A universal energy accuracy tradeoff in nonequilibrium cellular sensing

Sarah E. Harvey,* Subhaneil Lahiri, and Surya Ganguli

Department of Applied Physics, Stanford University, Stanford, California, 94305

(Dated: February 26, 2020)

We combine stochastic thermodynamics, large deviation theory, and information theory to derive fundamental limits on the accuracy with which single cells can detect external chemical concentrations through arbitrarily complex, energy consuming nonequilibrium cell-surface receptors. We show that if estimation is performed by an ideal observer of the entire trajectory of receptor states, then *no* energy consuming non-equilibrium receptor that can be divided into two pools of bound signaling and unbound non-signaling states can outperform a simple equilibrium two-state receptor. Moreover, we derive an energy accuracy tradeoff for such general non-equilibrium receptors when the estimation is performed by a simple observer of the duration the receptor is bound. Our tradeoff reveals that the simple observer can only attain the performance of the ideal observer in the limit of large receptor energy consumption and size. Our derivations generalize the classic 1977 Berg-Purcell limit on cellular chemosensation along multiple dimensions, and also yield a novel thermodynamic uncertainty relation for the time a physical system spends in a pool of states. This relation itself is of independent interest, with applications not only to cellular chemosensation, but also to the reliability of other biological processes, like clocks and motors, as a function of energy consumption.

Single celled organisms possess extremely sensitive mechanisms for detecting external chemical concentrations through the binding of individual molecules to cell-surface receptors (fig. 1 (a)) [1, 2]. This remarkable capacity may require the consumption of energy, and raises important questions about fundamental limits on the statistical accuracy of cellular chemosensation, both as a function of the energy consumed by arbitrarily complex nonequilibrium receptors, and the computational sophistication of downstream observers of these dynamics.

A seminal line of work initiated by Berg and Purcell in 1977 [3–5] addressed this issue for the simple case of equilibrium receptors comprising two states, bound and unbound, in which the binding transition rate is proportional to external concentration c (fig. 1 (b)). Berg and Purcell studied the accuracy of a concentration estimate $\hat{c}$ computed by a simple observer (SO) that only has access to the fraction of time the receptor is bound over a time T, finding a fundamental lower bound on the error of this estimate, in

---

*harveys@stanford.edu

terms of the fractional uncertainty:

$$\epsilon_{\hat{c}}^2 \equiv \frac{\left\langle (\delta\hat{c})^2 \right\rangle}{c^2} \geq \frac{2}{\overline{N}}. \tag{1}$$

Here $\left\langle (\delta\hat{c})^2 \right\rangle$ is the variance of the estimate $\hat{c}$, and $\overline{N}$ is the mean number of binding events in time T.

For over 30 years eq. (1) was thought to constitute a fundamental physical limit to the accuracy of cellular chemosensation. However, recent work focusing on highly specific receptor models [6–9] revealed this bound could be circumvented in two qualitatively distinct ways. First, in the simple case of a two-state receptor, an ideal observer (IO) that has access to the entire receptor trajectory of binding and unbinding events, could outperform the SO by performing maximum-likelihood estimation (MLE), obtaining a fractional error $\epsilon_{\hat{c}}^2 = \frac{1}{\overline{N}}$ [7]. The IO in this case outperforms the SO by a factor of two, by simply employing the mean duration of unbound intervals, and ignoring the duration of bound intervals as these simply contribute spurious noise because the transition rate out of the bound state is independent of c. Second, even when the estimate is performed by a SO, the Berg-Purcell limit can be overcome

by energy-consuming non-equilibrium receptors with more than two states (fig. 1 (d)), reflecting different receptor conformations or phosphorylation states [6, 7, 9]. Notably, [9] numerically observed a tradeoff between estimation fractional error and energy consumption for a very specific class of receptor models with states arranged in a ring topology.

While these more recent works demonstrate circumventions of the Berg-Purcell limit in highly specific models, they leave open several foundational theoretical questions about the general interplay between cellular sensing accuracy and energy consumption across the large space of all possible complex non-equilibrium receptors. For example, could adding more states and consuming energy in a cell-surface receptor reduce the fractional error of an IO to *less than* $\epsilon_{\hat{c}}^2 = \frac{1}{N}$? Moreover, for such complex non-equilibrium receptors, what function of the receptor trajectory would the IO have to compute? Returning to the SO, which may be implemented in a more biologically plausible manner, can we derive a general *analytic* relationship between energy and accuracy applicable to large classes of non-equilibrium receptors?
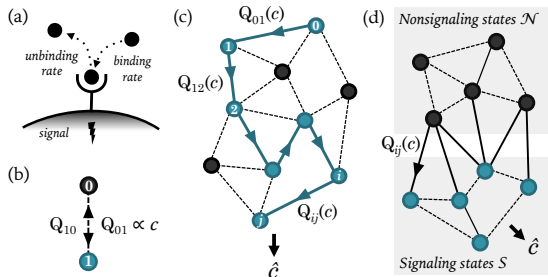

FIG. 1. Single receptors as continuous-time Markov processes. (a) Cartoon of a single receptor. (b) Single receptor modeled as a two-state continuous-time Markov process. (c-d) Generalizing to many-state processes. (c) An ideal observer which uses full Markov process trajectory to form an estimate $\hat{c}$ of the signal c, which modulates the receptor's transition rates. (d) A simple observer which only has access to the fraction of time the receptor spends in a subset of states ('signaling states') to form an estimate of the concentration $\hat{c}$.

tors? Moreover, can we derive exact formulae for accuracy in this general setting?

In this letter, we answer all of these questions, by combining and extending the methods of stochastic thermodynamics [10–15] and large deviation theory of Markov chains [16–22]. Along the way we derive a novel thermodynamic uncertainty relation governing fluctuations in the time a stochastic process occupies a subset of states, and the energy spent by that process. This uncertainty relation itself is of independent interest to the field of non-equilibrium statistical mechanics [21–24] and could have applications not only to cellular chemosensation, as discussed here, but also to understanding relations between energy and accuracy in other biological processes, like cellular motors and biological clocks [25–27].

Overall framework. A general non-equilibrium receptor can be modeled as a continuous-time Markov process [9, 28], with n different conformational or signaling states indexed by $i = 0, \ldots, n - 1$. The transition rate from state i to j is $Q_{ij} \geq 0$, and we assume some subset of these rates are proportional to concentration c. Over an observation time T, the receptor then moves stochastically through a sequence of states $\{x_0, x_1, \ldots, x_m\}$ with transitions out of each state $x_k \in \{0, \ldots, n-1\}$ occurring at random times $0 \leq t_k \leq T$, yielding a random trajectory $x(t)$ with $x(t) = x_k$ for $t_{k-1} \leq t < t_k$. An IO that has access to the entire trajectory $x(t)$ (fig. 1 (c)) can compute an optimal estimate $\hat{c}$ of the concentration c via MLE:

$$\hat{c} = \text{argmax}_c \log \mathbb{P}[x(t)|c], \qquad (2)$$

where $\mathbb{P}[x(t)|c]$ denotes the probability distribution over receptor state trajectories at a given concentration c. However, this MLE computation of the IO may be difficult to implement in a biologically plausible manner. Therefore we also consider receptor models in which the states can be divided into two sets: unbound, nonsignaling states, $\mathcal{N}$, and bound, signaling states, $\mathcal{S}$, with only the transition rates from nonsignaling to signaling states proportional to the external concentration, c (fig. 1 (d)). We as-

sume that a SO can easily access the fraction of time the receptor spends in the signaling states by counting the number of downstream signaling molecules generated while the receptor is in those states, and we will compute the performance of the SO below. We note these general assumptions are consistent with previous work [7, 9], though they do exclude receptors with intermediate states that are bound but not signaling [29].

Receptor energy consumption and the ideal observer. Due to the properties of Markov processes [30, §B B], the log probability of an entire trajectory $x(t)$ reduces to

$$\log \mathbb{P}[x(t)|c] = -T \sum_{i \neq j} \left[p_i^T Q_{ij} - \phi_{ij}^T \log Q_{ij} \right], \quad (3)$$

where $p_i^T \equiv \frac{1}{T} \int_0^T \mathrm{d}t \, \delta_{x(t)\,i}$ is the empirical density, or fraction of time the trajectory $x(t)$ spends in state i, and ϕ_{ij}^T is the empirical flux, or the number of transitions from state i to state j divided by T. Maximizing eq. (3) w.r.t. c yields the IO estimate $\hat{c}$ in eq. (2). When only transition rates from $\mathcal{N}$ to $\mathcal{S}$ are proportional to c as in fig. 1 (d),

$$\hat{c} = \frac{R^T}{R^p\big|_{c=1}}, \quad (4)$$

where $R^T \equiv \sum_{i \in \mathcal{N}, j \in \mathcal{S}} \phi_{ij}^T$ is the receptor's empirical binding rate along trajectory $x(t)$, and $R^p \equiv \sum_{i \in \mathcal{N}, j \in \mathcal{S}} p_i^T Q_{ij}$, the expected binding rate conditioned on the empirical density p_i^T. For the two state case, eq. (4) reduces to the inverse of the mean duration of unbound intervals, consistent with [7]. However, this result goes beyond [7] to reveal what function of the receptor trajectory $x(t)$ an optimal IO of arbitrarily complex receptors, as in fig. 1 (d), must compute to estimate concentration.

The fractional error, $\epsilon_{\hat{c}}^2$, of the IO can be lower bounded by the Fisher information J_c the receptor trajectory $x(t)$ contains about the external concentration c through the Cramér-Rao bound [31]. A simple calculation [30, §II A] yields

$$J_c = J_c^0 + T \sum_{i \neq j} \pi_i Q_{ij} \left[\partial_c \log Q_{ij} \right]^2. \quad (5)$$

Here π_i is the steady state probability of state i, and $J_c^0 = \sum_i \pi_i (\partial_c \log \pi_i)^2$ is the Fisher information the initial state observation contains about c. The second term grows with T and indicates additional information obtainable from observing the entire trajectory. Note only transition rates Q_{ij} modulated by c contribute information. This expression holds for arbitrary receptors as in fig. 1 (c), but simplifies to

$$J_c = J_c^0 + \frac{T}{c^2} R^\pi. \quad (6)$$

for receptors of the form shown in fig. 1 (d). Here $R^\pi = \sum_{i \in \mathcal{N}, j \in \mathcal{S}} \pi_i Q_{ij}$ is the expected steady-state binding rate. Then, employing the Cramér-Rao bound, we find for large T,

$$\epsilon_c^2 \equiv \frac{\left\langle (\delta \hat{c})^2 \right\rangle}{c^2} \geq \frac{1}{J_c \, c^2} = \frac{1}{T \, R^\pi} = \frac{1}{\overline{N}}. \quad (7)$$

Here $\overline{N} = R^\pi T$ is the expected number of binding events. Moreover, we directly calculate the variance of the IO concentration estimate in eq. (4) and demonstrate its fractional error does indeed saturate the bound in eq. (7) [30]. This result thus generalizes [7] from simple equilibrium two state receptors to arbitrarily complex nonequilibrium receptors of the form in fig. 1 (d). Remarkably, our result reveals that *any* energy consuming non-equilibrium receptor that can be divided into two pools of bound signaling and unbound non-signaling states *cannot* outperform a simple equilibrium two-state receptor, as long as a downstream ideal observer is used to estimate concentration.

Fluctuations and gain determine simple observer performance. The IO concentration estimate in eq. (4) requires computing a rather complex function of the receptor trajectory $x(t)$, which may not be biologically plausible in general. We therefore turn our attention to the SO, which must estimate concentration using only the fraction of time the receptor is bound, which we denote by $q^T = \sum_{i \in \mathcal{S}} p_i^T$. Due to randomness in the trajectory $x(t)$, the observable q^T fluctuates about its mean $q^\pi = \sum_{i \in \mathcal{S}} \pi_i$, where the steady state probabilities π_i and therefore q^π depend on the true concentration c. Given

the observable q^T, one can then estimate $\hat{c}$ by solving $q^\pi(\hat{c}) = q^T$. The standard rules for error propagation then yield

$$\epsilon_{\hat{c}}^2 = \frac{\langle(\delta\hat{c})^2\rangle}{c^2} = \left[c\,\frac{\mathrm{d}q^\pi}{\mathrm{d}c}\right]^{-2}\langle(\delta q^T)^2\rangle. \qquad (8)$$

Thus a larger variance $\langle(\delta q^T)^2\rangle$ in the time spent bound increases the error $\epsilon_{\hat{c}}^2$, while a larger gain $\left|\frac{\mathrm{d}q^\pi}{\mathrm{d}c}\right|$ in the mean time spent bound decreases the error. We next compute and bound this variance and gain.

A thermodynamic uncertainty relation for signaling density. We first derive a lower bound on the variance $\langle(\delta q^T)^2\rangle$ using ideas from stochastic thermodynamics and the large deviation theory [16] of Markov processes. Indeed, a random trajectory $x(t)$ of duration T in a general Markov process will yield an empirical density p_i^T and an empirical current $j_{ij}^T = \phi_{ij}^T - \phi_{ji}^T$, which corresponds to the *net* number of transitions from i to j divided by T. As $T \to \infty$, these random variables will converge to their mean values, corresponding to the steady state probabilities $\pi_i = \lim_{T\to\infty} p_i^T$ and steady state currents $j_{ij}^\pi \equiv \pi_i Q_{ij} - \pi_j Q_{ji} = \lim_{T\to\infty} j_{ij}^T$. However, at large but finite T, p^T and j^T fluctuate about their means, and their joint distribution takes the form $\mathbb{P}(p^T = p, j^T = j) \sim \mathrm{e}^{-TI(p,j)}$ [20, 30]. Here $I(p,j)$ is a large deviation rate function that achieves its minimum at $p = \pi$ and $j = j^\pi$, and describes how fluctuations in p^T and j^T away from these means are suppressed. This rate function is [17–20]

$$I(p,j) = \sum_{i<j} j_{ij}\left(\mathrm{arcsinh}\,\frac{j_{ij}}{a_{ij}} - \mathrm{arcsinh}\,\frac{j_{ij}^p}{a_{ij}}\right) \\ -\left(\sqrt{a_{ij}^2 + j_{ij}^2} - \sqrt{a_{ij}^2 + j_{ij}^{p\,2}}\right), \qquad (9)$$

where $a_{ij} \equiv 2\sqrt{p_i p_j Q_{ij} Q_{ji}}$ and $j_{ij}^p \equiv p_i Q_{ij} - p_j Q_{ji}$.

Similarly, at large but finite T, the distribution of the fraction of time time spent bound, namely $q^T = \sum_{i\in\mathcal{S}} p_i$, takes the form $\mathbb{P}(q^T = q) = \mathrm{e}^{-TI(q)}$. Here the large deviation rate function $I(q)$ achieves its minimum at the mean value

$q^\pi \equiv \sum_{i\in\mathcal{S}} \pi_i$, and describes how deviations in q^T from its mean are suppressed. Indeed the variance of q^T is given by $1/(TI''(q^\pi))$ [16], so any upper bound on $I''(q^\pi)$ will yield a lower bound on the variance of q^T.

$I(q)$ can be obtained from the more general rate function $I(p,j)$ through the contraction principle [16], which states that $I(q) = \inf_{p,j} I(p,j)$, subject to the constraints $\sum_j j_{ij} = 0 \,\forall\, i$, $\sum_i p_i = 1$ and $\sum_{i\in\mathcal{S}} p_i = q$. Instead of calculating this directly, the infimum can be bounded by evaluating $I(p,j)$ for any choice of $j = j^*(q)$ that satisfies current conservation, $\sum_j j_{ij}^* = 0 \,\forall\, i$, and $p = p^*(q)$, such that $\sum_{i\in\mathcal{S}} p_i^* = q$. Hence,

$$I(q) \le I(p^*, j^*) \qquad (10)$$

As we show in [30, §III], we can choose a $p^*(q)$ and a $j^*(q)$ that satisfies $j_{ij}^*(q^\pi) = j_{ij}^\pi$ and $p_i^*(q^\pi) = \pi_i$, ensuring that the inequality eq. (10) is saturated at the minimum, $q = q^\pi$.

Using eq. (9) and our choice of p^*, j^*, we can find an explicit upper bound on $I''(q^\pi)$ in terms of the total energy consumption rate and the total binding rate. Defining

$$\Sigma^\pi = \sum_{i<j} j_{ij}^\pi \log\frac{\phi_{ij}^\pi}{\phi_{ji}^\pi} \qquad (11)$$

as the steady state energy consumption rate of the receptor (in units of $k_\mathrm{B}T$) [10], we therefore find a lower bound on the variance of q [30, §III],

$$\langle(\delta q)^2\rangle\left[T\Sigma^\pi + 4\overline{N}\right] \ge 8\left[q^\pi(1 - q^\pi)\right]^2. \qquad (12)$$

Equation (12) can be thought of as a new, general thermodynamic uncertainty relation which implies that the more energy $T\Sigma^\pi$ a system consumes, the more reliably it can achieve a given time spent in a pool of states. This can be compared to another thermodynamic uncertainty relation connecting increased energy consumption to a reduction in current fluctuations in general stochastic processes [22]. Thus by extending the class of observables (i.e. from currents to pooled state occupancies) for which thermodynamic uncertainty relations can be generally proven, our result in eq. (12) may be of independent interest in nonequilibrium thermodynamics.

An energy accuracy tradeoff for the simple observer. Now the gain $c \frac{\mathrm{d}q^\pi}{\mathrm{d}c}$ in eq. (8) can be calculated for arbitrary nonequilibrium processes using the theory of first passage times [30]. Our formulae in [30, §IV] simplify to $c \frac{\mathrm{d}q^\pi}{\mathrm{d}c} = q^\pi(1 - q^\pi)$ for arbitrary nonequilibrium receptors of the form in fig. 1 (d) with only one nonsignaling state. Combining this result for gain with the thermodynamic uncertainty relation in eq. (12) for variance, and inserting both into eq. (8), we obtain a general lower bound on fractional error in terms of energy consumption $T\Sigma^\pi$ and the expected number of binding events $\overline{N}$:

$$\epsilon_{\hat{c}}^2 \geq \frac{8}{T\Sigma^\pi + 4\overline{N}}. \tag{13}$$

Clearly, eq. (13) recovers the Berg-Purcell limit eq. (1) at zero energy consumption. But overall, eq. (13) can be thought of as a fundamental generalization of the Berg-Purcell limit to general energy consuming non-equilibrium receptors of the form in fig. 1 (d), with one nonsignaling state, but an arbitrary network of signaling states.

Exact estimation error for the simple observer. Equation (13) only provides a lower bound on the fractional error because our choice of p^*, j^* in eq. (10) does not achieve the infimum of the true contraction. When the true contraction is expanded as a Taylor series in $(q - q^\pi)$, one can compute the leading terms exactly [30, §V C]. Under the same assumption of only one nonsignaling state, we find

$$\epsilon_{\hat{c}}^2 = \frac{2}{\overline{N}} \frac{T_{\mathrm{unbind}}}{T_{\mathrm{hold}}}, \tag{14}$$

where T_{unbind} is the mean time until the next unbinding event given that the receptor is currently in a signaling state, $\sum_{i\in\mathcal{S}} \mathbf{T}_{i\mathcal{N}} \mathbb{P}(x(t) = i | x(t) \text{ in } \mathcal{S})$, T_{hold} is the mean duration of a full journey through the signaling states (starting right when the receptor first enters the signaling states), $\sum_{i\in\mathcal{S}} \mathbf{T}_{i\mathcal{N}} \mathbb{P}(x(t) = i | x(t) \text{ just entered } \mathcal{S})$, and $\mathbf{T}_{i\mathcal{N}}$ is the mean first passage time from state i to the single nonsignaling state. In [30, §V C] we

have numerically verified this formula by comparing it with Monte-Carlo simulations. Moreover, in a two state system $T_{\mathrm{unbind}} = T_{\mathrm{hold}}$ because the unbinding process is memoryless, and so eq. (14) reduces to the Berg-Purcell result in eq. (1). Thus eq. (14) can be thought of as another generalization of Berg-Purcell to general energy consuming non-equilibrium receptors of the form in fig. 1 (d), with one nonsignaling state, but an arbitrary network of signaling states. While eq. (14) gives an exact formula for the fractional error, our lower bound in eq. (13) makes manifest a connection between accuracy and energy.

Numerical verification. Figure 2 (a) compares the theoretical fractional error bounds in eqs. (7) and (13) to results of numerical optimization of eq. (14) and Monte Carlo simulations of random receptors with 5 states. Our lower bounds are respected by all models studied (to within error bars for the simulations). Receptors with five states can saturate the bound at low, but not high energy consumption. Figure 2 (i–iii) depicts the optimal receptors obtained at three different levels of energy consumption, and are representative of the three qualitative forms that the optimal receptors move through as the energy consumption is increased (see supplementary video). At low energy consumption the optimal receptor is equivalent to a two-state receptor with all the signaling states behaving like one coarse-grained state (see [30, §VI A 1]). At intermediate energy consumption the optimal receptor behaves roughly like a three state system, with 'inner' states of $\mathcal{S}$ being highly short-lived. At high energy consumption, the optimal receptor is a uniform ring that becomes more asymmetric at higher energies.

Figure 2 (b) provides a comparison of the performance of numerically optimized receptors of various sizes and unconstrained connectivity. As the number of states n in the receptor is increased, the performance at high energy consumption becomes closer to the bound.

Example: the uniform ring receptor. We can understand this last observation by noting from fig. 2 (a) that a uniform ring receptor becomes
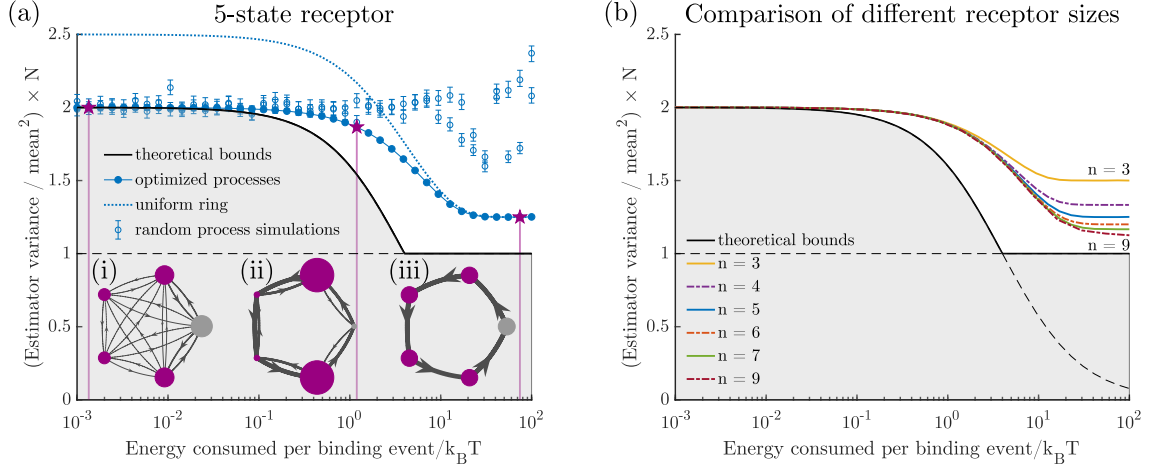
FIG. 2. (a) Numerical results for 5-state receptors. Solid black curve shows the upper envelope of the theoretical bounds eqs. (7) and (13). Gray shaded regions are forbidden by our bounds. Solid blue circles show the fractional error achieved by optimal 5-state receptors, obtained by numerically minimizing eq. (14) (see [30, §V]) subject to an energy consumption constraint. Dotted blue line shows the performance of a ring receptor with uniform transition rates in each direction. Open blue circles show the performance of 5-state receptors with randomly chosen transition rates (subject to an energy consumption constraint), obtained by Monte-Carlo simulation. (i-iii) Three optimal receptors found at three given energy consumption level indicated by the three magenta vertical lines. Magenta (gray) nodes in the diagrams represent signaling (nonsignaling) states. Node radii are proportional to the steady state probabilities π_i and edge widths are proportional to the steady state fluxes between nodes. (b) Fractional error of optimal receptors with different numbers of states n, obtained by numerically minimizing eq. (14) subject to an energy consumption constraint.

optimal at large energy consumption rate Σ^π. We therefore analytically compute all mean first passage times, T_{unbind}, T_{hold}, and $\epsilon_{\hat{c}}^2$ as a function of n, Σ^π and $\overline{N}$ [30, §V D] for the uniform ring. We find as $\Sigma^\pi \to \infty$, $\epsilon_{\hat{c}}^2 \to \frac{1}{\overline{N}} \frac{1}{1-1/n}$, indicating that an SO observing an energy consuming ring receptor can indeed approach the estimation of performance in eq. (7) of an IO observing *any* receptor, as the number of states n becomes large. However, a large uniform ring can be highly suboptimal under a SO if it does not consume energy: as $\Sigma^\pi \to 0$ we find $\epsilon_{\hat{c}}^2 \to \frac{2}{\overline{N}} \frac{n(n+1)}{6(n-1)}$. This reproduces Berg-Purcell in eq. (1) for $n = 2, 3$, but is strictly worse when $n > 3$. Thus to take advantage of a larger number of states n, a ring receptor must consume more energy.

Discussion. In summary, we have derived several general results (eqs. (4), (5), (7), (13) and (14)) delineating fundamental performance limits of cellular chemosensation using arbitrarily complex energy consuming nonequilibrium cell-surface receptors, as a joint function of observation time, energy consumption rate, and the computational complexity of the downstream observer. Along the way we have also derived a general thermodynamic uncertainty relation (eq. (12)) for signaling density, which reveals one must pay a universal energetic cost for reliable occupation time in any physical process. We hope these analytic relations between time, energy and accuracy will find further applications in myriad biological and physical processes [26, 32–38].

[1] C. U. M. Smith, *Biology of sensory systems*, 2nd ed. (Wiley-Blackwell, 2008).

[2] V. Sourjik and N. S. Wingreen, Responding to chemical gradients: bacterial chemotaxis, Curr. Opin. Cell Biol. **24**, 262 (2012).

[3] H. C. Berg and E. M. Purcell, Physics of chemoreception, Biophys. J. **20**, 193 (1977).

[4] W. Bialek and S. Setayeshgar, Physical limits to biochemical signaling, Proc. Natl. Acad. Sci. U.S.A. **102**, 10040 (2005).

[5] K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P. R. ten Wolde, The Berg-Purcell limit revisited., Biophys. J. **106**, 976 (2014).

[6] T. Mora and N. S. Wingreen, Limits of Sensing Temporal Concentration Changes by Single Cells, Phys. Rev. Lett. **104**, 248101 (2010).

[7] R. G. Endres and N. S. Wingreen, Maximum Likelihood and the Single Receptor, Phys. Rev. Lett. **103**, 158101 (2009).

[8] P. Mehta and D. J. Schwab, Energetic costs of cellular computation, Proc. Natl. Acad. Sci. U.S.A. **109**, 17978 (2012).

[9] A. H. Lang, C. K. Fisher, T. Mora, and P. Mehta, Thermodynamics of Statistical Inference by Cells, Phys. Rev. Lett. **113**, 148103 (2014).

[10] U. Seifert, Stochastic thermodynamics: principles and perspectives, Eur. Phys. J. B **64**, 423 (2008).

[11] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[12] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics, Vol. 799 (Springer, 2010).

[13] X.-J. Zhang, H. Qian, and M. Qian, Stochastic theory of nonequilibrium steady states and its applications. Part I, Phys. Rep. **510**, 1 (2012).

[14] H. Ge, M. Qian, and H. Qian, Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics, Phys. Rep. **510**, 87 (2012).

[15] U. Seifert, Stochastic thermodynamics, fluctuation theorems, and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[16] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. **478**, 1 (2009), arXiv:0804.0327.

[17] C. Maes and K. Netočný, Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states, Europhys. Lett. **82**, 30003 (2008).

[18] L. Bertini, A. Faggionato, and D. Gabrielli, Large deviations of the empirical flow for continuous time Markov chains, arXiv:1210.2004 [math] (2012).

[19] L. Bertini, A. Faggionato, and D. Gabrielli, Flows, currents, and cycles for Markov Chains: large deviation asymptotics, (2014), arXiv:1408.5477.

[20] A. C. Barato and R. Chetrite, A Formal View on Level 2.5 Large Deviations and Fluctuation Relations, J. Stat. Phys. **160**, 1154 (2015), arXiv:1408.5033.

[21] A. C. Barato and U. Seifert, Dispersion for two classes of random variables: General theory and application to inference of an external ligand concentration by a cell, Phys. Rev. E **92**, 032127 (2015).

[22] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Dissipation Bounds All Steady-State Current Fluctuations, Phys. Rev. Lett. **116**, 120601 (2016).

[23] A. C. Barato and U. Seifert, Thermodynamic Uncertainty Relation for Biomolecular Processes, Phys. Rev. Lett. **114**, 158101 (2015).

[24] J. M. Horowitz and T. R. Gingrich, Thermodynamic uncertainty relations constrain nonequilibrium fluctuations, Nature Physics **16**, 15 (2020).

[25] G. Li and H. Qian, Kinetic Timing: A Novel Mechanism That Improves the Accuracy of GTPase Timers in Endosome Fusion and Other Biological Processes, Traffic **3**, 249 (2002).

[26] H. Qian, Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological Functions, Annual Review of Physical Chemistry **58**, 113 (2007).

[27] R. Marsland, W. Cui, and J. M. Horowitz, The thermodynamic uncertainty relation in biochemical oscillations, J. R. Soc. Interface **16**, 10.1098/rsif.2019.0098 (2019), arXiv:1901.00548.

[28] J. M. Horowitz and M. Esposito, Thermodynamics with Continuous Information Flow, Phys. Rev. X **4**, 031015 (2014).

[29] M. Skoge, S. Naqvi, Y. Meir, and N. S. Wingreen, Chemical Sensing by Nonequilibrium Cooperative Receptors, Phys. Rev. Lett. **110**, 248102 (2013).

[30] See Supplementary Information.

[31] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. (Wiley-Interscience, 2006) p. 748.

[32] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, The energy-speed-accuracy trade-off in sensory adaptation, Nat. Phys. **8**, 422 (2012).

[33] P. Mehta, A. H. Lang, and D. J. Schwab, Landauer in the Age of Synthetic Biology: Energy Consumption and Information Processing in Biochemical Networks, Journal of Statistical Physics **162**, 1153 (2016), arXiv:1505.02474.

[34] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[35] S. Lahiri, J. Sohl-Dickstein, and S. Ganguli, A universal tradeoff between power, precision and speed in physical communication, arXiv:1603.07758 [cond-mat] (2016).

[36] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, Thermodynamics of Prediction, Phys. Rev. Lett. **109**, 120604 (2012).

[37] A. B. Boyd, D. Mandal, P. M. Riechers, and J. P. Crutchfield, Transient dissipation and structural costs of physical information transduction, Phys. Rev. Lett. **118**, 220602 (2017), arXiv:1612.08616.

[38] S. E. Marzen and J. P. Crutchfield, Prediction and Dissipation in Nonequilibrium Molecular Sensors: Conditionally Markovian Channels Driven by Memoryful Environments, Bulletin of Mathematical Biology **82**, 1 (2020).

Supplementary Information

CONTENTS

I. OVERVIEW

In this supplement we provide complete derivations of several results presented in the main text, as well as background material on Markov processes and large deviation theory for a general physics audience.

In §II below we derive eq. (4) and eq. (5) of the main text. Equation (4) reveals the the computation the ideal observer (IO) must make to construct an optimal estimate of the concentration from knowledge of the entire trajectory of receptor states over a given time interval. Correspondingly, eq. (5) describes the Fisher information that the entire receptor state trajectory contains about the external concentration. The reciprocal of this Fisher information bounds the error of the IO estimate through the Cramér-Rao bound.

In sections III and IV we derive the main text's eqs. (12) and (13). Equation (12) describes a thermodynamic uncertainty relation revealing that energy must be spent to reduce fluctuations in the time a physical process spends in a subset of states. Equation (13) describes how this thermodynamic uncertainty relation, when combined with a computation of the receptor gain, yields a lower bound on estimation error in terms of energy consumption.

In §V we use large deviation theory to derive exact formulae for the fractional error for the ideal observer (IO) and the simple observer (SO), including the special case of a uniform ring receptor. The second formula (main text eq. (14)) was used for the numerical optimization of SO performance over the space of receptors in fig. 2 of the main text.

In §VI we provide details for the numerical computations in the main text fig. 2.

Finally, to make this supplement self contained, we provide appendices A and B with brief reviews of the theory of continuous-time Markov processes and the large deviation theory for empirical density, flux and current.

II. INFORMATION AND ESTIMATION ACCURACY FOR THE IDEAL OBSERVER

A. Fisher information in a Markovian trajectory

In this section we derive the Fisher information from an extended observation of a system with Markovian dynamics, eq. (5) in the main text. We first consider a discrete time Markov process, and will later take the limit as the size of the discrete time steps Δt become vanishingly small.

The discrete-time transition matrix is given by M, where M_{ij} is the probability of transition from state i to state j if the system is in state i at a particular time step. For a set of states labeled by i, we define π_i as the steady state distribution, which satisfies $\boldsymbol{\pi}\mathbf{M} = \boldsymbol{\pi}$ and has elements which sum to 1. The matrix $\mathbf{M}$ can be expanded in terms of the continuous time transition rate matrix $\mathbf{Q}$, which has elements Q_{ij} and obeys $\sum_j Q_{ij} = 0$ (see eq. (82) in §A A).

$$\mathbf{M} = e^{\mathbf{Q}\Delta t} = \mathbf{I} + \mathbf{Q}\Delta t + \dots \tag{15}$$

We would like to consider the general probability of a Markovian trajectory from a state x_0 at $t = 0$ to state x_n at $t = n\Delta t$.[1] Assuming that the system begins in the

———

[1] In the main text we described a trajectory by the sequence of states visited, $\{x_0, \dots, x_m\}$ and

steady state distribution, the probability of this trajectory in discrete time is given by $\mathbb{P}(x_0, \ldots, x_n) = \pi_{x_0} M_{x_0 x_1} M_{x_1 x_2} \ldots M_{x_{n-1} x_n}$. We can now directly calculate the Fisher information matrix for this distribution with respect to the parameter λ_μ, using the notation $\partial_\mu \equiv \frac{\partial}{\partial \lambda_\mu}$:

$$
\begin{aligned}
J_{\mu\nu} = \sum_{x_0, \ldots, x_n} \pi_{x_0} M_{x_0 x_1} \ldots M_{x_{n-1} x_n} \Big[\partial_\mu \log(\pi_{x_0} M_{x_0 x_1} \ldots M_{x_{n-1} x_n}) \Big] \\
\times \Big[\partial_\nu \log(\pi_{x_0} M_{x_1 x_2} \ldots M_{x_{n-1} x_n}) \Big].
\end{aligned}
\tag{16}
$$

We now recognize that the Fisher information matrix eq. (16) can be rewritten as (using the fact $\sum_k M_{jk} = 1$ and $\sum_j \pi_j M_{jk} = \pi_k$)

$$
\begin{aligned}
J_{\mu\nu} = \sum_{x_0} \pi_{x_0} \partial_\mu \log \pi_{x_0} \partial_\nu \log \pi_{x_0} \\
+ \sum_{x_0 \ldots x_n} \pi_{x_0} M_{x_0 x_1} M_{x_1 x_2} \ldots M_{x_{n-1} x_n} \Big[\partial_\mu \big(\log M_{x_0 x_1} + \cdots + \log M_{x_{n-1} x_n} \big) \Big] \\
\times \Big[\partial_\nu \big(\log M_{x_0 x_1} + \cdots + \log M_{x_{n-1} x_n} \big) \Big].
\end{aligned}
\tag{17}
$$

or,

$$
J_{\mu\nu} = J_{\mu\nu}^0 + \sum_{x_0 \ldots x_n} \pi_{x_0} M_{x_0 x_1} M_{x_1 x_2} \cdots M_{x_{n-1} x_n} \Big[\partial_\mu \sum_{k=0}^{n-1} \log M_{x_k x_{k+1}} \Big] \Big[\partial_\nu \sum_{j=0}^{n-1} \log M_{x_j x_{j+1}} \Big]
\tag{18}
$$

where $J_{\mu\nu}^0$ is the Fisher information matrix for a random variable representing a single observation of the system state, and the indices k and j index the time steps in the measurement interval. The expression eq. (18) simplifies if we write the sums inside the brackets on the right hand side out term-by-term. For $k = 0$, $j = 0$, the second term on the right hand side simplifies to

$$
\sum_{x_0 \ldots x_n} \pi_{x_0} M_{x_0 x_1} \ldots M_{x_{n-1} x_n} \Big[\partial_\mu \log M_{x_0 x_1} \Big] \Big[\partial_\nu \log M_{x_0 x_1} \Big]
$$
$$
= \sum_{x_0, x_1} \pi_{x_0} M_{x_0 x_1} \Big[\partial_\mu \log M_{x_0 x_1} \Big] \Big[\partial_\nu \log M_{x_0 x_1} \Big]. \tag{19}
$$

Similarly, for $k = 0$, $j = 1$, we have

$$
\sum_{x_0 \ldots x_n} \pi_{x_0} M_{x_0 x_1} \ldots M_{x_{n-1} x_n} \Big[\partial_\mu \log M_{x_0 x_1} \Big] \Big[\partial_\nu \log M_{x_1 x_2} \Big]
$$
$$
= \sum_{x_0, x_1, x_2} \pi_{x_0} M_{x_0 x_1} M_{x_1 x_2} \Big[\partial_\mu \log M_{x_0 x_1} \Big] \Big[\partial_\nu \log M_{x_1 x_2} \Big]
\tag{20}
$$
$$
= \sum_{x_0, x_1, x_2} \pi_{x_0} \partial_\mu M_{x_0 x_1} \partial_\nu M_{x_1 x_2} = 0.
$$

---

the transition times $\{t_0, \ldots, t_m\}$. For *this section only*, we find it more convenient to describe the trajectory by the identity of the state occupied at each of a discrete set of time steps, $x_k = x(k\Delta t), k = 0, \ldots, n$. In the continuum limit, $\Delta t \to 0, n \to \infty$, the two descriptions contain the same information.

In the same fashion, all $k = j$ terms in eq. (18) give an expression similar to eq. (19) and all $k \neq j$ terms vanish as in eq. (20). Our expression for the Fisher information matrix then becomes:

$$J_{\mu\nu} = J_{\mu\nu}^0 + \sum_{k=1}^{n-1} \sum_{x_k, x_{k+1}} \pi_{x_k} M_{x_k, x_{k+1}} \, \partial_\mu \log M_{x_k, x_{k+1}} \, \partial_\nu \log M_{x_k, x_{k+1}}. \tag{21}$$

Given that M is not changing in time, after relabeling $x_k \to i$ and $x_{k+1} \to j$ all terms in the sum over k are identical. We therefore find:

$$J_{\mu\nu} = J_{\mu\nu}^0 + n \sum_{i,j} \pi_i M_{ij} \, \partial_\mu \log M_{ij} \, \partial_\nu \log M_{ij}. \tag{22}$$

Lastly, we take the continuous time limit by sending $\Delta t \to 0$. For infinitesimal Δt, $M_{ij} = Q_{ij} \Delta t$ for $i \neq j$ and $M_{ii} = 1 + Q_{ii} \Delta t$. We can then rewrite eq. (22) as

$$\begin{aligned} J_{\mu\nu} = J_{\mu\nu}^0 &+ n \sum_{i \neq j} \pi_i Q_{ij} \Delta t \, \partial_\mu \log(Q_{ij} \Delta t) \partial_\nu \log(Q_{ij} \Delta t) \\ &+ n \sum_{i=j} \pi_i (1 + Q_{ii} \Delta t) \, \partial_\mu \log(1 + Q_{ii} \Delta t) \partial_\nu \log(1 + Q_{ii} \Delta t). \end{aligned} \tag{23}$$

In the limit $\Delta t \to 0$, $\Delta t \log \Delta t \to 0$, and $\log(1 + Q_{ii} \Delta t) \approx Q_{ii} \Delta t$. Defining $T \equiv n \Delta t$, we therefore find

$$J_{\mu\nu} = J_{\mu\nu}^0 + T \sum_{i \neq j} \pi_i Q_{ij} \, \partial_\mu (\log Q_{ij}) \partial_\nu (\log Q_{ij}), \tag{24}$$

where we have recognized that the $i = j$ terms from eq. (23) all vanish in the limit $\Delta t \to 0$.

We then assume that the signal to be estimated is a scalar denoted by c, which could represent an external concentration of some ligand. For a scalar parameter, the Fisher information of the entire trajectory then becomes:

$$J_c = J_c^0 + T \sum_{i \neq j} \pi_i Q_{ij} \left[\partial_c \log Q_{ij} \right]^2, \tag{25}$$

which is eq. (5) in the main text.

If we specialize to the models of receptors studied in the main text, the only off-diagonal transition rates that depend on c are those along the edges in the set $\vec{\mathcal{B}}$: the edges that start in the set $\mathcal{N}$ and end in $\mathcal{S}$. As those transition rates are proportional to c, eq. (25) reduces to:

$$J_c = J_c^0 + T \sum_{ij \in \vec{\mathcal{B}}} \frac{\pi_i Q_{ij}}{c^2} = J_c^0 + \frac{T R^\pi}{c^2}. \tag{26}$$

This is eq. (6) in the main text. This leads to a lower bound on the uncertainty of any unbiased estimate of c, via the Cramér-Rao bound [1, 2].

B.   Maximum likelihood estimation for the ideal observer

In the previous section we computed the Fisher information for the ideal observer, which leads to a lower bound on the uncertainty of any estimate of c. In general, the maximum likelihood estimator saturates the Cramér-Rao bound asymptotically, in the limit of a large number of independent observations [3]. We compute this estimator in this section. We will postpone calculating its variance to §V A.

In §B B we see that, when the duration of observation is large, the likelihood of any single trajectory collapses to a function of certain summary statistics: the empirical density, p_i^T, the fraction of time spent in state i, and the empirical flux, ϕ_{ij}^T, the rate at which transitions from state i to j occur (see §B A for precise definitions). In eq. (105) we see that the likelihood is:

$$\log \mathbb{P}(x(t)|c) = -T \sum_{i \neq j} \left[p_i^T Q_{ij} - \phi_{ij}^T \log Q_{ij} \right], \tag{27}$$

where Q_{ij} is the source of dependence on c.

If we use the notation $\phi_{ij}^p = p_i^T Q_{ij}$, the maximum of this function must satisfy

$$\frac{\partial \log \mathbb{P}(x(t)|c)}{\partial c} = T \sum_{i \neq j} \left(\phi_{ij}^T - \phi_{ij}^p \right) \frac{\partial \log Q_{ij}}{\partial c} = 0. \tag{28}$$

Now we can specialize to the models of receptors studied in the main text, where the only off-diagonal transition rates that depend on c are those along the edges in the set $\vec{\mathcal{B}}$. As those transition rates are proportional to c, eq. (25) reduces to:

$$\frac{\partial \log \mathbb{P}(x(t)|c)}{\partial c} = \frac{T}{c} \left(R^T - R^p \right) = 0. \tag{29}$$

Because R^p is proportional to c, the maximum likelihood estimator is

$$\hat{c} = \frac{R^T}{R^p\big|_{c=1}}. \tag{30}$$

We will compute the variance of this estimator for large T in §V A.

III.   A THERMODYNAMIC UNCERTAINTY PRINCIPLE FOR DENSITY

Here we present a derivation of eq. (12) in the main text, which constitutes a thermodynamic uncertainty relation connecting fluctuations in the fraction of time a physical process spends in a pool of states to the energy consumption rate of that process. This uncertainty relation reveals that one cannot reduce fluctuations in total occupation time without paying an energy cost.

We make use of a known result that the empirical density and currents for continuous-time Markov processes obey a large deviation principle with a known joint rate function. The large deviation rate function $I(p, j)$ describes both fluctuations in the quantities p and j around their steady states and highly unlikely large

deviations [4]. This rate function is known to take the following form [5] (see also §B D):

$$I(p,j) = \sum_{i<j} \Psi(j_{ij}, j^p_{ij}, a^p_{ij}) \tag{31}$$

with (dropping the state indices i, j for notational simplicity)

$$\Psi(j, j^p, a) = j\left(\operatorname{arcsinh}\frac{j}{a} - \operatorname{arcsinh}\frac{j^p}{a}\right) - \left(\sqrt{a^2 + j^2} - \sqrt{a^2 + j^{p\,2}}\right) \tag{32}$$

where $a^p_{ij} \equiv 2\sqrt{p_i p_j Q_{ij} Q_{ji}}$ and $j^p_{ij} \equiv p_i Q_{ij} - p_j Q_{ji}$. We also require that the probability current is conserved, $\sum_j j_{ij} = 0$ for all nodes indexed by i.

For the purposes of this sensing problem, we are interested in the rate function of the density in a subset of states we call the signaling states, $q = \sum_{i \in \mathcal{S}} p_i$. In the main text, we argued that we can bound this rate function by repeated application of the contraction principle, such that

$$I(q) \le I_b(q) = \sum_{i<j} \Psi(j^*_{ij}, j^{p^*}_{ij}, a^{p^*}_{ij}). \tag{33}$$

is given by where j^* and p^* are arbitrary choices for j and p in place of evaluating the infimum.

As discussed in the main text, we are interested in the variance of the signaling density q, which is given by $1/(TI''(q^\pi))$ [4]. Therefore, we are interested in bounding the quantity $I''(q^\pi)$,

$$I''(q^\pi) \le I''_b(q^\pi) = \sum_{i<j} \left.\frac{\mathrm{d}^2 \Psi(j^*_{ij}, j^{p^*}_{ij}, a^{p^*}_{ij})}{\mathrm{d}q^2}\right|_{q=q^\pi}. \tag{34}$$

For any choice of $j^*_{ij}(q)$ and $p^*_i(q)$ that satisfy $j^*_{ij}(q^\pi) = j^\pi_{ij}$ and $p^*_i(q^\pi) = \pi_i$,[2] the second derivative of the rate function is given by

$$I''_b(q^\pi) = \sum_{i<j} \frac{1}{\phi^\pi_{ij} + \phi^\pi_{ji}} \left.\left[\frac{\mathrm{d}}{\mathrm{d}q}(j^*_{ij} - j^{p^*}_{ij})\right]^2\right|_{q=q^\pi}. \tag{35}$$

This sum can be split into the following three contributions:

$$I''(q^\pi) \le I''_{b,\vec{\mathcal{S}}}(q^\pi) + I''_{b,\vec{\mathcal{N}}}(q^\pi) + I''_{b,\vec{\mathcal{B}}}(q^\pi), \tag{36}$$

where $\vec{\mathcal{S}}$ is the set of transitions between signaling states, $\vec{\mathcal{N}}$ the transitions between non-signaling states, and $\vec{\mathcal{B}}$ the transitions from non-signaling to signaling states.

––––––

[2] This ensures that the inequality (33) is saturated at the minimum when $q = q^\pi$. Otherwise the second derivative of the bound would not necessarily be a bound on the second derivative.

Our choice of j^* must satisfy the condition $\sum_j j_{ij}^* = 0$, and our choice of p^* must satisfy the conditions $\sum_i p_i^* = 1$ and $\sum_{i \in \mathcal{S}} p_i^* = q$. We also require $j_{ij}^*(q^\pi) = j_{ij}^\pi$ and $p_i^*(q^\pi) = \pi_i$. With the benefit of hindsight, we can then choose:

$$j_{ij}^*(q) = \left[\frac{q(1-q) + q^\pi(1-q^\pi)}{2q^\pi(1-q^\pi)} \right] j_{ij}^\pi,$$

$$p_i^*(q) = \begin{cases} \frac{q}{q^\pi} \pi_i & i \in \mathcal{S}, \\ \frac{1-q}{1-q^\pi} \pi_i & i \in \mathcal{N}. \end{cases} \tag{37}$$

Defining

$$\Sigma_{\mathcal{X}}^\pi \equiv \sum_{\substack{i<j \\ (i,j) \in \mathcal{X}}} \sigma_{ij}^\pi = \sum_{\substack{i<j \\ (i,j) \in \mathcal{X}}} j_{ij}^\pi \log \frac{\phi_{ij}^\pi}{\phi_{ji}^\pi} \tag{38}$$

as the steady state energy consumption rate (in units of $k_\text{B}T$) due to transitions along edges in the sets $\mathcal{X} = \{\vec{\mathcal{S}}, \vec{\mathcal{N}}, \vec{\mathcal{B}}\}$, and

$$R^\pi \equiv \sum_{\substack{i \in \mathcal{S} \\ j \in \mathcal{N}}} \phi_{ij}^\pi = \sum_{\substack{i \in \mathcal{S} \\ j \in \mathcal{N}}} \phi_{ji}^\pi \tag{39}$$

as the flux due to transitions from the signaling states to the non-signaling states, we find

$$I_{b,\vec{\mathcal{S}}}''(q^\pi) = \sum_{i<j} \frac{\phi_{ij}^\pi - \phi_{ji}^\pi}{4\left[q^\pi(1-q^\pi)\right]^2} \tanh\left[\frac{1}{2} \log \frac{\phi_{ij}^\pi}{\phi_{ji}^\pi} \right], \tag{40}$$

where we made use of the identity $\frac{\phi_{ij}^\pi - \phi_{ji}^\pi}{\phi_{ij}^\pi + \phi_{ji}^\pi} = \tanh\left[\frac{1}{2} \log \frac{\phi_{ij}^\pi}{\phi_{ji}^\pi} \right]$. In general the inequality $\tanh\left[\frac{1}{2} \log \frac{\phi_{ij}^\pi}{\phi_{ji}^\pi} \right] \leq \frac{1}{2} \log \frac{\phi_{ij}^\pi}{\phi_{ji}^\pi}$ holds, becoming an approximate equality for receptors near equilibrium.[3] Applying this inequality to eq. (40) and using eq. (38), we find,

$$I_{b,\vec{\mathcal{S}}}''(q^\pi) \leq \frac{\Sigma_{\vec{\mathcal{S}}}^\pi}{8\left[q^\pi(1-q^\pi)\right]^2}. \tag{41}$$

By the same arguments, we also find that

$$I_{b,\vec{\mathcal{N}}}''(q^\pi) \leq \frac{\Sigma_{\vec{\mathcal{N}}}^\pi}{8\left[q^\pi(1-q^\pi)\right]^2}. \tag{42}$$

For the term contributed by transitions between the signaling and non-signaling states, we find that

$$I_{b,\vec{\mathcal{B}}}''(q^\pi) = \sum_{\substack{i \in \mathcal{S} \\ j \in \mathcal{N}}} \frac{\phi_{ij}^\pi + \phi_{ji}^\pi}{4\left[q^\pi(1-q^\pi)\right]^2} = \frac{R^\pi}{2\left[q^\pi(1-q^\pi)\right]^2}. \tag{43}$$

---

[3] When $\phi_{ij}^\pi < \phi_{ji}^\pi$ the inequality is reversed, but the factor of $\phi_{ij}^\pi - \phi_{ji}^\pi$ in eq. (40) is negative in such cases.

Plugging eqs. (41) to (43) into eq. (36), we arrive at our final bound for the second derivative of the rate function of q evaluated at q^π:

$$I''(q^\pi) \le \frac{\Sigma_{\vec{\mathcal{S}}}^\pi + \Sigma_{\vec{\mathcal{N}}}^\pi + 4R^\pi}{8\left[q^\pi(1-q^\pi)\right]^2}, \tag{44}$$

which implies that

$$I''(q^\pi) \le \frac{\Sigma_{\text{tot}}^\pi + 4R^\pi}{8\left[q^\pi(1-q^\pi)\right]^2}. \tag{45}$$

Where $\Sigma_{\text{tot}}^\pi \equiv \Sigma_{\vec{\mathcal{S}}}^\pi + \Sigma_{\vec{\mathcal{N}}}^\pi + \Sigma_{\vec{\mathcal{B}}}^\pi$. Therefore, we find that the uncertainty in q is bounded by the energy consumption and flux:

$$\text{var}(q) \ge \frac{8\left[q^\pi(1-q^\pi)\right]^2}{T\left[\Sigma_{\text{tot}}^\pi + 4R^\pi\right]}. \tag{46}$$

This is eq. (12) in the main text.

IV. COMPUTING THE RECEPTOR GAIN

Equation (8) from the main text shows that we need an expression for $\frac{dq}{d\hat{c}}$, the rate of change of the signaling density, q, with respect to the concentration estimate, $\hat{c}$. Here we present the derivation of the expression used in the main text for systems with only one non-signaling state. As discussed in the main text, this receptor gain plays a role in the estimation error of the simple observer (SO), with larger gain leading to smaller error.

Given an empirical observation of the signaling density, we can estimate the concentration by asking the question: "For what value of c would this value of q be typical?". For any value of c, the typical q is the one determined by the steady-state distribution: $q^\pi(c) = \sum_{i \in \mathcal{S}} \pi_i(c)$, with π_i varying with c via the transition rates $Q_{ij} \propto c$ for $i \in \mathcal{N}$, $j \in \mathcal{S}$. Thus, the concentration estimate, $\hat{c}(q)$ is the solution to the equation $q^\pi(\hat{c}) = q$, and therefore $\frac{dq}{d\hat{c}} = \frac{dq^\pi}{dc}\big|_{c=\hat{c}}$.

Using the result from [6] (see eq. (93), §A D) the effect of a perturbation to the rate matrix, $\mathbf{Q}$, on the steady-state distribution $\boldsymbol{\pi}$ is related to the mean first-passage times, $\overline{\mathbf{T}}$, as follows:

$$\frac{d\pi_k}{dc} = \sum_{i \neq j} \pi_i \frac{dQ_{ij}}{dc}\left(\overline{T}_{ik} - \overline{T}_{jk}\right)\pi_k, \tag{47}$$

where $\overline{T}_{ij}$ is the mean first passage time from state i to state j for $i \neq j$ and 0 for $i = j$ (see §A C).

We are interested in the gradient of $q^\pi = \sum_{k \in \mathcal{S}} \pi_k$. Furthermore, the only off-diagonal transition rates that depend on c are $Q_{ij} \propto c$ for $i \in \mathcal{N}$ and $j \in \mathcal{S}$. Therefore:

$$\frac{dq^\pi}{dc} = \frac{1}{c}\sum_{i \in \mathcal{N}}\sum_{j,k \in \mathcal{S}} \pi_i Q_{ij}\left(\overline{T}_{ik} - \overline{T}_{jk}\right)\pi_k. \tag{48}$$

From [7] (see eq. (90), §A C), we note that $\sum_j Q_{ij}\overline{T}_{jk} = \delta_{ik}/\pi_i - 1$. Then we can write

$$
\begin{aligned}
c\frac{\mathrm{d}q^\pi}{\mathrm{d}c} &= \sum_{i\in\mathcal{N}}\sum_j\sum_{k\in\mathcal{S}}\pi_i\,Q_{ij}\left(\overline{T}_{ik}-\overline{T}_{jk}\right)\pi_k - \sum_{i,j\in\mathcal{N}}\sum_{k\in\mathcal{S}}\pi_i\,Q_{ij}\left(\overline{T}_{ik}-\overline{T}_{jk}\right)\pi_k \\
&= \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{S}}\pi_i\pi_k - \sum_{i,j\in\mathcal{N}}\sum_{k\in\mathcal{S}}\pi_i\,Q_{ij}\left(\overline{T}_{ik}-\overline{T}_{jk}\right)\pi_k \\
&\equiv q^\pi\left(1-q^\pi\right) - A,
\end{aligned}
\tag{49}
$$

where A is defined by this equation.

When detailed balance is satisfied, $\pi_i Q_{ij}$ is symmetric in i,j, whereas $\left(\overline{T}_{ik}-\overline{T}_{jk}\right)$ is antisymmetric, so $A = 0$. Similarly, when there is only one non-signalling state, the sum over i and j consists of one term with $i = j$, which gives zero. More generally, $A = 0$ if there are no transitions between non-signaling states with nonzero rates and unbalanced fluxes. Therefore, all detailed balanced systems *and* all systems with only one non-signalling state have

$$
\hat{c}\frac{\mathrm{d}q}{\mathrm{d}\hat{c}} = c\frac{\mathrm{d}q^\pi}{\mathrm{d}c} = q^\pi(1-q^\pi) \qquad\Longrightarrow\qquad q^\pi(c) = \frac{1}{1+(K_d/c)} \qquad\Longrightarrow\qquad \hat{c}(q) = \frac{q\,K_d}{1-q},
\tag{50}
$$

where K_d is the dissociation constant, the concentration at which $q^\pi = \frac{1}{2}$.

V.   EXACT ERROR FORMULAE FOR THE IDEAL AND SIMPLE OBSERVERS

In this section we compute the fractional error, for the ideal observer and show that it saturates the Cramér-Rao bound (eq. (7) in the main text). We will then derive the expression for the simple observer's fractional error that we used in numerical optimization (eq. (14) and fig. 2 in the main text).

A.   Error in estimating concentration: the ideal observer

In §II B we saw that the maximum likelihood estimator could be written in terms of the empirical densities and fluxes as

$$
\hat{c} = \frac{R^T}{R^p|_{c=1}}.
$$

In §B C we see that the empirical densities and fluxes obey a large deviation principle. Therefore, the concentration estimate also obeys a large deviation principle described by the contraction:

$$
I_{\hat{c}}(\hat{c}) = \inf_{\mathbf{p},\boldsymbol{\phi}} I(\mathbf{p},\boldsymbol{\phi}) \qquad \text{subject to} \qquad \sum_i p_i = 1, \quad \sum_{(ij)\in\vec{\mathcal{B}}}\phi_{ij} = \hat{c}\sum_{(ij)\in\vec{\mathcal{B}}}\frac{p_i Q_{ij}}{c}, \quad \sum_{j\neq i}\phi_{ij} = \sum_{j\neq i}\phi_{ji} \ \ \forall i,
$$

$$
p_i \geq 0 \ \ \forall i, \quad \phi_{ij}\geq 0 \ \ \forall i\neq j.
$$

We have a constrained optimization problem for each possible value of $\hat{c}$, so we have a Lagrangian for each value of $\hat{c}$:

$$\mathcal{L}(\hat{c}) = I(p, \phi) + \alpha(\hat{c}) \left[\sum_i p_i - 1 \right] + \beta(\hat{c}) \sum_{(ij) \in \vec{\mathcal{B}}} \left[\phi_{ij} - \frac{\hat{c}}{c} p_i Q_{ij} \right] + \sum_{i \neq j} \gamma_i(\hat{c}) \left(\phi_{ij} - \phi_{ji} \right) \quad (51)$$
$$- \sum_i \mu_i(\hat{c}) p_i - \sum_{i \neq j} \nu_{ij}(\hat{c}) \phi_{ij}.$$

where α, β, γ_i are Lagrange multipliers and μ_i, ν_{ij} are Karush-Kuhn-Tucker multipliers, satisfying $\mu_i \geq 0$, and $\mu_i \partial \mathcal{L} / \partial \mu_i = 0$, the allowed region being $\partial \mathcal{L} / \partial \mu_i \leq 0$, and similar for the ν_{ij}. The Lagrange/KKT multipliers will take different values for each $\hat{c}$ as well.

The conditions for the infimum are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_i} &= \frac{\partial I}{\partial p_i} + \alpha(\hat{c}) - \frac{\beta(\hat{c})\hat{c}}{c} e_i^{\mathcal{N}} [\mathbf{Q}\mathbf{e}^{\mathcal{S}}]_i - \mu_i(\hat{c}) \quad = 0, \\
\frac{\partial \mathcal{L}}{\partial \phi_{ij}} &= \frac{\partial I}{\partial \phi_{ij}} + \beta(\hat{c}) e_i^{\mathcal{N}} e_j^{\mathcal{S}} + \gamma_i(\hat{c}) - \gamma_j(\hat{c}) - \nu_{ij}(\hat{c}) = 0,
\end{aligned} \quad (52)$$

where $\mathbf{e}^{\mathcal{S}}$ is a vector of ones for states in $\mathcal{S}$ and zero elsewhere, and $\mathbf{e}^{\mathcal{N}}$ is the reverse.[4]

To calculate the variance of $\hat{c}$ for large observation time, T, we only need the second derivative of the rate function at $\hat{c} = c$. So we Taylor expand the minimizers of eq. (51) in $(\hat{c} - c)$ as

$$p_i = \pi_i + p_i'(\hat{c}-c) + \frac{p_i''}{2}(\hat{c}-c)^2 + \mathcal{O}(\hat{c}-c)^3, \qquad \phi_{ij} = \phi_{ij}^{\pi} + \phi_{ij}'(\hat{c}-c) + \frac{\phi_{ij}''}{2}(\hat{c}-c)^2 + \mathcal{O}(\hat{c}-c)^2.$$

Because the zeroth order parts of p are nonzero, and we are only considering infinitesimal fluctuations, the inequality constraints will be loose and we can set the μ_i to zero. Some of the ϕ_{ij}^{π} could be zero, but as we shall see below, those components do not receive any corrections and we can set the ν_{ij} to zero as well.

The Taylor series expansion of $I(p, \phi)$ begins at second order:

$$I(p, \phi) = \sum_{i \neq j} \frac{\left(\phi_{ij}' - p_i' Q_{ij} \right)^2}{2\phi_{ij}^{\pi}} (\hat{c} - c)^2 + \mathcal{O}(\hat{c} - c)^3. \quad (53)$$

Therefore, expanding eq. (52) to zeroth order gives

$$\alpha(c) - \beta(c) e_i^{\mathcal{N}} [\mathbf{Q}\mathbf{e}^{\mathcal{S}}]_i = 0, \qquad \beta(c) e_i^{\mathcal{N}} e_j^{\mathcal{S}} + \gamma_i(c) - \gamma_j(c) = 0.$$

If we multiply the second equation by Q_{ij}, sum over $j \neq i$, and add the result to the first equation, we find $\mathbf{Q}\boldsymbol{\gamma} = \alpha \mathbf{e}$. The only solutions are $\alpha(c) = 0$ and $\gamma_i(c) =$ constant. The original equations then imply that $\beta(c) = 0$.

---

[4] We use the following notation: given a vector $\mathbf{v}$ and set of states $\mathcal{X}$, the vector $\mathbf{v}^{\mathcal{X}}$ has components
$$v_i^{\mathcal{X}} = \begin{cases} v_i & \text{if } i \in \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases}$$

Then the Taylor series expansion of eq. (51) is

$$\mathcal{L} = \left[\sum_{i \neq j} \frac{\left(\phi'_{ij} - p'_i Q_{ij}\right)^2}{2\phi^\pi_{ij}} + \alpha' \sum_i p'_i + \beta' \sum_{(ij) \in \vec{\mathcal{B}}} \left(\phi'_{ij} - p'_i Q_{ij} - \frac{\phi^\pi_{ij}}{c}\right) + \sum_{i \neq j} \gamma'_i \left(\phi'_{ij} - \phi'_{ji}\right)\right](\hat{c} - c)^2$$
$$+ \mathcal{O}(\hat{c} - c)^3.$$

$$(54)$$

If we minimize this expression with respect to p'_i and ϕ'_{ij}, we find

$$\sum_{j \neq i} \frac{(p'_i Q_{ij} - \phi'_{ij})Q_{ij}}{\phi^\pi_{ij}} + \alpha' - \beta' e^{\mathcal{N}}_i [\mathbf{Q} \mathbf{e}^{\mathcal{S}}]_i = 0, \qquad \frac{\phi'_{ij} - p'_i Q_{ij}}{\phi^\pi_{ij}} + \beta' e^{\mathcal{N}}_i e^{\mathcal{S}}_j + \gamma'_i - \gamma'_j = 0.$$

We can see that $\alpha' = 0$ and $\gamma'_i =$constant with the same method used for the zeroth order parts. This leaves us with $\phi'_{ij} - p'_i Q_{ij} = \beta' e^{\mathcal{N}}_i \phi^\pi_{ij} e^{\mathcal{S}}_j$. To determine β' we can look at the β' constraint in eq. (54). It shows that, to first order in $(\hat{c} - c)$, we require $R' - R^{p'} = \frac{R^\pi}{c}$ and therefore $\beta' = \frac{1}{c}$.

We can substitute these results into eq. (54) to find

$$I''_{\hat{c}}(c) = \frac{R^\pi}{c^2} \qquad \Longrightarrow \qquad \frac{\langle \delta c^2 \rangle}{c^2} = \frac{1}{N}.$$

$$(55)$$

This saturates the Cramér-Rao bound, eq. (7) in the main text.

B. Exact variance of the signaling density

First we compute the variance of the signaling density q, which is the fraction of time the receptor is bound and signalling along a single trajectory, by solving the contraction to leading order in $(q - q^\pi)$. The contraction of the rate function from empirical density, p, and current, j, to empirical signalling density, q is

$$I_q(q) = \inf_{\mathbf{p}, \mathbf{j}} I(\mathbf{p}, \mathbf{j}) \qquad \text{subject to} \qquad \sum_i p_i = 1, \quad \sum_{i \in \mathcal{S}} p_i = q, \quad \sum_j j_{ij} = 0, \quad p_i \geq 0.$$

We can find the infimum by minimizing the following Lagrangian:

$$\mathcal{L} = I(p, j) + \alpha(q) \left[\sum_i p_i - 1\right] + \beta(q) \left[\sum_{i \in \mathcal{S}} p_i - q\right] + \sum_{ij} \gamma_i(q) j_{ij} - \sum_i \mu_i(q) p_i.$$

where α, β, γ_i are Lagrange multipliers and μ_i are Karush-Kuhn-Tucker multipliers. As we have an optimization problem for each possible q, there will be different values of the Lagrange/KKT multipliers for each q as well. The contraction is then determined by

$$\frac{\partial I}{\partial p_i} = -\alpha(q) - \beta(q) e^{\mathcal{S}}_i + \mu_i(q), \qquad \frac{\partial I}{\partial j_{ij}} = \gamma_j(q) - \gamma_i(q),$$

$$(56)$$

where $\mathbf{e}^{\mathcal{S}}$ is a vector of ones for states in $\mathcal{S}$ and zero elsewhere.

We assume that at $q = q^\pi$, we have $p_i = \pi_i$ and $j_{ij} = j_{ij}^\pi = \pi_i Q_{ij} - \pi_j Q_{ji}$ (see eq. (34)). We also assume that the solution lies in the interior of the allowed region where $p_i > 0$ and $\mu_i = 0$ (for an ergodic process, all π_i are nonzero, and for infinitesimal $(q - q^\pi)$ the same will be true of p_i). From the series expansion of $I_q(q)$ about $q = q^\pi$ and eq. (35) we can see that

$$I_q(q) = \frac{(q - q^\pi)^2}{2} \sum_{i<j} \frac{1}{\phi_{ij}^\pi + \phi_{ji}^\pi} \left[\frac{\mathrm{d}}{\mathrm{d}q} (j_{ij} - j_{ij}^p) \right]_{q=q^\pi}^2 + \mathcal{O}(q - q^\pi)^3. \tag{57}$$

Therefore, we only need the expansion of the optimal p, j to first order in $(q - q^\pi)$, whose coefficients we denote by p', j'.[5] Then we can expand eq. (56) to first order to find

$$\sum_{j \neq i} \frac{Q_{ij}(j'_{ij} - j_{ij}^{p'})}{\phi_{ij}^\pi + \phi_{ji}^\pi} = \alpha' + \beta' e_i^\mathcal{S}, \qquad \frac{j'_{ij} - j_{ij}^{p'}}{\phi_{ij}^\pi + \phi_{ji}^\pi} = \gamma'_j - \gamma'_i. \tag{58}$$

The constraints on p and j ($\mathbf{pe} = 1$, $\mathbf{pe}^\mathcal{S} = q$, $\mathbf{je} = 0$) imply that

$$\mathbf{p'e} = 0, \qquad \mathbf{p'e}^\mathcal{S} = 1, \qquad \mathbf{j'e} = 0. \tag{59}$$

We can solve these equations with some tools from §A. First, we can solve for $j'_{ij} - j_{ij}^{p'}$ in the second equation of (58) and insert the result into the first equation of (58):

$$j'_{ij} - j_{ij}^{p'} = (\phi_{ij}^\pi + \phi_{ji}^\pi)(\gamma'_j - \gamma'_i) \qquad \Longrightarrow \qquad \sum_{j \neq i} Q_{ij}(\gamma'_j - \gamma'_i) = \alpha' + \beta' e_i^\mathcal{S}. \tag{60}$$

The γ'_i term supplies the missing $j = i$ term from the sum. So we can rewrite the second part of eq. (60) as

$$\mathbf{Q}\boldsymbol{\gamma'} = \alpha' \mathbf{e} + \beta' \mathbf{e}^\mathcal{S}. \tag{61}$$

If we premultiply by $\boldsymbol{\pi}$, we find that $\alpha' = -q^\pi \beta'$. If we premultiply by the Drazin pseudoinverse, $\mathbf{Q}^\mathcal{D}$ (see eq. (86), §A B), we find that $(\mathbf{I} - \mathbf{e}\boldsymbol{\pi})\boldsymbol{\gamma'} = \beta' \mathbf{Q}^\mathcal{D} \mathbf{e}^\mathcal{S}$. Looking at eq. (58), we only care about differences of the γ'_i, so we can shift γ'_i by an arbitrary constant and choose to set $\boldsymbol{\pi}\boldsymbol{\gamma'} = 0$. Then

$$\boldsymbol{\gamma'} = \beta' \mathbf{Q}^\mathcal{D} \mathbf{e}^\mathcal{S} = \beta' (\mathbf{I} - \mathbf{e}\boldsymbol{\pi}) \overline{\mathbf{T}} \boldsymbol{\Pi} \mathbf{e}^\mathcal{S}, \tag{62}$$

where $\Pi_{ij} = \pi_i \delta_{ij}$ and $\overline{T}_{ij}$ is the mean first-passage-time from state i to j (see eq. (91), §A C).

Now we go back to the first part of eq. (60) and sum over j:

$$\sum_j p'_j Q_{ji} = \sum_j (\pi_i Q_{ij} \gamma'_j + \pi_j \gamma'_j Q_{ji}),$$

---

[5] The choice made in §III, eq. (37) would give $\mathbf{p'} = \frac{\boldsymbol{\pi}^\mathcal{S}}{q^\pi} - \frac{\boldsymbol{\pi}^\mathcal{N}}{1-q^\pi}$ and $\mathbf{j'} = \frac{\mathbf{j}^\pi}{2} \left(\frac{1}{q^\pi} - \frac{1}{1-q^\pi} \right)$.

or, using the natural definition of the adjoint (see eq. (95), §A E):

$$\mathbf{p}'\mathbf{Q} = (\mathbf{Q}\boldsymbol{\gamma}')^{\dagger} + \boldsymbol{\gamma}'^{\dagger}\mathbf{Q} = \boldsymbol{\gamma}'^{\dagger}(\mathbf{Q} + \mathbf{Q}^{\dagger}).$$

Substituting in eq. (62) and postmultiplying by $\mathbf{Q}^{\mathcal{D}}$:

$$\mathbf{p}' = \beta'\boldsymbol{\pi}^{\mathcal{S}}\mathbf{Q}^{\mathcal{D}\dagger}(\mathbf{Q} + \mathbf{Q}^{\dagger})\mathbf{Q}^{\mathcal{D}} = \beta'\boldsymbol{\pi}^{\mathcal{S}}(\mathbf{Q}^{\mathcal{D}} + \mathbf{Q}^{\mathcal{D}\dagger}) = \boldsymbol{\gamma}'^{\dagger} + \bar{\boldsymbol{\gamma}}'^{\dagger}, \quad \text{or:} \quad p_i' = \pi_i(\gamma_i' + \bar{\gamma}_i') \tag{63}$$

where we defined $\bar{\boldsymbol{\gamma}}' = \beta'\mathbf{Q}^{\mathcal{D}\dagger}\mathbf{e}^{\mathcal{S}}$, i.e. the quantity $\boldsymbol{\gamma}'$ but computed for the time-reversed process. We can then determine the Lagrange multiplier β' using the normalization constraints, eq. (59):

$$\mathbf{p}'\mathbf{e}^{\mathcal{S}} = 1 \quad \Longrightarrow \quad \boldsymbol{\pi}^{\mathcal{S}}\boldsymbol{\gamma} = \boldsymbol{\pi}^{\mathcal{S}}\bar{\boldsymbol{\gamma}} = \frac{1}{2}$$

$$\Longrightarrow \quad \beta' = \frac{1}{2\boldsymbol{\pi}^{\mathcal{S}}\mathbf{Q}^{\mathcal{D}}\mathbf{e}^{\mathcal{S}}} = \frac{1}{2\sum_{ij}\left[(1 - q^{\pi})\pi_i^{\mathcal{S}} - q^{\pi}\pi_i^{\mathcal{N}}\right]\overline{T}_{ij}\pi_j^{\mathcal{S}}}. \tag{64}$$

Now we can determine $\mathbf{j}'$ using the first part of eq. (60),

$$j_{ij}' = (\bar{\gamma}_i' + \gamma_j')\phi_{ij}^{\pi} - (\gamma_i' + \bar{\gamma}_j')\phi_{ji}^{\pi},$$

although we do not actually need this quantity.

Instead, we note that eq. (57) depends only on $j_{ij}' - j_{ij}^{p'}$. By eq. (60), this can be rewritten in terms of the ϕ_{ij}^{π} and γ_i'. We can then substitute eqs. (62) and (64) into eq. (57), to find:

$$\begin{aligned} I_q''(q^{\pi}) &= \sum_{i<j}(\phi_{ij}^{\pi} + \phi_{ji}^{\pi})\left(\gamma_i' - \gamma_j'\right)^2 = \sum_{ij}\phi_{ij}^{\pi}\left(\gamma_i' - \gamma_j'\right)^2 \\ &= -2\sum_{ij}\phi_{ij}^{\pi}\gamma_i'\gamma_j' = -2\beta'^2\boldsymbol{\pi}^{\mathcal{S}}\mathbf{Q}^{\mathcal{D}\dagger}\mathbf{e}^{\mathcal{S}} = -\beta'. \end{aligned} \tag{65}$$

In going from the first to second line, we made use of the fact that $\sum_i \phi_{ij}^{\pi} = \sum_j \phi_{ij}^{\pi} = 0$ when we include the diagonal terms.

The variance in the signaling density q is given by $1/(TI''(q^{\pi}))$ [4], where T is the total observation time, so from eq. (64) we have

$$\left\langle(\delta q)^2\right\rangle = \frac{2\sum_{ij}\left[q^{\pi}\pi_i^{\mathcal{N}} - (1 - q^{\pi})\pi_i^{\mathcal{S}}\right]\overline{T}_{ij}\pi_j^{\mathcal{S}}}{T}. \tag{66}$$

We can rewrite this in terms of set-to-point mean first-passage times

$$\overline{T}_{\mathcal{X}j} = \frac{\sum_{i\in\mathcal{X}}\pi_i\overline{T}_{ij}}{\sum_{i\in\mathcal{X}}\pi_i}, \tag{67}$$

where each term is weighted by the conditional probability of being in state i conditional on being in the set $\mathcal{X}$, $\mathbb{P}(x(t) = i|x(t) \in \mathcal{X})$ for any nonspecific time t.

Then eq. (66) reads as

$$\left\langle(\delta q)^2\right\rangle = \frac{2q^{\pi}(1 - q^{\pi})}{T}\sum_{j\in\mathcal{S}}\left(\overline{T}_{\mathcal{N}j} - \overline{T}_{\mathcal{S}j}\right)\pi_j = \frac{2q^{\pi}(1 - q^{\pi})}{T}\sum_{j\in\mathcal{N}}\left(\overline{T}_{\mathcal{S}j} - \overline{T}_{\mathcal{N}j}\right)\pi_j, \tag{68}$$

where we used eq. (92), §A C, which implies that $\sum_{j \in \mathcal{N}} \overline{T}_{\mathcal{X}j} \pi_j + \sum_{j \in \mathcal{S}} \overline{T}_{\mathcal{X}j} \pi_j = \eta$, a constant independent of the initial set $\mathcal{X}$.

This expression simplifies dramatically when there is only one non-signalling state, so that the sum collapses to a single term

$$\left\langle (\delta q)^2 \right\rangle = \frac{2 q^\pi (1 - q^\pi)^2 \overline{T}_{\mathcal{S}0}}{T}.$$

We can interpret this result physically if we rewrite it as follows:

$$\left\langle (\delta q)^2 \right\rangle = \frac{2 \left[q^\pi (1 - q^\pi) \right]^2}{\overline{N} T_{\mathrm{hold}} / T_{\mathrm{unbind}}}, \qquad \text{where:} \quad T_{\mathrm{hold}} = \frac{q^\pi T}{\overline{N}}, \qquad T_{\mathrm{unbind}} = \overline{T}_{\mathcal{S}0}. \tag{69}$$

Here T_{hold} is *holding time*, the mean time spent in bound states during one bound interval. Also, when there is only one non-signaling state, the set-to-set mean first-passage time $\overline{T}_{\mathcal{S}\mathcal{N}} = \overline{T}_{\mathcal{S}0}$ so T_{unbind} is the mean time until the next unbinding event given that the receptor is currently bound.

Note that the quantity T_{unbind} is not the same as T_{hold}. In the case of T_{hold}, we would condition on the receptor having entered the bound state at the particular time, t_0, from which we measure the holding time. The states would then be weighted by $\mathbb{P}(x(t_0) = i | \text{bound at } t_0)$, the probability that the binding transition was to state i.

$$\begin{aligned} T_{\mathrm{hold}} &= \sum_{i \in \mathcal{S}} \overline{T}_{i0} \ \mathbb{P}(x(t_0) = i | \text{bound at } t_0), \\ T_{\mathrm{unbind}} &= \sum_{i \in \mathcal{S}} \overline{T}_{i0} \ \mathbb{P}(x(t) = i | x(t) \in \mathcal{S}). \end{aligned} \tag{70}$$

In eq. (67), by using the steady-state distribution we effectively average over the length of time since the last binding event, whereas if we were to calculate the holding time we would condition on it being zero. It is always the case that $q^\pi T = \overline{N} T_{\mathrm{hold}}$, and therefore:

$$\left\langle (\delta q)^2 \right\rangle = \frac{2 \left[q^\pi (1 - q^\pi) \right]^2}{\overline{N}} \frac{T_{\mathrm{unbind}}}{T_{\mathrm{hold}}}. \tag{71}$$

When looking at the definitions of T_{unbind} and T_{hold}, one might think that $T_{\mathrm{hold}} \geq T_{\mathrm{unbind}}$. This is not the case, due to the difference in the probability distribution of the initial state. We will look at an illustrative example in §V D.

C.   Exact error for the simple observer

To find the fractional error of $\hat{c}$, we note that at the minimum of the large deviation rate function:

$$\left\langle (\delta \hat{c})^2 \right\rangle = \left(T \frac{\mathrm{d}^2 I}{\mathrm{d} \hat{c}^2} \right)^{-1} = \left(T \frac{\mathrm{d}^2 I}{\mathrm{d} q^2} \left[\frac{\mathrm{d} q}{\mathrm{d} \hat{c}} \right]^2 \right)^{-1} = \frac{\left\langle (\delta q)^2 \right\rangle}{\left[\frac{\mathrm{d} q}{\mathrm{d} \hat{c}} \right]^2}.$$
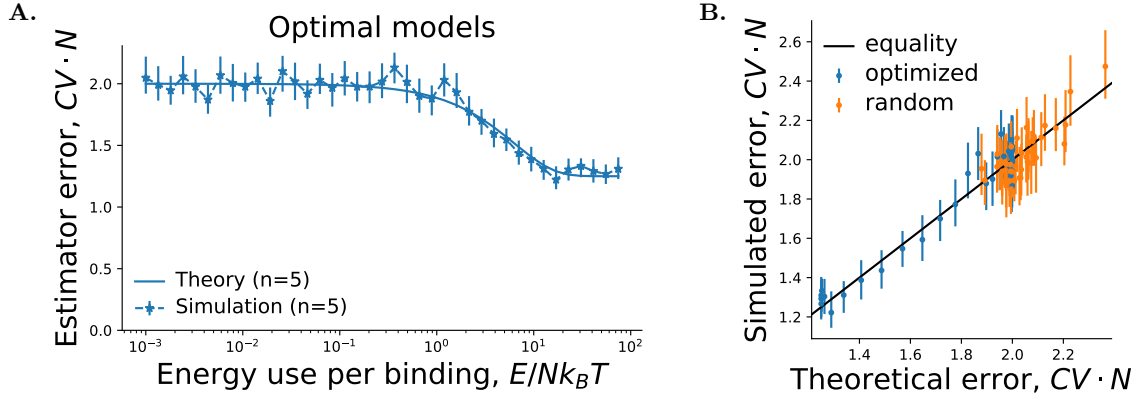
FIG. 3. Monte Carlo validation of eq. (73). **A.** Comparison of the analytic expression for fractional error with Monte Carlo simulations for models with numerically optimized CV at fixed E (see fig. 2 of the main text) and $n = 5$ states, one of which is non-signaling. **B.** Comparison of the analytic expression for fractional error, eq. (73), with Monte Carlo simulations for numerically optimized models from A. and randomly generated models. Simulations were performed with $T = 4000$ in units of each model's typical holding time, and 1600 repeats. Error bars indicate 95% confidence intervals from 100 bootstrap resamples.

With only one non-signaling state, we can use eq. (50) for the jacobian between c and q. Thus:

$$\frac{\langle (\delta \hat{c})^2 \rangle}{c^2} = \frac{2}{\overline{N}} \frac{T_{\text{unbind}}}{T_{\text{hold}}}. \tag{72}$$

This is eq. (14) in the main text. In the case of a two-state process (or one that is lumpable to a two-state process, see [8]), T_{unbind} and T_{hold} have the same distribution. When the holding time has an exponential distribution, the time until the next unbinding is independent of the time since the last binding. For such receptors, eq. (69) reduces to the Berg-Purcell result [9], $\frac{\langle (\delta \hat{c})^2 \rangle}{c^2} = \frac{2}{\overline{N}}$.

In general, we expect the fractional error to grow with the mixing time of the receptor, as the effective number of independent observations of the receptor scales $\propto T/T_{\text{mix}}$ due to autocorrelation. We would expect that, in most cases, a long unbinding time implies a long mixing time.

When there is more than one non-signaling state, using eq. (66) and the jacobian from eq. (48), the long time limit of the fractional error is:

$$\overline{N} \frac{\langle (\delta \hat{c})^2 \rangle}{c^2} = \frac{2R^{\pi} \left[\sum_{ijk} \pi_i^{\mathcal{N}} \pi_j^{\mathcal{S}} \pi_k^{\mathcal{S}} \left(\overline{T}_{ik} - \overline{T}_{jk} \right) \right]}{\left[\sum_{ijk} \phi_{ij}^{\mathcal{NS}} \pi_k^{\mathcal{S}} \left(\overline{T}_{ik} - \overline{T}_{jk} \right) \right]^2}, \tag{73}$$

where $\overline{N} = R^{\pi} T$ and $\phi_{ij}^{\mathcal{NS}}$ is ϕ_{ij}^{π} for $i \in \mathcal{N}, j \in \mathcal{S}$ and zero otherwise. This is the formula that we used in numerical optimization for fig. 2 in the main text.

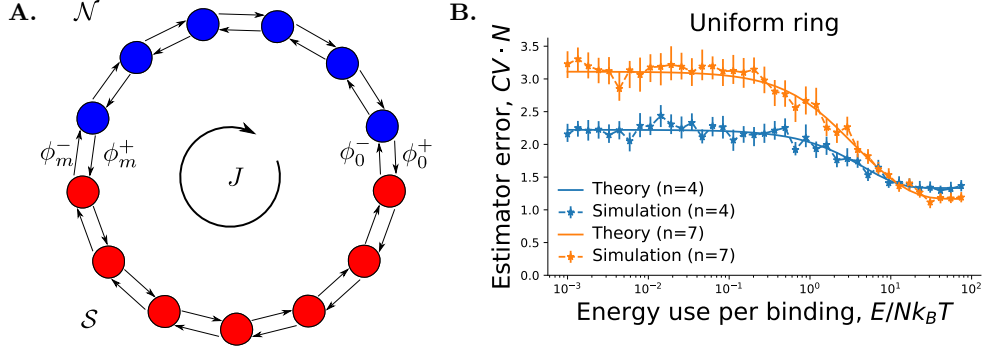We can validate eq. (73) with Monte Carlo simulations, as shown in fig. 3.

FIG. 4. Ring receptors. **A.** A receptor with a ring topology. **B.** Comparison of the analytic expression for fractional error of a uniform ring, eq. (77) with Monte Carlo simulations. Simulations were performed with $T = 4000$ in units of each model's typical holding time, and 1600 repeats. Error bars indicate 95% confidence intervals from 100 bootstrap resamples.

D. Exact first passage times and error in a uniform ring receptor

In this section we apply eq. (73), the fractional error for a general receptor, to the case of a uniform ring receptor. We consider receptors of the type depicted in fig. 4 A, but with only one non-signaling state labeled as state 0. The transition rates are given by

$$Q_{ij} = Q_+ \left(\delta_{i+1,j} - \delta_{ij} \right) + Q_- \left(\delta_{i-1j} - \delta_{ij} \right), \tag{74}$$

where the indices are to be interpreted modulo n, the total number of states. It will be convenient to parameterize these models with the energy consumed in one full circuit of the ring (in units of $k_\mathrm{B}T$): $\sigma = n \ln \left[Q_+/Q_- \right]$.

We can determine the mean first-passage-times to the nonsignaling state using the recursion relation eq. (90)

$$\mathbf{Q}\overline{\mathbf{T}} = \mathbf{\Pi}^{-1} - \mathbf{e}\mathbf{e}^\mathrm{T}, \quad \text{or} \quad Q_+(\overline{T}_{i+10} - \overline{T}_{i0}) + Q_-(\overline{T}_{i-10} - \overline{T}_{i0}) = \begin{cases} -1, & i \neq 0, \\ q^\pi/(1 - q^\pi), & i = 0, \end{cases} \tag{75}$$

whose solution is

$$\overline{T}_{i0} = \frac{1}{Q_+ - Q_-} \left[n \left(\frac{1 - \mathrm{e}^{-i\sigma/n}}{1 - \mathrm{e}^{-\sigma}} \right) - i \right].$$

Furthermore, the conditional probabilities in eq. (70) are

$$\mathbb{P}(x(t_0) = i | \text{bound at } t_0) = \begin{cases} (1 + \mathrm{e}^{-\sigma/n})^{-1}, & i = 1, \\ (1 + \mathrm{e}^{\sigma/n})^{-1}, & i = n - 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\mathbb{P}(x(t) = i | x(t) \in \mathcal{S}) = \frac{1}{n-1}, \quad i = 1, \ldots, n-1.$$

If we substitute these equations into eq. (70), we find

$$T_{\text{hold}} = \frac{n-1}{Q_+ + Q_-}, \qquad T_{\text{unbind}} = \frac{n\left(n\coth\left[\frac{\sigma}{2}\right] - \coth\left[\frac{\sigma}{2n}\right]\right)}{2(n-1)(Q_+ - Q_-)}, \tag{76}$$

Substituting these expressions into eq. (72), we find

$$\overline{N}\,\frac{\langle\langle(\delta\hat{c})^2\rangle\rangle}{c^2} = \frac{n\coth\left[\frac{\sigma}{2n}\right]\left(n\coth\left[\frac{\sigma}{2}\right] - \coth\left[\frac{\sigma}{2n}\right]\right)}{(n-1)^2}. \tag{77}$$

As $\sigma \to \pm\infty$, this expression asymptotes to $\frac{n}{n-1}$. As $\sigma \to 0$, it becomes $2 + \frac{(n-3)(n-2)}{3(n-1)}$.

With the same parametrization, the energy consumption per binding is given by

$$\frac{\Sigma^\pi}{R^\pi} = \sigma\tanh\left[\frac{\sigma}{2n}\right]. \tag{78}$$

In fig. 4 B. we have verified eq. (77) with Monte-Carlo simulations.

Looking at eq. (76) we see that for large σ, $T_{\text{hold}} > T_{\text{unbind}}$. For small σ this is reversed, $T_{\text{hold}} < T_{\text{unbind}}$. We can understand how this happen by looking at the mean first-passage-times, as in fig. 5. In each case, T_{unbind} is the arithmetic mean of the first-passage-times in fig. 5 A.

When σ is large, the ring is approximately uni-directional. The probability distribution of the state immediately after binding is concentrated at state 1. This is where the first-passage-time $\overline{T}_{i0}$ is largest, as it must go through all of the other states before reaching 0. Therefore T_{hold} is above-average and $T_{\text{hold}} > T_{\text{unbind}}$.

When σ is small, the ring is symmetric between both directions. The probability distribution of the state immediately after binding is equally concentrated in states 1 and $n-1$. This is where the first-passage-time $\overline{T}_{i0}$ is smallest, as it has a 50% chance of going straight back to 0. Therefore T_{hold} is below-average and $T_{\text{hold}} < T_{\text{unbind}}$.
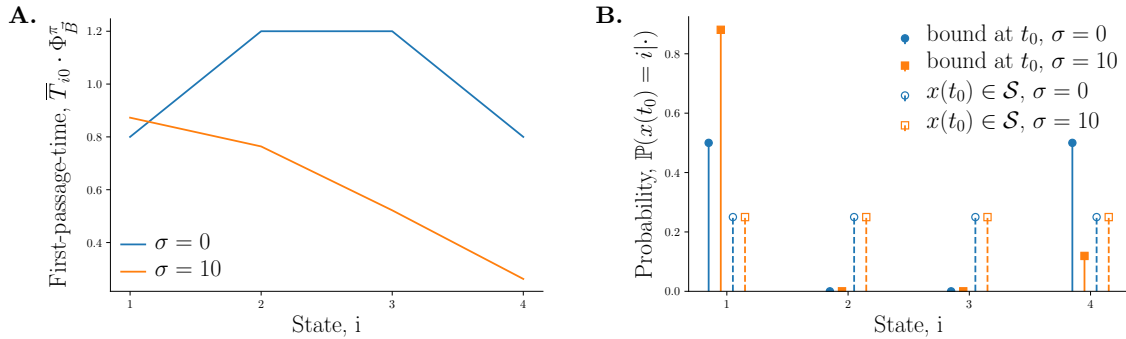

FIG. 5. First-passage-times for uniform ring objects with $n = 5$. **A.** Mean first-passage-time from each state to state 0, the only state in $\mathcal{N}$. **B.** Initial probability distribution (solid) immediately after the ligand has bound, (dashed) at a generic time after binding.

VI. NUMERICAL METHODS

Here we explain in detail how we obtained the results of Figure 2 in the main paper, which contains numerical results falling into two categories: results for optimized networks, and results of directly simulating randomly generated networks.

A. Numerical optimization of receptors

In order to validate our theoretical bounds (eq. (7) and eq. (13) in the main paper), we numerically generate networks that minimize the the exact formula (73) subject to an energy consumption constraint. The optimization problem is then:

$$\text{minimize } \overline{N}\,\frac{\langle \delta c^2 \rangle}{c^2} = \frac{2R^\pi \left[\sum_{ijk} \pi_i^\mathcal{N} \pi_j^\mathcal{S} \pi_k^\mathcal{S} \left(\overline{T}_{ik} - \overline{T}_{jk} \right) \right]}{\left[\sum_{ijk} \phi_{ij}^{\mathcal{NS}} \pi_k^\mathcal{S} \left(\overline{T}_{ik} - \overline{T}_{jk} \right) \right]^2},$$

$$\text{subject to } \frac{\Sigma^\pi}{R^\pi} = \text{constant} \tag{79}$$

Receptors of a given number of states and division between signaling and nonsignaling states were optimized using the MATLAB built-in nonlinear optimizing function `fmincon` [10]. The interior-point algorithm was used to minimize the objective function in (79) starting from randomly initialized transition rates in a complete graph. At each energy consumption constraint, the data point presented in fig. 2 is the network found that gives the minimum error out of 400 optimizations starting from different random initializations.

1. Lumpability of optimized networks

Lumpability [8] is a property of certain continuous-time Markov chains which indicates that the size of the state space can be reduced by 'lumping' together states according to a certain partitioning. A lumped state, which represents some subset of original states, will obey the same exponentially distributed holding time as the original subset. Let a continuous-time Markov chain with states $\mathcal{V}$ have a partitioning of states into n disjoint subsets $\{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n\}$. The Markov chain is *lumpable* with respect to the partitioning if the transition rates Q_{ij} from state i to state j obey the following:

$$\sum_{j \in \mathcal{A}_\ell} Q_{ij} = \sum_{j \in \mathcal{A}_\ell} Q_{kj}, \ \ \forall i, k \in \mathcal{A}_m \tag{80}$$

for any pairs of subsets in the partitioning (values of ℓ and m). Under this condition, due to the memoryless nature of the exponential distribution, the probability of transition out of a subset $\mathcal{A}_m$ is independent of the microscopic details of which state in $\mathcal{A}_m$ the system occupies. The lumped chain formed by the partitioning is

then also a Markov chain with transition rate between $\mathcal{A}_m$ and $\mathcal{A}_\ell$ given by $\sum_{j \in \mathcal{A}_\ell} Q_{ij}$ for $i \in \mathcal{A}_m$.

It is potentially interesting to consider whether the optimal networks for concentration estimation are lumpable to two state processes, along the partitioning into nonsignaling and signaling states. To measure the lumpability, we calculate the variance over the mean squared (uncertainty or CV) of the unbinding rates Q_{i0}, where 0 indicates the one nonsignaling state. If the process is perfectly lumpable, this uncertainty will be 0. For an n-state uni-directional cyclic process with uniform transition rates, the CV will be $n-2$. Figure fig. 6 shows the CV of unbinding rates for n-state Markov processes found to be optimal for concentration estimation (as in fig. 2 of the main text), as a function of energy dissipation per binding event. All processes are approximately lumpable to a two-state system until a critical dissipation level, where they separate, eventually saturating at $n-2$ as the optimal processes are all uniform rings.

B. Direct Monte-Carlo simulations of random receptors

To independently validate our calculations, we also performed direct simulations of processes found by the optimization procedure described in section VI A, as well as randomly generated networks. We find that the error associated with direct simulations of networks found during the optimization procedure agree with the analytical
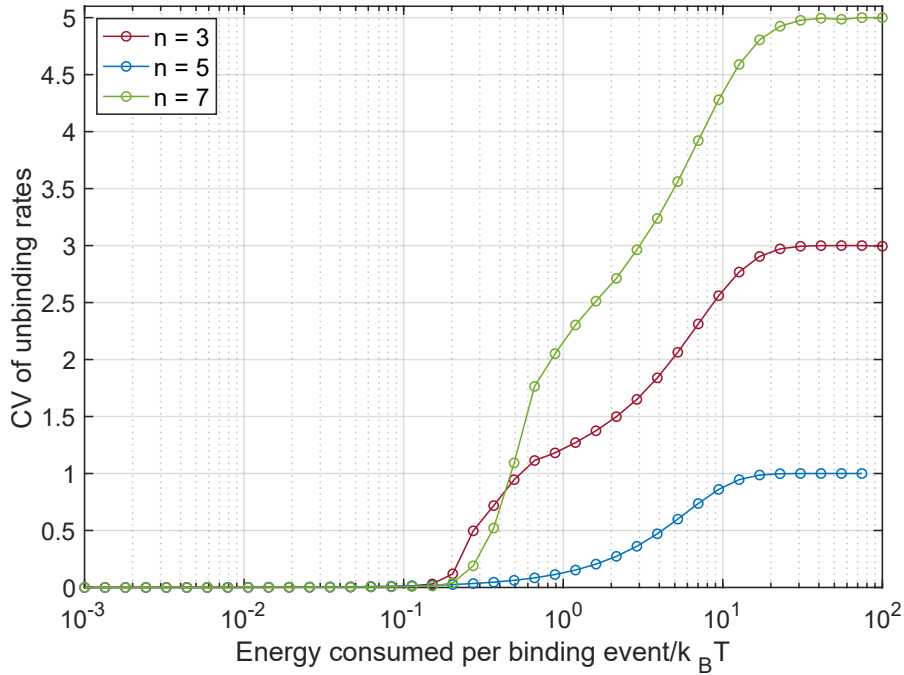

FIG. 6. CV of the unbinding rates for n-state optimized networks as a function of energy dissipation per binding event. Networks that are lumpable to a two-state system would have CV = 0.

formula (73), and the randomly generated networks lie above our derived theoretical bounds.

A given network is defined by a set of transition rates (matrix Q) and the nonsignaling/signaling state assignment. To randomly generate networks that comply with a energy consumption constraint (as in figure 2 in the main text), we use the same nonlinear optimization procedure described in the previous section, but replace the objective function $\overline{N} \frac{\langle \delta c^2 \rangle}{c^2}$ with a constant. Starting from a random initialization, the algorithm will then generate a matrix Q which complies with the nonlinear constraint $\frac{\Sigma^\pi}{R^\pi}$ = constant but without minimizing the fractional error.

Whether the transition rate matrices are randomly generated or the result of an optimization procedure, we can simulate the Markov process that it represents by generating sample trajectories of the process over some time interval $T \gg T_{\mathrm{hold}}$. The ith empirical trajectory spends time $t_{\mathcal{S}}^i$ in the signaling states, and therefore produces an empirical density in the signaling states, $q_{\mathrm{emp}}^i = \frac{t_{\mathcal{S}}^i}{T}$ and number of binding transitions, N^i. Each q_{emp}^i can be converted to a concentration estimate $\hat{c}$ using a numerical Jacobian look-up table. This look-up table is generated by scaling the binding transition rates in the matrix Q by concentrations c, while keeping all other transition rates fixed. We can then calculate the stationary distribution in the signaling states for that scaled Q matrix. The pairs of concentrations and resultant signaling state densities (c, q^π) populate a look-up table describing the Jacobian between the concentration and the signaling state density. Given an empirically observed q_{emp}^i, we can then interpolate the corresponding $\hat{c}$ from the look-up table. Therefore, given a particular matrix of transition rates Q, we generate a distribution of $\hat{c}$ observed during the simulations. From this distribution and the set of N^i we can calculate $\overline{N} \frac{\langle \delta c^2 \rangle}{c^2}$ as desired.

The error bars for the direct simulation data presented in figure 2 are generated by resampling the joint distribution of (q_{emp}, N). The upper and lower error bars then represent the 95th and 5th percentiles respectively of the distribution of the errors generated by resampling.

Appendix A   PRIMER ON CONTINUOUS-TIME MARKOV PROCESSES

In this appendix we provide a quick summary of those aspects of the theory of Markov processes in continuous time that are used in this supplement.

In the following sections we describe the transition rate matrix, its Drazin pseudoinverse, its relation to mean first-passage times, the relation between first passage times and perturbations of the steady-state distribution, and the natural definition of the inner product and adjoint for vectors on the Markov chain state-space.

A Master equation and the transition rate matrix

We consider a Markov process on a discrete set of n states indexed by $i = 1, \ldots, n$. We describe this process by a set of *transition rates* Q_{ij} denoting the probability per unit time that the system jumps to state j given that it is currently in state i. The probability that the system is in state i at time t, $p_i(t)$, evolves according to the *master equation*:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \sum_{j \neq i} \left[p_j(t)Q_{ji} - p_i(t)Q_{ij} \right]. \tag{81}$$

The master equation can be written in matrix form if we let p_i be the components of a row vector $\mathbf{p}(t)$ and we define the diagonal elements of the *transition rate matrix* as

$$Q_{ii} = -\sum_{j \neq i} Q_{ij} \equiv -\lambda_i, \qquad \frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = \mathbf{p}(t)\mathbf{Q}. \tag{82}$$

The quantity λ_i is the probability per unit time that the system jumps to any other state given that it is currently in state i. The holding time, or amount of time spent in any individual visit to state i, follows an exponential distribution with mean $1/\lambda_i$. The probability that the next state visited by the Markov process is state j, given that it is currently in state i, is given by $P_{ij} = \frac{Q_{ij}}{\lambda_i}$.

The definition of the diagonal elements in eq. (82) imply that the sum of matrix elements across any row of $\mathbf{Q}$ is zero. If we define $\mathbf{e}$ to be a column vector of ones, we can express the row sums as $\mathbf{Qe} = \mathbf{0}$.

The *steady-state distribution* $\boldsymbol{\pi}$ is the solution of $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{0}$, and thus obeys:

$$\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}, \qquad \boldsymbol{\pi}\mathbf{e} = 1. \tag{83}$$

For an ergodic process $\boldsymbol{\pi}$ is uniquely defined by eq. (83) and is strictly positive in every state.

For future use, it will be helpful to define diagonal matrices, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$, with

$$\Lambda_{ij} = \lambda_i \delta_{ij}, \qquad \Pi_{ij} = \pi_i \delta_{ij}. \tag{84}$$

Then the matrix of next-state probabilities, P_{ij}, can be written as:

$$P_{ij} = \mathbb{P}(x_{r+1} = j | x_r = i) = \begin{cases} 0, & i = j, \\ \frac{Q_{ij}}{\sum_k Q_{ik}}, & \text{otherwise,} \end{cases} \qquad \text{or} \qquad \mathbf{P} = \mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{Q}. \tag{85}$$

B Drazin pseudoinverse

The transition rate matrix $\mathbf{Q}$ of an ergodic Markov process has a one dimensional null-space (because $\boldsymbol{\pi}\mathbf{Q}$ and $\mathbf{Qe}$ are both zero). Therefore the rate matrix is not

invertible. However there are several ways of defining a pseudoinverse [11]. The most useful one for our purposes is the *Drazin pseudoinverse* of $\mathbf{Q}$, defined by

$$\mathbf{Q}^{\mathcal{D}} = \tau \mathbf{e}\boldsymbol{\pi} - \left(\frac{\mathbf{e}\boldsymbol{\pi}}{\tau} - \mathbf{Q}\right)^{-1}, \qquad \mathbf{Q}^{\mathcal{D}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathcal{D}} = \mathbf{I} - \mathbf{e}\boldsymbol{\pi}, \tag{86}$$

where τ is an arbitrary timescale. The Drazin pseudoinverse, $\mathbf{Q}^{\mathcal{D}}$, has the same left and right eigenvectors and null spaces as $\mathbf{Q}$, but with nonzero eigenvalues inverted. In particular, $\mathbf{Q}^{\mathcal{D}}\mathbf{e} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{Q}^{\mathcal{D}} = \mathbf{0}$.

C   Mean first-passage times

We define the *mean first-passage time*, T_{ij}, as the mean time it takes the process to reach state j for the first time, starting from state i. The diagonal elements, T_{ii}, are defined to be the mean time it takes the process to leave and then return to state i. It will be convenient to additively decompose the mean first-passage-time matrix into its diagonal and off-diagonal parts: $\mathbf{T} = \mathbf{T}^{\mathrm{dg}} + \overline{\mathbf{T}}$.

To compute the mean first passage times, consider the first time the process leaves state i. On average, this will take time λ_i^{-1}. With probability P_{ij}, it will go directly to j, so the conditional mean time would be λ_i^{-1}. On the other hand, if it goes to some other state, k, with probability P_{ik}, the conditional mean time would be $\lambda_i^{-1} + T_{kj}$. Combining these, we get the recursion relation

$$\begin{aligned}
T_{ij} &= \sum_{k \neq j} P_{ik}(\lambda_i^{-1} + T_{kj}) + P_{ij}\lambda_i^{-1} \\
&= \sum_{k \neq j} P_{ik}T_{kj} + \sum_{k} P_{ik}\lambda_i^{-1} \\
&= \sum_{k} P_{ik}\overline{T}_{kj} + \lambda_i^{-1}, \qquad \text{or} \quad \mathbf{T} = \mathbf{P}\overline{\mathbf{T}} + \boldsymbol{\Lambda}^{-1}\mathbf{e}\mathbf{e}^{\mathrm{T}},
\end{aligned} \tag{87}$$

where $\mathbf{e}\mathbf{e}^{\mathrm{T}}$ is the matrix of all ones. Remembering eq. (85) (that $\mathbf{P} = \mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{Q}$), we can write eq. (87) as

$$\boldsymbol{\Lambda}\mathbf{T}^{\mathrm{dg}} = \mathbf{Q}\overline{\mathbf{T}} + \mathbf{e}\mathbf{e}^{\mathrm{T}}. \tag{88}$$

The recurrence times are given by

$$T_{ii}^{\mathrm{dg}} = \frac{1}{\pi_i \lambda_i}. \tag{89}$$

This can be proved by pre-multiplying eq. (88) by $\boldsymbol{\pi}$ and employing $\boldsymbol{\pi}\mathbf{Q} = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$

We can substitute eq. (89) into eq. (88) to get a recursion relation for the off diagonal part (see [7]):

$$\mathbf{Q}\overline{\mathbf{T}} = \boldsymbol{\Pi}^{-1} - \mathbf{e}\mathbf{e}^{\mathrm{T}}. \tag{90}$$

Because we require that $\overline{\mathbf{T}}$ is zero on the diagonal, and the only null vector of $\mathbf{Q}$ is nonzero everywhere, eq. (90) has a unique solution given by

$$\overline{T}_{ij} = \frac{Q_{ij}^{\mathcal{D}} - Q_{jj}^{\mathcal{D}}}{\pi_j}. \tag{91}$$

This equation can also be written as $\mathbf{Q}^{\mathcal{D}} = (\mathbf{I} - \mathbf{e}\boldsymbol{\pi})\overline{\mathbf{T}}\boldsymbol{\Pi}$.

This expression for $\overline{\mathbf{T}}$ leads to *Kemeney's constant* η given by [8]:

$$\eta = \sum_j \overline{T}_{ij}\pi_j. \tag{92}$$

That η is indeed a constant reflects the remarkable fact that the quantity $\sum_j \overline{T}_{ij}\pi_j$ is actually *independent* of the initial state i. If we substitute eq. (91) in, we find that $\eta = -\operatorname{Tr}\mathbf{Q}^{\mathcal{D}}$.

D Perturbations of the steady state distribution

Suppose the transition rate matrix $\mathbf{Q}$ is a function of some parameter α. By eq. (83), $\boldsymbol{\pi}$ will also be a function of α. If α is changed by a small amount, $\boldsymbol{\pi}$ will also change. This change can be expressed in terms of first-passage-times, as shown in [6]

$$\frac{\mathrm{d}\pi_k}{\mathrm{d}\alpha} = \sum_{i \neq j} \pi_i \frac{\mathrm{d}Q_{ij}}{\mathrm{d}\alpha}\left(\overline{T}_{ik} - \overline{T}_{jk}\right). \tag{93}$$

This result can be proved by expanding $\frac{\mathrm{d}}{\mathrm{d}\alpha}(\boldsymbol{\pi}\mathbf{Q}) = 0$, postmultiplying by $\mathbf{Q}^{\mathcal{D}}$ and using the identities from appendices A B and A C. Note that the summand vanishes for $i = j$, so we could drop the restriction $i \neq j$ from the range of the sum.

E Inner products and adjoints

It is useful to define a natural inner product and associated norm on the space of functions over Markov chain states. To motivate this, it is useful to first consider inner products of functions over infinite or continuous state spaces. The constant function, corresponding in the discrete case to the vector of all ones, $\mathbf{e}$, plays such a fundamental role that it is important that its norm, $\|\mathbf{e}\|$, be finite. In order to achieve any such finite norm for a constant function over an infinite space, one requires a distribution against which to integrate the function, or compute inner products.

Returning from continuous state-spaces to discrete state-spaces, functions over continuous space correspond to column vectors over discrete states and distributions over continuous spaces correspond to row vectors over discrete states. In the context of Markov processes, a natural such distribution is the steady-state distribution corresponding to the row vector $\boldsymbol{\pi}$. We thus define the *inner product* $\langle \mathbf{u}, \mathbf{v} \rangle$ over a pair of column vectors $\mathbf{u}$ and $\mathbf{v}$ using the natural distribution $\boldsymbol{\pi}$:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_i \pi_i u_i^* v_i = (\mathbf{u}^*)^{\mathrm{T}}\boldsymbol{\Pi}\mathbf{v}, \tag{94}$$

where u_i^* is the complex conjugate of u_i and $\mathbf{\Pi} = \operatorname{diag}(\boldsymbol{\pi})$. This inner product defines an associated norm $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ and under this norm the constant function $\mathbf{e}$ has norm $\|\mathbf{e}\| = 1$.

The *adjoints* $(\cdot)^\dagger$ of column vectors $\mathbf{u}$, row vectors $\boldsymbol{\xi}$ and operators $\mathbf{M}$ are defined by

$$\mathbf{u}^\dagger \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle \quad \forall \mathbf{v}, \qquad \left\langle \boldsymbol{\xi}^\dagger, \mathbf{u} \right\rangle = \boldsymbol{\xi}\mathbf{u} \quad \forall \mathbf{u}, \qquad \left\langle \mathbf{M}^\dagger \mathbf{u}, \mathbf{v} \right\rangle = \langle \mathbf{u}, \mathbf{M}\mathbf{v} \rangle \quad \forall \mathbf{u}, \mathbf{v}$$

$$\left(\mathbf{u}^\dagger\right)_i = \pi_i u_i^*, \qquad\qquad \left(\boldsymbol{\xi}^\dagger\right)_i = \frac{\xi_i^*}{\pi_i}, \qquad\qquad \left(\mathbf{M}^\dagger\right)_{ij} = \frac{\pi_j M_{ji}^*}{\pi_i}. \tag{95}$$

Note that $\mathbf{e}^\dagger = \boldsymbol{\pi}, \boldsymbol{\pi}^\dagger = \mathbf{e}$ and the adjoint of a transition matrix is its time-reversal, which we next explain. In discrete time, Bayes rule states that the probability of the previous state given the current state is

$$\mathbb{P}(x_r | x_{r+1}) = \frac{\mathbb{P}(x_{r+1} | x_r)\, \mathbb{P}(x_r)}{\mathbb{P}(x_{r+1})}.$$

For a system in its steady state $\mathbb{P}(x_{r+1} = i) = \mathbb{P}(x_r = i) = \pi_i$. If the transition probabilities are given by $\mathbb{P}(x_{r+1} = j | x_r = i) = M_{ij}$, the time-reversed process obtained via Bayes rule then has transition probabilities $M_{ji}^\dagger$, defined in eq. (95).

The equivalent statement in continuous time follows from $\exp(\mathbf{Q}t)^\dagger = \exp(\mathbf{Q}^\dagger t)$. This can be seen by noticing that this adjoint obeys the usual product rule $(\mathbf{AB})^\dagger = \mathbf{B}^\dagger \mathbf{A}^\dagger$, implying that $(\mathbf{A}^n)^\dagger = (\mathbf{A}^\dagger)^n$, and computing the matrix exponential from its Taylor series.

One can then show that a reversible process (one that satisfies detailed balance and has zero net currents in its steady-state) has a transition matrix that is self-adjoint under eq. (95), and therefore has real eigenvalues with eigenvectors that are orthogonal under the inner product in eq. (94).

Appendix B   PRIMER ON LARGE DEVIATION THEORY FOR MARKOV PROCESSES

Here, for the convenience of the general physicist reader, we outline the derivation of the level 2.5 rate function that is used as a starting point in the main text and §III of the supplement, at a physical level of rigor. Much more elaborate proofs can be found in more mathematical works [5, 12–14], which take great care in dealing with subtle issues regarding the existence, uniqueness and convexity of large deviation rate functions. In our exposition below, we simply present the essential steps and physical intuition, without delving into these mathematical subtleties. We hope this provides a straightforward introduction to the large deviation theory of Markov processes for the general physicist. Below, we first derive the large deviation rate function for empirical densities and fluxes in §B C. Then in §B D we use the contraction principle to find the rate function for empirical densities and currents.

A Empirical densities, fluxes, and currents

Given a continuous time ergodic Markov process with transition rates Q_{ij} and unique stationary distribution π_i, we can imagine observing a particular realization of a trajectory $x(t)$ through a sequence of states $x_0, x_1, \ldots, x_m$ with corresponding transition times $t_0, t_1, \ldots, t_m$. Each $x_r \in \{1, \ldots, n\}$ denotes one of n possible occupied states of the Markov process. We can also describe the trajectory in terms of the sequence of states and the holding time in each state, $\tau_r = t_{r+1} - t_r$. This collection of states and holding times $\{x, \tau\}$ defines a trajectory, or path $x(t)$ of the Markov process. The empirical density for state i that we would observe over the course of this path is defined (as in the main text) as

$$p_i^T \equiv \frac{1}{T} \int_0^T \mathrm{d}t \, \delta_{x(t)\,i} = \frac{\sum_r \delta_{x_r i} \tau_r}{T}, \tag{96}$$

where T is the duration of the trajectory. Thus p_i^T is simply the fraction of time spent in state i along the trajectory $x(t)$. Similarly, the empirical flux ϕ_{ij}^T is defined as the empirically measured rate at which transitions from state i to state j occur over the duration T of the trajectory $x(t)$, which we can write as

$$\phi_{ij}^T = \frac{1}{T} \int_0^T \mathrm{d}t \, \delta_{x(t^-)i} \, \delta_{x(t^+)j}, = \frac{\sum_r \delta_{x_r i} \delta_{x_{r+1} j}}{T}, \tag{97}$$

where $x(t^-)$ and $x(t^+)$ are the states before and after a transition. The empirical *current* from state i to j is defined as

$$j_{ij}^T = \phi_{ij}^T - \phi_{ji}^T, \tag{98}$$

i.e. the *net* empirical rate of transitions from i to j minus j to i.

All of these observables p_i^T, ϕ_{ij}^T, and j_{ij}^T can be measured along a *single* trajectory of the Markov process, and will approach stationary values as the trajectory duration $T \to \infty$. The empirical density will converge to the stationary distribution, $p_i^T \to \pi_i$, the empirical flux will converge to the steady state flux, $\phi_{ij}^T \to \phi_{ij}^\pi$, where $\phi_{ij}^\pi = \pi_i Q_{ij}$, and the empirical current will converge to the steady state current, $j_{ij}^T \to j_{ij}^\pi$, where $j_{ij}^\pi = \pi_i Q_{ij} - \pi_j Q_{ji}$.

We would now like to find the joint distribution $\mathbb{P}(p^T, \phi^T)$ over empirical densities p^T and empirical fluxes ϕ^T, from which we can find the joint distribution $\mathbb{P}(p^T, j^T)$ over empirical densities p^T and empirical currents j^T. This will enable us to study the fluctuations and large deviations of these observables around their most likely values of π, ϕ^π, and j^π in the large T limit. In this limit, these joint distributions are well characterized by a large deviation rate functions $I(p^T, \phi^T)$ and $I(p^T, j^T)$, defined implicitly by the relations

$$\mathbb{P}\left(p^T, \phi^T\right) \approx \mathrm{e}^{-TI(p^T, \phi^T)}, \qquad \mathbb{P}\left(p^T, j^T\right) \approx \mathrm{e}^{-TI(p^T, j^T)}. \tag{99}$$

For an ergodic process Q, $I(p^T, \phi^T)$ is expected to be a non-negative convex function of p^T and ϕ^T that achieves a global minimum value of 0 at the unique point $p_i^T = \pi_i$

and $\phi_{ij}^T = \phi_{ij}^\pi = \pi_i Q_{ij}$ [12]. Similarly, $I(p^T, j^T)$ is expected to be a non-negative convex function of p^T and j^T that achieves a global minimum value of 0 at the unique point $p_i^T = \pi_i$ and $j_{ij}^T = j_{ij}^\pi = \pi_i Q_{ij} - \pi_j Q_{jj}$ [5, 13]. Thus eq. (99) implies that the probability of any large deviation in the empirical density p_i^T, empirical flux ϕ_{ij}^T, and empirical current j_{ij}^T, from their respective, typical stationary values π_i, $\pi_i Q_{ij}$, and $\pi_i Q_{ij} - \pi_j Q_{ji}$, is exponentially suppressed in the duration T of the trajectory. We present these rate functions in eqs. (109) and (111) below. Readers who are simply interested in the form of this function, but not the ideas underlying its derivation, can safely skip the remainder of this appendix.

To derive the rate functions in eq. (99), we follow the approach of [12]. The essential idea is to use a common tilting method from large deviation theory. This method involves comparing the probability of a particular path under our given true Markov process with the probability of that same path under a perturbed Markov process in which transition rates are *tilted* to make a particular large deviation more likely. In particular, we compare the probability of the observed trajectory $x(t)$ under the given Markov process with transition rates Q_{ij}, with the probability of the same trajectory under a fictitious process with tilted transition rates $\widehat{Q}_{ij}$. This fictitious process is the one that would produce the observed trajectory $x(t)$ as a 'typical' realization. That is, it is a Markov process with stationary distribution $\widehat{\pi}_i = p_i^T$ and transition rates $\widehat{Q}_{ij} = \phi_{ij}^T / \widehat{\pi}_i$. In essence, empirically densities p_i^T and fluxes ϕ_{ij}^T that might be rarely observed under Q_{ij} are typical under $\widehat{Q}_{ij}$.

B   Trajectory probabilities

Now, the probability of a particular trajectory $x(t)$ under the original Markov process Q_{ij} is

$$\mathbb{P}\left[x(t)\right] = \mathbb{P}\left(\{x, \tau\}\right) = \mathbb{P}\left(\tau_m \geq T - \sum_{r=0}^{m-1} \tau_r \middle| x_m\right)\left[\prod_{r=0}^{m-1} \mathbb{P}(x_{r+1}|x_r)\,\mathbb{P}(\tau_r|x_r)\right]\mathbb{P}\left(x_0\right), \quad (100)$$

with

$$\mathbb{P}(x_{r+1}|x_r) = \frac{Q_{x_r x_{r+1}}}{\sum_j Q_{x_r j}} \qquad\qquad (101)$$

and

$$\mathbb{P}(\tau_r|x_r) = \lambda_{x_r} e^{-\lambda_{x_r}\tau_r}, \qquad\qquad (102)$$

where $\lambda_{x_r} = \sum_{j \neq x_r} Q_{x_r j}$ (see eq. (84) in §A A). Equation (101) is equivalent to eq. (85) which follows from the definition of the transition rates of a continuous time Markov process, and eq. (102) is the exponential distribution with parameter λ_{x_r} describing the holding time in each state.

For large T the boundary effects at $t = 0, T$ will be negligible. We will neglect those factors in eq. (100) from here on. The joint probability for the trajectory $x(t)$ is then given by

$$\mathbb{P}\left(\{x, \tau\}\right) = \prod_r \mathbb{P}(x_{r+1} | x_r) \, \mathbb{P}(\tau_r | x_r) \; = \; \prod_r e^{-\lambda_{x_r} \tau_r} \, Q_{x_r \, x_{r+1}}$$

$$= \exp\left(-\sum_{r=0}^{m} \lambda_{x_r} \tau_r + \sum_{r=0}^{m-1} \log Q_{x_r x_{r+1}}\right). \tag{103}$$

We can now split up the sums over states and transitions indexed by r into two contributions. For the first term in the sum in eq. (103), we will perform an inner sum over all instances in the trajectory in which the Markov process is in state i, which corresponds to summing over r such that $x_r = i$, and then we will perform an outer sum over all of states i of the Markov process. Similarly, we will break the second term in eq. (103) into an inner sum over all of the transitions from state i to state j in a trajectory (i.e. transitions in which $x_r = i, x_{r+1} = j$), and then an outer sum over all pairs of states i and j in the Markov process. These groupings of sums yield

$$\mathbb{P}\left(\{x, \tau\}\right) = \exp\left(-\sum_i \sum_{r:x_r=i} \lambda_{x_r} \tau_r + \sum_{i \neq j} \sum_{\substack{r:x_r=i \\ x_{r+1}=j}} \log Q_{x_r x_{r+1}}\right). \tag{104}$$

We can now simplify this expression, recognizing that the sum of the occupancy times of state i during this trajectory is $\sum_{r:x_r=i} \tau_r = T p_i^T$, and the total number of transitions from state i to j is $\sum_{\substack{r:x_r=i \\ x_{r+1}=j}} 1 = T \phi_{ij}^T$:

$$\mathbb{P}\left(\{x, \tau\}\right) = \exp\left(-\sum_i \lambda_i \sum_{r:x_r=i} \tau_r + \sum_{i \neq j} \log Q_{ij} \sum_{\substack{r:x_r=i \\ x_{r+1}=j}} 1\right)$$

$$= \exp\left(-\sum_i \lambda_i \, T p_i^T + \sum_{i \neq j} T \phi_{ij}^T \log Q_{ij}\right) \tag{105}$$

$$= \exp\left(-T \sum_{i \neq j} \left[p_i^T Q_{ij} - \phi_{ij}^T \log Q_{ij}\right]\right).$$

Thus the probability density assigned to any *individual* trajectory $x(t)$ of duration T depends on that trajectory *only* through two types of observables: the empirical densities p_i^T and empirical fluxes ϕ_{ij}^T. This dramatic simplification of the distribution over trajectories singles out empirical densities and fluxes as uniquely important order parameters in Markovian non-equilibrium processes.

C   Rate function for densities and fluxes

Next, to go from a probability distribution over trajectories $x(t)$ to a joint distribution over empirical densities and fluxes (p^T, ϕ^T), we would need to integrate over

all possible trajectories $x(t)$ that produce any given empirical density and flux pair (p^T, ϕ^T). This direct integration would result in adding a difficult to compute entropic contribution to eq. (105). Instead, we compute the ratio of probabilities for the *same* path under two different processes, Q and the fictitious process with tilted rates $\widehat{Q}$:

$$\frac{\mathbb{P}\left(\{x,\tau\}\right)}{\widehat{\mathbb{P}}\left(\{x,\tau\}\right)} = \exp\left(-T\sum_{i\neq j}\left[p_i^T\left[Q_{ij} - \widehat{Q}_{ij}\right] + \phi_{ij}^T\log\left(\frac{\widehat{Q}_{ij}}{Q_{ij}}\right)\right]\right). \tag{106}$$

Noting that this ratio depends on the trajectory $x(t)$ only though the observables (p^T, ϕ^T), we can find a computable relation between the distribution $\mathbb{P}\left(p^T, \phi^T\right)$ of these observables under the process Q, and the distribution $\widehat{\mathbb{P}}\left(p^T, \phi^T\right)$ of these observables under the tilted process $\widehat{Q}$:

$$\begin{aligned}
\mathbb{P}\left(p^T, \phi^T\right) &= \int\limits_{\{x,\tau\}\to(p^T,\phi^T)} \{\mathrm{d}x,\mathrm{d}\tau\}\,\mathbb{P}\left(\{x,\tau\}\right) = \int\limits_{\{x,\tau\}\to(p^T,\phi^T)} \{\mathrm{d}x,\mathrm{d}\tau\}\,\widehat{\mathbb{P}}\left(\{x,\tau\}\right)\frac{\mathbb{P}\left(\{x,\tau\}\right)}{\widehat{\mathbb{P}}\left(\{x,\tau\}\right)} \\
&= \frac{\mathbb{P}\left(\{x,\tau\}\right)}{\widehat{\mathbb{P}}\left(\{x,\tau\}\right)} \int\limits_{\{x,\tau\}\to(p^T,\phi^T)} \{\mathrm{d}x,\mathrm{d}\tau\}\,\widehat{\mathbb{P}}\left(\{x,\tau\}\right) = \frac{\mathbb{P}\left(\{x,\tau\}\right)}{\widehat{\mathbb{P}}\left(\{x,\tau\}\right)}\,\widehat{\mathbb{P}}\left(p^T, \phi^T\right),
\end{aligned} \tag{107}$$

where $x(t) \to (p^T, \phi^T)$ indicates integration over the set of trajectories that lead to a given empirical density and flux.

For large T, the distributions $\mathbb{P}\left(p^T, \phi^T\right)$ and $\widehat{\mathbb{P}}\left(p^T, \phi^T\right)$ are both characterized by their large deviation rate functions $I(p^T, \phi^T)$ and $\widehat{I}(p^T, \phi^T)$, respectively. Thus eqs. (106) and (107) yield a relation between these two rate functions:

$$I(p^T, \phi^T) = \widehat{I}(p^T, \phi^T) + \sum_{i\neq j}\left[p_i^T\left[Q_{ij} - \widehat{Q}_{ij}\right] + \phi_{ij}^T\log\left(\frac{\widehat{Q}_{ij}}{Q_{ij}}\right)\right]. \tag{108}$$

So, if we knew the rate function for the fictitious process $\widehat{Q}$, we could find the rate function for Q.

Now we note that we have constructed the fictitious process $\widehat{Q}$ *specifically* so that its rate function evaluated at (p^T, ϕ^T) is exceedingly simple. Indeed we have chosen the transition rates

$$\widehat{Q}_{ij} = \frac{\phi_{ij}^T}{p_i^T}$$

in order to make the empirically observed densities and fluxes (p^T, ϕ^T) typical. Thus, because the observed (p^T, ϕ^T) are typical under $\widehat{Q}$, we must have $\widehat{I}(p^T, \phi^T) = 0$, so that the probability of these observed values is not exponentially suppressed in T for large T under $\widehat{Q}$. This implies that eq. (108) simplifies to yield the sought after expression for the large deviation rate function for joint densities and fluxes under the true process Q,

$$I(p^T, \phi^T) = \sum_{i\neq j}\left[\phi_{ij}^p - \phi_{ij}^T + \phi_{ij}^T\log\left(\frac{\phi_{ij}^T}{\phi_{ij}^p}\right)\right], \tag{109}$$

where we have written $\phi_{ij}^p = p_i^T Q_{ij}$.

We note that our notation singles out three types of fluxes: (a) a completely empirical flux ϕ_{ij}^T computed via eq. (97) along a single trajectory of duration T under the process Q; (b) a mixed empirical/true flux $\phi_{ij}^p = p_i^T Q_{ij}$ computed via the empirical density p_i^T along a trajectory via eq. (96), but multiplied by the *true* transition rates Q_{ij}; (c) the true stationary fluxes of the process Q which can be written as $\phi_{ij}^\pi = \pi_i Q_{ij}$. As expected, the large deviation rate function in eq. (109) is 0 if and only if the empirically observed density p_i^T equals the equilibrium distribution π_i of Q for all states i, and the empirically observed fluxes ϕ_{ij}^T equal the stationary fluxes $\pi_i Q_{ij}$ for all pairs of transitions from i to j. In this situation, all three fluxes are the same: $\phi_{ij}^T = \phi_{ij}^p = \phi_{ij}^\pi = \pi_i Q_{ij}$. However, according to eq. (99) and eq. (109), the probability of any discrepancy between any pair of these three fluxes is exponential suppressed in time T, for large T.

D Rate function for densities and currents

The empirical *current* from state i to j is defined as $j_{ij}^T = \phi_{ij}^T - \phi_{ji}^T$, i.e. the net empirical rate of transitions from i to j minus j to i. To find the joint large deviation rate function for empirical density p_i^T and current j_{ij}^T, we can apply the contraction principle [4] to the joint rate function for the density and flux in eq. (109):

$$I(p^T, j^T) = \inf_{\phi^T} I(p^T, \phi^T) \quad \text{subject to:} \ \phi_{ij}^T - \phi_{ji}^T = j_{ij}^T \text{ and } \phi_{ij}^T \geq 0. \qquad (110)$$

The infimum can be found by writing ϕ_{ij}^T and ϕ_{ji}^T in terms of j_{ij}^T and f_{ij}^T, where we define $f_{ij}^T \equiv \phi_{ij}^T + \phi_{ji}^T$. We can then minimize with respect to f_{ij}^T subject only to $f_{ij}^T \geq |j_{ij}^T|$. The infimum is achieved at $f_{ij}^T = \sqrt{(j_{ij}^T)^2 + (a_{ij}^p)^2}$, where $a_{ij}^p = 2\sqrt{\phi_{ij}^p \phi_{ji}^p}$. If we substitute the flux back into eq. (109) we find

$$
\begin{aligned}
I(p^T, j^T) &= \sum_{i<j} \left[\frac{j_{ij}}{2} \log\left(\frac{\left[f_{ij}^T + j_{ij}^T \right]\left[f_{ij}^p - j_{ij}^p \right]}{\left[f_{ij}^T - j_{ij}^T \right]\left[f_{ij}^p + j_{ij}^p \right]} \right) - \left(f_{ij}^T - f_{ij}^p \right) \right] \\
&= \sum_{i<j} \left[j_{ij} \left(\operatorname{arcsinh} \frac{j_{ij}^T}{a_{ij}^p} - \operatorname{arcsinh} \frac{j_{ij}^p}{a_{ij}^p} \right) - \left(\sqrt{j_{ij}^{T\,2} + a_{ij}^{p\,2}} - \sqrt{j_{ij}^{p\,2} + a_{ij}^{p\,2}} \right) \right].
\end{aligned}
\qquad (111)
$$

This is the rate function for large deviations of the empirical density and current, eq. (9) in the main text.

[1] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, 1945) p. 575.
[2] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. **37**, 81 (1945).
[3] J. B. S. Haldane and S. M. Smith, The sampling distribution of a maximum-likelihood estimate, Biometrika **43**, 96 (1956).

[4] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. **478**, 1 (2009), arXiv:0804.0327.

[5] A. C. Barato and U. Seifert, Dispersion for two classes of random variables: General theory and application to inference of an external ligand concentration by a cell, Phys. Rev. E **92**, 032127 (2015).

[6] G. Cho and C. Meyer, Markov chain sensitivity measured by mean first passage times, Linear Algebra Appl. **316**, 21 (2000).

[7] D. D. Yao, First-Passage-Time Moments of Markov Processes, J. Appl. Probab. **22**, 939 (1985).

[8] J. G. Kemeny and J. L. Snell, *Finite markov chains* (Springer, 1960) p. 226.

[9] H. C. Berg and E. M. Purcell, Physics of chemoreception, Biophys. J. **20**, 193 (1977).

[10] *MATLAB version 9.3.0.713579 (R2017b)*, The Mathworks, Inc., Natick, Massachusetts (2017).

[11] J. Hunter, A survey of generalized inverses and their use in stochastic modelling, Res. Lett. Inf. Math. Sci. **1**, 25 (2000).

[12] C. Maes and K. Netočný, Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states, Europhys. Lett. **82**, 30003 (2008).

[13] L. Bertini, A. Faggionato, and D. Gabrielli, Flows, currents, and cycles for Markov Chains: large deviation asymptotics, (2014), arXiv:1408.5477.

[14] L. Bertini, A. Faggionato, and D. Gabrielli, Large deviations of the empirical flow for continuous time Markov chains, Ann. I. H. Poincaré - Pr. **51**, 867 (2015).