
Supplementary Materials for: An Information Theoretic Framework for Eukaryotic Gradient Sensing

Joseph M. Kimmel* and **Richard M. Salter†**
joekimmel@uchicago.edu, rms@cs.oberlin.edu
Computer Science Program
Oberlin College
Oberlin, Ohio 44074

Peter J. Thomas‡
peter.j.thomas@case.edu
Departments of Mathematics, Biology and Cognitive Science
Case Western Reserve University
Cleveland, Ohio 44106

In this supplement we detail the derivations of analytical formulae given in the main text, we exhibit an additional figure excluded from the main text due to space limitations, and we provide pseudocode detailing the direction and entropy estimation algorithms.

1 Methods: supplemental

1.1 Analysis

1.1.1 Mutual information of the receptors

Let the random variables $\{B_j\}_{j=1}^N$ represent the states of the N cAMP receptors on the cell surface; $B_j = 1$ if the receptor is bound to a molecule of cAMP, otherwise $B_j = 0$. Each receptor is taken to lie on the surface of a sphere of radius $R = 7.5\mu\text{m}$, with displacement from the center of the cell in a direction given by a unit vector $\vec{x} \in \mathbb{S}^2$. We set the center of the sphere to be the origin of our coordinate system. Invoking the assumption of gradient linearity, we take the equilibrium concentration of cAMP at \vec{x} to be $c(\vec{x}|\vec{g}) = a + b(\vec{x} \cdot \vec{g})$ where $\vec{g} \in \mathbb{S}^2$ is a unit vector in the direction of the gradient. The parameter a is the mean concentration over the cell surface, and $b = R|\vec{\nabla}c|$ is half the drop in concentration from one extreme on the cell surface to the other. Before the stimulus occurs, the gradient direction is undefined.

At equilibrium, given a particular gradient direction \vec{g} , the probability of finding the receptor at \vec{x}_j in the bound state is

$$P(B_j = 1|\vec{g}) = h(c(\vec{x}_j|\vec{g})) \triangleq \frac{c(\vec{x}_j|\vec{g})}{c(\vec{x}_j|\vec{g}) + K_{\text{eq}}},$$

where $K_{\text{eq}} = k_{\text{off}}/k_{\text{on}}$ is the equilibrium constant for the ligand-receptor binding interaction, and $h(c) = c/(c + K_{\text{eq}})$ is a first-order Hill function with half-saturation at $c = K_{\text{eq}}$.

*Current address: Computational Neuroscience Graduate Program, The University of Chicago.

†Oberlin Center for Computation and Modeling, <http://occam.oberlin.edu/>.

‡To whom correspondence should be addressed. <http://www.case.edu/artsci/math/thomas/thomas.html>;
Oberlin College Research Associate.

Letting $\Phi[p]$ denote the binary channel entropy function,

$$\Phi[p] = \begin{cases} -(p \log_2(p) + (1-p) \log_2(1-p)), & 0 < p < 1 \\ 0, & p = 0 \text{ or } 1 \end{cases},$$

we can identify the uncertainty in the state of the j^{th} receptor, located at \vec{x}_j , as $\Phi[h(c(\vec{x}_j|\vec{g}))]$. Invoking independence and summing over all N receptors, the total entropy of the receptor states $\{B_j\}$ approaches an integral over the sphere's surface as N becomes large:

$$H[\{B_j\}|\vec{g}] \sim N \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \Phi \left[\frac{a + b \cos(\theta)}{a + b \cos(\theta) + K_{\text{eq}}} \right] \frac{\sin(\theta)}{4\pi} d\phi d\theta \quad (\text{as } N \rightarrow \infty). \quad (1)$$

To estimate the mutual information at the receptors we wish to compare the entropy of the receptor states when the stimulus direction \vec{g} has been specified, with the entropy when it remains indeterminate. This does *not* mean we compare the entropy of receptor states before and after stimulus onset. Indeed, if the stimulus onset is accompanied by an increase in mean concentration (as is often the case experimentally) the entropy of the receptors may initially be quite low (*e.g.*, almost all receptors may be in the unbound state) compared with the entropy after the applied gradient has approached equilibrium. Instead we must compare the receptor activity under an imposed gradient with the *ensemble* of receptor activities when all gradients are considered equally likely. With the gradient indeterminate (\vec{g} is uniformly distributed over \mathbb{S}^2) the probability of a given receptor being bound is

$$P(B_j) = \int_{\vec{g} \in \mathbb{S}^2} P(B_j|\vec{g})P(\vec{g}) d\Omega = \int_{\vec{g} \in \mathbb{S}^2} h(c(\vec{x}|\vec{g})) \frac{1}{4\pi} d\Omega,$$

where $d\Omega$ represents the surface area element on the sphere, $d\Omega = \sin(\theta)d\theta d\phi$. By rotational symmetry, when the gradient direction is indeterminate all receptors have the same *a priori* probability of being bound or not. Hence with respect to the ensemble of possible gradient directions, the uncertainty in the state of the receptors is:

$$\begin{aligned} H[\{B_j\}] &\sim N\Phi[P(B_j)] \quad (\text{as } N \rightarrow \infty) \\ &\sim N\Phi \left[\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \left(\frac{a + b \cos(\theta)}{a + b \cos(\theta) + K_{\text{eq}}} \right) \frac{\sin(\theta)}{4\pi} d\phi d\theta \right]. \end{aligned} \quad (2)$$

In both equation (1) and (2), the argument of Φ is a probability taking values $0 \leq p \leq 1$. In (1) the values of Φ are averaged over the sphere; in (2) Φ is evaluated after averaging probabilities. Because $\Phi[p]$ is convex for $0 \leq p \leq 1$, the integral in equation 1 cannot exceed that in equation 2. Therefore the mutual information

$$MI[\{B_j\}; \vec{g}] \triangleq H[\{B_j\}] - H[\{B_j\}|\vec{g}] \geq 0,$$

as expected upon receiving the signal. The analytic solution for equation (1) involves the polylogarithm function. For the parameters shown in the simulation ($a = 1.078$ nMol, $b = .512$ nMol, $K_{\text{eq}} = 25$ nMol), the mutual information with 980 receptors is 2.16 bits. As one would expect, the mutual information peaks when the mean concentration is close to the K_{eq} of the receptor, exceeding 16 bits when $a = 25$, $b = 12.5$ and $K_{\text{eq}} = 25$ (nMol).

1.1.2 Dimension reduction

As an alternative to the approximations necessary when treating the mutual information of the 980-dimensional set of receptor states and the 3-dimensional gradient direction $\vec{g} \in \mathbb{R}^3$, we construct a directional estimator \hat{g} from the spatiotemporal pattern of receptor binding events. Let $\{\vec{x}_j\}_{j=1}^N$ be the spatial locations of the N receptors on the cell's surface. With each vector we associate a weight w_j . Whenever the j^{th} receptor binds a cAMP molecule, w_j is incremented by one; otherwise w_j decays with time constant τ . We construct an instantaneous estimate of the gradient direction from the linear combination of receptor positions

$$\hat{g}_\tau(t) = \sum_{j=1}^N w_j(t) \vec{x}_j.$$

This procedure may be thought of as reflecting the release and accumulation of intracellular second messengers from the cell membrane, and their dissipation or recycling on a given time scale. We examine the evolution of the ensemble of directional estimates for time scales ranging over $20 \leq \tau \leq 2000$ (msec).

Before the stimulus is applied, the weighted directional estimates \hat{g} are small in absolute magnitude, with direction uniformly distributed on \mathbb{S}^2 . In order to determine the information gained as the estimate vector evolves after stimulus application, we wish to determine the change in entropy in an ensemble of such estimates. The uniform distribution on \mathbb{S}^2 has entropy $H_0 = \log_2(4\pi) \approx 3.6515$ bits. As the cell gains information about the direction of the gradient signal from its receptors, the entropy of the estimate should decrease, leading to a rise in mutual information.

By repeating multiple runs ($M = 600$) of the simulation we obtain samples from the ensemble of direction estimates, given a particular stimulus direction, \vec{g} . When a probability distribution on \mathbb{R} is not known exactly, its entropy may be efficiently approximated from a set of samples of the distribution. In the method of Kozachenko and Leonenko [1], adapted for the analysis of neural spike train data by Victor [2] (“KLV method”), the cumulative distribution function is approximated directly from the observed samples, and the entropy is estimated *via* a change of variables transformation (see below). This method may be formulated in vector spaces \mathbb{R}^d for $d > 1$ ([1]), but it is not guaranteed to be unbiased in the multivariate case [3] and has not been extended to curved manifolds such as the sphere. In the present case however, we may exploit the symmetries inherent in the model (rotational invariance and axial equivariance assumptions (see main text)) to reduce the empirical entropy estimation problem to one dimension.

Suppose that a probability distribution on \mathbb{S}^2 , $P(\theta, \phi)$, depends only on θ . For example, if we select a true gradient direction \vec{g} , we may parametrize the sphere with θ representing the colatitude with respect to \vec{g} , and ϕ representing the longitude. Then by axial symmetry (see Assumptions in main text) the resulting expected receptor activities would depend only on θ . The area element on the sphere in this coordinate system is $d\Omega = (\sin(\theta) d\phi)(d\theta)$, and the marginal distribution in θ is by definition

$$P_\theta(\theta) \triangleq \int_{\phi=0}^{2\pi} P(\theta, \phi) \sin(\theta) d\phi.$$

The assumption of axial symmetry is therefore tantamount to setting $P(\theta, \phi) = \frac{P_\theta(\theta)}{2\pi \sin(\theta)}$. The entropy of $P(\theta, \phi)$ is minus the expected value of the log likelihood:

$$H = - \int_{\mathbb{S}^2} P \log_2(P) d\Omega = -\mathbf{E}_\theta \left[\log_2 \left(\frac{P_\theta(\theta)}{2\pi \sin(\theta)} \right) \right], \quad (3)$$

where $\mathbf{E}_\theta[f] = \int_{\theta=0}^{\pi} f(\theta) P_\theta(\theta) d\theta$ is expectation with respect to the marginal P_θ .

Given a set of M observed samples from a distribution $P(\theta, \phi)$ which we know *a priori* to be symmetric with respect to ϕ , we use the observed colatitudes $\{\theta_1, \dots, \theta_M\}$ to estimate the expectations in equation 3. Following the KLV method, we approximate $\mathbf{E}_\theta \log_2(P_\theta)$ *via* the invertible change in variables: $y(\theta) \triangleq \int_{t=0}^{\theta} P_\theta(t) dt$. Because $dy = P_\theta(\theta) d\theta$, the entropy integral becomes

$$H = - \int_{y=0}^1 \log_2 \left(\frac{P_\theta(\theta)}{2\pi \sin(\theta)} \right) dy \sim - \frac{1}{M} \sum_{k=1}^M \log_2 \left(\frac{P_\theta(\theta_k)}{2\pi \sin(\theta_k)} \right) \text{ (as } M \rightarrow \infty \text{)}. \quad (4)$$

This approximation amounts to replacing the true cumulative distribution function $y(\theta)$ with the piecewise constant cumulative distribution given by the samples $\{y(\theta_k) = k/M\}$. Next, we sort the sample data into a monotonic sequence and approximate $P_\theta(\theta_k)$ by relating the probability density near θ_k to the expected value of the nearest neighbor distance, $\lambda_k \triangleq \min(|\theta_k - \theta_{k\pm 1}|)$ [2]. The resulting estimate for the entropy from the sample data is:

$$H \sim \frac{1}{M} \sum_{k=1}^M \left(\log_2(\lambda_k) + \log_2(2(M-1)) + \frac{\gamma}{\log_e(2)} + \log_2(2\pi) + \log_2(\sin(\theta_k)) \right) \text{ (as } M \rightarrow \infty \text{)} \quad (5)$$

where γ is the Euler-Mascheroni constant. As shown in Figure 2, this approximation agrees with the analytic result for the uniform distribution, $H_{\text{unif}} = \log_2(4\pi) \approx 3.651$.

2 Results: supplemental

See Supplemental Figure 1.

3 Direction and entropy estimation algorithms

3.1 Direction estimation

The algorithm we use for creating an estimated direction vector at each timestep is to first calculate for each receptor a `Weights` vector that is 0 before the receptor binds, 1 at the timestep when the receptor binds, and decays exponentially after a binding. The weight is cumulative, so that a receptor can have a weight above 1. After assembling the weights we then assemble a `Directions` vector for the run which is created by taking a weighted average across all receptor locations. Note that we do not normalize our weighted average, leaving it simply as a weighted sum. The weights decay at a rate given by $\text{rate} = 1/\tau$.

```
foreach Receptor R
  create a zero vector, W, of length timesteps
  foreach binding event B
    W[B.time] += 1
    foreach timestep t, from B.time+1 to timesteps
      W[t] *= (1-rate*dt)

create a zero vector of Points in  $\mathbb{R}^3$ , D, of length timesteps
foreach timestep t
  foreach Receptor R
    D[t].x += R.W[t] * R.location.x-component
    D[t].y += R.W[t] * R.location.y-component
    D[t].z += R.W[t] * R.location.z-component
```

3.2 Entropy estimation

To compute the entropy, first, for each run, we construct the `Direction` vector as outlined above, then construct from that an `Angles` vector giving the angle between the estimate at each time t from the correct direction. Then, for each timestep, take the collection of angles from all runs, and apply our adaptation of the method presented by Victor, Kozachenko, and Leonenko (see main text; “KLV method”).

```
create a matrix M with height num_runs and width num_timeSteps
foreach simulation run
  create the direction vector, d, as described in §3.1
  fill in the corresponding row in M with  $\text{acos}(\cos(d, [-1, 0, 0]))$ 
To Estimate Entropy for each timestep
  take from M the corresponding column, C
  initialize a count of the entropy, Ent = 0
  sort C so that each entry will be next to its nearest neighbor
  foreach entry in C,  $\theta$ , at index  $i$ 
    Dist =  $\min(\text{abs}(\theta - C[i+1]), \text{abs}(\theta - C[i-1]))$ 
    Ent +=  $\log_2(\text{Dist})$ 
    Ent +=  $\log_2(\sin \theta)$ 
  Ent = Ent * length(C) +  $\log_2(2\pi) + \log_2(2 \times (\text{length}(C)-1)) + \frac{\gamma}{\log_e(2)}$ 
```

As in the main text, γ is the Euler-Mascheroni constant. Algorithm implemented in C++ using the GNU STL.

References

- [1] L.F. Kozachenko and N.N. Leonenko. *Probl. Peredachi Inf. [Probl. Inf. Transm.]*, 23(9):95, 1987.

- [2] Jonathan D. Victor. Binless strategies for estimation of information from neural data. *Physical Review E*, 66:051903, Nov 11 2002.
- [3] Marc M. Van Hulle. Edgeworth approximation of multivariate differential entropy. *Neural Computation*, 17:1903–1910, 2005.
- [4] Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. John Wiley & Sons, West Sussex, England, 2000.

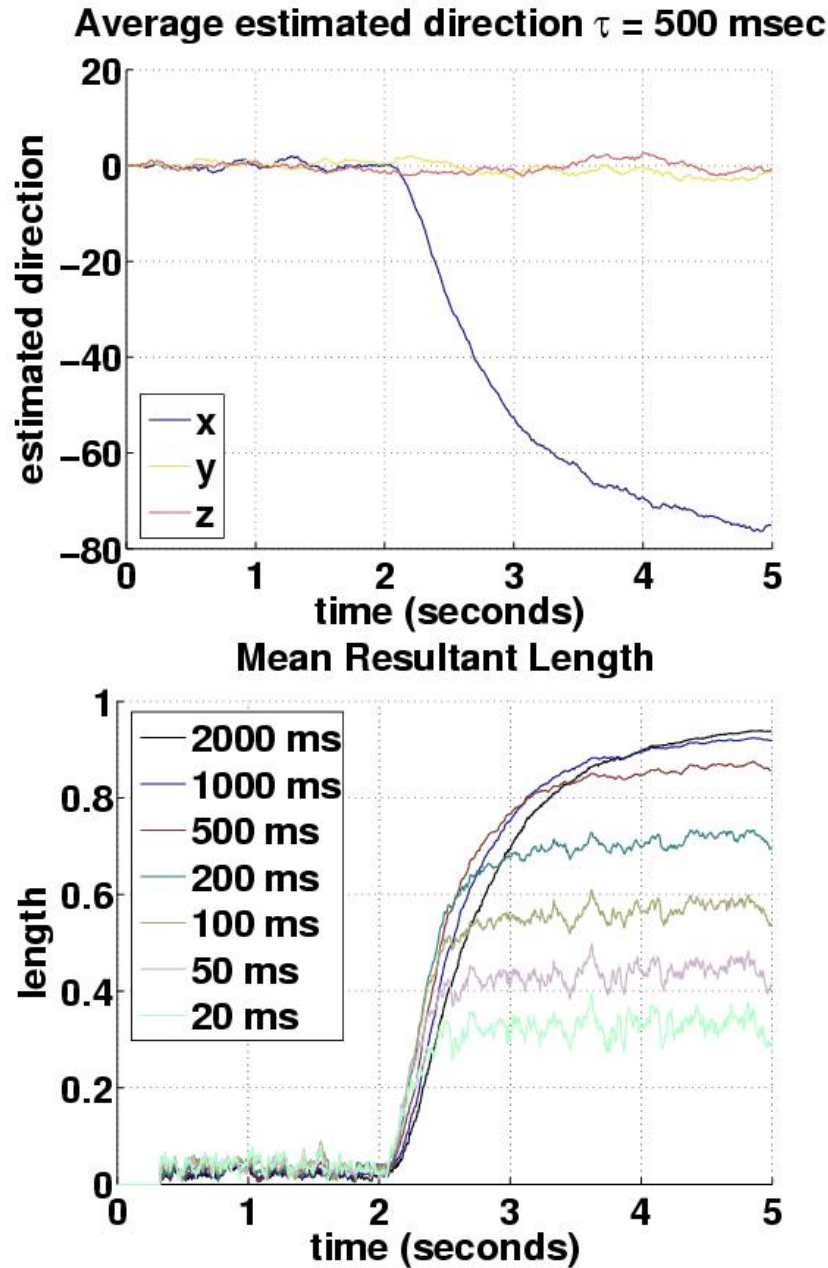


Figure 1: Population timecourse of the directional estimate vector.

Top: Average of $M = 600$ directional estimate vectors \hat{g}_τ , $\tau = 500$ msec. The average length is close to zero when the cAMP signal is uniformly distributed; after the stimulus is applied the average estimate grows in the $-x$ direction (descending trace) but not in the y or z directions, indicating that \hat{g}_τ provides an unbiased estimate of the gradient direction. Similar results obtain for each value of τ tested.

Bottom: Mean resultant length [4] versus time. The mean resultant length is the average of the unit vectors $\hat{g}_\tau/|\hat{g}_\tau|$ obtained by normalizing the directional estimates \hat{g}_τ , and has length bounded between zero (when \hat{g} are uniformly distributed) and one (when \hat{g} are highly clustered in direction). The timecourse of the mean resultant length mirrors that of the mutual information.