
[11] A. Mugler, F. Tostevin, and P. R. Ten Wolde, Proc. Natl. Acad. Sci. U. S. A. **110**, 5927 (2013).
[12] L. Song, S. M. Nadkarni, H. U. Bödeker, C. Beta, A. Bae, C. Franck, W. J. Rappel, W. F. Loomis, and E. Bodenschatz, Eur. J. Cell Biol. **85**, 981 (2006).
[13] P. J. Van Haastert and M. Postmay, Biophys. J. **93**, 1787 (2007).
[14] F. Tostevin, P. R. Ten Wolde, and M. Howard, PLoS Comput. Biol. **3**, 763 (2007).
[15] W. J. Rappel and H. Levine, Phys. Rev. Lett. **100**, 6 (2008).
[16] R. G. Endres and N. S. Wingreen, Proc. Natl. Acad. Sci. **105**, 15749 (2008).
[17] R. G. Endres and N. S. Wingreen, Prog. Biophys. Mol. Biol. **100**, 33 (2009).
[18] G. Aquino, L. Tweedy, D. Heinrich, and R. G. Endres, Sci. Rep. **4**, 1 (2014), 1406.7673.
[19] T. Strünker, L. Alvarez, and U. B. Kaupp, Curr. Opin. Neurobiol. **34**, 110 (2015).
[20] V. Lakhani and T. C. Elston, PLoS Comput. Biol. **13**, 1 (2017).
[21] H. C. Berg and E. M. Purcell, Biophys. J. **20**, 193 (1977).
[22] W. Bialek and S. Setayeshgar, Proc. Natl. Acad. Sci. **102**, 10040 (2005).
[23] U. Dobramysl and D. Holcman, Sci. Rep. **8**, 1 (2018).
[24] U. Dobramysl and D. Holcman, J. Comput. Phys. **355**, 22 (2018).
[25] B. W. Andrews and P. A. Iglesias, PLoS Comput. Biol. **3**, 1489 (2007).
[26] J. M. Kimmel, R. M. Salter, and P. J. Thomas, Adv. Neural Inf. Process. Syst. (2007).
[27] See Supplemental Material at [URL] for details on the asymptotic splitting probabilities and source location inference simulations, including Refs. [42–47].
[28] I. M. Nemenman and A. S. Silbergleit, J. Appl. Phys. **86**, 614 (1999).
[29] A. Tikhonov and A. Samarskii, *Equations of Mathematical Physics* (Pergamon Press Ltd., Oxford, 1963) p. 800.
[30] B. A. Luty, J. A. Mccammon, and H. X. Zhou, J. Chem. Phys. **97**, 5682 (1992).
[31] C. Hwang and M. Mascagni, J. Appl. Phys. **95**, 3798 (2004).
[32] C. G. Bowsher and P. S. Swain, Curr. Opin. Biotechnol. **28**, 149 (2014).
[33] R. Swinbank and R. J. Purser, Quarterly Journal of the Royal Meteorological Society **132**, 1769 (2006).
[34] A. S. Perelson and G. Weisbuch, Rev. Mod. Phys. **69**, 1219 (1997).
[35] C. Bouzigues, D. Holcman, and M. Dahan, PLOS ONE **5**, 1 (2010).
[36] We expect that for a large enough N, the curves for fixed f would approach the limit. This could be investigated with fast numerical schemes such as [48].
[37] A. E. Lindsay, A. J. Bernoff, and M. J. Ward, Multiscale Model. Simul. **15**, 74 (2017).
[38] B. Hu, W. Chen, H. Levine, and W. J. Rappel, J. Stat. Phys. **142**, 1167 (2011).
[39] R. G. Endres and N. S. Wingreen, Phys. Rev. Lett. **103**, 1 (2009).
[40] T. Mora and I. Nemenman, Phys. Rev. Lett. **123**, 198101 (2019), 1908.04057.
[41] C. Gourieroux and A. Monfort, *Statistics and Econometric Models*, 1st ed. (Cambridge University Press, 1995) p. 524.

[42] I. N. Sneddon, *Elements of Partial Differential Equations* (McGraw-Hill Book Company, Inc., 1957).
[43] I. N. Sneddon, *Mixed boundary value problems in potential theory* (North-Holland, 1966).
[44] D. G. Duffy, *Mixed Boundary Value Problems* (Chapman and Hall/CRC, 2008).
[45] P. Novák and J. Novák, Proc. SPIE **8789**, 878913 (2013).
[46] A. J. Bernoff and A. E. Lindsay, SIAM J. Applied Math. **78**, 266 (2018).
[47] Á. González, Mathematical Geosciences **42**, 49 (2009).
[48] J. Kaye and L. Greengard, Journal of Computational Physics , 100047 (2019).

Supplementary Material for "Receptor organization determines the limits of single-cell source location detection"

Sean D. Lawley,[1] Alan E. Lindsay,[2] and Christopher E. Miles[3,*]

[1] *Department of Mathematics, University of Utah, Salt Lake City, UT 84112 USA*

[2] *Department of Applied & Computational Mathematics & Statistics, University of Notre Dame, South Bend, IN 46556 USA*

[3] *Courant Institute of Mathematical Sciences, New York University, New York, NY 10005 USA*

(Dated: June 12, 2020)

CONTENTS

MATCHED ASYMPTOTIC ANALYSIS FOR THE SPLITTING PROBABILITIES

In this section, we use a matched asymptotic approach to derive an asymptotic expansion for the splitting probabilities. To simplify notation, we solve for $p(\mathbf{x})$ where it is assumed that the distinguished receptor is $j = 1$. The problem to be solved is then

$$\Delta p = 0, \quad \mathbf{x} \in \mathbb{R}^3 \backslash \Omega; \qquad \partial_\nu p = 0, \quad \mathbf{x} \in \partial\Omega \backslash \{\cup_{j=1}^N \partial\Omega_j\}; \tag{S1a}$$

$$p = \delta_{1j}, \quad \mathbf{x} \in \partial\Omega_j, \quad j = 1, \ldots, N; \qquad p \to 0, \quad \text{as} \quad |\mathbf{x}| \to \infty. \tag{S1b}$$

To establish the correct expansion of (S1) as $\varepsilon \to 0$, the first step is to analyze the solution in the $\mathcal{O}(\varepsilon)$ neighborhood near each receptor, hereafter referred to as the "inner" region. The region away from the receptors, in which $|\mathbf{x} - \mathbf{x}_k| = \mathcal{O}(1)$ for $k = 1, \ldots, N$, is referred to as the "outer" region. A key part of the analysis is knowledge of $G(\mathbf{x}; \mathbf{x}_0)$, the surface Green's function exterior to the unit sphere that satisfies the system

$$\Delta G = 0, \quad r = |\mathbf{x}| > 1, \quad \theta \in (0, \pi), \quad \phi \in (0, 2\pi); \tag{S2a}$$

$$-\partial_r G|_{r=1} = \frac{1}{\sin\theta_0}\delta(\varphi - \varphi_0)\delta(\theta - \theta_0), \quad \theta \in (0, \pi), \quad \phi \in (0, 2\pi); \tag{S2b}$$

$$G \sim \frac{1}{2\pi|\mathbf{x}|} \quad \text{as} \quad |\mathbf{x}| \to \infty; \tag{S2c}$$

where $\mathbf{x}_0 = (\sin\theta_0\cos\varphi_0, \sin\theta_0\sin\varphi_0, \cos\theta_0)^T$. The solution of (S2) for $|\mathbf{x}_0| = 1$ is given explicitly (cf. [S1]) by

$$G(\mathbf{x};\mathbf{x}_0) = \frac{1}{2\pi}\Big[\frac{1}{|\mathbf{x}-\mathbf{x}_0|} - \frac{1}{2}\log\Big(\frac{1-\mathbf{x}\cdot\mathbf{x}_0+|\mathbf{x}-\mathbf{x}_0|}{|\mathbf{x}|-\mathbf{x}\cdot\mathbf{x}_0}\Big)\Big]. \tag{S3}$$

We first develop the solution to (S1) in a boundary layer in the vicinity of the receptor centered at $\mathbf{x}_k$. The first step is to introduce the local coordinate system

$$\mathbf{z} = \frac{\mathbf{x}-\mathbf{x}_k}{\varepsilon}, \qquad \eta = \frac{r-1}{\varepsilon}, \qquad s_1 = \sin\theta_k\frac{\varphi-\varphi_k}{\varepsilon}, \qquad s_2 = \frac{\theta-\theta_k}{\varepsilon}, \tag{S4}$$

where $r = |\mathbf{x}|$ and η is a rescaled measure of distance to Ω. Here $\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2$ represents, for $\eta = 0$, an approximate surface cartesian coordinate system near the kth receptor. Defining the operators

$$\mathcal{L} := \partial_{\eta\eta} + \partial_{s_1s_1} + \partial_{s_2s_2},$$

$$\mathcal{A} := -2\eta(\partial_{s_1s_1} + \partial_{s_2s_2}) + \cot\theta_k(\partial_{s_2} - 2s_2\partial_{s_1s_1}) + 2\partial_\eta,$$

equation (S1a) is transformed to

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial p}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial\phi^2}$$

$$\sim \varepsilon^{-2}\mathcal{L}p + \varepsilon^{-1}\mathcal{A}p + \mathcal{O}(1) = 0. \tag{S5}$$

The key ingredient required to establish the correct asymptotic expansion of (S1) as $\varepsilon \to 0$, is the local behavior of the Green's function $G(\mathbf{x};\mathbf{x}_k)$ (S3) as $\mathbf{x} \to \mathbf{x}_k$. In terms of the rescaled local coordinates (S4), this behavior is given by

$$G = \frac{1}{2\pi}\left[\frac{1}{\varepsilon|\mathbf{z}|} + \frac{1}{2}\log(\varepsilon/2) + \frac{1}{2}\log[\eta+|\mathbf{z}|] + \mathcal{O}(\varepsilon)\right]. \tag{S6}$$

The next step is to use the following two-term expansion (cf. Lemma A.1 of [S2]) for $1/|\mathbf{z}|$ in terms of the local coordinates (s_1, s_2, ρ), where $|\mathbf{s}| := (s_1^2+s_2^2)^{\frac{1}{2}}$ and $\rho := (\eta^2+|\mathbf{s}|^2)^{\frac{1}{2}}$,

$$\frac{1}{|\mathbf{z}|} = \frac{1}{\rho} - \frac{\varepsilon}{2\rho^3}\left[\eta|\mathbf{s}|^2 + s_1^2 s_2\cot\theta_k\right] + \mathcal{O}(\varepsilon^2).$$

Substituting this into (S6) yields the local behavior of $G(\mathbf{x};\mathbf{x}_k)$ as $\mathbf{x} \to \mathbf{x}_k$ is

$$G = \frac{1}{2\pi}\left[\frac{1}{\varepsilon\rho} + \frac{1}{2}\log(\varepsilon/2) + \frac{1}{2}\Big(\log[\eta+\rho] - \frac{1}{\rho^3}\big(\eta|\mathbf{s}|^2 + s_1^2 s_2\cot\theta_k\big)\Big)\right] + \mathcal{O}(\varepsilon). \tag{S7}$$

In view of (S7), and the fact that $p_k = \mathcal{O}(1)$ on $\eta = 0$, we expand the inner solution near the kth receptor, in terms of the inner variables (ρ, s_1, s_2), as

$$p \sim w_0^k + \varepsilon \log(\varepsilon/2) w_1^k + \varepsilon w_2^k + \cdots, \quad \text{as } \varepsilon \to 0. \tag{S8}$$

Plugging (S8) into (S1) and using (S5), we have for $n = 0, 1, 2$, and $k = 1, \ldots, N$,

$$\mathcal{L} w_n^k = -\delta_{1k}\delta_{2n} \mathcal{A} w_0^1, \quad \eta > 0, \quad \mathbf{s} \in \mathbb{R}^2; \tag{S9a}$$

$$w_n^k = \delta_{1k}\delta_{0n}, \quad \eta = 0, \quad |\mathbf{s}| < a_k; \qquad \partial_\eta w_n^k = 0, \quad \eta = 0, \quad |\mathbf{s}| \geq a_k. \tag{S9b}$$

The problems (S9) can be solved exactly for $n = 0, 1, 2$ and $k = 1, \ldots, N$. To obtain these solutions, we introduce (denoted by w_c) the planar electrified disk problem defined on the tangent plane to the sphere at $\mathbf{x} = \mathbf{x}_k$,

$$\mathcal{L} w_c = 0, \quad \eta > 0, \quad \mathbf{s} \in \mathbb{R}^2; \quad w_c \to 0 \quad \text{as} \quad \rho \to \infty, \tag{S10a}$$

$$w_c = 1, \quad \eta = 0, \quad |\mathbf{s}| < a_k; \qquad \partial_\eta w_c = 0, \quad \eta = 0, \quad |\mathbf{s}| \geq a_k, \tag{S10b}$$

The exact solution (see [S2, S3]) to (S10) is

$$w_c = \frac{2}{\pi} \sin^{-1}(a_k/L), \quad L := \frac{1}{2}\Big[\sqrt{(|\mathbf{s}| + a_k)^2 + \eta^2} + \sqrt{(|\mathbf{s}| - a_k)^2 + \eta^2}\Big]. \tag{S11a}$$

In terms of the capacitance c_k of the kth receptor, the far-field behavior is

$$w_c \sim c_k \Big(\frac{1}{\rho} + \frac{\pi^2 c_k^2}{24}\Big(\frac{1}{\rho^3} - \frac{3\eta^2}{\rho^5}\Big) + \cdots\Big) \quad \text{as } \rho \to \infty; \quad c_k := \frac{2a_k}{\pi}. \tag{S11b}$$

The solution of problems (S9) may all be represented in terms of w_c. For example, we have that

$$w_0^k = \delta_{1k} w_c. \tag{S12}$$

In light of the fact that $w_0^k = 0$ for $k \neq 1$, we additionally have that

$$w_1^k = A_k(1 - w_c), \quad k = 1, \ldots, N;$$

$$w_2^k = B_k(1 - w_c), \quad k = 2, \ldots, N,$$

for constants $\{A_1, \ldots, A_N\}$ and $\{B_2, \ldots, B_N\}$ to be found from matching with the outer expansion. To determine the problem for the last equation w_2^1, we note that equations

(S12), (S9), together with $\mathcal{L}w_c = 0$ in (S10) imply

$$\mathcal{L}w_2^1 = -2(\partial_\eta w_c + \eta\partial_{\eta\eta} w_c) - \cot\theta_1(\partial_{s_2} w_c - 2s_2\partial_{s_1 s_1} w_c), \quad \eta > 0, \quad \mathbf{s} \in \mathbb{R}^2; \tag{S13a}$$

$$w_2^1 = 0, \quad \eta = 0, \quad |\mathbf{s}| < a_1; \qquad \partial_\eta w_2^1 = 0, \quad \eta = 0, \quad |\mathbf{s}| \geq a_1. \tag{S13b}$$

The process of matching solutions of (S13) to the outer expansion requires that we determine all monopoles in the far-field. For this reasons we identify such terms in (S13) by first decomposing

$$w_2^1 = B_1(1 - w_c) - w_{2o} - w_{2e}, \tag{S14}$$

where w_{2o} satisfies

$$\mathcal{L}w_{2o} = \cot\theta_1(\partial_{s_2} w_c - 2s_2\partial_{s_1 s_1} w_c), \quad \eta > 0, \quad \mathbf{s} \in \mathbb{R}^2; \tag{S15a}$$

$$w_{2o} = 0, \quad \eta = 0, \quad |\mathbf{s}| < a_1; \qquad \partial_\eta w_{2o} = 0, \quad \eta = 0, \quad |\mathbf{s}| \geq a_1; \tag{S15b}$$

$$w_{2o} \sim \frac{c_1}{2\rho^3}(s_1^2 s_2 \cot\theta_1), \quad \text{as} \quad \rho \to \infty, \tag{S15c}$$

and w_{2e} satisfies

$$\mathcal{L}w_{2e} = 2(\partial_\eta w_c + \eta\partial_{\eta\eta} w_c), \quad \eta > 0, \quad \mathbf{s} \in \mathbb{R}^2; \tag{S16a}$$

$$w_{2e} = 0, \quad \eta = 0, \quad |\mathbf{s}| < a_1; \qquad \partial_\eta w_{2e} = 0, \quad \eta = 0, \quad |\mathbf{s}| \geq a_1; \tag{S16b}$$

$$w_{2e} \sim -\frac{c_1}{2}\log[\eta + \rho] + \frac{c_1}{2\rho^3}\eta|\mathbf{s}|^2, \quad \text{as} \quad \rho \to \infty. \tag{S16c}$$

Clearly w_2^1 in (S14) has a monopole arising from the w_c term. To determine whether other monopole terms arise from equations (S15) and (S16), we use the exact solution (see [S2, Lemma B.2]) of (S15)

$$w_{2o} = \cot\theta_1\left(\frac{s_1^2}{2}\partial_{s_2} w_c - s_2 s_1 \partial_{s_1} w_c\right). \tag{S17}$$

Using the solution (S17) and evaluating the limit $\rho \to \infty$, it can be shown that w_{2o} does not give rise to a monopole in the far field. However, w_{2e} does exhibit a monopole in its far field. Specifically, it was shown in [S2, Lemma B.1] that

$$w_{2e} \sim -\frac{c_1}{2}\log[\eta + \rho] + \frac{c_1}{2\rho^3}\eta|\mathbf{s}|^2 + \frac{c_1 b_1}{\rho}, \quad \text{as } \rho \to \infty, \quad b_1 := \frac{c_1}{2}\Big[\log(4a_1) - \frac{3}{2}\Big]. \tag{S18}$$

Combining equations (S14) and (S18), we obtain the needed far behavior

$$w_2^1 \sim B_1\left(1 - \frac{c_1}{\rho}\right) - \frac{c_1 b_1}{\rho} + \frac{c_1}{2}\log[\eta + \rho] + \mathcal{O}(\rho^{-3}), \qquad \rho \to \infty. \tag{S19}$$

Matching and solving outer problems

In this section, we develop the solution in the outer region away for the receptors and this solution to the inner through a matching process. In light of the local solution expansion (S8), the correct expansion of the outer solution to (S1) is

$$p(\mathbf{x}) = \varepsilon p_0 + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 + \cdots, \tag{S20}$$

where each p_j satisfies

$$\Delta p_j = 0, \quad |\mathbf{x}| > 1, \quad \partial_r p_j = 0, \quad \mathbf{x} \in \partial\Omega \backslash \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}, \tag{S21}$$

subject to certain asymptotic behaviors as $\mathbf{x} \to \mathbf{x}_k$ for $k = 1, \ldots, N$ that are to be determined by matching. For each of these problems, we show below that the solution will either be a zero constant (since $p_j \to 0$ as $|\mathbf{x}| \to \infty$) or a superposition of Green's functions, where each receptor effectively introduces a Coulomb source of a certain strength.

The matching condition is that as $\mathbf{x} \to \mathbf{x}_k$ and $\rho \to \infty$.

$$\varepsilon p_0 + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 + \cdots \sim w_0^k + \varepsilon \log(\varepsilon/2) w_1^k + \varepsilon w_2^k + \cdots. \tag{S22}$$

Plugging (S12) into (S22) and using (S11b) and the leading order behavior,

$$\rho \sim \varepsilon^{-1} |\mathbf{x} - \mathbf{x}_k|,$$

we determine the leading order matching condition

$$\varepsilon p_0 \sim \delta_{1k} \frac{\varepsilon c_k}{|\mathbf{x} - \mathbf{x}_k|}. \tag{S23}$$

Hence, we have that p_0 solves the problem

$$\Delta p_0 = 0, \quad |\mathbf{x}| > 1; \qquad \partial_r p_0 = 0, \quad \mathbf{x} \in \partial\Omega \backslash \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}; \tag{S24a}$$

$$p_0 \sim \frac{c_1}{|\mathbf{x} - \mathbf{x}_1|}, \quad \text{as} \quad \mathbf{x} \to \mathbf{x}_1; \qquad p_0 \to 0, \quad \text{as} \quad |\mathbf{x}| \to \infty. \tag{S24b}$$

A comparison of (S2) and (S3) reveals that the solution of (S24) can be expressed as

$$p_0(\mathbf{x}) = 2\pi c_1 G(\mathbf{x}; \mathbf{x}_1). \tag{S25}$$

We now seek the higher order corrections to (S25) by matching higher order terms to the expansion (S22). Now, if $k \neq 1$, the matching condition as $\mathbf{x} \to \mathbf{x}_k$ becomes

$$\begin{aligned} &\varepsilon p_0(\mathbf{x}_k) + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 + \cdots \\ &\quad \sim \varepsilon \log(\varepsilon/2) A_k \Big(1 - \frac{\varepsilon c_k}{|\mathbf{x} - \mathbf{x}_k|}\Big) + \varepsilon B_k \Big(1 - \frac{\varepsilon c_k}{|\mathbf{x} - \mathbf{x}_k|}\Big) + \cdots, \\ &\quad \sim \varepsilon \log(\varepsilon/2) A_k + \varepsilon B_k - \varepsilon^2 \log(\varepsilon/2) A_k \frac{c_k}{|\mathbf{x} - \mathbf{x}_k|} - \varepsilon^2 B_k \frac{c_k}{|\mathbf{x} - \mathbf{x}_k|} + \cdots. \end{aligned} \tag{S26}$$

This suggests that we need to modify the outer expansion by adding a constant logarithmic switchback term $\varepsilon \log(\varepsilon/2)\chi$. However, since p must vanish as $|\mathbf{x}| \to \infty$, any constant solution must be trivial so that $\chi = 0$ and thus $A_k = 0$ for $k = 2, \ldots, N$. Hence, the matching condition (S26) reduces to

$$\varepsilon p_0(\mathbf{x}_k) + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 + \cdots \sim \varepsilon B_k - \varepsilon^2 B_k \frac{c_k}{|\mathbf{x} - \mathbf{x}_k|} + \cdots. \tag{S27}$$

From (S27), it follows that

$$B_k = p_0(\mathbf{x}_k) = 2\pi c_1 G(\mathbf{x}_1; \mathbf{x}_k), \qquad p_2 \sim -\frac{B_k c_k}{|\mathbf{x} - \mathbf{x}_k|}, \quad \text{as} \quad \mathbf{x} \to \mathbf{x}_k, \quad k \neq 1. \tag{S28}$$

The term p_1 will match to terms of $\mathcal{O}(\varepsilon^2 \log(\varepsilon/2))$ in the higher order corrections of (S8). We now examine the local behavior near the distinguished receptor as $\mathbf{x} \to \mathbf{x}_1$. From the local behavior of the Green's function (S7), we have that as $\mathbf{x} \to \mathbf{x}_1$,

$$\varepsilon p_0 \sim c_1 \Big[\frac{1}{\rho} + \frac{\varepsilon}{2} \log(\varepsilon/2) + \frac{\varepsilon}{2}\Big(\log[\eta + \rho] - \frac{1}{\rho^3}\big(\eta |\mathbf{s}|^2 + s_1^2 s_2 \cot\theta_k\big)\Big)\Big] + \mathcal{O}(\varepsilon^2). \tag{S29}$$

The matching condition (S22) as $\mathbf{x} \to \mathbf{x}_1$, $\rho \to \infty$, using (S29), (S14), and (S18), is then

$$\begin{aligned} &c_1 \Big[\frac{1}{\rho} + \frac{\varepsilon}{2} \log(\varepsilon/2) + \frac{\varepsilon}{2}\Big(\log[\eta + \rho] - \frac{1}{\rho^3}\big(\eta |\mathbf{s}|^2 + s_1^2 s_2 \cot\theta_k\big)\Big)\Big] + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 \\ &\quad \sim \frac{c_1}{\rho} + \varepsilon \log(\varepsilon/2) A_1 \big(1 - \frac{c_1}{\rho}\big) \\ &\quad + \varepsilon \Big[B_1 \big(1 - \frac{c_1}{\rho}\big) - \frac{c_1}{2\rho^3}(s_1^2 s_2 \cot\theta_1) + \frac{c_1}{2} \log[\eta + \rho] - \frac{c_1}{2\rho^3} \eta |\mathbf{s}|^2 - \frac{c_1 b_1}{\rho} \Big] + \cdots, \end{aligned}$$

Canceling out the terms that match automatically, this condition reduces to

$$\begin{aligned} &c_1 \frac{\varepsilon}{2} \log(\varepsilon/2) + \varepsilon^2 \log(\varepsilon/2) p_1 + \varepsilon^2 p_2 + \cdots \\ &\quad \sim \varepsilon \log(\varepsilon/2) A_1 \Big(1 - \frac{c_1}{\rho}\Big) + \varepsilon \Big[B_1 \Big(1 - \frac{c_1}{\rho}\Big) - \frac{c_1 b_1}{\rho} \Big] + \cdots, \quad \text{as} \quad \mathbf{x} \to \mathbf{x}_1, \quad \rho \to \infty. \end{aligned}$$

To establish the local behavior of p_1 and p_2, we return to the outer variable using $\rho \sim \varepsilon^{-1}|\mathbf{x}-\mathbf{x}_1|$ to obtain as $\mathbf{x}\to\mathbf{x}_1$,

$$\begin{aligned} &c_1\frac{\varepsilon}{2}\log(\varepsilon/2)+\varepsilon^2\log(\varepsilon/2)p_1+\varepsilon^2 p_2+\cdots \\ &\quad\sim \varepsilon\log(\varepsilon/2)A_1-c_1\varepsilon^2\log(\varepsilon/2)A_1\frac{1}{|\mathbf{x}-\mathbf{x}_1|}+\varepsilon B_1-\frac{\varepsilon^2 c_1}{|\mathbf{x}-\mathbf{x}_1|}(B_1+c_1)+\cdots. \end{aligned}$$

This matching condition yields that

$$A_1=c_1/2,\quad B_1=0.$$

It then follows that p_1 and p_2 have the local behaviors

$$p_1\sim\frac{-c_1^2/2}{|\mathbf{x}-\mathbf{x}_1|},\quad \text{as } \mathbf{x}\to\mathbf{x}_1,\qquad p_2\sim\frac{-c_1b_1}{|\mathbf{x}-\mathbf{x}_1|},\quad \text{as } \mathbf{x}\to\mathbf{x}_1.$$

The full specification of problem p_1 is now

$$\Delta p_1=0,\quad |\mathbf{x}|>1;\qquad \partial_r p_1=0,\quad \mathbf{x}\in\partial\Omega\backslash\{\mathbf{x}_1,\dots,\mathbf{x}_N\}; \tag{S30a}$$

$$p_1\sim-\frac{1}{2}\frac{c_1^2}{|\mathbf{x}-\mathbf{x}_1|},\quad \text{as}\quad \mathbf{x}\to\mathbf{x}_1;\qquad p_1\to 0,\quad \text{as}\quad |\mathbf{x}|\to\infty. \tag{S30b}$$

while the problem for p_2 is

$$\Delta p_2=0,\quad |\mathbf{x}|>1;\qquad \partial_r p_2=0,\quad \mathbf{x}\in\partial\Omega\backslash\{\mathbf{x}_1,\dots,\mathbf{x}_N\}; \tag{S31a}$$

$$p_2\sim-\frac{c_1b_1}{|\mathbf{x}-\mathbf{x}_1|},\quad \text{as}\quad \mathbf{x}\to\mathbf{x}_1;\qquad p_2\sim-\frac{c_1B_k}{|\mathbf{x}-\mathbf{x}_k|},\quad \text{as}\quad \mathbf{x}\to\mathbf{x}_k,\quad k\neq 1; \tag{S31b}$$

$$p_2\to 0,\quad \text{as}\quad |\mathbf{x}|\to\infty. \tag{S31c}$$

The solutions of these problem are represented in terms of $G(\mathbf{x};\mathbf{x}_0)$ as

$$p_1(\mathbf{x})=-\pi c_1^2 G(\mathbf{x};\mathbf{x}_1), \tag{S32a}$$

$$p_2(\mathbf{x})=-2\pi b_1c_1G(\mathbf{x};\mathbf{x}_1)-4\pi^2\sum_{k=2}^{N}c_1c_kG(\mathbf{x}_k;\mathbf{x}_1)G(\mathbf{x};\mathbf{x}_k), \tag{S32b}$$

where the value of b_1 is defined in (S18). By combining (S25) and (S32) into expansion (S20), we have the final expression for the splitting probability in the limit as $\varepsilon\to 0$,

$$\begin{aligned} p\sim{}&\varepsilon 2\pi c_1G(\mathbf{x};\mathbf{x}_1)-\varepsilon^2\log(\varepsilon/2)\pi c_1^2G(\mathbf{x};\mathbf{x}_1)\\ &-2\pi\varepsilon^2c_1\left[b_1G(\mathbf{x};\mathbf{x}_1)+2\pi\sum_{k=2}^{N}c_kG(\mathbf{x}_k;\mathbf{x}_1)G(\mathbf{x};\mathbf{x}_k)\right]+\cdots. \end{aligned} \tag{S33}$$

The result (S33) is valid when the $k = 1$ receptor is distinguished. Transferring the distinguished receptor to the jth, we recover the main result, namely the splitting probability $p_j(\mathbf{x})$ has the limiting behavior

$$p_j(\mathbf{x}) \sim \varepsilon 2\pi c_j G(\mathbf{x};\mathbf{x}_j) - \varepsilon^2 \log(\varepsilon/2)\pi c_j^2 G(\mathbf{x};\mathbf{x}_j) \\ - 2\pi\varepsilon^2 c_j \left[b_j G(\mathbf{x};\mathbf{x}_j) + 2\pi \sum_{\substack{k=1 \\ k\neq j}}^{N} c_k G(\mathbf{x}_k;\mathbf{x}_j) G(\mathbf{x};\mathbf{x}_k) \right] + \cdots . \tag{S34}$$

as $\varepsilon \to 0$.

As one validation of this new result, we compare with the asymptotic result for the capacitance problem, derived in [S2]. We consider $P(\mathbf{x})$, the probability that the diffusing particle reaches any receptor, i.e.

$$P(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}(\text{absorbed at cell starting from } \mathbf{x}). \tag{S35}$$

From linearity, we have that $P(\mathbf{x}) = \sum_{j=1}^{N} p_j(\mathbf{x})$ and so we have that

$$P(\mathbf{x}) = 1 + C_0 v(\mathbf{x}). \tag{S36}$$

where $v(\mathbf{x})$ is that "escape from capture problem" satisfying

$$\Delta v = 0, \quad \mathbf{x} \in \mathbb{R}^3 \backslash \Omega; \qquad v \sim \frac{1}{|\mathbf{x}|} - \frac{1}{C_0}, \quad \text{as} \quad |\mathbf{x}| \to \infty; \tag{S37a}$$

$$\partial_\nu v = 0, \quad \mathbf{x} \in \partial\Omega \backslash \{\cup_{k=1}^{N} \partial\Omega_j\}; \qquad v = 0, \quad \mathbf{x} \in \partial\Omega_j, \quad k = 1, \ldots, N. \tag{S37b}$$

The capacitance C_0 was determined in [S2] to have expansion as $\varepsilon \to 0$,

$$\frac{1}{C_0} \sim \frac{\pi}{N\varepsilon}\left[1 + \frac{\varepsilon}{\pi}\Big(\log(2\varepsilon) - \frac{3}{2} + \frac{4}{N}\sum_{j=1}^{N}\sum_{k=j+1}^{N} \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|} + \frac{1}{2}\log\left[\frac{|\mathbf{x}_j - \mathbf{x}_k|}{2 + |\mathbf{x}_j - \mathbf{x}_k|}\right]\Big)\right]$$

As a check on Principal Result 1, we use $P(\mathbf{x}) = \sum_{j=1}^{N} p_j(\mathbf{x})$ and result (S21) for $p_j(\mathbf{x})$, to find that

$$P(\mathbf{x}) = \sum_{j=1}^{N} p_j(\mathbf{x}) \sim \varepsilon 2\pi \sum_{j=1}^{N} c_j G(\mathbf{x};\mathbf{x}_j) - \varepsilon^2 \log(\varepsilon/2)\pi \sum_{j=1}^{N} c_j^2 G(\mathbf{x};\mathbf{x}_j) \\ - \varepsilon^2 \sum_{j=1}^{N} \left[2\pi b_j c_j G(\mathbf{x};\mathbf{x}_j) + 4\pi^2 \sum_{\substack{k=1 \\ k\neq j}}^{N} c_j c_k G(\mathbf{x}_j;\mathbf{x}_k) G(\mathbf{x};\mathbf{x}_k) \right] + \cdots . \tag{S38}$$

After algebra and simplification, we find that that (S36) is exactly satisfied.

NUMERICAL VALIDATION OF ASYMPTOTIC SPLITTING PROBABILITIES

In this section, we validate the asymptotic formulas for the splitting probabilities (S34) by comparing to numerical results obtained from a spectral boundary integral method solution of the exterior mixed Neumann-Dirichlet boundary value problem (S1). The computational method is a linear integral equation relating the surface potential, $f(\mathbf{x}) = p|_{\partial\Omega}$, to the surface flux, $q(\mathbf{x}) = \partial_\nu p|_{\partial\Omega} = -\partial_r p|_{\partial\Omega}$. An application of Lagrange's identity to $p(\mathbf{x})$ and $G(\mathbf{x};\mathbf{y})$, the Green's function satisfying (S2), yields the integral equation

$$p(\mathbf{x}) = \int_{\mathbf{y}\in\Omega} G(\mathbf{y};\mathbf{x}) q(\mathbf{y})\, dS, \qquad \mathbf{x} \in \mathbb{R}^3 \setminus \Omega. \tag{S39}$$

The surface flux, $q(\mathbf{x})$, is non-zero only on the receptors, Γ_a. When restricting to the surface of the sphere with specified Dirichlet data $u(\mathbf{x})|_{\partial\Omega} = f(\mathbf{x})$, the following linear integral equation is found

$$f(\mathbf{x}) = \mathcal{A}\left[q(\mathbf{y})\right] \equiv \frac{1}{2\pi} \int_{\mathbf{y}\in\Gamma_a} g\left(|\mathbf{y}-\mathbf{x}|\right) q(\mathbf{y})\, dS, \qquad \mathbf{x} \in \Omega\,, \tag{S40a}$$

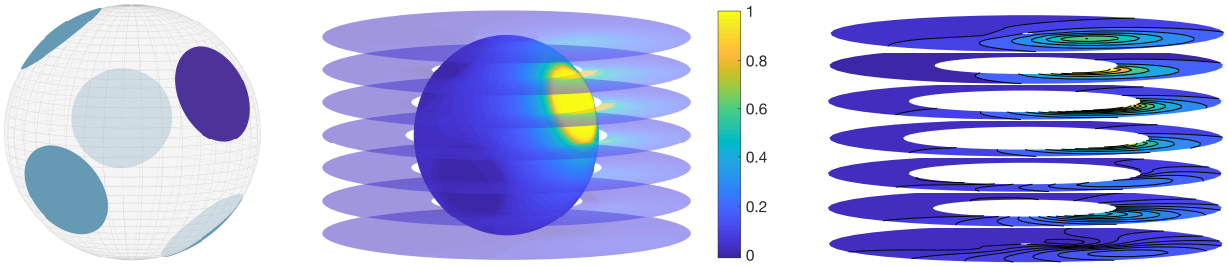
where the kernel of the integral operator is defined by the Green's function (S3) restricted to the sphere

$$G(\mathbf{x};\mathbf{y}) = \frac{1}{2\pi} g\left(|\mathbf{x}-\mathbf{y}|\right), \quad \text{for} \quad \mathbf{x},\mathbf{y} \in \Omega; \qquad g(d) \equiv \frac{1}{d} + \frac{1}{2}\log\left(\frac{d}{2+d}\right). \tag{S40b}$$

The key challenge in the accurate numerical solution of (S40) is the divergence of the surface flux $q(\mathbf{x})$ along $\partial\Gamma_a$ - a notorious feature of mixed Neumann-Dirichlet boundary value problems [S4, S5]. This issue can be resolved by a careful choice of basis functions for the surface potential, $f(\mathbf{y})$, and the surface flux, $q(\mathbf{y})$ in terms of the Zernike polynomials [S6] which are a complete basis for square integrable function supported on circular geometries. By mimicking the known inverse square root singularity structure of the flux, as observed from the solution of the electrified disk problem (S11), equation (S40) can be solved pseudo-spectrally to high accuracy.

This approach, together with full implementation details, has been validated for a wide variety of surface receptor configurations and is effective for the case of thousands [S7] and even hundreds of thousands [S8] of receptors. When the order of the method is fixed, accuracy is reduced when receptors boundaries are nearly touching. The accuracy of the numerical solution can be improved in such cases by taking additional Zernike modes in

the expansion of the solution. In the following test cases, we focus on using this tool to numerically validate the asymptotic formula (S34) for the splitting probabilities. In Fig. S1 we plot the 3D numerical solution of (S1) in a test problem featuring 5 receptors.



(a)Configuration. (b)Solution heat map. (c)Solution contours.

FIG. S1. Numerical solution of (S1) in a 5 receptor case with a single distinguished (purple) receptor. The positions of the receptors are given by (S43) and each have common radius $\varepsilon = 0.4$. We remark that the solution drops to zero quite rapidly away from the sphere.

In the following, we validate the accuracy of the asymptotic formula in three scenarios: a simple three receptor configuration, a clustered configuration, and a homogeneous covering. The accuracy is quantified by considering the relative error

$$\mathcal{E}_{\text{rel}} = \left| \frac{p_{\text{num}} - p_{\text{asy}}}{p_{\text{num}}} \right|. \tag{S41}$$

In each of the examples, we find that the asymptotic expansion has the expected convergence rate as $\varepsilon \to 0$ and accurately predicts the splitting probabilities, provided the receptors are well separated.

Three receptor test case

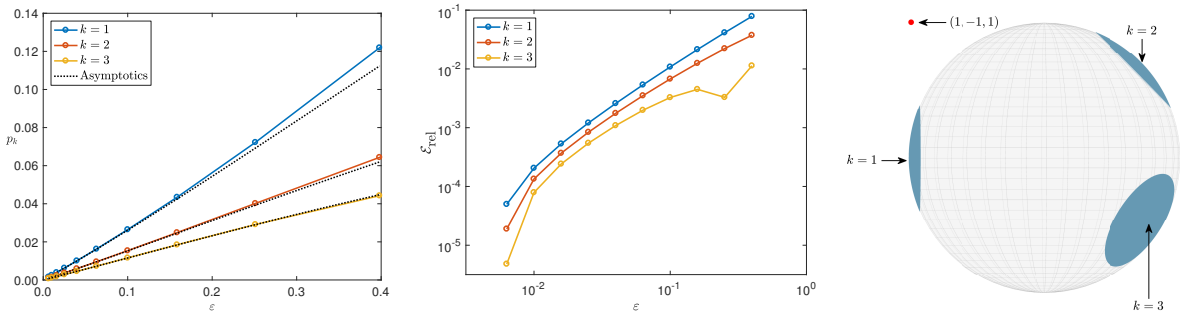
This example benchmarks the asymptotic formula for a simple configuration with three receptors located at

$$\mathbf{x}_k = (\sin\theta_k \cos\varphi_k, \sin\theta_k \sin\varphi_k, \cos\theta_k), \quad k = 1, 2, 3; \tag{S42a}$$

$$(\theta_1, \varphi_1) = \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \qquad (\theta_2, \varphi_2) = \left(\frac{\pi}{4}, \frac{\pi}{2}\right) \qquad (\theta_3, \varphi_3) = \left(\frac{2\pi}{3}, \frac{\pi}{3}\right). \tag{S42b}$$

For the fixed point $\mathbf{x} = (1, -1, 1)$, we demonstrate in Fig. S2 the accuracy of the asymptotic splitting probabilities in the limit as the common receptor radius $\varepsilon \to 0$. In this simple con-

figuration, the splitting probabilities associated with each receptor are remarkable accurate, even at large receptor radii.



(a)Asymptotic splitting probabilities.

(b)Relative error versus receptor radius.

(c)Three receptor configuration for $\varepsilon = 0.4$.

FIG. S2. Comparison of asymptotic and numerical splitting probabilities for the three receptor configuration described in (S42). All curves shown for splitting probabilities evaluated at location $\mathbf{x} = (1, -1, 1)$.

Homogeneous covering case

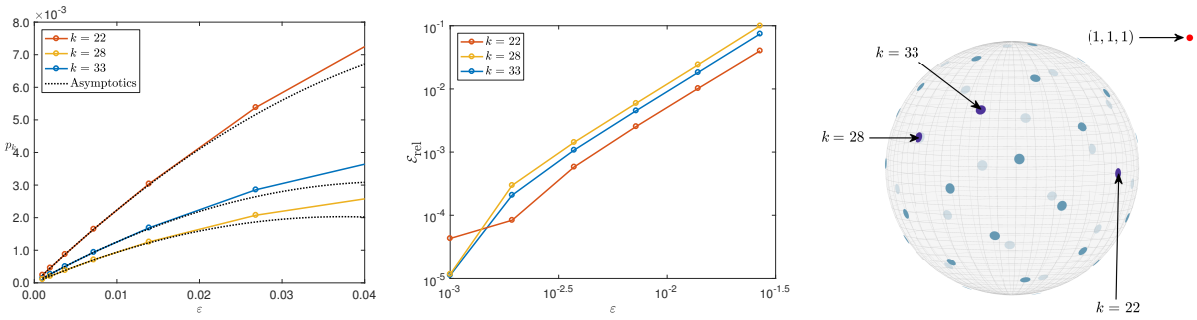
To generate homogeneous coverings of the sphere, we use a simple and effective set known as the Fibonacci spiral points [S9, S10]. Given an integer M, the position of the $N = 2M+1$ spiral points are given by

$$\mathbf{x}_k = (\sin\theta_k \cos\varphi_k, \sin\theta_k \sin\varphi_k, \cos\theta_k); \tag{S43a}$$

$$\sin\theta_k = \frac{2k}{N}, \qquad \varphi_k = \frac{2\pi k}{\Phi}, \qquad k = 1, \ldots, N, \tag{S43b}$$

and $\Phi = 1 + \Phi^{-1} = (1+\sqrt{5})/2 \approx 1.618$ is the golden ratio. We calculate the splitting probabilities for a configuration of $N = 43$ points given by (S43) and display the results in Fig. S3 for three distinguished receptors.

The asymptotic approximation is accurate, provided the individual receptor radius is small and the resulting configuration is well separated.



(a)Asymptotic splitting probabilities

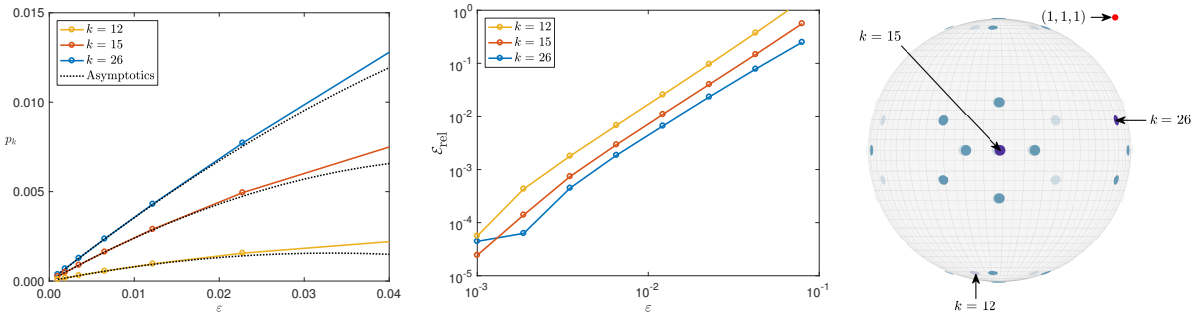
(b)Relative error versus receptor radius.

(c)Homogeneous receptor configuration for $\varepsilon = 0.04$.

FIG. S3. Validation of the asymptotic splitting probabilities for a homogeneous covering of $N = 43$ receptors from the Fibonacci spiral points (S43). The asymptotic is accurate provided the receptor radius is small enough to keep the configuration well-separated. All splitting probabilities evaluated at the point $\mathbf{x} = (1, 1, 1)$.

Clustered Test Case

In this example, we benchmark the asymptotic splitting probabilities on an example with a total of $N = 42$ surface receptors partitioned into 6 clusters of 7 receptors each.



(a)Asymptotic splitting probabilities

(b)Relative error versus receptor radius.

(c)Clustered receptor configuration for $\varepsilon = 0.04$.

FIG. S4. Validation of the asymptotic splitting probabilities for a configuration of $N = 42$ receptors into 6 clusters of 7 each. The asymptotic description is accurate, provided the receptor radius is small enough to keep the configuration well-separated. All splitting probabilities evaluated at the point $\mathbf{x} = (1, 1, 1)$.

MAXIMUM LIKELIHOOD ESTIMATE ERROR SCALING

In FIG. 5 in the main text, we show spatial histograms of the maximum likelihood estimated position $\hat{\mathbf{x}}_{\mathrm{MLE}}$ as a function of the source distance $\|\mathbf{x}\| = R$ and number of particles arriving to receptors M. In the following figures, we plot the average error $\langle \|\hat{\mathbf{x}}_{\mathrm{MLE}} - \mathbf{x}\| \rangle$ as a function of M, R.

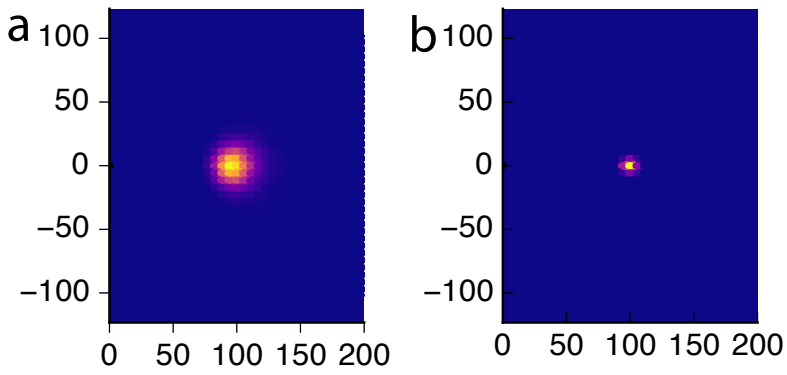


FIG. S5. *Continuation of* FIG. 5 *in main text for* $R = 100$ *and* $M = 10^6, 10^7$. *MLE estimate approaches the true answer as the sample size* M *increases even for far away sources.*

In more detail: in each of the following parameter scenarios, we ran $50,000$ trials. In each trial, we draw counts c_n for $n = 1, \ldots, N$ from a multinomial distribution with M particles so $\sum_n c_n = M$ with the nth receptor having probability $q_n(\mathbf{x})$. The likelihood function $L(x)$ (see main text) was computed with $z = 0$ fixed and the maximizer $\mathbf{x}_{\mathrm{MLE}}$ was found numerically. For each of these trials, a single error is computed, $\|\hat{\mathbf{x}}_{\mathrm{MLE}} - \mathbf{x}\|$ and the averaged quantities $\langle \|\hat{\mathbf{x}}_{\mathrm{MLE}} - \mathbf{x}\| \rangle$ are therefore averaged over these trials, corresponding to the bias of the estimate.

As claimed in the main text, and seen Fig S6, the average error scales $\sim M^{-1/2}$ and $\sim R^2$asymptotically. The scaling on M is expected from the central limit theorem and asymptotic normality of the MLE estimator. The scaling on R persists from the infinite data scenarios described in the main text. Since these are asymptotic scalings, deviations occur if M is small.

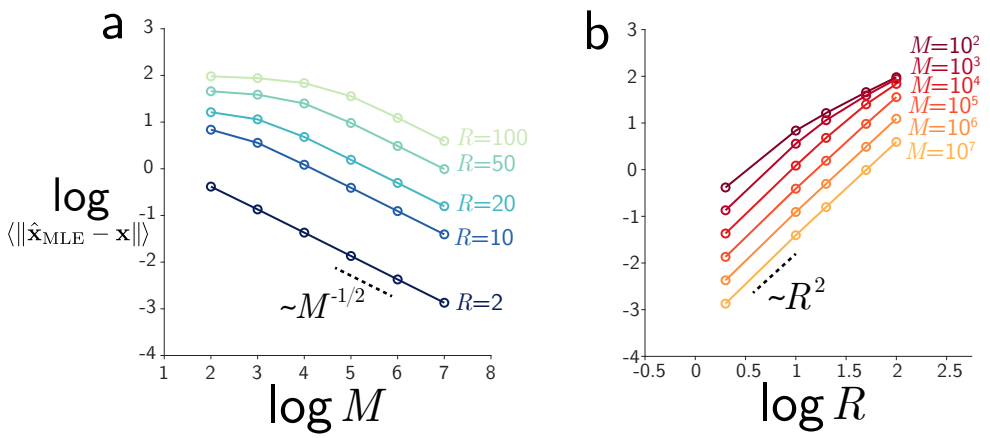


FIG. S6. *Mean error scaling as a function of the number of arriving particles M and source distance R.* Numerical values corresponding to demonstrations in Fig. 5 in the main text.

MEAN ERROR UNIFORM AND CLUSTERED CONFIGURATIONS AND VARIED M, R

In FIG. 6 in the main text, a scenario is depicted demonstrating that clustered configurations can provide smaller average errors $\langle\|\hat{\mathbf{x}}_{\mathrm{MLE}} - \mathbf{x}\|\rangle$. However, this behavior is not general. In the following figures, we report other scenarios where the error is approximately the same or worse for the clustered configuration. Denote e_u the mean error for the uniform configuration and e_c mean the error for the clustered configuration. Sweeping over the source distance R and number of arriving particles M for close sources, we find the following mean errors.

e_u	$R = 1.25$	$R = 1.5$	$R = 2$	$R = 2.5$	$R = 3$
$M = 50$	0.179	0.330	0.721	1.226	1.779
$M = 100$	0.128	0.221	0.478	0.813	1.254
$M = 250$	0.080	0.137	0.293	0.494	0.738
$M = 500$	0.056	0.096	0.206	0.343	0.512
$M = 1000$	0.040	0.067	0.145	0.241	0.358

TABLE S1. Mean error in the MLE estimated locations for the clustered receptor configuration $\mathbf{y}_{\mathrm{unif}}$ in the main text.

From these values, we see that clustered configurations *can* provide lower mean error when

e_c	$R = 1.25$	$R = 1.5$	$R = 2$	$R = 2.5$	$R = 3$
$M = 50$	0.131	0.264	0.625	1.075	1.621
$M = 100$	0.092	0.186	0.449	0.760	1.143
$M = 250$	0.059	0.118	0.287	0.491	0.727
$M = 500$	0.042	0.084	0.205	0.349	0.516
$M = 1000$	0.030	0.060	0.145	0.247	0.367

TABLE S2. Mean error in the MLE estimated locations for the clustered receptor configuration $\mathbf{y}_{\text{clust}}$ in the main text.

$e_u - e_c$	$R = 1.25$	$R = 1.5$	$R = 2$	$R = 2.5$	$R = 3$
$M = 50$	0.049	0.065	0.095	0.151	0.157
$M = 100$	0.036	0.035	0.029	0.053	0.112
$M = 250$	0.021	0.018	0.006	0.003	0.011
$M = 500$	0.014	0.011	0.002	-0.006	-0.004
$M = 1000$	0.010	0.008	-0.001	-0.006	-0.009

TABLE S3. Difference in the mean error in the MLE estimates for the uniform and clustered configuration. Positive values indicate that the error in the uniform configuration is *larger*, indicating worse performance for uniform.

M, R are small. As M and R become sufficiently large that the benefits of the clustered configurations become diminished. This frontier is highlighted in Tables S3 and S4.

* christopher.miles@cims.nyu.edu

[S1] I. M. Nemenman and A. S. Silbergleit, J. Appl. Phys. **86**, 614 (1999).

[S2] A. E. Lindsay, A. J. Bernoff, and M. J. Ward, Multiscale Model. Simul. **15**, 74 (2017).

[S3] I. N. Sneddon, *Elements of Partial Differential Equations* (McGraw-Hill Book Company, Inc., 1957).

[S4] I. N. Sneddon, *Mixed boundary value problems in potential theory* (North-Holland, 1966).

[S5] D. G. Duffy, *Mixed Boundary Value Problems* (Chapman and Hall/CRC, 2008).

[S6] P. Novák and J. Novák, Proc. SPIE **8789**, 878913 (2013).

$100(e_u - e_c)/R$	$R = 1.25$	$R = 1.5$	$R = 2$	$R = 2.5$	$R = 3$
$M = 50$	3.897	4.347	4.753	6.030	5.239
$M = 100$	2.863	2.364	1.435	2.101	3.723
$M = 250$	1.657	1.222	0.324	0.123	0.362
$M = 500$	1.147	0.754	0.089	-0.244	-0.131
$M = 1000$	0.796	0.513	-0.026	-0.250	-0.293

TABLE S4. Relative percentage difference in the mean error in the MLE estimates for the uniform and clustered configuration, scaled by the distance of the target R. Positive values indicate that the error in the uniform configuration is *larger*, indicating worse performance for uniform.

[S7] A. J. Bernoff and A. E. Lindsay, SIAM J. Applied Math. **78**, 266 (2018).

[S8] J. Kaye and L. Greengard, Journal of Computational Physics , 100047 (2019).

[S9] Á. González, Mathematical Geosciences **42**, 49 (2009).

[S10] R. Swinbank and R. J. Purser, Quarterly Journal of the Royal Meteorological Society **132**, 1769 (2006).