Supplementary Text for "Physical limits on bacterial navigation in dynamic environments"

by A. M. Hein, D. R. Brumley, F. Carrara, R. Stocker, and Simon A. Levin

Ramp rate and concentration slope estimation

Estimating the ramp rate from a series of molecule absorptions. Here we discuss constraints on the estimation of the ramp rate c_1 and the concentration slope g. Following [1, 2, 3], we approximate a cell as an idealized measuring device: a sphere of radius a that absorbs all molecules that come in contact with its surface.

We begin by recalling relevant results of [3]. The average flux of molecules (molecules $\times$ time^{-1}) arriving at the surface of the sphere at position $\mathbf{x}$, time t is $\langle I(\mathbf{x},t)\rangle = 4\pi DaC(\mathbf{x},t)$ [1, 3] , which is equivalent in our notation to $\langle I(t)\rangle = 4\pi Dac(t)$. Again, the far-field chemoattractant concentration in the medium surrounding the sphere $c(t)$ can be linearized to $c(t) \approx c_0 + c_1(t-t_0)$. The question discussed in [3] is: given a series of absorption times $\{t_i\},\ i = 1, 2, ..., n$ measured during the interval $t_i \in (t_0 - T/2, t_0 + T/2)$, what is the minimum variance in the estimate of c_1 that a cell could possibly achieve? A natural tool for answering this question is the statistical framework known as Maximum Likelihood. Given a set of data (the time series $\{t_i\}$) and a generating model for those data – in this case, that $c(t) = c_0 + c_1(t-t_0)$, and absorptions are Poisson with rate $\langle I(t)\rangle = 4\pi Dac(t)$ – one seeks values of the parameters of the generating model (c_0 and c_1) that maximize the probability, or "likelihood", of the data. Maximum likelihood estimates are optimal in the sense that, as the number of observations becomes large, the variance of these estimators approaches a theoretical minimum variance for any unbiased estimator, which is given by the Cramér-Rao theorem [4]. This lower bound can be used to establish a bound on the accuracy with which cells can measure changes in concentration.

In the context considered in [3] and in our study, molecule absorptions are assumed to be independent Poisson events. Let $t = 0$ be the time at which the pulse appears, and $t = t_0$ be a reference time marking the midpoint of the measurement interval $(t_0 - T/2, t_0 + T/2)$. The lengths of the time intervals between molecule arrivals, $\sigma_i = t_i - t_{i-1}$ [5] obey

$$\mathbb{P}(\sigma_i) = \langle I(t_i)\rangle \exp\left[-\int_{t_{i-1}}^{t_i} \langle I(s)\rangle ds\right], \tag{S1}$$

where σ_1 is defined as $t_1 - (t_0 - T/2)$ and the integral in Equation (S1) is taken from $t_0 - T/2$ (the start of the measurement interval) to t_1 for σ_1. The probability of observing the set $\{t_i\}$ is

$$\mathbb{P}(\{t_i\}) = \mathbb{P}(\{\sigma_i\}) = \prod_{i=1}^{n} \langle I(t_i)\rangle e^{-\int_{t_0-T/2}^{t_0+T/2}\langle I(t)\rangle dt}. \tag{S2}$$

By solving $\partial \log(\mathbb{P}(\{t_i\}))/\partial \hat{c}_0 = 0$ and $\partial \log(\mathbb{P}(\{t_i\}))/\partial \hat{c}_0 = 0$, one can show that the values of $\hat{c}_0$ and $\hat{c}_1$ that maximize Eq. (S2) are:

$$\hat{c_0} = \frac{n}{4\pi DaT} \tag{S3}$$

and

$$\hat{c_1} = \hat{c_0}\frac{\sum_i (t_i - t_0)}{\sum_i (t_i - t_0)^2}, \tag{S4}$$

where n is the number of molecules absorbed during the observation interval. Note that $\langle n \rangle \approx 4\pi Dac_0T$ as the measurement interval becomes long, and as n becomes large, $\sum_i (t_i - t_0) \approx \langle \sum_{t_i} (t_i - t_0) \rangle = 4\pi Da \int_{t_0-T/2}^{t_0+T/2} [c_0 + c_1(t-t_0)](t-t_0)dt = \pi Dac_1T^3/3$ and $\sum_{t_i} (t_i - t_0)^2 \approx \langle \sum_{t_i} (t_i - t_0)^2 \rangle = 4\pi Da \int_{t_0-T/2}^{t_0+T/2} [c_0 + c_1(t-t_0)](t-t_0)^2 dt = \pi Dac_0T^3/3$, indicating that maximum likelihood estimators, $\hat{c_0}$ and $\hat{c_1}$, are asymptotically unbiased, i.e.,

$$\hat{c_0} \to c_0 \qquad \text{and} \qquad \hat{c_1} \to c_1 \qquad \text{for large } n. \tag{S5}$$

We derive Eq. (1) in the Main Text by calculating a lower bound on the variance of the ramp rate estimator, $\hat{c_1}$. The Cramér-Rao theorem states that the variance of $\hat{c_1}$ is bounded by the relation [4]:

$$\text{var}(\hat{c_1}) \geq -\mathbb{E}\left[\frac{\partial^2 \log(\mathbb{P}(\{t_i\}; c_1))}{\partial {c_1}^2}\right]^{-1} = \mathbb{E}\left[\sum_{t_i} \frac{(t_i - t_0)^2}{[c_0 + c_1(t_i - t_0)]^2}\right]^{-1}. \tag{S6}$$

Employing the assumption that $c_0 \gg c_1 T$, and using $\sum_{t_i} [t_i - t_0]^2 \approx \pi Dac_0T^3/3$ as the number of absorptions becomes large implies

$$\text{var}(\hat{c_1}) \gtrsim \frac{c_0^2}{\sum_{t_i} [t_i - t_0]^2} \approx \frac{3c_0}{\pi DaT^3}, \tag{S7}$$

which is the relation given in Eq. (1) of the Main Text.

Estimating the concentration slope in static and dynamic concentration fields. We are concerned with cells that use their estimate of the ramp rate $\hat{c_1}$ to estimate the spatial gradient in chemical concentration, which we refer to as the concentration slope. We also assume that the concentration field C changes over time as a pulse spreads. In this setting, the cell must estimate the concentration slope g along its path using some estimator $\hat{g}$ that can be computed from a series of observed absorption times. The concentration experienced by the cell can still be written $c(t) = c_0 + c_1(t - t_0)$, but now $c_1 = gv + \partial C/\partial t$. If we begin by assuming $\partial C/\partial t = 0$, the maximum likelihood estimator for g follows from the estimator for c_1:

$$\hat{g} = \hat{c_0} \frac{\sum_i (t_i - t_0)}{v \sum_i (t_i - t_0)^2}. \tag{S8}$$

In the limit of many molecule absorptions, $\sum_i (t_i - t_0) \approx \pi Davg T^3/3$ and $\sum_i (t_i - t_0)^2 \approx \pi Dac_0T^3/3$ (using the assumption that $c_0 \gg c_1 T$), which imply that $\hat{g}$ approaches the true concentration slope g as the number of molecule absorptions becomes large (i.e., $\hat{g}$ is asymptotically unbiased).

When $\partial C/\partial t$ is not equal to zero and the cell is swimming at speed $v > 0$, the absorption time series $\{t_i\}$ does not contain the information necessary to estimate both g and $\partial C/\partial t$. This can be shown by combining Eq. (S4) and (S5):

$$\hat{c_1} = \frac{n \sum_i (t_i - t_0)}{4\pi DaT \sum_i (t_i - t_0)^2} \to c_1 = gv + \partial C/\partial t. \tag{S9}$$

Equation (S9) is clearly underdetermined; an infinite number of g and $\partial C/\partial t$ value pairs can satisfy Eq. (S9). Without additional information, any estimator of the concentration slope g will be biased. For instance, a cell could implement the maximum likelihood estimator $\hat{g}$ defined above to estimate the concentration slope in a dynamic environment. For $\partial C/\partial t \neq 0$, the sum $\sum_i (t_i - t_0) \approx \pi Da(gv + \partial C/\partial t)T^3/3$, which implies that

$$\hat{g} \to g + (\partial C/\partial t)/v. \tag{S10}$$

Equation (S10) illustrates two important points: the bias in the concentration slope estimate (second term on the r.h.s. of Eq. (S10)) is reduced as swimming speed increases; and this bias does not depend on measurement time T. One could propose alternative estimators for the concentration slope g that satisfy Eq. (S9), but these basic conclusions are unchanged.

Dynamics of the outer boundary

Derivation of the outer boundary, r_o. To derive the time-scaling of the outer boundary r_o, we consider a cell that is traveling directly toward the origin of the pulse at speed v. We define r_o as the largest radius, r that satisfies Eq. (4) in the Main Text. To approximate this value, note that temporal changes in concentration are described by

$$\frac{\partial C}{\partial t} = \left[\frac{r^2}{4Dt^2} - \frac{N}{2t}\right] C, \tag{S11}$$

for the concentration profile studied in the Main Text, which implies that near $r = \sqrt{2NDt}$, temporal changes in the concentration field are small. We assume r_o is in this region and therefore neglect contributions of $\partial C/\partial t$ to the ramp rate measured by a swimming cell. This implies that the condition for the signal-to-noise ratio to rise above δ_0 is $-v\frac{\partial C}{\partial r}C^{-1/2} = vrC^{1/2}/(2Dt) \geq \delta$. Solving for r gives:

$$r = \sqrt{-4Dt\, W(-16Dtd^2)}, \tag{S12}$$

where $d = \delta(4\pi Dt)^{N/4}(4\sqrt{M}v)^{-1}$, and $W(\cdot)$ is the product log function. In general, M will be large so the argument of the product log function will be negative and close to zero (because d is small). An approximation for the branch of the product log function that corresponds to r_o in this regime [6] is $W(x) \approx \ln(-x) - \ln(-\ln(-x))$, which yields the approximation for r_o given by Eq. (5) in the Main Text.

Derivation of the time when chemotaxis ceases, t^*. The signal-to-noise ratio takes its maximum value at $r = \sqrt{4Dt}$. Near this radius the contribution of temporal changes in the chemical field to the cell's perceived ramp rate are small and, again, the signal-to-noise ratio is approximately proportional to $-v\frac{\partial C}{\partial r}C^{-1/2}$. Solving for the time at which the maximum signal-to-noise ratio falls below threshold yields the expression for t^* given by Eq. (6) in the Main Text.

Dynamics of the inner boundary.

The inner boundary r_i is given implicitly by

$$\left| vg(r,t) + \frac{\partial C(r,t)}{\partial t} \right| C(r,t)^{-1/2} - \delta = 0, \tag{S13}$$

with $g = \partial C/\partial r$. Eq. (S13) has zero, one, or two positive roots. When this expression has no positive roots, cells traveling down the concentration gradient experience a signal-to-noise ratio of the ramp rate estimator that is below threshold δ_0 everywhere. When this expression has one positive and one negative root, there is a maximum distance, beyond which cells traveling down the concentration gradient typically fail to detect a signal that is resolvable above noise, but any cell within this outer radius can typically resolve the ramp rate (Fig. 3 of the Main Text, early time). When Eq. (S13) has two positive roots, there exists an inner boundary, $r_i > 0$ within which, cells cannot resolve the ramp rate. This latter case is shown in Fig. 2a (dashed blue curve) and Fig. 3 inset in the Main Text.

References

[1] Berg JC, Purcell EM (1977) Physics of chemoreception. *Biophys. J.* 20:193–219.

[2] Endres RG, Wingreen NS (2008) Accuracy of direct gradient sensing by single cells. *Proc. Natl. Acad. Sci. U.S.A.* 105:15749–15754.

[3] Mora T, Wingreen NS (2010) Limits of sensing temporal concentration changes by single cells. *Phys. Rev. Lett.* 104:248101.

[4] Rice J (1995) *Mathematical Statistics and Data Analysis* (Duxbury, Belmont), 2nd edition.

[5] Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis* (Springer, New York, NY USA).

[6] Abramowitz M, Stegun I (1964) *Handbook of Mathematical Functions* (US Government Printing Office, Washington, DC, USA).