Limits on the accuracy of contact inhibition of locomotion

Wei Wang (汪巍)[1] and Brian A. Camley[1,2]

[1]*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA*
[2]*Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA*

Cells that collide with each other repolarize away from contact, in a process called contact inhibition of locomotion (CIL), which is necessary for correct development of the embryo. CIL can occur even when cells make a micron-scale contact with a neighbor—much smaller than their size. How precisely can a cell sense cell-cell contact and repolarize in the correct direction? What factors control whether a cell recognizes it has contacted a neighbor? We propose a theoretical model for the limits of CIL where cells recognize the presence of another cell by binding the protein ephrin with the Eph receptor. This recognition is made difficult by the presence of interfering ligands that bind nonspecifically. Both theoretical predictions and simulation results show that it becomes more difficult to sense cell-cell contact when it is difficult to distinguish ephrin from the interfering ligands, or when there are more interfering ligands, or when the contact width decreases. However, the error of estimating contact position remains almost constant when the contact width changes. This happens because the cell gains spatial information largely from the boundaries of cell-cell contact. We study using statistical decision theory the likelihood of a false-positive CIL event in the absence of cell-cell contact, and the likelihood of a false negative where CIL does not occur when another cell is present. Our results suggest that the cell is more likely to make incorrect decisions when the contact width is very small or so large that it nears the cell's perimeter. However, in general, we find that cells have the ability to make reasonably reliable CIL decisions even for very narrow (micron-scale) contacts, even if the concentration of interfering ligands is ten times that of the correct ligands.

I. INTRODUCTION

More than half a century ago, researchers first observed contact inhibition of locomotion (CIL), when two colliding fibroblasts changed direction rapidly and migrated away from the collision [1–5]. During the following years, different variants of CIL were found in a large number of healthy and cancerous cell types [2]. The outcomes of cell-cell collisions that can be called CIL may include both cells reversing or a single cell reversing, leading to cell train formation, depending on the cell environment and cell type [5–10]. The outcomes of cell-cell collisions are also known to be stochastic in nature [5–7]. We will think of CIL as the phenomenon of a cell actively changing its polarization upon encountering—and specifically recognizing—another cell. CIL can be classified into homotypic (occurring between cells of the same type) and heterotypic CIL (occurring between cells of different types), which is usually lost in cancer cells when confronted with normal cells [2]. CIL is now a well-known characteristic of normal cells and plays an important role in regulating cell motility [4,5], tissue growth [11], and development [12–14]. For example, experiments revealed that *Drosophila* macrophages (haemocytes) require CIL for their uniform embryonic dispersal [14,15]. CIL can happen—though rarely—even when cells are crawling on isolated nanofibers, leading to very small (∼μm) contact regions [5]. How can cells reliably observe the presence of another cell just by such a small contact?

CIL is regulated by the Eph-ephrin signaling pathway [12,14,16,17]. Contact detected by Eph-ephrin or cadherin binding [18,19] reorganizes the actin cytoskeleton [20] by modulating the main elements of the Rho GTPase family [21,22], resulting in the increase of actomyosin contractility and the inhibition of lamellipodia formation at the cell-cell contact [12,23]. Eph receptors are a family of transmembrane receptor tyrosine kinases (RTK) [12,16] that regulate tissue boundary formation in development [24] and tissue organization maintenance in adult organisms [12]. Their ligands, known as ephrins, are normally anchored to the cell surface [12,13,16]—Eph-ephrin signaling occurs when cells come into contact with one another. Ephrin cues can be used to reorganize and repolarize cells. Ephrin-coated beads lead to CIL when cells collide with them [13], and micropatterned ephrin spatial cues can be used to align intestinal crypt cells [12].

Cells can use receptors on their membrane to estimate external ligand concentrations [25–27] and also to sense spatial [28–33] or temporal [34] gradients. The accuracy limits of these estimations have been extensively studied [26,28–31], including related measurements of spatially localized extracellular signals [29]. For a cell to undergo CIL, it must detect the presence and location of the cell-cell contact—akin to measuring a spatially localized chemical. The cell's ability to make this detection will be challenged if ligands other than ephrin, i.e., spurious ligands, bind nonspecifically to the Eph receptors [35–38]. Potential candidates for the spurious ligands are soluble ephrins [39], ephrins on extracellular vesicles [40], or ligands for other RTKs, e.g., growth factors [41]. In scenarios where there is only a small patch of contact [5], such
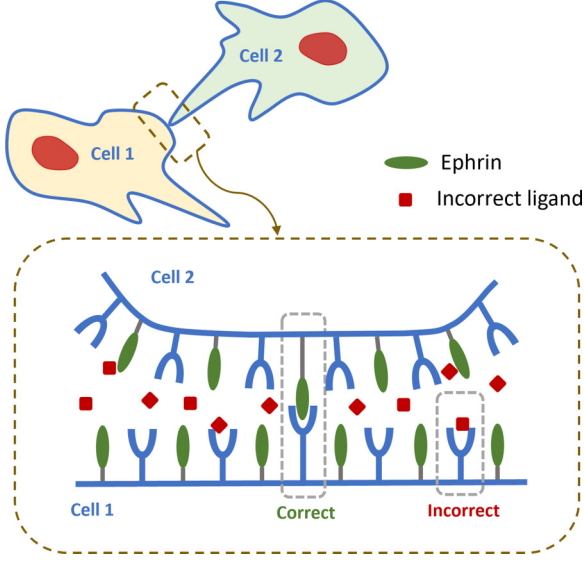
FIG. 1. Illustration of two cells in contact where they can sense each other through Eph-ephrin signaling. Both Eph receptors (blue receptors) and their ligands—ephrins (green ellipses), are distributed on the cell membranes. Besides the cognate ligands, other spurious ligands (red squares) can also bind to the Eph receptors. These incorrect binding events mixed with the correct Eph-ephrin binding will affect the sensing accuracy of contact inhibition of locomotion.

interference from other ligands poses a significant threat to the precision of cell sensing.

In this paper, we develop a theory for sensing cell-cell contact via Eph receptor-ephrin contact in the presence of spurious ligands using a multiligand model [26,35,36,42]. To analyze the optimal performance of such a sensing mechanism, we adopt the method of maximum likelihood estimation (MLE) and derive the fundamental sensing limits by using the Cramér-Rao bound. We find that the accuracy of detecting the location of the other cell is degraded when the binding affinity of the spurious ligands is close to the true ephrin's affinity, or when the fraction of correct ligands is small. Surprisingly, accuracy does not depend strongly on the contact width. These results are also supported by Monte Carlo simulations. We use statistical decision theory to study the problem of whether the cell can effectively detect the presence of another cell via a small cell-cell contact, and show how cells can trade off false negatives, where they fail to respond to another cell, and false positives, where they react even in the absence of a contacting cell. We find that cell-cell contact can be highly reliably detected even at large levels of spurious ligand concentrations.

II. MODEL

Consider two cells coming into contact as shown in Fig. 1, where both cells can sense each other through Eph-ephrin binding [12,13]. We study the sensing accuracy for cell 1 detecting cell 2, as the mirror problem is equivalent.

As an initial model for the fundamental sensing limits of contact inhibition of locomotion, we work in an effectively one-dimensional model, with Eph receptors characterized by their location x around the perimeter of cell 1. We assume
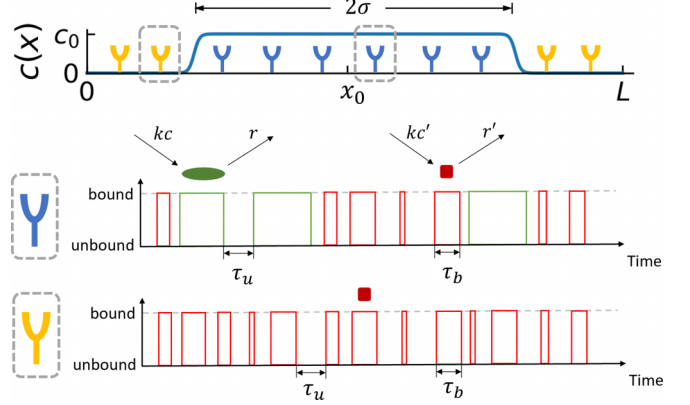

FIG. 2. Extracting information from bound-unbound time series of receptors. Top: The ephrin concentration is a smooth rectangular function $c(x)$ that is high at the region of contact ($\sim \pm \sigma$ of the contact center at x_0). Bottom: The temporal record of the binding states of example receptors inside and outside the contact region. Receptors within the contact region (blue) can bind to both ephrins (green) and incorrect ligands (red) while receptors outside (yellow) can only bind to the spurious ligands.

ephrins from cell 2 can bind to the Eph receptors on cell 1 with a rate $kc(x)$, i.e., a base rate k times an effective ephrin concentration $c(x)$. The rate $kc(x)$ combines both the area density of the ephrin on the membrane with the reduced likelihood of binding for ephrins which are further away from the membrane of cell 1—so if we treat k as constant, then we expect $c(x)$ to vary from its maximum value at the cell-cell contact to essentially zero far away from the contact. The details of the shape of $c(x)$ will depend on the details of the cell-cell junction and the cell-cell contact, but we make an initial hypothesis that $c(x)$ is effectively constant over the cell-cell contact width and then smoothly decreases to zero (Fig. 2, top). We use a smooth rectangular function for the effective ephrin concentration that cell 1 senses on cell 2's surface:

$$c(x) = c_0 S(x - (x_0 - \sigma)) S((x_0 + \sigma) - x), \quad (1)$$

where $S(x) = [1 + \tanh(x/\omega)]/2$, and ω controls the steepness at the transition between 0 and c_0. x_0 is the center of the region of cell-cell contact and 2σ is the contact width. Receptors bound with the correct ligand (ephrin) will unbind with rate r. Receptors can also bind the incorrect (spurious) ligand with rate kc' and unbind with rate r' (Fig. 2, bottom), where we assume the effective concentration of incorrect ligands c' is constant.

In our model, we adopt the assumption of Ref. [35], where the only distinction between the two ligands is their binding kinetics, neglecting features like potential ligand bias [43,44]. This means that the cell only knows whether a receptor is bound, not what ligand is bound to it. How, then, can the cell detect the presence of another cell and repolarize away from it? First, even if ligands had identical unbinding rates, there would be more bound receptors at the cell-cell contact. Second, if the correct and incorrect ligands have different affinities ($r \neq r'$), then the cell can discriminate between them statistically. Ligands with a larger unbinding

rate remain bound to receptors for a shorter time compared to ligands with a smaller unbinding rate. Thus, the cell can discern the mixture of two different ligands if the detailed occupancy history of each receptor is available [35]. This record can be summarized by the times the receptor spends unbound $\tau_{u,i}$ and the times it spends bound $\tau_{b,i}$ [34,35,45], e.g., $\{\tau_{u,1}, \tau_{b,1}, \tau_{u,2}, \tau_{b,2}, \cdots\}$ (as shown in Fig. 2, bottom). Given the receptor record, $c(x)$ can be estimated which allows the cell to identify the location of the cell-cell contact x_0.

The probability densities to have a bound interval of length τ_b and an unbound interval of length τ_u for a receptor seeing concentration c of correct ligand are [30,35]

$$P(\tau_u) = k(c + c')e^{-k(c+c')\tau_u}, \tag{2}$$

$$P(\tau_b) = \frac{c}{c + c'}re^{-r\tau_b} + \frac{c'}{c + c'}r'e^{-r'\tau_b}, \tag{3}$$

where the factor $(c + c')$ in Eq. (2) is the total concentration, and the factors $c/(c + c')$ and $c'/(c + c')$ in Eq. (3) are the probabilities of the binding event being the correct ligand and incorrect ligand, respectively. We assume that there are N receptors evenly spaced over the cell perimeter L, each with position x_n so the nth receptor has ephrin concentration $c(x_n)$.

We assume that the cells observe the binding and unbinding events over a "measuring time" T and use this to make a CIL decision—estimating the position of the cell-cell contact. The probability of a particular time record $\{\tau_{u,1}, \tau_{b,1}, \tau_{u,2}, \tau_{b,2}, \cdots\}$ for the nth receptor is $P_n = \prod_{i=1}^{M_n} P(\tau_{u,i}^{(n)}) \cdot P(\tau_{b,i}^{(n)})$, where M_n is the number of pairs of binding-unbinding events during the time T. Therefore, given the three parameters $\boldsymbol{\theta} \equiv (c_0, c', x_0)$ in our model, the total probability of a trajectory of bound and unbound times arising from N independent receptors is $P(\{\tau\}|\boldsymbol{\theta}) = \prod_{n=1}^{N} P_n$.

To derive the fundamental sensing limits, we use the method of maximum likelihood estimation (MLE) [28,45,46]. The log-likelihood function is $\ln \mathcal{L}(\boldsymbol{\theta}|\{\tau\}) = \ln P(\{\tau\}|\boldsymbol{\theta})$. The MLE estimate of the parameters $\hat{\boldsymbol{\theta}}$ is obtained by maximizing the log-likelihood, e.g., by solving $\partial \ln \mathcal{L}/\partial \theta_\mu|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$, where the index $\mu = 1, 2, 3$, or by numerical maximization of $\ln \mathcal{L}$. The error between any unbiased estimate of the parameters that the cell can make, $\hat{\boldsymbol{\theta}}$, and the true parameters is limited by the Fisher information matrix and the Cramér-Rao bound [Eq. (4)]. We can compute the Fisher information matrix $I(\boldsymbol{\theta})_{\mu\nu} = -\langle \partial^2 \ln \mathcal{L}/\partial\theta_\mu \partial\theta_\nu \rangle$ for our model in a partially analytic way. After taking the derivatives of the log-likelihood necessary to compute the Fisher information matrix, we get an equation that is a complicated function of the bound and unbound times $\{\tau_b, \tau_u\}$. This reduces the problem to quadrature—we can calculate the average $\langle \cdots \rangle$ by numerical integration using the probability densities in Eqs. (2) and (3) (see Appendix B for details). The Cramér-Rao bound (CRB) then gives the lower bound for the variance of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of parameters [31]:

$$\text{Var}(\hat{\theta}_\mu) \geqslant [I(\boldsymbol{\theta})^{-1}]_{\mu\mu}, \tag{4}$$

where $I(\boldsymbol{\theta})^{-1}$ is the matrix inverse. Equation (4) sets the best precision with which a cell could measure, e.g., the location of the cell-cell contact, given the unavoidable stochasticity arising from the ligand-receptor interactions.

III. RESULTS

How precisely can a cell detect the position of cell-cell contact, or the concentrations of ligands? This is set by Eq. (4)—the best possible unbiased estimates we can make of our parameters. We will display our results in terms of the variables $\boldsymbol{\theta}' \equiv (c_t, \chi, x_0)$, where $c_t \equiv c_0 + c'$ is the total concentration, i.e., the concentration that sets the total binding rate at the cell-cell contact and $\chi = c_0/c_t$ is the fraction of correct ligands. When $\chi \ll 1$, most bindings are the incorrect ligand, and we expect the cell to struggle to perform CIL. We also define the unbinding rate ratio of the two ligands as $\alpha \equiv r/r'$. The correct ligands are expected to have a higher affinity to the receptors and thus take longer to unbind, i.e., $r \leqslant r'$, so we assume $\alpha \leqslant 1$.

In Fig. 3, we plot both our calculations of the minimal standard deviations required from the Cramér-Rao bound in Eq. (4) as well as simulation results from Monte Carlo (MC) simulations where we generate stochastic receptor trajectories according to Eqs. (2) and (3), and then numerically maximize the likelihood to determine the maximum-likelihood estimators $\hat{c}_t$, $\hat{\chi}$, and $\hat{x}_0$ (Appendix C). Broadly, we see good agreement between simulation and theory, except for a few data points at small patch sizes when $\sigma \sim 1$–$2\,\mu m$, which we discuss further below.

A. How do the sensing limits change when it becomes difficult to distinguish the two ligands?

The unbinding rate ratio $\alpha = r/r'$ quantifies the difference in receptor-ligand affinity between the two ligands. In other words, it characterizes how hard it is for the cell to discriminate the two ligands. When α is close to 0, i.e., in the easy discrimination regime, the standard deviations of all three estimated parameters ($\hat{c}_t$, $\hat{\chi}$, $\hat{x}_0$) are small. When α gets closer to 1, it becomes harder to distinguish the two ligands. If $\alpha = 1$, then both ligands have the same unbinding rate and are thus indistinguishable for the cell within our assumptions. The difficulty to tell the two ligands apart leads to a rise in sensing errors for all parameters as α nears 1, i.e., the smaller the discrepancy between ligands, the less accurately the cell can sense the contact [Figs. 3(a)–3(f)]. The standard deviations of all three parameters are small compared to their relevant scales, which are for Fig. 3 total concentration $c_t = 1000\,\mu m^{-2}$, correct ligand fraction $\chi = 0.1$, and the cell perimeter $L = 100\,\mu m$. We also see that the errors relative to the mean are small (Fig. 9 in Appendix D). This means that the limits imposed by stochastic noise of ligand-receptor binding may not be that stringent in practice. For example, if the cell can locate the source of cell-cell contact to $\pm\text{SD}(\hat{x}_0)$ of ~ 0.4 microns [the worst case of Fig. 3(f)], then the cell has an angular uncertainty of the direction to go which is $2\pi\,\text{SD}(\hat{x}_0)/L$ of $\sim 1.5°$. This angular uncertainty is much smaller than the typical angular range of acceleration vectors observed in CIL on two-dimensional substrates [14], which indicates that cells likely have more positional information than they really use in CIL—unless the concentration of spurious ligands is extraordinarily high ($\chi \ll 1$) or our model is incomplete or there is additional noise due to one of the other factors we raise later in Sec. IV.
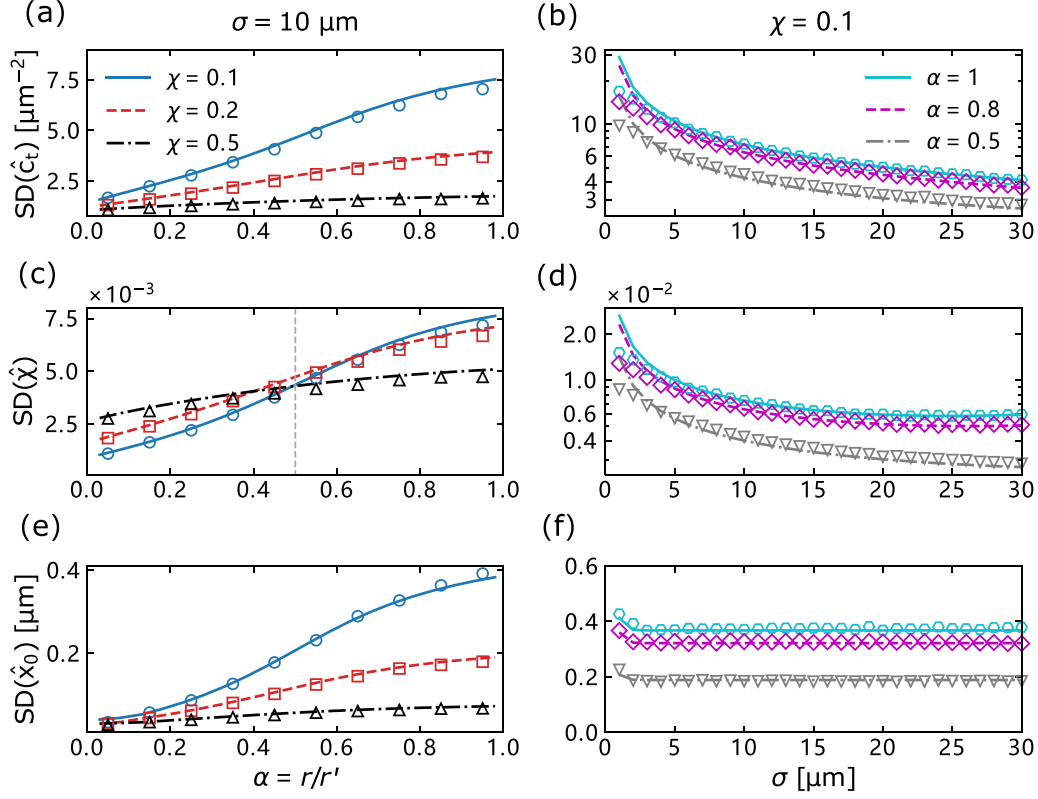
FIG. 3. Errors of maximum likelihood estimate (MLE) influenced by receptor-ligand affinity, fraction of correct ligands, and contact width. Panels (a), (b) show the standard deviation (SD) of the total concentration $\hat{c}_t$, panels (c), (d) show the SD of correct ligand fraction $\hat{\chi}$, and panels (e), (f) display the SD of contact center position $\hat{x}_0$. In all figures, the lines (solid, dashed, and dash-dot) represent theoretical results obtained from Eq. (4), while the points are obtained from Monte Carlo simulations (the standard deviations are computed from 10^4 simulations for each point); Measuring time is $T = 10$ s. Corresponding relative errors (SD/average) are shown in Fig. 9 in Appendix D.

In Figs. 3(a), 3(c), and 3(e), we also show the standard deviations for different correct ligand fractions $\chi = c_0/c_t$. We expect that, as the fraction of correct ligands decreases, the error in estimating each parameter should increase. We do see that errors associated with estimating the total concentration c_t and the contact center position x_0 decrease as χ increases [Figs. 3(a) and 3(e)]. However, estimating the correct ligand fraction χ itself has a more complex behavior [Fig. 3(c)]. Estimating χ in our model is akin to a spatially dependent multireceptor model of the question asked by Ref. [35]—how precisely can a single receptor detect a rare ligand binding? Our model, though, shows qualitative differences compared to the single receptor model of Ref. [35]. For a single receptor in the easy discrimination regime when the cell is confident about what type of ligand is bound ($\alpha < 1/2$), the uncertainty of sensing χ increases with χ as SD($\hat{\chi}$) $\sim \chi^{\beta/2}$ with $\beta = 1 - \alpha/(1 - \alpha)$, though the relative error SD($\hat{\chi}$)/χ always decreases with χ [35], while in the hard discrimination regime ($\alpha > 1/2$), SD($\hat{\chi}$) is roughly independent of χ, and diverges as $\alpha \to 1$, where a single receptor cannot discriminate between ligands. In our spatial, multireceptor model, we find that when $\alpha < 1/2$ the error in χ increases with χ, as seen in Ref. [35]. However, in the hard discrimination regime ($\alpha > 1/2$), we find that SD($\hat{\chi}$) has the opposite behavior, increasing when χ decreases. We argue that this occurs because there are two ways for the cell to estimate χ: first, by discriminating between the two ligands by their binding times, and

second, by comparing the total amount of binding observed at the cell-cell contact (depending on $c_0 + c'$) to the binding outside the contact (depending only on c'). The second mechanism does not require that the cell can discriminate between the two ligands, and will still function as $\alpha \to 1$. Hence, we see that as $\alpha \to 1$, the error of sensing χ decreases as χ increases just like the error of the total concentration c_t. This source of error is essentially arising because at small χ, the difference between $c_0 + c'$ and c_0 will shrink. By contrast, for $\alpha < 1/2$, the estimation of χ is similar to the single-molecule discrimination of Ref. [35] and the error in χ increases with χ.

B. How does contact width limit CIL?

As the contact width σ expands, more receptors come inside the contact region and the cell can obtain more information about the concentration of the true ligands—we expect that the accuracy of estimating our parameters should generally increase as the contact width increases. In Figs. 3(b) and 3(d), we observe that both the standard deviation of total concentration SD($\hat{c}_t$) and the standard deviation of the fraction of correct ligands SD($\hat{\chi}$) decrease with increasing contact width σ. Surprisingly, we find the error in estimating the contact position SD($\hat{x}_0$) remains almost constant as σ ranges from 3–30 μm [Fig. 3(f)]. Why? We argue that the cell's information about the contact location is largely coming from where the edges of the contact are, i.e., the transitions

in $c(x)$. To better illustrate this, consider a simpler model that only involves one snapshot in time and one species of ligands, which we denote with a capital $C(x)$. In this toy model, given concentration $C(x)$, the occupancy for each receptor is given by $p_n = C(x_n)/[C(x_n) + K_D]$, with the dissociation constant $K_D = r/k$ [32]. The joint probability for the state of all receptors is $P(\{Z_n\}) = \mathcal{L}^{\text{toy}}$ with

$$\mathcal{L}^{\text{toy}} = \prod_n p_n^{Z_n}(1 - p_n)^{1-Z_n}, \tag{5}$$

where $Z_n = 0$ indicates the receptor is unbound and $Z_n = 1$ indicates it is bound. In this toy model, we can compute the Fisher information:

$$I_{\mu\nu}^{\text{toy}} = -\left\langle \frac{\partial^2 \ln \mathcal{L}^{\text{toy}}}{\partial \theta_\mu \partial \theta_\nu} \right\rangle = \sum_n \frac{K_D}{[C(x_n) + K_D]^2 C} \frac{\partial C(x_n)}{\partial \theta_\mu} \frac{\partial C(x_n)}{\partial \theta_\nu}. \tag{6}$$

If the concentration $C(x)$ takes the form of a smooth rectangular function [Eq. (1)], and we compute the x_0-x_0 component of the Fisher information matrix, then the only terms that are nonzero in the sum in Eq. (6) are those where the derivative $\partial C(x)/\partial x_0$ is nonzero—which are the edges of the contact zone. By contrast, if we were going to estimate the concentration level of the contact, then you would get a nonzero term at every receptor within the contact, and we would expect the error to decrease with contact size.

In Fig. 3(f) the uncertainty in measuring the contact center x_0 does increase for small enough σ, in the range of $\sigma = 1$–$2\,\mu\text{m}$. This happens when the contact width σ is similar to the scale $\omega = 1\,\mu\text{m}$ over which $c(x)$ is varying—so then $\partial c/\partial x_0$ will be nonzero over the whole contact patch. If we increase ω, then the general behaviors of the errors will not change, but the increase in SD$(\hat{x}_0)$ at small σ will occur at a larger σ, as the concentration profile becomes strongly dependent on x when $\sigma \sim \omega$ [see Fig. 6(f) in Appendix D].

However, the derivative $\partial c/\partial x_0$ depends significantly on the chosen concentration profile—it is zero except at the boundaries because we are using a smooth rectangular function [Eq. (1)] here. If we choose a concentration profile where $\partial c/\partial x_0$ is nonzero throughout, such as a Gaussian function, then all receptors within the contact region will contribute to the accuracy of position sensing, and the dependence of the estimation errors of $\hat{x}_0$ shown in Fig. 3(f) will be different. See Sec. IV and Fig. 10 (see Appendix D) for more details.

For most of Figs. 3(a)–3(f), we see good agreement between our bound from the Fisher information in Eq. (4) and the variance of the maximum likelihood estimator computed from a stochastic simulation. However, at the smallest σ, we do see a larger deviation between Monte Carlo simulations and Fisher information results in Figs. 3(b), 3(d), and 3(f). This discrepancy likely arises from the difficulty of the maximization process when σ is small—when we do numerical optimization to find the maximum-likelihood estimator $\hat{\boldsymbol{\theta}}$, there might be some deviations from the true global maximum point—if the optimizer becomes trapped in a local maximum, or if the number of iterations provided is too few, etc. These numerical problems are more likely to happen for small contact size σ and $\alpha \to 1$, because in such cases, there is less difference between a cell with a very small contact patch and one with essentially no contact ($\chi = 0$). In fact, the reasons

that make it difficult for us to do this numerical optimization also make it difficult for the cell to determine if another cell is present, which we address in the next section. In other words, we suspect that any effective biochemical optimization process would also be difficult in cases where numerical optimization fails.

C. How reliably can the cell detect contact at all?

In many situations, determining the presence of another cell is more useful for the cell than precisely locating the contact point. For instance, a recent experiment [5] studied CIL for cells on suspended nanofibers, which are used to mimic the extracellular matrix. In this biologically relevant case—and in other experiments of CIL on micropatterns [7,13]—cells only have two possible directions. Thus, the precision in estimating the location of contact x_0 is not so important, but deciding whether there is contact or not is very important.

We study the problem of how precisely a cell can detect the presence of a neighbor from a localized signal by using tools from statistical decision theory. First, if we regard the cell as a binary classifier, then we can pose a question to the cell: "Is there another cell in proximity?" This question can be formally translated into the task of distinguishing between two hypotheses: the presence versus absence of another cell. In our model, if we denote the three-dimensional parameter space where $\boldsymbol{\theta}$ lives as Θ, then the two hypotheses—presence or absence [35,42,47]—can be written as

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \equiv \{(c_0, c', x_0)|c_0 = 0\}, \tag{7}$$

$$H_1 : \boldsymbol{\theta} \in \Theta_0^C \equiv \Theta \backslash \Theta_0, \tag{8}$$

where H_0 (called the null hypothesis) states $\boldsymbol{\theta}$ lives within a subspace Θ_0, which corresponds to the plane of $c_0 = 0$, i.e., the other cell is not there. And H_1 (called the alternative hypothesis) asserts $\boldsymbol{\theta}$ lives in Θ_0^C, the complement space of Θ_0, i.e., $c_0 \neq 0$ and there is another cell nearby.

This question can be addressed by employing a statistical method known as the likelihood-ratio test (Wilks test) [47] to help the cell make decisions. The basic concept of this test is to calculate the ratio of the maximum likelihood under the two competing hypotheses:

$$\lambda = \ln \frac{\sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}}{\sup_{\boldsymbol{\theta} \in \Theta_0} \mathcal{L}} = \sup_{\boldsymbol{\theta} \in \Theta}(\ln \mathcal{L}) - \sup_{\boldsymbol{\theta} \in \Theta_0}(\ln \mathcal{L}) \geqslant 0, \tag{9}$$

where the sup notation refers to the supremum. Therefore, if λ is large, then the alternative hypothesis H_1 (presence) is more competitive. In contrast, the null hypothesis H_0 (absence) becomes a more compelling choice if λ is small. In general, the ratio λ will tune how much evidence is needed to prefer hypothesis H_1 over hypothesis H_0. Then we can set an adjustable threshold value λ_c for the cell. If $\lambda > \lambda_c$, then the cell believes it is contacting another cell. Conversely, if $\lambda < \lambda_c$, then the cell concludes that there is no other cell nearby.

To gain deeper insight into the problem, we want to know: How likely is it for the cell to fail in detecting the presence of another cell when it is indeed there? Or how likely is it for the cell to claim the detection of another cell but it is actually absent? In statistics, the former probability is defined as the false-negative rate (FNR), and the latter is referred to as the false-positive rate (FPR). In our model, the false-negative rate
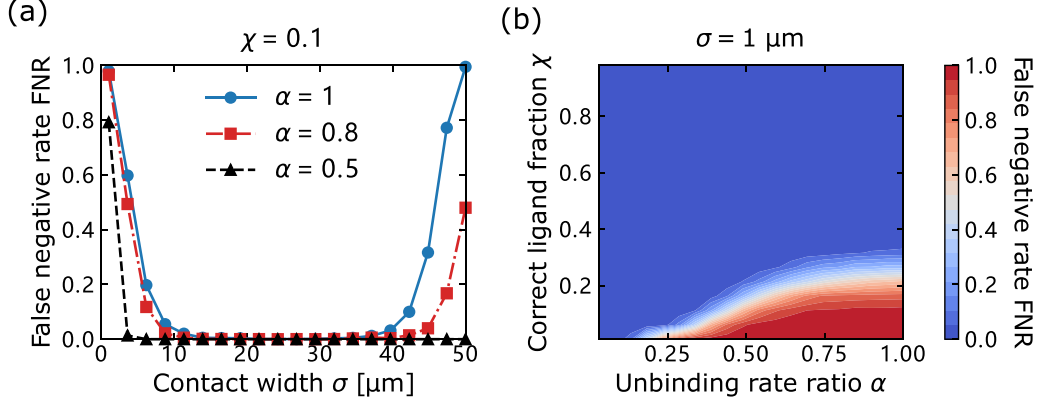
FIG. 4. False-negative rates showing how the probability of a cell missing the presence of a contacting cell is affected by contact width σ, unbinding rate ratio α, and the fraction of correct ligands χ. Panel (a) illustrates the change in FNR as contact width varies, and the phase diagram in panel (b) shows a sharp transition from low α, high χ values to high α, low χ values. We conduct $\sim 10^5$ simulations for each data point. The threshold parameter is set to $\lambda_c = 7$ is chosen somewhat arbitrarily to show when the maximum value of FNR can be close to 1 in these figures.

can be computed as [35]

$$\text{FNR} = \langle H(\lambda_c - \lambda) \rangle_+, \tag{10}$$

where $H(x)$ is the Heaviside step function. The symbol $\langle \cdots \rangle_+$ denotes the average over all allowed configurations of the time series, and the positive subscript $+$ indicates the average is taken under the condition of another cell being there. Only when the cell misses the presence of the other cell, i.e., $\lambda < \lambda_c$, the step function is nonzero. Thus, Eq. (10) gives exactly the probability of the cell making a false-negative decision, where it erroneously asserts the absence of the other cell. To compute the FNR, we choose a value λ_c, use our Monte Carlo simulation to generate a set of receptor occupancies over time given the presence of a second cell (Appendix C), then numerically maximize the likelihood under the two hypotheses (presence and absence of the contacting cell) to compute λ from Eq. (9). We then repeat this many times to compute the FNR from Eq. (10).

Figure 4(a) shows the Monte Carlo simulation results of FNR for varying contact widths. At small σ, when there are relatively few receptors where the true concentration $c(x)$ is nonzero, the false-negative rate is large. Essentially, in this case, the probability of small numbers of correct ligands binding from cell-cell contact being mistaken for a coincidence is large. The FNR decreases as the contact width is increased, leading more receptors to come into the contact region. However, when the contact width 2σ becomes very large and approaches the cell perimeter L, the false-negative rate starts to rise again, particularly when $\alpha = 1$. This $2\sigma \to L$ case could happen if the cell is nearly engulfed by another cell, or is in contact with many cells. FNR will increase as the contact region approaches L only in the hard discrimination regime (α approaches 1), when the two ligands are indistinguishable. When 2σ approaches the cell perimeter, $c(x)$ becomes nearly constant $c(x) \approx c_0$—equivalent to a simple change in the background concentration c', if the cell cannot discriminate between the two ligands. The rise in FNR at large σ disappears as the cell becomes increasingly capable of distinguishing between the two ligands, i.e., when $\alpha < 1$.

We then perform a sweep across the unbinding rate ratio α and the correct ligand fraction χ, both of which significantly impact the false-negative rate. We show a phase diagram for when cells might fail to recognize a contact in Fig. 4(b), which shows a sharp transition in FNR as we go from low α, high χ values to high α, low χ values. Notably, the FNR exhibits nonzero values only when α approaches 1 and $\chi \ll 1$, i.e., the cell is more prone to making mistakes in the hard discrimination regime, where the ligands are nearly indistinguishable, and when there is a higher presence of interfering background ligands. These findings align with the previous results in Figs. 3(a), 3(c), and 3(e).

The false-negative rates are influenced by the adjustable threshold parameter λ_c we chose. We can always increase the FPR and decrease the false-negative rate by adjusting the threshold parameter λ_c to a small value, and vice versa. To compare the ability to discriminate independent of the arbitrary λ_c, we use the detection error tradeoff (DET) graph [35,48,49]. First, we can compute the false-positive rate in a similar manner to FNR:

$$\text{FPR} = \langle H(\lambda - \lambda_c) \rangle_-, \tag{11}$$

where $\langle \cdots \rangle_-$ indicates taking the average with no second cell present, i.e., setting the ephrin concentration $c_0 = 0$. We visualize the trade-off between false positives and false negatives by varying the threshold parameter λ_c over the whole range of λ observed in our simulations, and then plotting the FPR found for each λ_c against the FNR, as shown in Figs. 5(a)–5(c). The figures in Figs. 5(a)–5(c) show how FPR varies as a function of FNR. For instance, if we look at Fig. 5(a), the contours indicate the best false-positive rate the cell can achieve given a fixed false-negative rate, or vice versa. For instance, if the cell can tolerate a false-positive rate of 10^{-3}, and the contact width is $\sigma = 20\,\mu\text{m}$, then the cell can achieve a low false-negative rate (down to $\sim 10^{-3}$, red line). However, either decreasing the contact to $\sigma = 10\,\mu\text{m}$, or expanding the contact to $\sigma = 40\,\mu\text{m}$ makes the FNR larger for this FPR—the cell is more likely to overlook the contact.

In Fig. 5(b), we see that the ability of the cell to reliably distinguish the presence of contact gets progressively worse
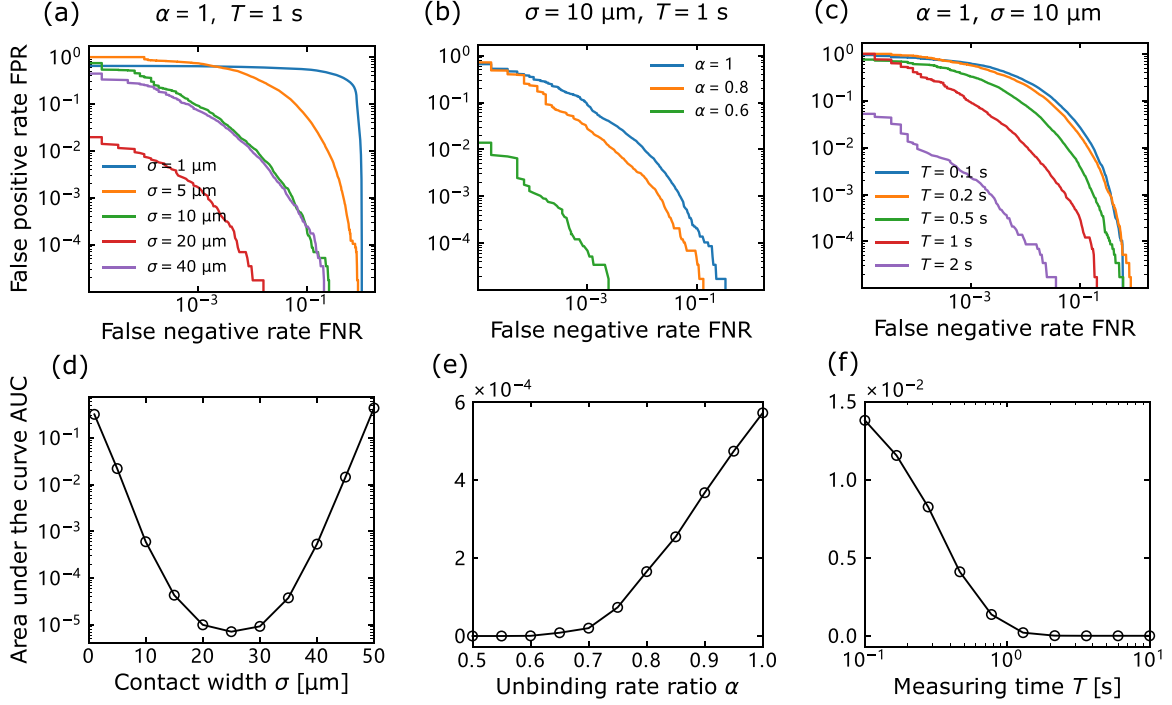
FIG. 5. Detection error tradeoff (DET) graphs and the associated area under the curve (AUC) showing how the cell's ability of making accurate decisions varies with the contact width σ, unbinding rate ratio α, and measuring time T. Panels (a) and (d) show that the optimal contact width is at some intermediate value. Panels (b) and (e) show that the cell is more likely to make incorrect decisions when it is difficult to distinguish the two ligands. Panels (c) and (f) show that the cell can more reliably detect the presence or absence of another cell with a longer measuring time T. Correct ligand fraction is set to $\chi = 0.1$ in all figures.

as $\alpha \to 1$—the FPR-FNR tradeoff curve gets further from the origin. Increasing the cell's measurement time T improves the accuracy with which the cell can sense cell-cell contact [Fig. 5(c)].

We can summarize the efficacy of the cell's ability to distinguish the presence of a contacting cell by computing the area under the curve (AUC) of the tradeoff graphs in Figs. 5(a)–5(c). A smaller AUC indicates that the cell exhibits higher accuracy in distinguishing between the two hypotheses H_0 and H_1—roughly, that the cell achieves a lower false-negative rate for a given false-positive rate. We note that tradeoffs of this sort are also often plotted on a receiver operating characteristic (ROC) curve, which plots true positive rate against false-positive rate. In our convention, a smaller area under the curve is more predictive, instead of larger; our AUC is $1 - \text{AUC}_{\text{ROC}}$ [48,49].

Figure 5(d) shows the AUC value as a function of contact width, which initially decreases and then increases as 2σ gets closer to L. This confirms the results of Fig. 4(a): The cell can better sense the existence or absence of the other cell when the contact width is at some intermediate value. Figs. 5(b) and 5(e) show how the DET curve and the associated AUC change with the unbinding rate ratio α. The area under the curve increases as $\alpha \to 1$, which means the cell is more likely to make incorrect decisions in the hard discrimination regime. We also observe that the AUC drops to approximately 0 in the easy discrimination regime when $\alpha \approx 0.5$, which is consistent with where the transition happens in the phase diagram of FNR [Fig. 4(b)]. As we saw above, accuracy in sensing also

depends on the fraction of ligands that are spurious. If we decrease the correct ligand fraction to $\chi = 0.01$ (Fig. 8), then the overall AUC can become much larger, reaching AUC ~ 0.4, i.e., it becomes much harder to make correct decisions. However, even with this extremely small correct ligand fraction χ, the cell can detect contact very accurately once the unbinding rate ratio reaches $\alpha \sim 0.25$ (Fig. 8). So far, we have kept the amount of time that the cell uses to make a decision constant. Increasing the measurement time T monotonically brings the DET curve toward the origin, decreasing the AUC [Figs. 5(c) and 5(f)]. Cells can more reliably detect the presence or absence of another cell by extending the measuring time T, as we would expect. AUC are mostly large when the measuring time is on the scale of seconds or less. This is a relatively short time compared to the time over which CIL takes place, which can be tens of minutes [7].

IV. DISCUSSION

Our multiligand model highlights the significant impact of various factors on the accuracy of sensing cell-cell contact, including the receptor-ligand affinity, the fraction of correct ligands, and the contact width. It becomes more challenging to sense the cell-cell contact when it is difficult for the cell to distinguish ephrins from spurious ligands, or when there are more interfering spurious ligands present. Moreover, our model reveals that as the contact width σ expands, it becomes easier for the cell to estimate the concentration and the fraction of correct ligands. However, the error in contact

position x_0 estimation remains constant while the contact width increases. We argue that this phenomenon occurs because the cell can only obtain spatial information from the two boundaries where the concentration changes. Furthermore, we investigate whether the cell can successfully detect the presence of another cell through a small cell-cell contact. Our finding indicates that the cell is more likely to make incorrect decisions when the contact width is either too small or too large. In addition, our results also suggest that the cell is more probable to make mistakes if it is difficult for the cell to distinguish ephrins from spurious ligands ($\alpha \to 1$). The cell can extend the time it takes to make a decision to improve decision-making robustness.

A core element of our model is the profile $c(x)$ in Eq. (1), which represents a cell in contact with another cell with a clear contact region of fixed size and a smoothed rectangular form. This neglects many potential complications—e.g., that the size of cell-cell contacts may evolve over time during CIL (see, e.g., the movies of Ref. [5]), how ligand-receptor binding could drive changes in the size and spacing of the cell-cell contact [50,51], and how cadherin proteins could facilitate the formation of contacting patches by zipping (or unzipping) regions of the contacting membrane, thereby leading the coalescence of a patch of defined size [52]. To check the robustness of some of our assumptions, we also explore the case where the ephrin concentration takes the form of a Gaussian function $c(x) = c_0 \exp(-(x-x_0)^2/2\sigma^2)$. This would arise if we assume that our effective binding rate $kc(x)$ is suppressed when there is a gap large enough to require ephrin to undergo a strain to bind. If there is a gap $h(x)$ between the two cells, then we would expect $kc \sim e^{-\frac{1}{2}\kappa h^2/k_B T}$ with κ an effective spring stiffness [53]. If $h(x)$ increases linearly away from the point of closest approach x_0, then we would get a Gaussian profile. Many of our core results are quite similar between the Gaussian and the smooth step profile for $c(x)$, see Figs. 10(a)–10(e) in Appendix D. However, in the scenario of a Gaussian concentration, we find that the error in contact position estimation SD($\hat{x}_0$) increases with increasing contact width [Fig. 10(f) in Appendix D]. Unlike the case of rectangular profiles, the cell now can obtain spatial information from receptors away from the contact edges, as $\partial c/\partial x_0 = c(x)(x-x_0)/\sigma^2$ is nonzero throughout the contact and depends systematically on σ—there is more spatial information at any given point as the width σ becomes smaller. Therefore, the error of estimating the contact position x_0 does not remain constant as σ expands; instead, it increases with increasing σ. Moreover, the concentration profile $c(x)$ may take on a more intricate form, such as comprising several small patches. Our model currently implicitly assumes a single patch. If a cell expecting a single patch had multiple patches in contact with it, then it would do its best to interpret this—either lumping together multiple patches or treating one patch as the "real" one and assuming the other was a coincidence. In either case, we expect similar factors would control the cell's ability to detect the presence of a second cell, but the localization error we study here might not be as directly applicable.

Can a cell reasonably reach the limits we suggest here? Reference [35] proposed a network for sensing the concentration of correct ligands so that biological systems can approach the physical limits given by the Cramér-Rao bound. Other papers have also suggested signal transduction networks to interpret sensing from ligands [36,54]. However, the current state of the art in this field only treats concentration sensing, and its extension to our spatially dependent model for position sensing, let alone decision-making, is not straightforward. We know that experimentally, CIL occurs through the regulation of Rho GTPase polarity [13]. We speculate that modeling patterning occurring via Rho GTPases, regulated by the measured local concentrations as proposed by Ref. [35], may serve as a sensible first model for spatial decision-making. However, the details of this model would be much more complex than the approach we take here.

The estimation errors shown in our model (Fig. 3) are generally small compared to their relevant scales, suggesting cells may have fairly precise knowledge of the cell-cell contact location and ephrin levels. This may reflect the large number of binding-unbinding events. With the total number of receptors $N = 10^4$ and typical receptor correlation times of $\sim$s (Appendix A), and the measuring time $T = 10$ s, there are $\sim 10^5$ binding-unbinding events during the time the cell makes its CIL measurement. Given our results, it is likely that, unless one of our assumptions is incorrect, or that the fraction of spurious ligands is extremely high (e.g., $\chi < 0.1$), stochastic ligand-receptor noise is not a crucial factor in CIL decisions. Where could our assumptions be incomplete? One possibility is that we have neglected details of the Eph receptor interactions. Experiments have found that receptor clustering and other biophysical complexities [16,55–57] can influence the functioning of receptors in binding kinetics in Eph receptors and other RTKs. For instance, EGF receptors dimerize while binding to their ligands EGF [55]. Moreover, for A549 lung cancer cells, it has been observed that the vast majority of EphA2 molecules exist in clusters [56]. As a result, receptor clustering could significantly reduce the number of distinct receptor locations in our model, thereby substantially increasing the errors when estimating the cell-cell contact. Additionally, the precise values of the binding constant k and unbinding rate r for ligands and receptors tethered to cell membranes, such as Eph-ephrin, are not entirely clear [53,58–60]. These parameters can also significantly influence the number of binding-unbinding events M that occur during the measuring time T. Sources of noise downstream of the initial ephrin binding, such as stochasticity in the polarization of the cell's Rho GTPases [8,61–63], might lead to significant additional noise in responses to CIL. Incorporating these factors into our model could provide a more comprehensive understanding of cell-cell interaction and response mechanisms.

APPENDIX A: CORRELATION TIME OF BINDING KINETICS

In our analytical model, we assume that the number of binding-unbinding events for each receptor is large. This requires that the time of the observation T is large in comparison to the timescales relevant for binding and unbinding. In this Appendix, we determine these timescales. We show that the probability of occupation of a single receptor relaxes to its steady state with two separate timescales $\tau_{\text{corr}}^{\pm}$, where the correlation time depends on the kinetics of binding (rates k, r, and r') and the effective concentrations c and c'. With the presence of the incorrect ligands, a receptor can be in an unbound state, a correct bound state, or an incorrect bound state. Thus, we can write the corresponding three-state master equations:

$$\frac{dp_{\text{correct}}}{dt} = (1 - p_{\text{correct}} - p_{\text{incorrect}})kc - p_{\text{correct}}r,$$

$$\frac{dp_{\text{incorrect}}}{dt} = (1 - p_{\text{correct}} - p_{\text{incorrect}})kc' - p_{\text{incorrect}}r',$$

where $1 - p_{\text{correct}}(t) - p_{\text{incorrect}}(t)$ is probability that the receptor is unbound. The steady-state solutions can be easily obtained by setting the time derivatives to zero:

$$p_{\text{correct}}^{\text{SS}} = \frac{kcr^{-1}}{1 + kcr^{-1} + kc'r'^{-1}}, \tag{A1}$$

$$p_{\text{incorrect}}^{\text{SS}} = \frac{kc'r'^{-1}}{1 + kcr^{-1} + kc'r'^{-1}}. \tag{A2}$$

If there is only one species of ligands, then the occupancy (aka fractional saturation) is given by $p = c/(c + K_D) = kcr^{-1}/(1 + kcr^{-1})$, where $K_D = r/k$ is the dissociation constant [32]. With the presence of the second species of ligands, the probability of being bound at all is [35]:

$$p = p_{\text{correct}}^{\text{SS}} + p_{\text{incorrect}}^{\text{SS}} = \frac{kcr^{-1} + kc'r'^{-1}}{1 + kcr^{-1} + kc'r'^{-1}}. \tag{A3}$$

However, to calculate the receptor correlation timescale, we have to solve the master equations over time. Because the final solutions should converge to the steady state, we assume the complete solutions are in the form of

$$p_{\text{correct}}(t) = Ae^{-t/\tau_1} + Be^{-t/\tau_2} + p_{\text{incorrect}}^{\text{SS}},$$

$$p_{\text{incorrect}}(t) = Ce^{-t/\tau_1} + De^{-t/\tau_2} + p_{\text{incorrect}}^{\text{SS}},$$

where $p_{\text{correct}} \to p_{\text{correct}}^{\text{SS}}$ and $p_{\text{incorrect}} \to p_{\text{incorrect}}^{\text{SS}}$ when $t \to \infty$. Here we are primarily interested in the two correlation times τ_1 and τ_2—which should not depend on the particular initial conditions. We then assume the receptor is bound to a correct ligand at $t = 0$, so the initial conditions are $p_{\text{correct}}(0) = 1$ and $p_{\text{incorrect}}(0) = 0$, i.e., $A + B = 1 - p_{\text{correct}}^{\text{SS}}$ and $C + D = -p_{\text{incorrect}}^{\text{SS}}$. Substituting the solutions back to the

master equations, we can find the four coefficients:

$$\begin{aligned} A &= -\frac{kc}{2\sqrt{\Delta}}p_{\text{incorrect}}^{\text{SS}} + \frac{1}{2}\left(1 - p_{\text{correct}}^{\text{SS}}\right) \\ &\quad + \frac{(kc + r) - (kc' + r')}{2\sqrt{\Delta}}\left(1 - p_{\text{correct}}^{\text{SS}}\right), \end{aligned}$$

$$\begin{aligned} B &= \frac{kc}{2\sqrt{\Delta}}p_{\text{incorrect}}^{\text{SS}} + \frac{1}{2}\left(1 - p_{\text{correct}}^{\text{SS}}\right) \\ &\quad - \frac{(kc + r) - (kc' + r')}{2\sqrt{\Delta}}\left(1 - p_{\text{correct}}^{\text{SS}}\right), \end{aligned}$$

$$C = \frac{1}{\tau_2}\frac{p_{\text{incorrect}}^{\text{SS}}}{\sqrt{\Delta}}, \quad D = -\frac{1}{\tau_1}\frac{p_{\text{incorrect}}^{\text{SS}}}{\sqrt{\Delta}},$$

and the two characteristic times:

$$\tau_1 = \frac{2}{kc + r + kc' + r' + \sqrt{\Delta}}, \tag{A4}$$

$$\tau_2 = \frac{2}{kc + r + kc' + r' - \sqrt{\Delta}}, \tag{A5}$$

where the discriminant-like term $\Delta = [k(c + c') + r - r']^2 - 4kc'(r - r')$ is larger than 0 under our assumption $r' \geqslant r$. We can also rewrite Δ in a more symmetric form:

$$\Delta = (kc + r + kc' + r')^2 - 4rr'(1 + kcr^{-1} + kc'r'^{-1}),$$

which ensures $\tau_1, \tau_2 > 0$. Moreover, when there is only one species of ligands, i.e., the correct ligand concentration $c = 0$, the two timescales degenerate to $\tau_2 = 1/r$ and $\tau_1 = 1/(kc' + r')$, which is exactly the correlation timescale for one species of ligands [65]. In brief, we can write the characteristic correlation times τ_1 and τ_2 as

$$\tau_{\text{corr}}^{\pm} = \frac{2}{kc + r + kc' + r' \pm \sqrt{\Delta}}, \tag{A6}$$

and expect that for our semi-analytical calculation of the Fisher information to be valid, we must be in the limit $T \gg \tau_{\text{corr}}^{\pm}$ for all receptors.

APPENDIX B: SEMIANALYTICAL CALCULATION OF FISHER INFORMATION

If the correct ligand concentration takes the form of $c(x) = c_0 g(x)$, and there are N independent receptors evenly distributed in the cell perimeter L, then the nth receptor will have a different value of the ephrin concentration $c(x_n)$ and a different number of binding-unbinding events M_n during a measuring time T. Then for the nth receptor, the probability to have a time record $\{\tau_b^{(n)}, \tau_u^{(n)}\}$ is

$$P_n = \prod_{i=1}^{M_n} P\left(\tau_{u,i}^{(n)}\right) \cdot P\left(\tau_{b,i}^{(n)}\right), \tag{B1}$$

where the unbinding and binding times are given by

$$P(\tau_u) = k(c + c')e^{-k(c+c')\tau_u}, \tag{B2}$$

$$P(\tau_b) = \frac{c}{c + c'}re^{-r\tau_b} + \frac{c'}{c + c'}r'e^{-r'\tau_b}, \tag{B3}$$

Therefore, given the three parameters $\boldsymbol{\theta} = (c_0, c', x_0)$ in our model, the total probability of all N receptors is $P(\{\tau\}|\boldsymbol{\theta}) = \prod_{n=1}^{N} P_n$, then the likelihood $\mathcal{L}(\boldsymbol{\theta}|\{\tau\}) = P(\{\tau\}|\boldsymbol{\theta})$ is given by

$$
\begin{aligned}
\mathcal{L} &= \prod_{n=1}^{N} \prod_{i=1}^{M_n} \left[k(c+c')e^{-k(c+c')\tau_{u,i}^{(n)}} \times \left(\frac{c}{c+c'} r e^{-r\tau_{b,i}^{(n)}} + \frac{c'}{c+c'} r' e^{-r'\tau_{b,i}^{(n)}} \right) \right] \\
&= \prod_{n=1}^{N} e^{-k(c+c')T_u^{(n)}} \prod_{i=1}^{M_n} k\left(rce^{-r\tau_{b,i}^{(n)}} + r'c'e^{-r'\tau_{b,i}^{(n)}} \right) \\
&= \prod_{n=1}^{N} e^{-kc_t[\chi g(x_n)+1-\chi]T_u^{(n)}} \prod_{i=1}^{M_n} kr'c_t e^{-r'\tau_{b,i}^{(n)}} \left[1 - \chi + \alpha\chi g(x_n) e^{(1-\alpha)r'\tau_{b,i}^{(n)}} \right] \\
&= \prod_{n=1}^{N} e^{-kc_t[\chi g(x_n)+1-\chi]T_u^{(n)} - r'T_b^{(n)}} \prod_{i=1}^{M_n} kr'c_t \left[1 - \chi + \alpha\chi g(x_n) e^{(1-\alpha)r'\tau_{b,i}^{(n)}} \right],
\end{aligned}
$$

where in the second and last steps we have defined the total amount of unbound and bound time for the nth receptor as $T_u^{(n)} = \sum_{i=1}^{M_n} \tau_{u,i}^{(n)}$ and $T_b^{(n)} = \sum_{i=1}^{M_n} \tau_{b,i}^{(n)}$, respectively. And in the penultimate step, we introduce a transformation from (c_0, c') to the total concentration $c_t = c_0 + c'$ and correct ligand fraction $\chi = c_0/c_t$, i.e., we substitute $c' = c_t(1 - \chi)$, $r = \alpha r'$, and $c(x) = c_t \chi g(x)$. After taking a logarithm, $\ln \mathcal{L}$ can be written as a sum of five independent contributions:

$$
\ln \mathcal{L}_0 = \sum_{n=1}^{N} \sum_{i=1}^{M_n} \ln kr' = \sum_{i=1}^{N} M_n \ln kr',
$$

$$
\ln \mathcal{L}_1 = -r' \sum_{n=1}^{N} \sum_{i=1}^{M_n} \tau_{b,i}^{(n)} = -r' \sum_{n=1}^{N} T_b^{(n)},
$$

$$
\ln \mathcal{L}_2 = \sum_{n=1}^{N} \sum_{i=1}^{M_n} \ln c_t = \sum_{n=1}^{N} M_n \ln c_t,
$$

$$
\ln \mathcal{L}_3 = \sum_{n=1}^{N} \sum_{i=1}^{M_n} \ln \left[1 - \chi + \alpha\chi g(na) e^{(1-\alpha)r'\tau_{b,i}^{(n)}} \right],
$$

$$
\ln \mathcal{L}_4 = -\sum_{n=1}^{N} kc_t[1 - \chi + \chi g(na)]T_u^{(n)}, \tag{B4}
$$

where a is the receptor spacing and $x_n = na$ gives the position of the nth receptor. Note that $\mathcal{L}_0$ and $\mathcal{L}_1$ are trivial terms since they do not depend on any parameter we are concerned with, and thus we only need to include the other three terms in the following computation. Calculating the derivatives $\partial^2 \ln \mathcal{L}/\partial\theta_\alpha \partial\theta_\beta$ is straightforward but not particularly informative. We include these results in a Mathematica notebook [66]. These derivatives depend on the bound and unbound times through terms like $T_u^{(n)}$, M_n, but also in more complex forms $f(\tau_{b,i}^{(n)})$ like $e^{(1-\alpha)r'\tau_{b,i}^{(n)}}$.

The complex part during the derivation of the Fisher information matrix $I(\boldsymbol{\theta})_{\mu\nu} = -\langle \partial^2 \ln \mathcal{L}/\partial\theta_\mu \partial\theta_\nu \rangle$ arises from computing the average $\langle \cdots \rangle$. To begin, we can calculate the average of the bound and unbound times for the nth receptor:

$$
\left\langle \tau_u^{(n)} \right\rangle = \int_0^\infty \tau P\left(\tau_u^{(n)} = \tau\right) d\tau = \frac{1}{k(c_n + c')},
$$

$$
\left\langle \tau_b^{(n)} \right\rangle = \int_0^\infty \tau P\left(\tau_b^{(n)} = \tau\right) d\tau = \frac{c_n}{r(c_n + c')} + \frac{c'}{r'(c_n + c')}.
$$

We then assume the average of the total binding-unbinding events M_n during the measuring time T can be approximated as

$$
\langle M_n \rangle = \frac{T}{\left\langle \tau_u^{(n)} \right\rangle + \left\langle \tau_b^{(n)} \right\rangle}, \tag{B5}
$$

which should only be valid when $T \gg \tau_{\text{corr}}^{\pm}$ (Appendix A). And the average of the total unbound time for the nth receptor is

$$
\left\langle T_u^{(n)} \right\rangle = (1 - p_n)T, \tag{B6}
$$

where p_n is the occupancy given by Eq. (A3). We can readily verify $\langle M_n \rangle = \langle T_u^{(n)} \rangle / \langle \tau_u^{(n)} \rangle$, which is expected to be true if $T \gg \tau_{\text{corr}}^{\pm}$. When computing a sum over all binding-unbinding events, e.g., $\langle \sum_{i=1}^{M_n} f(\tau_{b,i}^{(n)}) \rangle$, we assume that the average of each term is equal, and as a result, it is equivalent to multiply the average of a single term by $\langle M_n \rangle$. This essentially treats the value of the number of binding-unbinding events M_n as fixed, and requires again that we have $T \gg \tau_{\text{corr}}^{\pm}$. To sum up, to go from the analytical derivatives of Eq. (B4) to computing the averages required for the Fisher information matrix, we apply the rules:

$$
\left\langle \sum_{n=1}^{N} M_n \right\rangle \rightarrow \sum_{n=1}^{N} \langle M_n \rangle,
$$

$$
\left\langle \sum_{n=1}^{N} T_u^{(n)} \right\rangle \rightarrow \sum_{n=1}^{N} \langle T_u^{(n)} \rangle,
$$

$$
\left\langle \sum_{n=1}^{N} \sum_{i=1}^{M_n} f\left(\tau_{b,i}^{(n)}\right) \right\rangle \rightarrow \sum_{n=1}^{N} \langle M_n \rangle \int_0^\infty d\tau \, P\left(\tau_b^n = \tau\right) f(\tau),
$$

$$
\tag{B7}
$$

where the last term is computed by numerical integration using Gaussian quadrature (scipy.integrate.quad method provided by SciPy [67]).

We can see from Eq. (B4) that all terms in the Fisher information matrix should contain $\langle M_n \rangle$ or $\langle T_u \rangle$—so we expect that the Fisher information time is thus proportional to the total measuring time T, at least in the limit of $T \gg \tau_{\text{corr}}$ where our analytic theory is appropriate. After obtaining all the terms in the matrix numerically, we can compute its inverse and take the diagonal elements as the lower bound of the variances, namely the Cramér-Rao bound in Eq. (4).

APPENDIX C: DETAILS OF STOCHASTIC SIMULATIONS

1. Simulation methods

In the Monte Carlo simulations, we generate a series of random bound and unbound times for each receptor from the probability distribution in Eqs. (2) and (3). To do this, we note that Eq. (3) is essentially a composite of two exponential distributions: We draw samples from $re^{-r\tau_b}$ with a probability of $c/(c + c')$ and from $r'e^{-r'\tau_b}$ with a probability of $c'/(c + c')$, which is equivalent to running a kinetic Monte Carlo simulation to generate these time series. Subsequently, we compute the log-likelihood function $\ln \mathcal{L}(\boldsymbol{\theta}'|\tau)$ using Eq. (B4) and numerically maximize it to find the optimal parameters $(\hat{c}_t, \hat{\chi}, \hat{x}_0)$. When doing the optimization, we use the modified Powell algorithm [69] (provided by SciPy [67]) to minimize $-\ln \mathcal{L}$ within the bounds $c_t \in [0, 10^8]\,\mu\text{m}^{-2}$, $\chi \in [0, 1]$, $x_0 \in [0, 100]\,\mu\text{m}$, and set the relative error in solutions acceptable for convergence to 10^{-7}. This gives us the maximum-likelihood estimators $(\hat{c}_t, \hat{\chi}, \hat{x}_0)$ for this individual trajectory. We then repeat this process many times for new trajectories ($\sim 10^4$), thus allowing us to compute distributions of the estimators. We then plot the standard deviations of these estimator distributions in, e.g., Fig. 3.

We have not been able to compute the false-negative rates in Eq. (10) and false-positive rates in Eq. (11) analytically. However, we can still calculate the FNR and FPR through Monte Carlo simulations. To compute FNR, we assume the other cell is present, i.e., $c_0 \neq 0$, and then generate a time series to compute the likelihood ratio by computing the maximum log-likelihood [Eq. (B4)] numerically under both the assumption of the other cell being present and it being absent. We then repeat this process many times to compute the FNR for a fixed value of λ. For the FPR, we do simulations by first assuming the other cell is absent (setting $c_0 = 0$), but keep all other parameters the same. We perform approximately $\sim 10^5$ simulations to do averages for each data point shown in Fig. 4, which allows us to obtain robust and reliable results for the analysis.

2. Parameters

The default parameter values in our model are shown in Table I. The typical radius of the *Drosophila* haemocytes cell is in the order of $10\,\mu\text{m}$ [14], which makes the cell perimeter roughly $10^2\,\mu\text{m}$. The concentration of EphA2 receptors on the cell surface is reported to be around 100–$1000/\mu\text{m}^2$ for CHO cells [70] and about $600/\mu\text{m}^2$ for A549 lung cancer cells [56]. As we're modeling the perimeter of the cell as a line, we set the contact center $x_0 = L/2$ to avoid boundary effects. Additionally, we choose a receptor spacing of $a = 0.01\,\mu\text{m}$ to

TABLE I. Table of simulation parameters.[a]

Parameter	Description	Dimension	Value
c_0	Ephrin concentration	L^{-2}	$10^2\,\mu\text{m}^{-2}$ [68]
χ	Fraction of correct ligands c_0/c_t	1	Varying
k	Eph-ephrin binding constant	L^2/T	$10^{-2}\,\mu\text{m}^2/\text{s}$
r	Eph-ephrin unbinding rate	T^{-1}	$1\,\text{s}^{-1}$
α	Ratio of unbinding rates r/r'	1	Varying
x_0	Contact center position	L	$50\,\mu\text{m}$
σ	Contact width	L	Varying
ω	Boundary steepness	L	$1\,\mu\text{m}$
T	Measuring time	T	$1\,\text{s}$
L	Cell perimeter	L	$100\,\mu\text{m}$
a	Eph receptor spacing	L	$0.01\,\mu\text{m}$
N	Number of receptors	1	$10\,000$

[a]These parameters are used throughout the paper; any deviation from them is explicitly noted.

ensure that there are approximately 100 receptors per square micron of the surface area on the cell, assuming a height in the z-direction around the order of $1\,\mu\text{m}$. This neglects many potential complications like the possible change in cell shape with time. In our model, we only considered CIL happening between two cells with a fixed height. Increasing the height of the contacting region is equivalent to increasing the number of receptors in the contact region, which would increase the sensing accuracy in general.

3. Consistency between stochastic simulation and semi-analytic theory

The fundamental assumption we utilize when deriving the Fisher information matrix in Appendix B is that $T \gg \tau_{\text{corr}}^{\pm}$ for all receptors on the cell. In this limit, it is reasonable to compare our theory, which has a fixed number of binding and unbinding events for each receptor, to the stochastic simulations, where we fix the measuring time T. This assumption leads to the theoretical Cramér-Rao bound scaling as $\sim 1/T$ (Appendix B). When we compare simulation and theory, we should check this dependence, to ensure we are exploring times $T \gg \tau_{\text{corr}}^{\pm}$. We find that the variances obtained from simulations do not depend strongly on T when $T \ll 1$ second and the discrepancy between simulation and theory grows with decreasing measuring time (Fig. 7). This makes sense: In the limit $T \ll \tau_{\text{corr}}$, the cell is effectively making a "snapshot" measurement, using the receptor state at one instant to estimate parameters, and the error in measurement will not depend strongly on the measurement time. The correlation timescale of about a second is consistent with our analysis in Appendix A. With the chosen binding rate kc and unbinding rate r in the order of $1\,\text{s}^{-1}$ (Table I), the correlation times τ_{corr} given by Eq. (A6) are $\tau_1 \sim 10^{-1}\,\text{s}$ and $\tau_2 \sim 1\,\text{s}$. Therefore, to effectively compare the simulation results with the theoretical Cramér-Rao bound in Eq. (4), we need to increase the measuring time to $T \sim 10\,\text{s}$ in Figs. 3 and 10 (see Appendix D).
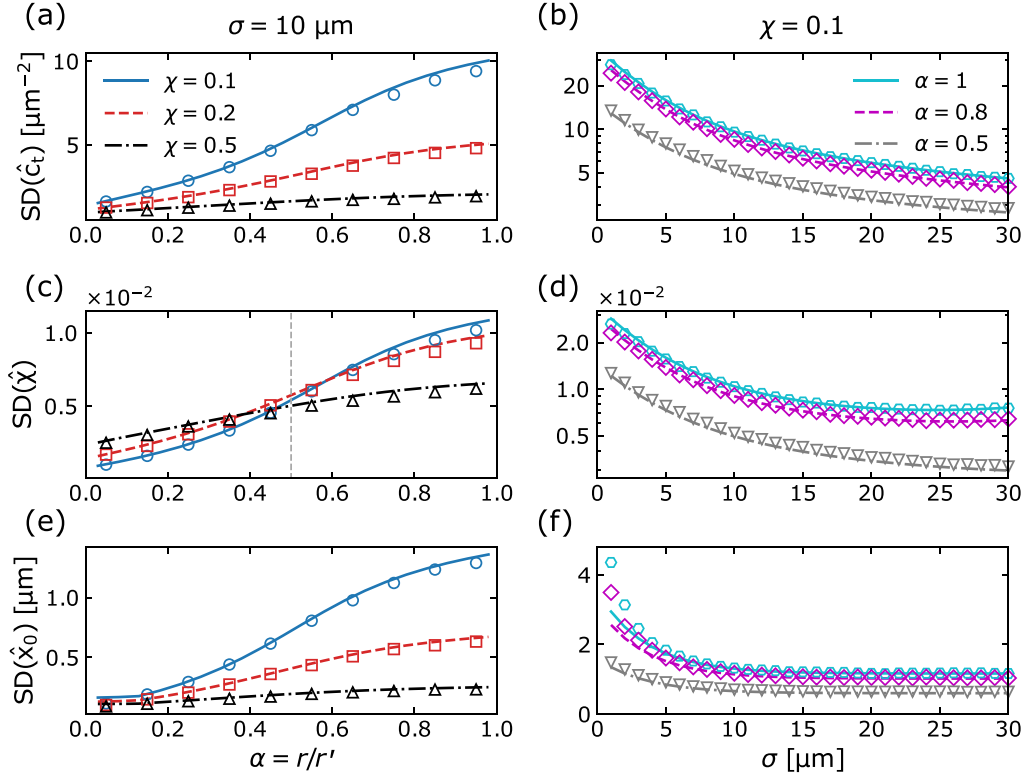
FIG. 6. Estimating errors of MLE when $\omega = 10\,\mu m$. We can see that all SDs have a similar pattern to Fig. 3 where $\omega = 1\,\mu m$. Besides, we note that the $SD(\hat{x}_0)$ also starts to rise as σ decreases to $\sim\omega$. In all figures, the lines (solid, dashed, and dash-dot) represent theoretical results obtained from Eq. (4), while the points are obtained from Monte Carlo simulations (the standard deviations are computed from 10^4 simulations for each point); measuring time is $T = 10\,s$.
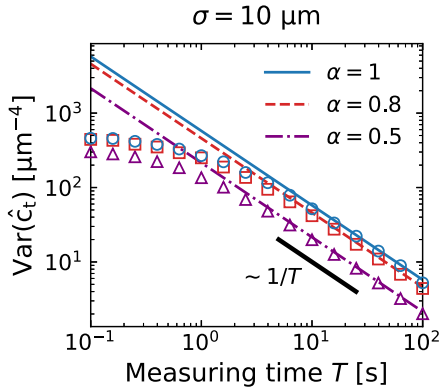
APPENDIX D: ADDITIONAL FIGURES

FIG. 7. Variances of MLE decrease with increasing measuring time T. The variances from the Cramér-Rao bound scales as $\sim 1/T$ in our derivation (Appendix B). Though the simulation results show a slower decrease compared to the theoretical predictions at small T, they still converge as the measuring time $T \gg \tau_{corr}$. Lines represent theoretical results obtained from Eq. (4), while the points are obtained from Monte Carlo simulations; correct ligand fraction is $\chi = 0.1$.
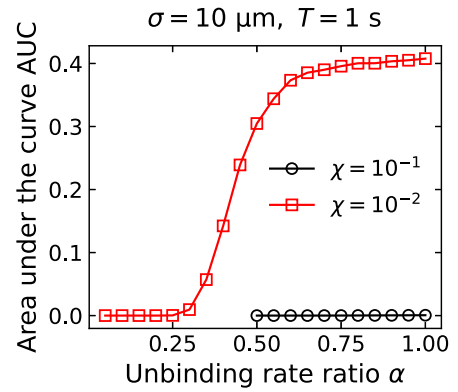

FIG. 8. Area under the DET curve changing with unbinding rate ratio α when the correct ligand fraction $\chi = 10^{-2}$. Black points are exactly the results in Fig. 5(e) when $\chi = 10^{-1}$.
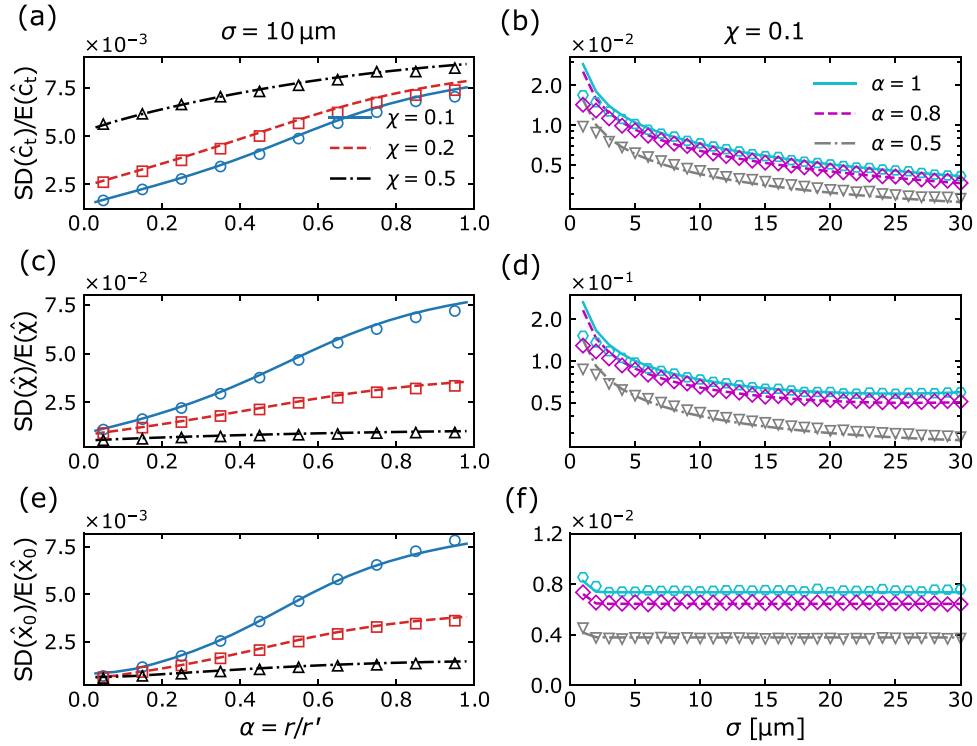
FIG. 9. Relative errors (SD/average) for Fig. 3 in the main text, which are generally small everywhere. In all figures, the lines (solid, dashed, and dash-dot) represent theoretical results obtained from Eq. (4), while the points are obtained from Monte Carlo simulations (the standard deviations are computed from 10^4 simulations for each point); measuring time is $T = 10\,\text{s}$.
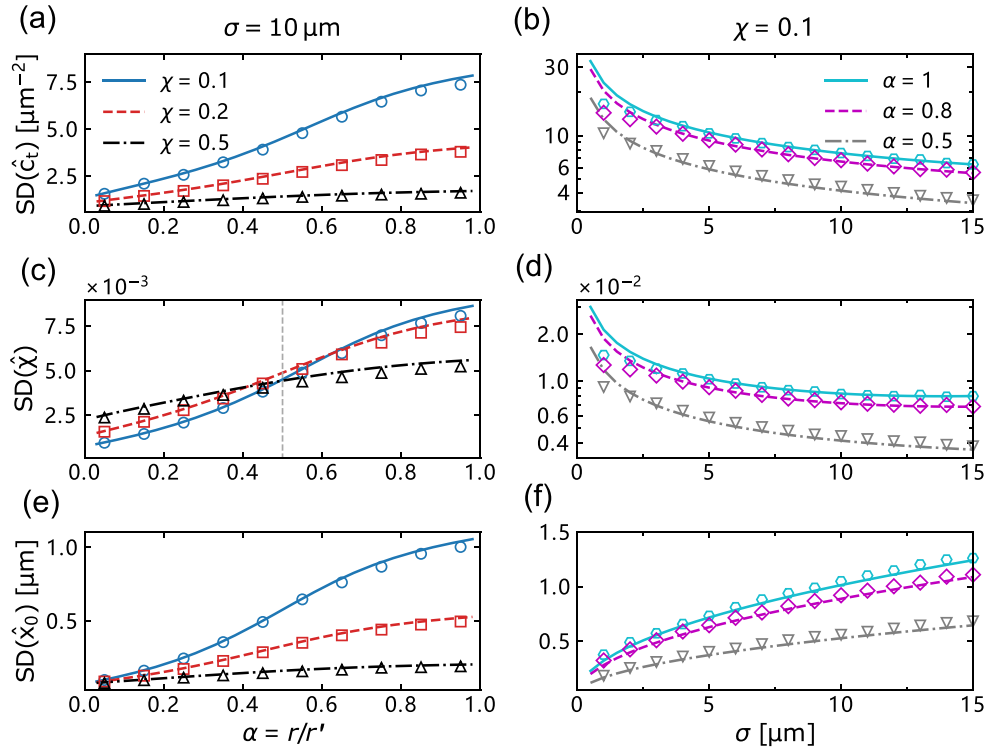

FIG. 10. Estimating errors of MLE when the correct ligand concentration $c(x)$ takes the form of a Gaussian function $c(x) = c_0 \exp(-(x - x_0)^2/2\sigma^2)$ rather than the rectangular function given in Eq. (1). Panels (a), (b) show the standard deviation (SD) of the total concentration $\hat{c}_t$, panels (c), (d) show the SD of correct ligand fraction $\hat{\chi}$, and panels (e), (f) display the SD of contact center position $\hat{x}_0$, where the estimating error of x_0 increases as the contact width σ expands. In all figures, the lines (solid, dashed, and dash-dot) represent theoretical results obtained from Eq. (4), while the points are obtained from Monte Carlo simulations (the standard deviations are computed from 10^4 simulations for each point); measuring time is $T = 10\,\text{s}$.

[1] M. Abercrombie, In Vitro **6**, 128 (1970).

[2] R. Mayor and C. Carmona-Fontaine, Trends Cell Biol. **20**, 319 (2010).

[3] M. Abercrombie, Nature (London) **281**, 259 (1979).

[4] P. Zadeh and B. A. Camley, Phys. Rev. E **106**, 054413 (2022).

[5] J. Singh, A. Pagulayan, B. A. Camley, and A. S. Nain, Proc. Natl. Acad. Sci. USA **118**, e2011815118 (2021).

[6] E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, and R. Mayor, Biol. Open **2**, 901 (2013).

[7] R. A. Desai, S. B. Gopal, S. Chen, and C. S. Chen, J. R. Soc. Interface. **10**, 20130717 (2013).

[8] D. A. Kulawiak, B. A. Camley, and W.-J. Rappel, PLoS Comput. Biol. **12**, e1005239 (2016).

[9] D. B. Brückner, N. Arlt, A. Fink, P. Ronceray, J. O. Rädler, and C. P. Broedersz, Proc. Natl. Acad. Sci. USA **118**, e2016602118 (2021).

[10] D. Li and Y.-l. Wang, Proc. Natl. Acad. Sci. USA **115**, 10678 (2018).

[11] J. Zimmermann, B. A. Camley, W.-J. Rappel, and H. Levine, Proc. Natl. Acad. Sci. USA **113**, 2660 (2016).

[12] E. Larrañaga, V. Fernández-Majada, S. Ojosnegros, J. Comelles, and E. Martinez, Adv. Mater. Interf. **9**, 2201301 (2022).

[13] B. Lin, T. Yin, Y. I. Wu, T. Inoue, and A. Levchenko, Nat. Commun. **6**, 6619 (2015).

[14] J. R. Davis, C.-Y. Huang, J. Zanet, S. Harrison, E. Rosten, S. Cox, D. Y. Soong, G. A. Dunn, and B. M. Stramer, Development **139**, 4555 (2012).

[15] B. Stramer, S. Moreira, T. Millard, I. Evans, C.-Y. Huang, O. Sabet, M. Milner, G. Dunn, P. Martin, and W. Wood, J. Cell Biol. **189**, 681 (2010).

[16] E. Zapata-Mercado, G. Biener, D. M. McKenzie, W. C. Wimley, E. B. Pasquale, V. Raicu, and K. Hristova, J. Biol. Chem. **298**, 102370 (2022).

[17] E. M. Lisabeth, G. Falivelli, and E. B. Pasquale, Cold Spring Harbor Perspect. Biol. **5**, a009159 (2013).

[18] A. Roycroft and R. Mayor, Cell. Mol. Life Sci. **73**, 1119 (2016).

[19] I. Noordstra, M. D. Hermoso, L. Schimmel, A. Bonfim-Melo, D. Currin-Ross *et al.*, Dev. Cell **58**, P1748 (2023).

[20] R. Klein, Development **139**, 4105 (2012).

[21] N. K. Noren and E. B. Pasquale, Cell. Signal. **16**, 655 (2004).

[22] D. Arvanitis and A. Davy, Genes Dev. **22**, 416 (2008).

[23] E. B. Pasquale, Cell **133**, 38 (2008).

[24] J. Cayuso, Q. Xu, and D. G. Wilkinson, Dev. Biol. **401**, 122 (2015).

[25] V. Singh and I. Nemenman, Phys. Rev. Lett. **124**, 028101 (2020).

[26] M. Carballo-Pacheco, J. Desponds, T. Gavrilchenko, A. Mayer, R. Prizak, G. Reddy, I. Nemenman, and T. Mora, Phys. Rev. E **99**, 022423 (2019).

[27] H. C. Berg and E. M. Purcell, Biophys. J. **20**, 193 (1977).

[28] B. Hu, W. Chen, W.-J. Rappel, and H. Levine, Phys. Rev. E **83**, 021917 (2011).

[29] V. H. Wasnik, P. Lipp, and K. Kruse, Phys. Rev. E **100**, 022401 (2019).

[30] A. Hopkins and B. A. Camley, Phys. Rev. Res. **2**, 043146 (2020).

[31] E. P. Ipiña and B. A. Camley, Phys. Rev. E **105**, 044410 (2022).

[32] B. Hu, W. Chen, W.-J. Rappel, and H. Levine, Phys. Rev. Lett. **105**, 048104 (2010).

[33] I. Nwogbaga, A. H. Kim, and B. A. Camley, Phys. Rev. E **108**, 064411 (2023).

[34] T. Mora and N. S. Wingreen, Phys. Rev. Lett. **104**, 248101 (2010).

[35] T. Mora, Phys. Rev. Lett. **115**, 038102 (2015).

[36] V. Singh and I. Nemenman, PLoS Comput. Biol. **13**, e1005490 (2017).

[37] O. Feinerman, R. N. Germain, and G. Altan-Bonnet, Mol. Immunol. **45**, 619 (2008).

[38] L. A. Chylek, D. A. Holowka, B. A. Baird, and W. S. Hlavacek, Front. Immunol. **5**, 172 (2014).

[39] J. Wykosky, E. Palma, D. Gibo, S. Ringler, C. Turner, and W. Debinski, Oncogene **27**, 7260 (2008).

[40] J. Gong, R. Körner, L. Gaitanos, and R. Klein, J. Cell Biol. **214**, 35 (2016).

[41] M. Minami, T. Koyama, Y. Wakayama, S. Fukuhara, and N. Mochizuki, Mol. Biol. Cell **22**, 3508 (2011).

[42] J.-B. Lalanne and P. François, Proc. Natl. Acad. Sci. USA **112**, 1898 (2015).

[43] K. Karl, M. D. Paul, E. B. Pasquale, and K. Hristova, J. Biol. Chem. **295**, 18494 (2020).

[44] M. Gomez-Soler, M. P. Gehring, B. C. Lechtenberg, E. Zapata-Mercado, A. Ruelos, M. W. Matsumoto, K. Hristova, and E. B. Pasquale, iScience **25**, 103870 (2022).

[45] R. G. Endres and N. S. Wingreen, Phys. Rev. Lett. **103**, 158101 (2009).

[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Hoboken, NJ, 1993).

[47] B. Li and G. J. Babu, *A Graduate Course on Statistical Inference* (Springer, Berlin, 2019).

[48] J. Navratil and D. Klusacek, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, Vol. 4 (IEEE, Piscataway, NJ, 2007), pp. IV–229.

[49] T. Fawcett, Pattern Recogn. Lett. **27**, 861 (2006).

[50] J. Hu, R. Lipowsky, and T. R. Weikl, Proc. Natl. Acad. Sci. USA **110**, 15283 (2013).

[51] K. T. DiNapoli, D. N. Robinson, and P. A. Iglesias, Biophys. J. **120**, 4905 (2021).

[52] C. L. Adams, W. J. Nelson, and S. J. Smith, J. Cell Biol. **135**, 1899 (1996).

[53] J. A. Janeš, C. Monzel, D. Schmidt, R. Merkel, U. Seifert, K. Sengupta, and A.-S. c. v. Smith, Phys. Rev. X **12**, 031030 (2022).

[54] C. C. Govern and P. R. ten Wolde, Phys. Rev. Lett. **109**, 218103 (2012).

[55] J. L. Macdonald and L. J. Pike, Proc. Natl. Acad. Sci. USA **105**, 112 (2008).

[56] D. R. Singh, P. Kanvinde, C. King, E. B. Pasquale, and K. Hristova, Commun. Biol. **1**, 1 (2018).

[57] M. Skoge, Y. Meir, and N. S. Wingreen, Phys. Rev. Lett. **107**, 178101 (2011).

[58] T. Bihr, S. Fenz, E. Sackmann, R. Merkel, U. Seifert, K. Sengupta, and A.-S. Smith, Biophys. J. **107**, L33 (2014).

[59] V. Zarnitsyna and C. Zhu, Phys. Biol. **9**, 045005 (2012).

[60] B. A. Bicknell, P. Dayan, and G. J. Goodhill, Nat. Commun. **6**, 7468 (2015).

[61] G. R. Walther, A. F. Marée, L. Edelstein-Keshet, and V. A. Grieneisen, Bull. Math. Biol. **74**, 2570 (2012).

[62] C. Copos and A. Mogilner, Mol. Biol. Cell **31**, 1637 (2020).

[63] E. P. Ipiña, J. D'Alessandro, B. Ladoux, and B. A. Camley, bioRxiv (2023).

[64] https://zenodo.org/doi/10.5281/zenodo.11094134.

[65] A. Kashyap, W. Wang, and B. A. Camley, Biophys. J. (2024).

[66] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.109.054408 for mathematica notebook for detailed derivations.

[67] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland *et al.*, Nature Methods **17**, 261 (2020).

[68] Q. Xu, W.-C. Lin, R. S. Petit, and J. T. Groves, Biophys. J. **101**, 2731 (2011).

[69] M. J. Powell, Comput. J. **7**, 155 (1964).

[70] D. Wirth, M. D. Paul, E. B. Pasquale, and K. Hristova, ChemSystemsChem **4**, e202200011 (2022).