Surprisingly high fidelity information processing in folic acid chemotaxis of Dictyostelium amoebae: Supplementary Information

Igor Segota, Surin Mong, Eitan Neidich, Archana Rachakonda, Catherine J. Lussenhop and Carl Franck
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853 (USA)

METHODS

Information measures

Shannon’s information theory frees the data analysis from being tied to any particular model (as an example of successful applications see e.g. [1] and [2]) – and in this case, from any particular details of signal transduction pathways, but still provides quantifiable relationships between inputs and outputs. The relevant quantities in information theory are defined as follows [3]. The information entropy of a random variable $X$, is measured in bits defined as $H(X) = -\int p(x) \log_2 p(x) dx$ (a definite integral defined over the entire range where $X$ is defined). It is a measure of “sharpness” of probability distribution $p(x)$; a perfectly sharp probability distribution has entropy zero, whereas a perfectly flat, uniform distribution gives the highest possible value for entropy $H(X)$. An alternative interpretation of information entropy is the number of bits or the amount of information required to describe the random variable $X$. Sharp probability distributions require fewer bits for their full description than flat probability distributions. Intuitively, for the former only a few values near the peak can be sufficient to describe most of the outcomes of $X$, while for the latter we need more information to achieve the same. For conditional probability distributions, the conditional entropy is measured in bits defined as $H(X|Y) = -\int dy p(y) \int dx p(x|y) \log_2 p(x|y)$. This measures how sharp $p(x|y)$ is, when averaged over all possible values of $y$. For some values $y$, $p(x|y)$ may be sharp, for some other values of $y$, $p(x|y)$ may not be so sharp, and the conditional entropy tells us on average what is the sharpness, when averaged over all possible $y$. The average gain in information about $x$, given $y$, is the difference between the two, called mutual information $I(X,Y) = H(X) - H(X|Y)$. This measure describes the increase in knowledge about $X$ after we have been given some value $y$, and then averaged over all possible $y$. In other words, $I(X,Y)$ describes how much on average $p(x|y)$ is sharper, when compared to $p(x)$. The sharper the probability distribution becomes, the more information we have acquired about a random variable $X$.

Application to gradient sensing

In this case, the sensing process essential to eukaryotic chemotaxis is depicted in Fig. 1a. Here we consider three random but conditionally dependent variables, the gradient direction $\theta_{\text{grad}}$, the receptor occupancy $\theta_{\text{rec}}$, and the cell response directions $\theta_{\text{res}}$. These variables are assumed to form a Markov chain (see Fig.1b), where the cell response is conditionally dependent on the distribution of occupied receptors; i.e. given the distribution of occupied receptors, the cell response is completely independent of the original direction of the gradient that caused this particular receptor occupancy. Due to noise, the same receptor occupancy distribution can occur for gradients pointing in different directions. Without any prior knowledge we assume the gradient is equally likely to be pointing in any direction. We will see how much information we can obtain about the gradient by either observing the cell response and by calculating the distribution of receptor occupancy, and then comparing the two gains. The mutual information $I_{\text{tot}}(\theta_{\text{grad}}, \theta_{\text{res}}) = H(\theta_{\text{grad}}) - H(\theta_{\text{grad}}|\theta_{\text{res}})$ quantifies the total amount of information cells gained about the gradient (or by how much the entropy of $\theta_{\text{grad}}$ is reduced); this is determined by observing their response (see Fig.1c). Therefore, $I_{\text{tot}}$ is the gain in information that includes all possible noise sources in the FA signal transduction sequence.
pathway.

In addition, the (external) mutual information (see Fig.1c) between the gradient direction and receptor occupancy

$$I_{ext}(\theta_{grad}, \theta_{rec}) = H(\theta_{grad}) - H(\theta_{grad}|\theta_{rec})$$

tells us the information gained about the gradient by knowing the distribution of receptors occupied with FA. Authors in [4] formulated a theory for computing this quantity and gave an analytical result applicable for shallow gradients. The assumptions behind this theory are: i) the steady state of the receptor-ligand binding process, ii) the first part of the Markov chain model shown in Fig.1b (receptor probability distribution is affected only by the local gradient), iii) cells of perfectly circular shapes and iv) uniform receptor distribution. While we have no direct way of confirming the plausible assumptions i) and ii) when sensing FA, *Dictyostelium* do have circular shapes and the distribution of FA receptors was previously measured as uniform [5]. This theory gives predictions for the external mutual information $I_{ext}$ using only two biochemical constants – the dissociation constant $K_d$ between FA and its receptor and the total receptor number per cell, $N$. Both have been measured previously and multiple receptor types/states have been discovered as is also the case for cAMP receptors (see main text for discussion). The dissociation constant and the total receptor number per cell, as well as the experimentally fixed FA concentration and its gradient in our devices are sufficient to predict the external mutual information $I_{ext}$. $I_{ext}$ provides the upper limit for the amount of information that can be acquired ($I_{tot}$), due to the data processing inequality: $I_{tot} \leq I_{ext}$ [7]. In other words, any kind of data processing can only destroy information. If the two quantities are roughly similar $I_{tot} \approx I_{ext}$, then the gain in information about $\theta_{grad}$ is about the same for both cases and the majority of the noise in the entire process comes from receptor-ligand binding events.

**Cell growth and preparation**

Cells of the well characterized axenic strain, AX4 (provided by Dictyostelium Stock Center, Northwestern University), were grown in shaken culture suspension at 150 RPM in Formedium HL5 (Formedium, Hunstanton, UK) with glucose culture medium up to the concentration of about 0.5 - 3 × 10^6 cells. Development Buffer (DB; DictyBase recipe: 5 mM Na₂HPO₄, 5 mM KH₂PO₄, 1 mM CaCl₂, 2 mM MgCl₂; pH 6.5) was chosen as the medium for FA chemotaxis experiments because it is a well-defined medium and is an approximation of a physiological environment due to its low ionic strength [8]. A negative aspect of using DB is cell starvation and progression into development after 6+ hours (depending on cell density) and eventual loss of FA chemotactic sensitivity [9]. This was circumvented by performing the experiment before the starvation response occurs, as indicated by cell morphology – cells still had circular shapes. Since it was shown that the HL5 medium already contains about 0.12 mg/l of FA [10] (~0.3 µM), the medium was diluted by factor $\geq 30,000 \times$, lowering the background FA concentration in the medium to at most 0.01 nM. This corresponds to about 1 molecule of FA per volume size of a Dictyostelium cell (100 µm³). Depending on the cell concentration, 1-5 ml of cell suspension was taken from the shaken culture and DB was added for a total volume of 10 ml (dilution $\geq 2 \times$). The cell suspension was then centrifuged for 40 seconds at 1000 RPM (200 g force), 9.8 ml of supernatant was removed, and 9.8 ml of DB was added to again have the final volume of 10 ml (dilution 50×); this was repeated once more (another dilution of 50×). 9.8 ml of supernatant was removed again and finally, 0.2 ml of 1µm diameter colloidal particles at concentration 10⁶ particles/ml (Polysciences, Inc.) in DB and 1-5 ml of DB was added, depending on the starting cell concentration (dilution $\geq 6 \times$). The colloidal particles allowed us to monitor unintended convection that could ruin the static gradient. The entire procedure took about 20-30 minutes after which the cells were immediately loaded into the microfluidic device with an already established gradient.

**Microfluidics design**

The microfluidic device was designed as an agarose gel containing 3 channels [11]: the static middle channel and two flowing side channels, that represent fixed boundary conditions, were separated by a layer of agarose gel and the gradient was formed by waiting for diffusion of FA to reach a steady state (see Fig.2 and Fig.3). Reservoirs were connected via Teflon tubing and the steady flow was supplied by a Harvard PHD 2000 syringe pump. The time to reach the steady state was checked by running a 2D diffusion simulation in COMSOL Multiphysics 3.5 (COMSOL, www.comsol.com) and analyzing the gradient in the middle of the channel (Fig.3). The microfluidic channel containing Dictyostelium cells, also contained 1µm-sized colloidal particles. These were used to monitor the flow rate in the static channel and the measured Peclet number $Lv/D$ (dimensionless number characterizing the ratio of advective versus diffusive transport) was always below 0.3, where $L$ is the channel height (250 µm), $D$ the diffusion constant of folic acid 194 µm²/s [12] and $v$ the measured average drift velocity of colloidal particles (0.04 to 0.23 µm/s). After loading the cells, the gradient in the middle channel was temporarily lost, however, the time-scale of diffusive refilling of that channel from the bulk of the material above is estimated to be only $t \sim L^2/D \approx 5$ minutes, an insignificant duration.

**Device preparation**

The 3% agarose gel was formed as follows. 0.300g of agarose was mixed with 10 ml of DB. The agarose mixture was heated and kept near the boiling point in a microwave oven for 40 seconds total. Agarose was molded by pouring the heated mixture over an inverted PDMS master, which was itself molded from an original Teflon master produced by conventional milling. After about 2 minutes the agarose solidified, the holes were punched and the chamber was secured between a plexiglas manifold and a glass microscope slide. In this experiment 3% agarose serves as an environment permeable to small molecules, such as water and folic acid, but not permeable to Dictyostelium. Dictyostelium are
FIG. 2. A schematic of the microfluidic device used here.
migrating naturally attached on the glass surface, with 250 µm of static liquid (DB+FA gradient) on top and around them. The agarose gel was sealed well enough that the cells were unable to crawl underneath it.

**Cell recording**

For each run, at t=0 hours: the gradient formation was started. At t=3 hours: the cells were loaded in the device. Since we noticed that cells were not very mobile when first introduced into the device, we allowed them to adjust to the new environment for about 3.3 hours to establish a good degree of mobility. At t=6.3 hours recording started. At t=9.3 hours: the recording stopped. This time was chosen based on the fact that this is the time when one would first observe morphological changes associated with cell-to-cell cAMP signaling during the starvation response (e.g. elongated cells and formation of streams) when the cell density was significantly (10x) higher. Cell motion was recorded using bright field time-lapse optical microscopy, using an Olympus IX71 inverted microscope and a Home Science Tools MI-DC5000 5.0 Megapixel camera. Snapshots were taken every 30 seconds and cell trajectories were later analyzed on a computer. The list of concentrations used in both channels is shown in Table I.

![Graph showing diffusion through agarose](image)

**FIG. 3.** Numerical 2D time-dependent simulation of diffusion through the agarose based microfluidic device used in this work, indicating a steady state gradient in the middle chamber. The top graph shows a FA concentration at the center of the channel 5 hours after the gradient started forming, in units of the FA concentration in the left channel, \(c_{\text{max}}\) (a slice through the middle of bottom figure). The bottom figure shows a concentration profile intensity of FA at the time of recording, 5 hours after the gradient started forming. Note the steady state has not been formed in the entire device, but only in the middle chamber.

**TABLE I.** List of experimentally used concentrations in the two channels of a microfluidic device, \(c_{\text{low}}\) and \(c_{\text{high}}\) with calculated gradient \(dc/dx\) and the mean concentration \(c_0\).

<table>
<thead>
<tr>
<th>combination</th>
<th>(dc/dx) (nM/µm)</th>
<th>(c_0) (nM)</th>
<th>(c_{\text{high}}) (nM)</th>
<th>(c_{\text{low}}) (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(3.20 \times 10^6)</td>
<td>(5.0 \times 10^4)</td>
<td>(1.00 \times 10^5)</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>(6.4 \times 10^6)</td>
<td>(1.0 \times 10^4)</td>
<td>(2.00 \times 10^4)</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>(3.2 \times 10^6)</td>
<td>(5.0 \times 10^4)</td>
<td>(1.00 \times 10^5)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>(1.6 \times 10^6)</td>
<td>(2.5 \times 10^4)</td>
<td>(5.00 \times 10^4)</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>(6.4 \times 10^{-1})</td>
<td>(1.0 \times 10^3)</td>
<td>(2.00 \times 10^4)</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>(3.2 \times 10^{-1})</td>
<td>(5.0 \times 10^2)</td>
<td>(1.00 \times 10^3)</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>(1.6 \times 10^{-1})</td>
<td>(2.5 \times 10^2)</td>
<td>(5.00 \times 10^4)</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>(3.2 \times 10^{-2})</td>
<td>(5.0 \times 10^1)</td>
<td>(1.00 \times 10^4)</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>(3.2 \times 10^{-3})</td>
<td>(5.0 \times 10^0)</td>
<td>(1.00 \times 10^5)</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>(3.2 \times 10^{-4})</td>
<td>(5.0 \times 10^{-1})</td>
<td>(1.00 \times 10^6)</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>2.5 \times 10^3</td>
<td>2.50 \times 10^4</td>
<td>2.50 \times 10^4</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1.0 \times 10^4</td>
<td>1.00 \times 10^5</td>
<td>1.00 \times 10^5</td>
</tr>
<tr>
<td>14</td>
<td>(3.2 \times 10^{-1})</td>
<td>(5.0 \times 10^3)</td>
<td>(5.50 \times 10^4)</td>
<td>(4.50 \times 10^4)</td>
</tr>
<tr>
<td>15</td>
<td>(1.6 \times 10^6)</td>
<td>(7.5 \times 10^3)</td>
<td>(1.00 \times 10^5)</td>
<td>(5.00 \times 10^4)</td>
</tr>
<tr>
<td>16</td>
<td>(3.2 \times 10^6)</td>
<td>(1.5 \times 10^4)</td>
<td>(2.00 \times 10^5)</td>
<td>(1.00 \times 10^4)</td>
</tr>
<tr>
<td>17</td>
<td>(3.2 \times 10^6)</td>
<td>(5.0 \times 10^4)</td>
<td>(5.50 \times 10^5)</td>
<td>(4.50 \times 10^4)</td>
</tr>
<tr>
<td>18</td>
<td>(3.2 \times 10^{-1})</td>
<td>(5.0 \times 10^4)</td>
<td>(5.05 \times 10^5)</td>
<td>(4.95 \times 10^4)</td>
</tr>
</tbody>
</table>

**Analysis of cell trajectories**

We used ImageJ (http://imagej.nih.gov/ij/) with ParticleTracker Plugin [13] for automated cell detection and tracking. Particle tracks were analyzed in a custom-made MATLAB (The MathWorks, Natick, MA) code, where the following filtering was applied: the cells that could not be tracked consistently for more than 6 minutes (3% of the total recording time) were discarded and points on the screen that did not move at all were discarded as well; the latter corresponding to dead cells or other artifacts on the glass surface or CCD. Each experimental run was repeated 3 to 11 times, until about 300 to 700 cell trajectories were gathered. A sample of such trajectories is shown in Fig.4. The distribution of trajectories was very broad with lengths of 260±220µm. Depending on the gradient, component of the velocity in gradient direction ranges from −0.15 µm/min to 0.51 µm/min.

**Analysis of different trajectory time lengths**

Here we experimentally check for the possibility that cells can integrate multiple gradient measurements over time scales longer than the pseudopod extension time (∼30 s). We calculated the chemotactic index (CI) as we progressively moved the end point of the cell trajectory from the one at frame 2 (30 seconds) to the one at frame 400 (3.3 hours).

The 30 second time interval between subsequent frames was chosen since the cell displacements were typically about 3 µm, which was at the limit for measuring displacements in our experiments.

If the cells were indeed integrating over more measurements as the time moved on, we would expect to see the CI
AVERAGING THE EXTERNAL MUTUAL INFORMATION

External mutual information $I_{\text{ext}}$ was averaged over the entire channel in the gradient direction, weighted by the fraction of cells in each spatial segment:

$$\langle I_{\text{ext}} \rangle = \sum_{i=1}^{M} p_i I_{\text{ext},i}$$

where $p_i$ is the fraction of cells in a segment $i$ of the microfluidic device (a sample of such distribution is shown in Fig.8) and $I_{\text{ext},i} = I_{\text{ext}}(\langle c(x_i) \rangle)$ is the external mutual information for the average concentration in segment $i$. If $I_{\text{ext}}$ is averaged assuming a perfectly uniform cell distribution:

$$\langle I_{\text{ext}} \rangle = \frac{1}{c_{\text{max}} - c_{\text{min}}} \int_{c_{\text{min}}}^{c_{\text{max}}} I_{\text{ext}}(c_0) dc_0$$

the analytical result is:

$$\langle I_{\text{ext}} \rangle = \frac{N}{4 \ln 2 (c_{\text{max}} - c_{\text{min}})} \times \left\{ \frac{1}{1 + c_{\text{max}}} - \frac{1}{1 + c_{\text{min}}} - \ln \left( \frac{1 + c_{\text{min}} c_{\text{max}}}{c_{\text{min}} (1 + c_{\text{max}})} \right) \right\}$$

which agrees to our estimate of $\langle I_{\text{ext}} \rangle$ to about 10% for our experiments.

EFFECTS OF FOLIC ACID DEGRADATION

Here we explore the possibility that most of the FA is degraded by cells themselves, and they were effectively sensing a lower FA concentration, closer to $K_d$. FA can be degraded by an extracellular form of FA deaminase protein and we estimate the extent to which the FA concentration can be
reduced by this process. Following up on the previous study of the level of deaminase secretion under the same conditions [14], we estimated the deaminase activity (defined as the amount of FA degraded per cell per unit time) for our system. The reported mean value for the deaminase activity from [14] is 35 pmol per 10⁶ cells per minute. Assuming a steady-state flat concentration profile of deaminase in our experiment of total volume of 0.15 ml, about 50 cells in total and about 5 hours the cells spent in the chamber (corresponding to the middle of our run), the amount of FA that could possibly be degraded by that time is 5.25 × 10⁻¹³ mol. On the other hand, the total amount of FA in this entire volume, at 2.5 μM mean concentration is 3.75 × 10⁻¹⁰ mol, so the degradation by FA deaminase could account for less than 0.1% of the expected amount of FA. This calculation is summarized in the Table II. This conclusion was verified experimentally by changing the cell density by a factor of four (from 7 cells/mm² to 30 cells/mm²) for the gradient where we observed peak response and noticing that the same results in terms of chemotactic index (0.10 ± 0.02 at lower vs 0.09 ± 0.01 at higher density) and total mutual information (0.14 ± 0.02 bits vs 0.14 ± 0.01 bits) was observed. Thus, we conclude that degradation of FA by FA deaminase cannot account for the violation of the data processing inequality.

**EFFECTS OF CELL POLARIZATION / BIAS**

The total mutual information with bias is defined by:

\[ I_{\text{tot}}^{\text{bias}} = H_{\text{bias}}(\theta_{\text{res}}) - H_{\text{bias}}(\theta_{\text{res}}|\theta_{\text{grad}}) \]  

with

\[ H_{\text{bias}}(\theta_{\text{res}}) = - \int p(\theta_{\text{res}}; K) \log_2 p(\theta_{\text{res}}; K) d\theta_{\text{res}} \]

where

\[ p(\theta_{\text{res}}; K) = \int p(\theta_{\text{res}}|\theta_{\text{grad}})p(\theta_{\text{grad}})d\theta_{\text{grad}} \]

which is calculated using the experimentally measured values for the distribution of the response given the gradient, \( p(\theta_{\text{res}}|\theta_{\text{grad}}) = p(\theta_{\text{res}} - \theta_{\text{grad}}) \). Since the measured values were discrete, we originally approximated the integral in Eq.2 (in the main text) with a discrete sum. However, here we calculated a more complicated integral and instead approximated a discrete distribution \( p(\theta_{\text{res}}|\theta_{\text{grad}}) \) with a continuous distribution using kernel density estimation [15]. Since this is a different method of estimating the total mutual information from the data, we first compared the results for non-biased total mutual information (corresponding to the case \( K = 0 \)) obtained using these two methods in Fig.9 and show they are very similar. We therefore used the kernel density estimation to compute the biased total mutual information, for various values of the biasing parameter \( K > 0 \).

Next, we numerically calculated the biased external mutual information \( I_{\text{ext}}^{\text{bias}} \) using the Eq.13 in [16]:

\[ I_{\text{bias}}^{\text{ext}} = I_{\text{ext}} - B(K) \]

where the term \( B(K) \) depends on the bias (see [16] for details).

\[ I_{\text{ext}}^{\text{bias}} = I_{\text{ext}} - \int p(\rho) h(\rho; K_p) d\rho \]

with:

\[ I_{\text{ext}} = \frac{1}{\ln 2} \left( \frac{\nu}{\sigma} \right)^2 - \int p(\rho) \log_2 I_0 \left( \frac{\rho \nu}{\sigma^2} \right) d\rho \]

\[ p(\rho; \nu, \sigma) = \frac{\rho}{\sigma^2} \exp \left[ \frac{-\rho^2 + \nu^2}{2\sigma^2} \right] I_0 \left[ \frac{\rho \nu}{\sigma^2} \right] \]

\[ \nu(N, c_0, \nabla c) = \frac{N}{2} \frac{\nabla c}{c_0 + 1} \]
FIG. 9. Comparison of two methods for calculating $I_{\text{tot}}$: first binning the data and approximating the integral for $I_{\text{tot}}$ with a sum and the second, approximating the discrete data with a continuous function obtained by kernel density estimation [15], showing that both methods give very similar results. Here we used $K_d = 150 \text{nM}$ and $R = 5 \mu \text{m}$.

\[
\sigma(N,c_0) = \sqrt{\frac{N}{2} \frac{c_0}{(c_0 + 1)^2}}
\]

\[
h(\rho;K_p) = \frac{1}{\ln 2} \left[ K \frac{I_1(K_p)}{I_0(K_p)} - \ln I_0 \left( \frac{\rho \sigma}{\nu} \right) \right]
\]

\[
K_p = \frac{K \nu \rho}{K \sigma^2 + \nu \rho}
\]

where $c_0$ is the local chemoattractant concentration in units of $K_d$, $\nabla c$ the gradient in units of $K_d/R$, $N$ the total number of receptors, $K$ the same biasing parameter and $I_0(K_p)$, $I_1(K_p)$ are the modified Bessel functions of the first kind of order zero and one, respectively. We computed both the total and external mutual information for different magnitudes of the bias, up to very sharp polarizations $K = 80$ (larger values require significantly higher numerical precision) and show the results in Fig.3c in the main text. These results show that the inclusion of this effect still results in the violation of the data processing inequality, and moreover, for a wide range of bias parameters, the violation is further increased.

**EFFECTS OF BINNING**

The total mutual information calculated using Eq.3 (main text) depends on the choice of number of bins $m$.

While there is no “best” number of bins, here the total number of bins chosen was 14 which gave similar results for all combinations of gradients and mean concentrations since we had roughly the same number of cells in each case (typically around 500). First, as stated in the main text, it correlates well with the CI (comparing Fig.1b and 2a in the main text). Second, $I_{\text{tot}}$ reaches a plateau in this bin range and becomes lower when we use too few bins (below $\approx 10$) or higher but with much larger uncertainty if we use too many bins (roughly 30 or more); see Fig.10. The plateau corresponds to the middle ground here where $I_{\text{tot}}$ does not change much if the bin number changes a little around the chosen value. Finally, this choice of 14 bins gave approximately the same results as the Kernel Density Estimate (Fig.9) used for data smoothing [15].