AN ABSTRACT OF THE DISSERTATION OF

Garrett D. Potter for the degree of Doctor of Philosophy in Physics presented on May 23, 2017.

Title: Information Processing in Multicellular Networks.

Abstract approved:

_____

Bo Sun

The ability for individual cells and multicellular networks to process information contained in perceived stimuli is of vital importance to their proper functioning and ultimately their cellular fate. The primary mechanism employed by cells to this end is the use of biochemical signaling in which a series of molecular events results in response to the stimuli. These series of molecular events can be highly dynamic in their behavior and recent work has suggested that these dynamics are important in the processing of information and resulting cellular response. Further, it has been shown that multicellular systems themselves exhibit dynamic, coordinated behavior that may act as an additional dimension to information processing in these systems.

To explore and make rigorous the information processing abilities of individual cells and multicellular networks, we focus on the store-operated calcium pathway in mouse fibroblast and human breast cancer cells stimulated by ATP. This pathway is well-known in its ability to create both spatial (blips, puffs, and waves) and temporal (oscillations) cellular calcium dynamics via a nonlinear feedback system. Using

microfluidics, fluorescence and confocal microscopy, stochastic modeling, and information theoretic techniques we probe how individual cells and multicellular networks utilize these dynamics to encode and decode environmental cues.

In the case of individual cells, we explore possible sampling strategies employed by cells to process dynamic biochemical signals. We find that sampling as frequently as possible leads to increased information but that this is a costly strategy and leads to increased redundant information. Instead an infrequent and well-separated sampling strategy arises as the optimal strategy, showing that cells cannot easily distinguish fast dynamics and thus act as a low-pass filter. Further, we find that information is more affected by extrinsic noise than intrinsic noise in the case of multi-time point or vector encoding.

For the case of multicellular networks, we probe how ATP stimuli are encoded in the calcium dynamics of varied network architectures. Our results suggest that the strength of stimuli is encoded in the magnitude of the calcium response in addition to the oscillation propensity of cells in the network. Modeling confirms this behavior and predicts that communication via gap junctions is additionally a vital part of encoding. We find that modulating communication of the network by adjusting cell density influences the oscillation propensity at moderate ATP levels. Further, modulating the communication of the network by increasing the fraction of cancer cells produces the same result at moderate ATP levels. We conclude that stimuli information is multiplexed in the oscillation propensity of the multicellular system, which is potentially beneficial in responding with specificity to a wide array of stimuli.

Information Processing in Multicellular Networks

by
Garrett D. Potter

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 23, 2017
Commencement June 2017

Doctor of Philosophy dissertation of Garrett D. Potter presented on May 23, 2017

APPROVED:

Major Professor, representing Physics

Chair of the Department of Physics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Garrett D. Potter, Author

ACKNOWLEDGEMENTS

I express my deepest gratitude to my partner Sera and my family for their unwavering love and support during this time; they have helped shape who I am today and without them this accomplishment would be impossible. I also express my sincerest thanks to my advisor Dr. Bo Sun whose exceptional guidance and tireless effort helped us overcome countless obstacles and taught me a great deal in our research efforts. A special thanks to my dog Apollo whose furry face and patient ear have helped me through these many years.

CONTRIBUTION OF AUTHORS

Drs. Andrew Mugler and Tommy Byrd provided the stochastic modeling of collective calcium dynamics and computational contributions to the Communication Shapes Sensory Response Multicellular Networks and Dynamic Sampling and Information Encoding in Biochemical Networks manuscripts. Rick Presley provided training and guidance for fabrication of photolithographic devices. Dr. Bo Sun provided guidance, motivation, and insight into all aspects of the work presented in this dissertation.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

CHAPTER 1

Introduction and Theory

1.1 Ca2+ Signaling Mechanism

The ion Ca^{2+} is a ubiquitous secondary messenger molecule capable of serving as a signaling agent for a host of higher order cell behavior including fertilization, embryonic development, cell proliferation and differentiation, and transcription factor activation[1]. Signaling begins when Ca^{2+} is released from its stores in the endoplasmic reticulum (ER) or, equivalently, the sarcoplasmic reticulum (SR) of muscle cells in response to stimuli such as chemical changes, mechanical stress, or radiation. These

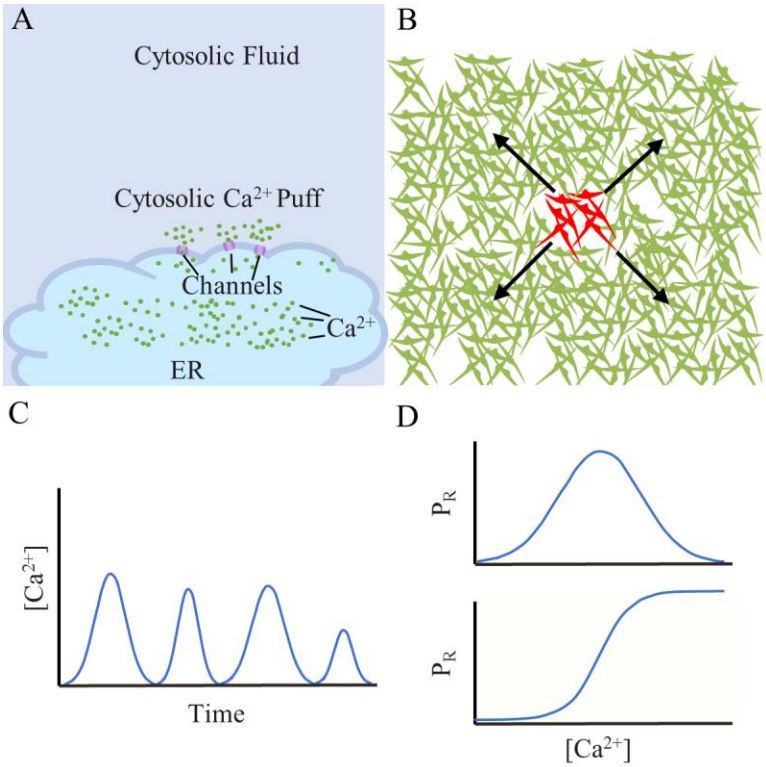

Figure 1.1 Ca^{2+} dynamics due to nonlinear feedback. (A) Localized channels on the ER/SR release Ca^{2+} in a small "puff" in response to perceived stimuli. (B) Ca^{2+}-induced-Ca^{2+}-release (CICR) effects may turn a puff into a pulse that triggers a wave of Ca^{2+} (Red) when Ca^{2+} is exchanged with neighboring cells in a network (Green). (C) Temporal oscillations are another possible outcome of the nonlinear feedback due to CICR. (D) Bell-curve (Top) and sigmoidal (Bottom) probabilities of Ca^{2+} release (P_R) as a function of increasing Ca^{2+} concentration in the cytosolic fluid.

signals have been shown to be highly varied, taking the form of spatial dynamics (blips, puffs, or waves[2–4], Fig. 1.1A and 1.1B) or temporal dynamics (oscillations[2,3], Fig. 1.1C) due primarily to nonlinear feedback from a process known as Ca^{2+}-induced-Ca^{2+}-release (CICR)[1,5]. CICR occurs when Ca^{2+} binds to specific receptor sites dotting the ER/SR causing the additional release of Ca^{2+} via channels that are opened upon binding (see Section 1.1.3 for more detail). This has traditionally been represented by a bell-curve where increasing levels of Ca^{2+} increase the probability of calcium release until the effect becomes inhibitory, however, recent work has suggested that the bell-curve may actually be sigmoidal[6] (Fig. 1.1D). As the spatial and temporal dynamics of Ca^{2+} are thought to encode necessary cellular information, a better understanding of the mechanisms that generate Ca^{2+} dynamics and how they may be used to communicate information is necessary.

1.1.1 Inositol 1,4,5-Trisphosphate Generation

Though vital to the CICR process, calcium alone is insufficient to generate the rich dynamics described above. Experiments have shown that specialized secondary messenger molecules working in conjunction with calcium are necessary for the opening of the ion channels on the ER and SR allowing for the release of calcium from their stores[7]. One such secondary messenger of primary importance is inositol 1,4,5-trisphosphate (IP_3). IP_3 generation can be broken into two major pathways: one involving G-proteins and the other receptor tyrosine kinases (RTKs).

The G-protein pathway begins when chemical or hormonal ligands bind to G-coupled (guanine nucleotide binding protein) protein receptors (GPCRs) lining the plasma membrane of the cell. GPCRs dot the cell membrane with a variety of sub-types

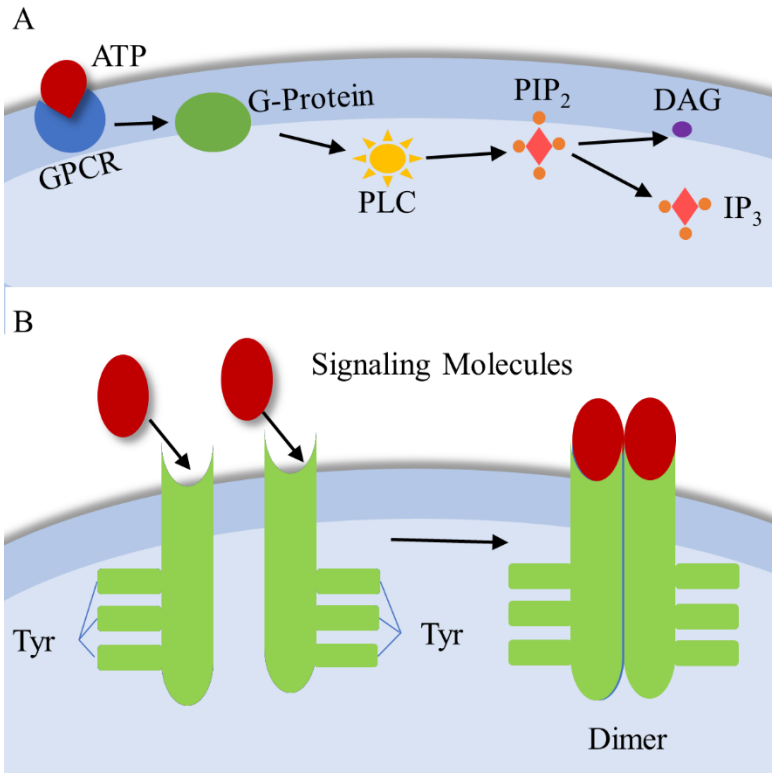

Figure 1.2 IP_3 generation in non-excitable cells. (A) Binding of ligand (ATP) to receptor site on GPCR causes generation of PLC$\boldsymbol{\beta}$ leading to the hydrolyzing of PIP_2 into the secondary messengers DAG and IP_3. IP_3 diffuses through cytosolic fluid to bind with IP_3 receptors located on ER. (B) Signaling molecules binding to the receptor sites of RTKs cause the dimerization of RTKs from which PLC$\boldsymbol{\gamma}$ is created. PLC$\boldsymbol{\gamma}$ acts like PLC$\boldsymbol{\beta}$ to hydrolyze PIP_2 creating IP_3 and DAG.

each exhibiting varying degrees of specificity to the type of ligand allowed to bind[8]. Binding of the ligand leads to the activation of the G-protein which subsequently activates the enzyme phosphoinositide phospholipase (PLC). A specific isotype of PLC, PLCβ, is responsible in this process for hydrolyzing phosphatidylinositol-4,5-bisphosphate (PIP_2)[5,9] which results in the formation of IP_3 and a byproduct called

diacylglycerol (DAG) as shown in Fig. 1.2A. DAG, though not the focus of this discussion, has other important secondary messenger roles and is linked with the translocation of protein kinase C (PKC) in the plasma membrane[5]. While IP_3 isn't the only secondary messenger molecule capable of causing the release of calcium from the ER stores it serves as the primary secondary messenger capable of doing so.

RTKs, like GPCRs, dot the cell membrane but are activated in response to the binding of growth factor signaling molecules such as insulin, fibroblast growth factor (FGF), or vascular endothelium growth factor (VEGF)[9]. As shown in Figure 1.2B, binding of the growth factor to the RTK causes dimerization to occur where two RTKs become bound by intermolecular forces. Dimerization starts an autophosphorylation process that adds phosphates from cytosolic ATP to the six tyrosine complexes on the bound RTKs. This new complex signals the activation other processes to generate the isomer PLCγ which, like PLCβ, goes on to hydrolyze PIP_2 to generate IP_3 and DAG.

1.1.2 Calcium Release Channels

There are two major channels in endothelial cells that are responsible for the pumping of Ca^{2+} from the ER to the surrounding cytosolic fluid: those linked with IP_3 receptors and those linked with ryanodine receptors (RyRs)[1,5,10]. The first of the two channels form the most studied and well understood mechanism of Ca^{2+} release from the ER. Diffusion of IP_3 through the cytosolic fluid after a triggering event leads to the binding of IP_3 to its IP_3 receptor located on the outside of the ER (Figure 1.3). There are three isoforms of IP_3 receptors (type 1, 2, and 3) with either one or all three of the

isoforms represented in all animal and human cells[5]. Current models of IP_3 and IP_3 receptor binding suggest that the IP_3 receptor is a tetramer composed of four identical subunits, with each of the subunits containing an IP_3 binding site that needs to be bound with IP_3 and Ca^{2+} for the release of calcium from the ER[11]. Originally, evidence showed

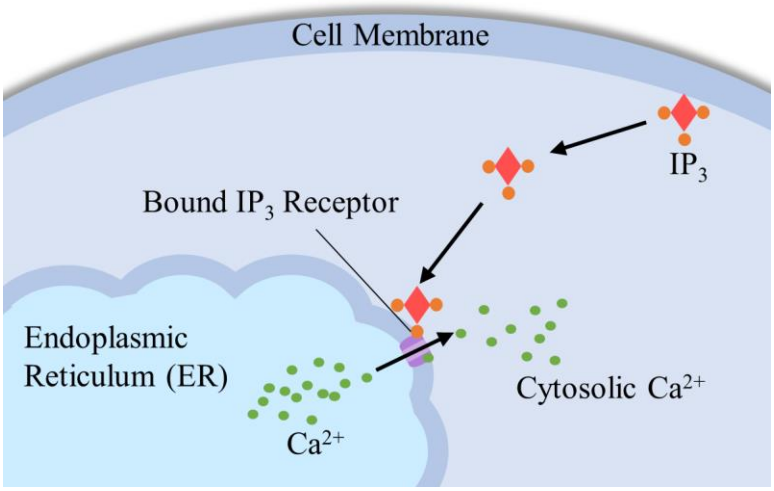

Figure 1.3 Ca^{2+} release from endoplasmic reticulum (ER) via secondary messenger IP_3. Binding of IP_3 and Ca^{2+} to stimulatory sites of the IP_3 receptor located on the ER causes a conformational change of the receptor allowing for stored Ca^{2+} to rush into the surrounding cytosolic fluid. A similar effect is achieved via ryanodine receptors (RyR) typically found in muscle cells, though the exact secondary messenger necessary to trigger the opening of the RyR is under debate.

that the binding of IP_3 to its receptor on the ER and in the presence of low concentrations of Ca^{2+} caused the channel to open and release Ca^{2+} into the cytosolic fluid; however, if the Ca^{2+} concentration were sufficiently high the effect became inhibitory causing the channel to close[7,12]. This bell-shaped response allowed for a unique positive/negative feedback control for Ca^{2+} release. More recent work suggests that this bell-shape in Ca^{2+} concentration may be more sigmoidal and that the presence

of IP_3 bound to the IP_3 receptor is the key factor, making the system sensitive to stimulatory binding of Ca^{2+} rather than inhibitory[1,6].

Ryanodine receptors share many similarities to IP_3 receptors. Like IP_3 receptors, RyRs are oligomeric with three known isoforms but unlike IP_3 receptors (which are found almost universally), RyRs are typically found in cardiac muscle, skeletal muscle, and in striated muscle cells[5]. The first isoform (RyR1) is a voltage gated channel that controls the release of Ca^{2+} while the exact mechanism of RyR2 and RyR3 control are debated. RyR2 and RyR3 behave similarly to IP_3 receptors in that they are sensitive to local Ca^{2+} concentrations, exhibiting the bell-like response curve, but the secondary messenger responsible for binding to RyR2 and RyR3 has not been confirmed. Some studies point to the messenger cyclic ADP-ribose (cADPR) but this has been contested[13].

The two channels presented here are not the only sources of Ca^{2+} release but rather dominate the discussion of Ca^{2+} release mechanisms. Mitochondria, another source of Ca^{2+} storage in cells, are able to release Ca^{2+} in response to a triggering event[5,8]. What is crucial to the above systems are the effects of CICR coupled with the bell-like or sigmoidal response curves of the release channels. This coupling stimulates or inhibits the release of Ca^{2+} into the cytosolic fluid leading to the rich spatial and temporal dynamics of Ca^{2+} signaling observed in all cells[2,5,13].

1.1.3 Calcium Uptake and Removal Channels

Release of Ca^{2+} from the ER must be followed by mechanisms that restore intracellular levels of Ca^{2+} to basal levels to promote the health and proper function of

the cell. Three main mechanisms of re-sequestering and removing of Ca^{2+} involve sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pumps, the plasma membrane Ca^{2+}-ATPase (PMCA), and the Na^{+}/Ca^{2+} exchangers (NCX)[1,2,11]. As their name suggests, the ATPase pumps utilize ATP by catalyzing their decomposition into ADP and energy that is used to exchange one Ca^{2+} (in the case of PMCA) or two Ca^{2+}

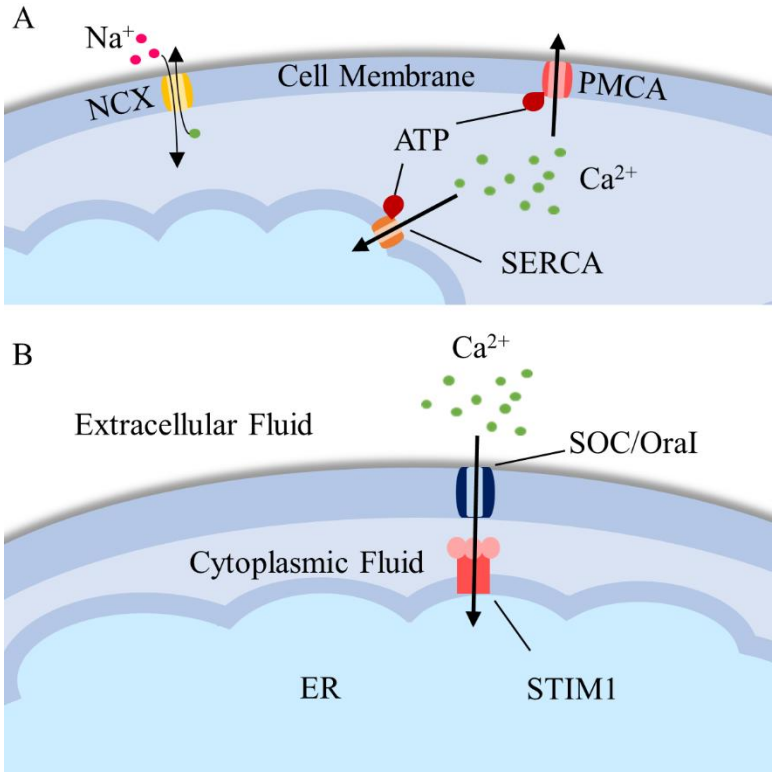

Figure 1.4 Storage and replenishment mechanisms of Ca^{2+}. (A) Hydrolysis of ATP into ADP for both plasma membrane Ca^{2+}-ATPase (PMCA) and sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pumps allows for either removal of Ca^{2+} to the extracellular fluid or storage of Ca^{2+} in the SR/ER respectively. (B) Depletion of Ca^{2+} levels in the SR/ER sensed by the STIM1 protein triggers entry of Ca^{2+} from the extracellular fluid through the store-operated channels (SOC) also known as OraI.

in the case of SERCA for protons (Fig. 1.4A)[11]. The NCX do not require ATP to complete the removal of Ca^{2+} but instead exchange three Na^{+} for each Ca^{2+}, demonstrated as well in Fig. 1.4A. PMCA and SERCA pumps have low capacities for

exchange making them suitable for maintenance of the low levels of Ca^{2+} whereas the NCX have very large capacities allowing them to exchange Ca^{2+} quite rapidly; in this way, the two systems complement each other in their ability to handle both the slow and fast Ca^{2+} dynamics in exhibited in all cell types[11].

In addition to the uptake and removal of cytosolic Ca^{2+} via the pumps and exchangers above, Ca^{2+} may enter the cell from the extracellular fluid through the store-operated channels (SOC) in the plasma membrane. Recent work has identified a protein dotting the ER/SR named STIM1 that acts as a sensor for Ca^{2+} levels inside the ER/SR. Upon store depletion, the STIM1 aggregate along the ER/SR membrane and move to within 25nm of channels on the plasma membrane (Fig. 1.4B)[14]. These channels, called OraI, have been identified as crucial for the Ca^{2+}-release activated current (CRAC) which is the signature of SOC entry of Ca^{2+} in cells[15]. The biochemical specifics behind SOC entry are beyond the scope of this dissertation, but Clapham does a nice overview in his review article in Cell[11].

1.1.4 Gap Junctions

The release and re-uptake mechanisms described above give rise to the rich set of calcium dynamics that is believed to be an information encoding strategy employed by cells. Perhaps equally important, however, is the ability to communicate that information to neighboring cells thereby ensuring survival. Intercellular calcium waves, which have been demonstrated in a variety of cell types including glial cells, neurons, endothelial cells, epithelial cells, and hepatocytes[16] provide an example of such communication. The propagation of these dynamics is achieved through the

diffusion of messenger molecules via small channels called gap junctions[16–19]. These channels allow for the exchange of molecules less than 1 kDa[17,20] and are important in the normal functioning of organs such as the heart[17].

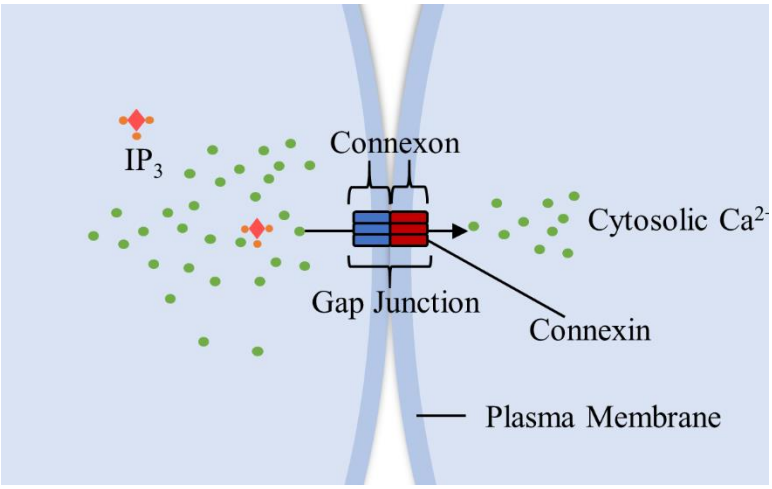

Figure 1.5 Gap junction communication. Connexins forming connexon on plasma membrane associate with neighboring connexon to create a gap junction between cells. Gap junctions allow the diffusion of small molecules such as IP_3 and Ca^{2+} between neighboring cells thereby allowing the propagation of intercellular calcium waves and information.

Gap junctions are formed when neighboring cells come into contact with one another. During this contact, connexons (a hexameric protein structure whose base unit is the connexin[17,18]) located on each cell's plasma membrane associate to form a gap junction (Fig. 1.5). This gap junction facilitates communication through diffusion. In the case of intercellular calcium waves, both IP_3 and Ca^{2+} can pass through the gap junction provided the channel is open. As gap junctions are voltage-gated channels, there has been evidence to show that the large increase in Ca^{2+} levels during a signaling event can restrict or close the channel[17]. Assuming the channel is open, IP_3 propagates

ahead of Ca^{2+} due to its larger diffusion coefficient and if the IP_3 and Ca^{2+} are sufficiently high they can trigger a Ca^{2+} release in the neighboring cell continuing the wave[21]. In this way the information encoded in the Ca^{2+} signaling dynamics of a single cell may be transmitted to neighboring cells to be decoded for further decision making.

1.2 Ca^{2+} Dynamics Modeling

1.2.1 Deterministic Modeling

To better understand the ability of the Ca^{2+} dynamics to encode information and serve as a mechanism of communication, we make use of the Tang and Othmer model of Ca^{2+} dynamics which is a deterministic model for a single cell[22]. The deterministic model used by Tang and Othmer focuses on Ca^{2+} dynamics triggered by the IP_3/Ca^{2+} pathway and assumes a single IP_3 binding site and two Ca^{2+} binding sites. If the IP_3 and one of the Ca^{2+} binding sites are occupied, the channel is opened, while if both Ca^{2+} sites are bound the channel is closed. The model also assumes the following reactions:

$$C_s \underset{\nu g_0}{\overset{g_0}{\rightleftharpoons}} C, \quad C_s + R_4 \underset{\nu\gamma_1}{\overset{\gamma_1}{\rightleftharpoons}} C + R_4, \quad C \xrightarrow{\nu f(c)} C_s, \tag{1.1}$$

$$R_2 \underset{k_{-1}}{\overset{k_1 I}{\rightleftharpoons}} R_3, \quad R_3 + C \underset{k_{-2}}{\overset{\lambda_2}{\rightleftharpoons}} R_4, \quad R_4 + C \underset{k_{-3}}{\overset{\lambda_3}{\rightleftharpoons}} R_5, \tag{1.2}$$

Eqn. 1.1 contains the reactions describing calcium transport between the ER and the cytoplasm, while Eqn. 1.2 contains the reactions describing the sequential (un)binding of molecules to receptors. In Eqn. 1.1, the first reaction describes leakage between the calcium in the ER (C_s) and calcium in the cytoplasm (C); ν is the ratio of the ER volume

V_s to the cytoplasm volume V. The second reaction describes transport of calcium through the receptor channels: receptors with both IP_3 and calcium bound (R_4) are active channels. The last reaction describes the pumping of calcium from the cytoplasm to the ER; f denotes the nonlinear pumping propensity. In Eqn. 1.2, the first reaction describes the binding of IP_3 to bare receptors (R_2), creating the complex R_3; I is the concentration of IP_3, which is treated as a parameter in their model (we later generalize this feature, making I a dynamic variable and its production rate α the control parameter, since this is more consistent with the experimental setup). The second reaction describes the subsequent binding of cytoplasmic calcium to the complex, which activates the complex (R_4). The last reaction describes the further binding of cytoplasmic calcium to another site on the complex, which deactivates the complex (R_5). Only receptors in the R_4 state serve as active channels.

The six reactions above are turned into a set of six differential equations, that, for the sake of brevity will not be expressed here. Using a series of simplifications (e.g. conservation of molecule numbers and receptors) brings the total number of equations down to four. A final realization by Tang and Othmer that two of the binding/unbinding reactions are much slower than the others permits a quasi-steady-state approximation bringing the final number of equations to two[22] shown below.

$$\epsilon \frac{dx}{d\tau} = \alpha_1(1-x) + \alpha_2(1-x)\frac{x(1-y)}{x+\beta_1[1+\beta_0(I)]} - \frac{x^2}{x^2+\alpha_3^2} \tag{1.3}$$

$$\frac{dy}{d\tau} = -y + \frac{\beta_2 x^2(1-y)}{x+\beta_1[1+\beta_0(I)]} \tag{1.4}$$

Greater detail of the simplification steps and parameter definitions can be found in the supplemental information of the manuscript presented in Chapter 2 of this dissertation[23]. Eqns. 1.3 and 1.4 exhibit four dynamic regimes. As shown in Fig. 1.6, the system can be (A) monostable, with low cytoplasmic calcium, (B) excitable, (C) oscillatory, or (D) monostable, with high cytoplasmic calcium. In the last regime (D), near the boundary, oscillations can also be supported for certain initial conditions. All four regimes are accessible by tuning the IP_3 concentration I (Fig. 1.6E). Transitions

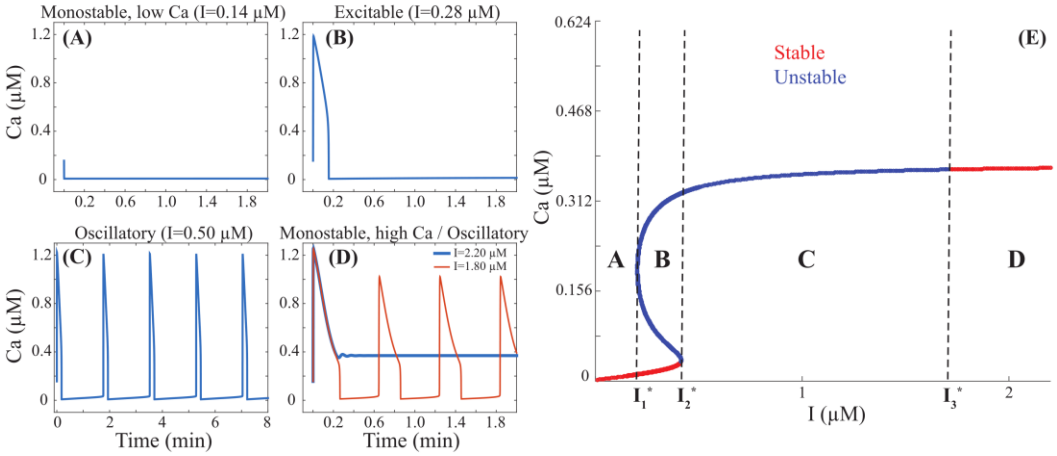

Figure 1.6 The deterministic model (Eqns. 3 and 4) exhibits four dynamic regimes: (A) monostable, with low cytoplasmic calcium, (B) excitable, (C) oscillatory, or (D) monostable, with high cytoplasmic calcium. In the last regime (D), near the boundary, oscillations can also be supported for certain initial conditions. (E) These regimes are accessible by tuning the IP_3 concentration I. The letters A-D in panel E correspond to the dynamics seen in panels A-D. The critical values I_1^*, I_2^*, and I_3^* separating the regimes (dashed lines) are determined by changes in the number or stability of fixed points, which follow from a standard linear stability analysis.

between regimes occur at the critical concentrations I_1^*, I_2^*, and I_3^*. Of particular interest is the transition from the excitable to the oscillatory regime (I_2^*), since these are the dynamics exhibited by cells in the experiments: transient pulsing or sustained oscillations.

1.2.2 Stochastic Modeling

The success of the deterministic model by Tang and Othmer in replicating Ca^{2+} dynamics measured in experiments provide insight into the key components necessary for Ca^{2+} signaling, however, reactions are rarely, if ever, deterministic in nature. Binding and unbinding of IP_3 and Ca^{2+} to their receptor sites are probabilistic and thus the extension of the Tang and Othmer model to include stochasticity is necessary.

In our model, stochasticity is implemented using an adaptive tau-leaping method[24–27] which is a more computationally efficient approximation of the exact Gillespie algorithm[28]. The simulation accounts for the fact that the number of molecules are integer values and that the reactions occur at random, exponentially distributed times. Noise is implemented in the model via the volume of the cell (which is experimentally accessible) and the number of receptors (which is estimated) in conjunction with the number of molecules. Large numbers of molecules correspond to low noise due to the low level of fluctuations and vice-versa. Our model also exchanges IP_3 concentration I for the number of IP_3 molecules which is a random variable controlled by the production rate α and the degradation rate of IP_3 given by μ. This modifies the first reaction in Eq. 1.2 resulting in

$$\underset{\alpha}{\overset{\mu}{I \rightleftharpoons \emptyset}} \quad R_2 + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R_3 \tag{1.5}$$

Typical Ca^{2+} dynamics for both the deterministic and stochastic model can be found in Figure 1.7.

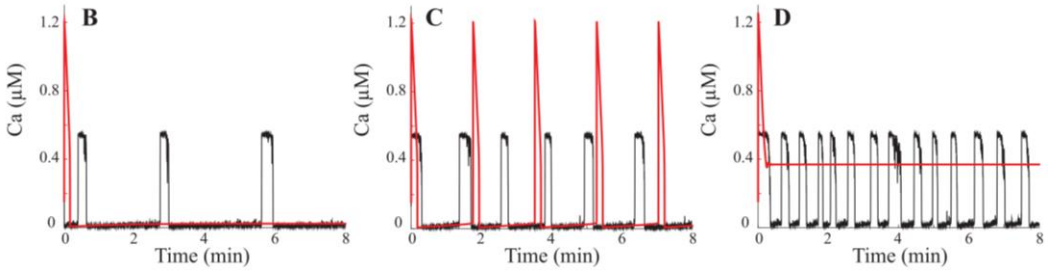

Figure 1.7 Noise expands the oscillatory regime. (B-D) Typical dynamics for the deterministic model (red) and stochastic model (black) are shown for the IP_3 levels corresponding to excitable, oscillatory, and the monostable/high Ca^{2+} dynamics described in Section 1.2.1. IP_3 levels corresponding to a monostable/high Ca^{2+} dynamic in the deterministic model are part of the expanded oscillatory regime in the stochastic model as shown in (D).

1.2.3 Communication

As we are interested in how Ca^{2+} signaling can be used as a mechanism for communication in multicellular networks, we simulate communication via gap junctions by allowing calcium molecules to diffuse between neighboring cells. Though recent work suggests this diffusion may be non-linear[21], we choose a model of diffusion based on Fick's Law in one dimension as a simple case

$$J = -D\frac{d\phi}{dx} \tag{1.6}$$

where J is the diffusion flux, D is the diffusion coefficient, ϕ is the concentration, and x is the position. Doing so introduces new reactions into the model

$$C_{i,j} \underset{h}{\overset{h}{\rightleftharpoons}} C_{i+1,j} \qquad C_{i,j} \underset{h}{\overset{h}{\rightleftharpoons}} C_{i,j+1} \tag{1.7}$$

where i, j index cell lattice sites, and h is the hopping rate of calcium ions from cell to cell. We estimate h from the experimentally measured diffusion coefficient $D \sim 1 \frac{\mu m^2}{s}$

(see Fig. 2.7B). Since in our model, ions are hopping between cells on a two-dimensional lattice, h^{-1} corresponds to the expected time $l^2/4D$ to diffuse a cell-to-cell distance l. Estimating a cell-to-cell distance of $10-20\mu m$ from typical cell densities in our devices, we find $h \sim 0.01-0.04\, s^{-1}$, and therefore we use $h = 0.025\, s^{-1}$.

Our calculations are done on a square lattice, where each lattice site is either empty or contains one cell, and therefore each cell can have up to four neighbors with which to communicate. Density is varied by changing the number of empty lattice sites. For each chosen density, we sample over individual realizations, in which cell locations are assigned randomly. Thus, the statistics encompass many possible spatial distributions of cells for each density.

1.3 Information Theory

To quantify information and communication in multicellular networks we implement the ideas of information theory in the analysis of the Ca^{2+} signaling dynamics. Biochemical signaling events are inherently noisy due to cell-to-cell variability and are often highly complicated in their reliability on multiple messengers, feedback loops, and multi-step processes. Information theory reduces this complexity of the signaling process to understanding its inputs and outputs and accounts for noise through utilization of Shannon entropy, referred to from here on simply as entropy[29]. Traditional approaches to handling noise such as calculations of the standard deviation or variance serve to quantify the amount of noise in the system but may not provide insight into the information stored in the relationships between inputs and outputs[29].

1.3.1 Entropy in Discrete Systems

Claude Shannon's work on information theory proved that entropy is the only mathematical quantity that satisfies three simple criteria of information: monotonicity, independence, and branching[30]. We expect in systems subject to uncertainty with multiple states that the information contained in them should increase monotonically with increasing number of possible outcomes. In this way, a flip of a coin contains less information than the roll of a die. The second criteria – independence – arises from the necessity that the total information contained in a system should be equal to the sum of the individual amounts of information contained in two independent constituent components. Finally, the branching criteria of information arises from the need of the total information contained in a system to be the weighted sum of the information gained from the multiple components of the system.

For a system with n discrete probabilities of its events occurring, we can write the entropy of the system as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i), \tag{1.8}$$

where H is the entropy, X is the random variable (a variable subject to probability, e.g. the flip of a coin), and $p(x_i)$ describes the probability of the i^{th} outcome of the n possible outcomes of X[29]. Taking the convention that $0 \log_2 0 = 0$ and keeping in mind that the $p(x_i)$ must necessarily be between 0 and 1, the entropy of the system must be positive. If we take, as a simple example, calculation of the entropy of a system with two equally probable outcomes, the entropy becomes

$$H(X) = -\sum_{i=1}^{2} p(x_i) \log_2 p(x_i)$$

$$H(X) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right]$$

$$= 1 \text{ bit}$$

indicating that a system with two equally likely possible outcomes has an entropy that is equal to one, with units of bits. The choice of base 2 in the logarithm was selected by Shannon for this reason – so that a system with two possible states contains 1 bit of entropy with N such systems containing N bits of entropy[31]. This is purely a statement of the monotonicity condition which guarantees that systems with greater numbers of possible outcomes have increased uncertainty, and indeed, if we perform the same calculation with systems that have 4, 8, and 16 equally probable states we find that we receive 2, 3, and 4 bits of entropy respectively.

In the case of unequal probabilities, consider the flip of an unfair coin which has $p_{heads} = \frac{3}{4}$ and $p_{tails} = \frac{1}{4}$. The entropy of this system becomes

$$H_{unfair\ coin}(X) = -\sum_{i=1}^{2} p(x_i) \log_2 p(x_i)$$

$$= -\left[\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right]$$

$$= 0.81 \text{ bits}$$

We see that the entropy of the system decreases in the case of the uneven coin as we might naturally believe it should; our uncertainty in such a system is diminished because we expect to receive heads more often than we do tails. In the extreme case where only heads are allowed, the entropy of the system becomes

$$H_{heads\ only} = -[1\log_2(1) + 0\log_2 0]$$

$$= 0 \text{ bits}$$

which is again consistent with our expectation that there should be no uncertainty in a system where only one result is guaranteed.

Addressing the properties of independence and branching (or as it is sometimes referred to as additivity) we select a system that contains a fair coin and die. The independence condition guarantees that a system containing M independent random variables will have a total entropy equal to the sum of their individual entropies. Taking X to represent the flip of the fair coin and Y to represent the roll of a fair die, we note that

$$H_{fair\ coin}(X) = 1 \text{ bit}$$

$$H_{fair\ die}(Y) = 2.58 \text{ bits}$$

$$H_{combined} = H(X) + H(Y) = 3.58 \text{ bits}$$

which must necessarily be true as the individual probabilities of the coin and die outcomes do not impact one another. The branching condition guarantees that the entropy is unaffected by the subdivision of the system into parts. Breaking the calculation of the entropy of a fair die roll into the entropy of rolling of a 1-3 and the entropy of rolling a 4-6, we find

$$p_1 = \frac{1}{6} = p_2 = \cdots = p_6 = \frac{1}{6}$$

$$H_{roll\ 1-3} = -\sum_{i=1}^{3} p_i \log_2 p_i = -\frac{1}{2}\log_2 \frac{1}{6} = 1.29 \text{ bits}$$

$$H_{roll\ 4-6} = -\sum_{i=4}^{6} p_i \log_2 p_i = -\frac{1}{2}\log_2 \frac{1}{6} = 1.29 \text{ bits}$$

$$H_{fair\ die} = 2.58 \text{ bits} = H_{roll\ 1-3} + H_{roll\ 4-6}$$

it is clear that the entropy of the system doesn't change just because we have partitioned the system in a particular way.

1.3.2 Entropy in Continuous Systems

The above discussion introduces the basic concepts and conditions of entropy but fails to address a large class of systems – those that cannot be described by discrete probabilities. Measurements of biochemical signals (of importance to this work) often take on a continuum of values and are thus ill-suited to being assigned discrete probabilities. To deal with this case, the entropy of a continuous system is defined as[30]

$$H(X) = -\int \rho(x) \log_2 \rho(x)\, dx, \tag{1.9}$$

where the discrete probabilities $p(x_i)$ have been replaced with their continuous analog the probability density $\rho(x)$ and have been integrated over all dx. It can be shown that the entropy in the case of a continuous probability distribution meets the criteria laid out in the beginning thus ensuring that nothing has been lost in the transition.

Two issues arise immediately from the representation of the entropy in the continuous case. The first comes from a practicality standpoint: a distribution of responses, whether they be biochemical in nature or not, aren't guaranteed to have an easily recognizable form. Defining probability densities for distributions that are Gaussian, Poisson, exponential, or some other well-known form are simple in that they

have unique parameters that define their shape and solving for these parameters becomes a relatively simple task. In cases where the form of the distribution is unclear, a non-parametric approach to determining the probability density becomes necessary. One such method, the k-nearest neighbors method (kNN) is implemented in the two manuscripts presented in this dissertation for its relative ease of implementation and unbiased estimation[32–34]. The second problem is that probability densities have units associated with them and taking logarithms of objects with units is inherently problematic[30,35]. If, for example, the random variable x has units of length we would expect that the probability density $\rho(x)$ would have units of inverse length. However, if differences of entropy are calculated, this problem goes away[30,35].

1.3.3 Mutual Information

The treatment thus far has focused on discussion and definition of entropy in the discrete and continuous cases, and – while entropy is a highly important and useful property of a system – it doesn't answer a question we are generally more interested in: what information does one random variable contain about another? Re-casting this in the context of biochemical signaling: what information does a set of responses contain about the set of stimuli used to achieve them? Intuitively, we expect that if the responses to one stimuli heavily overlap with the responses to a second stimuli we cannot easily determine which stimuli was applied. We effectively know less information about the stimuli in this situation than when the sets of responses only slightly overlap (as shown in Fig. 1.8). In the perfect case where the sets of responses

don't overlap at all, maximal information has been achieved because we can uniquely point to a unique stimulus after measuring a response.

The information can be quantified using the ideas of entropy that were laid out in section 1.3.1. Given a random variable X (which we will take to be the applied stimuli) with states x we can describe the likelihood of obtaining the states with the probability distribution $P_X(x)$ and define its entropy as $H[P_X(x)]$ which is often

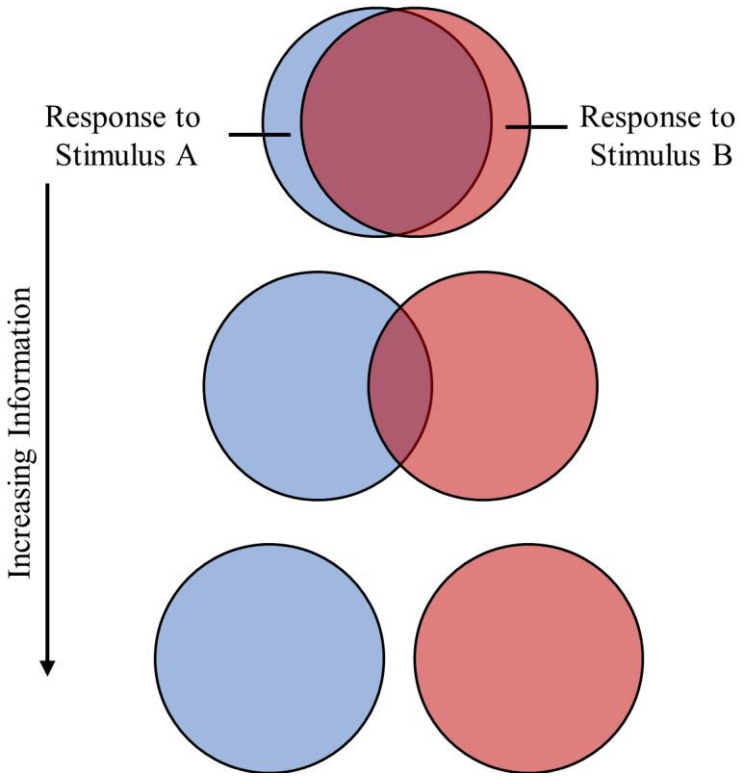

Figure 1.8 Information as a function of response separation. Given two sets of responses to two different stimuli, the amount of information one set of responses gives about its stimulus depends directly on the amount of overlap between the two sets of responses. Heavily overlapping responses do not provide as much information about the stimulus received as weakly overlapping responses. The perfect scenario, where the set of responses do not overlap at all, provide maximal information about the stimuli received.

shortened to $H(x)$[30]. Now, ideally, we have the case that we can make observations y (our biochemical responses) that provide information about x. Naturally, these

observations y can only be taken from the set of Y that describes the observations and they too can be described a probability distribution $P_Y(y)$ with entropy $H[P_Y(y)] = H(y)$. If useful information is to be gained we expect that making an observation y under the condition x, described by the conditional probability $P(x|y)$, will have an entropy $H(x|y)$ that is lower than the entropy $H(x)$. This is intuitive as we hope that by making some observation about the system will *lower* the total entropy of the system; or, put another way, results in a *gain* in information. This can be written mathematically as

$$I(y;x) = H(x) - H(x|y) \tag{1.10}$$

and is known as the mutual information[30,36]. Averaging over all possible observations y we can describe the average information y provides about x[30]

$$\langle I(y;x)\rangle = H(x) - \sum_y P_Y(y)H(x|Y) \tag{1.11}$$

Or, if it is convenient, we can make a series of substitutions to express this in a more compact form involving the joint distribution $P(x,y) = P(x|y)P_Y(y)$

$$\langle I(y;x)\rangle = -\sum_x P_X(x)\log_2 P_X(x) - \sum_y P_Y(y)\left[-\sum_X P(x|y)\log_2 P(x|y)\right]$$

$$= -\sum_x\sum_y P(x,y)\log_2 P_X(x) + \sum_x\sum_y P(x,y)\log_2(P(x|y)$$

$$= \sum_x \sum_y P(x,y) \log_2 \left(\frac{P(x|y)}{P_X(x)} \right)$$

$$= \sum_x \sum_y P(x,y) \log_2 \left(\frac{P(x,y)}{P_X(x) P_Y(y)} \right) \tag{1.12}$$

and for the continuous variable case we get[35]

$$\langle I(y;x) \rangle = \int dy \int P(x,y) \log_2 \left(\frac{P(x,y)}{P_X(x) P_Y(y)} \right) dx \tag{1.13}$$

These results are on average, symmetric in x and y[30,35,36] which indicates that we have some freedom in how we pose the question: how much information does y contain about x or how much information does x contain about y? Generally, however, we only care about the former question from an experimental standpoint as it is the observed y under the various conditions x that we have experimental access to. From a biochemical signaling standpoint, this would be equivalent to having measured a series of responses under several stimuli and finding out how much (average) information the responses give about the stimuli. Calculating the mutual information, then, becomes an extremely useful tool in characterizing information encoding in cell networks.

One item of importance is that in the case of continuous probability systems it is possible to obtain a negative mutual information. In this situation, the uncertainty in the conditioned measured responses is larger than the uncertainty in the conditions themselves, resulting in an information loss or more uncertainty overall.

CHAPTER 2

Communication Shapes Sensory Response in Multicellular Networks

Garrett Potter, Tommy Byrd, Andrew Mugler, and Bo Sun

Proc. Natl. Acad. Sci. **113,** 10334–10339 (2016).

Collective sensing by interacting cells is observed in a variety of biological systems, and yet, a quantitative understanding of how sensory information is collectively encoded is lacking. Here, we investigate the ATP-induced calcium dynamics of monolayers of fibroblast cells that communicate via gap junctions. Combining experiments and stochastic modeling, we find that increasing the ATP stimulus increases the propensity for calcium oscillations, despite large cell-to-cell variability. The model further predicts that the oscillation propensity increases with not only the stimulus, but also the cell density due to increased communication. Experiments confirm this prediction, showing that cell density modulates the collective sensory response. We further implicate cell–cell communication by coculturing the fibroblasts with cancer cells, which we show act as "defects" in the communication network, thereby reducing the oscillation propensity. These results suggest that multicellular networks sit at a point in parameter space where cell–cell communication has a significant effect on the sensory response, allowing cells to simultaneously respond to a sensory input and the presence of neighbors.

2.1 Introduction

Decoding the cellular response to environmental perturbations, such as chemosensing, photosensing, and mechanosensing, has been of central importance in our understanding of living systems. To date, most studies of cellular sensation and response have focused on single isolated cells or population averages. An emerging picture from these studies is the set of design principles governing cellular signaling pathways: these pathways are organized into an intertwined, often redundant network with architecture that is closely related to the robustness of cellular information processing[37,38]. However, many examples suggest that collective sensing by many interacting cells may provide another dimension for the cells to process environmental cues[39]. Examples, such as quorum sensing in bacterial colonies[40], olfaction in insects[41] and mammals[42], glucose response in the pancreatic islet[43], and the visual processing of retinal ganglion cells[44], suggest a fundamental need to revisit cellular information processing in the context of multicellular sensation and response, because even weak cell-to-cell interaction may have strong impact on the states of multicellular network dynamics[45]. In particular, we seek to examine how the sensory response of cells in a population differs from that of isolated cells and whether we can tune between these two extremes by controlling the degree of cell-cell communication.

Previously, we described the spatial-temporal dynamics of collective chemosensing of a mammalian cell model system[19,46]. In this system, high-density mouse fibroblast cells (NIH 3T3) form a monolayer that allows nearest neighbor communications through gap junctions[17]. When extracellular ATP is delivered to the monolayer, store-operated calcium dynamics is mediated by the second messenger

IP_3[47]. The dynamics are complicated by nonlinear feedback between Ca^{2+} and the ion channel opening probability, which leads to rich behaviors, such as cytosolic calcium oscillations[4]. In the situation of collective ATP sensing, we have found that gap junction communications dominate intercellular interactions[46]. Furthermore, these short-range interactions propagate and turn the cell monolayer into a percolating network[19]. These characteristics make the system ideal for studying how sensory response is modulated by communication in multicellular networks.

Here, we use this model system to examine how cell–cell communication affects collective chemosensing. Combining experiments with stochastic modeling, we find that cells robustly encode the ATP stimulus strength in terms of their propensity for calcium oscillations, despite significant cell-to-cell variability. The model further predicts that the oscillation propensity depends on not only the stimulus but also the density of cells, and that denser monolayers have narrower distributions of oscillation frequencies. We confirm both predictions experimentally. To verify that the mechanism behind the density dependence is the modulation of cell–cell communication, we introduce cancer cells (MDA-MB-231) into the fibroblast cell monolayer. As we show, MDA-MB-231 cells act as "defects" in the multicellular network, because they have distinct calcium dynamics compared with the fibroblasts caused by reduced gap junction communication[48–50]. We find that the oscillation propensity of the fibroblasts decreases as the fraction of cancer cells increases, confirming that the sensory response is directly affected by the cell–cell communication.

2.2 Materials and Methods

2.2.1 Fabrication of Microfluidic Devices

The organic elastomer polydimethylsiloxane (PDMS, Sylgard 184, Dow-Corning) used to create the microfluidic devices is comprised of a two-part mixture - a base and curing agent - that is mixed in a 10:1 ratio, degassed, and poured over a stainless-steel mold before curing at 65°C overnight. Once cured, the microfluidic devices are cut from the mold, inlet/outlet holes are punched, and the device is affixed to a No. 1.5 coverslip via corona treatment (Fig. 2.1A).

2.2.2 Cell Culture and Sample Preparation

NIH 3T3 and MDA-MB-231 cells were cultured in standard growth mediums (Dulbeccos modified Eagle medium (DMEM) supplemented with 10% bovine calf serum and1% penicillin and DMEM supplemented with 10% fetal bovine serum, 1% penicillin, and 1% non-essential amino acids respectively). To prepare samples, cells were detached from culture dishes using TrypLE Select (Life Technologies) and suspended in growth mediums before pipetted into the microfluidics devices and allowed to form monolayers (Fig. 2.1C). If MDA-MB-231 cells were the dominant species (a fraction greater than 50% of all cells), they were first allowed to attach the glass bottom of the microfluidics devices. Red fluorescent tag (CellTracker, Life Technologies) was then applied and subsequently washed with growth medium. Finally, NIH 3T3 cells were injected into the device so that the desired cell density ($\sim$1000 cells/mm^2) was reached. If NIH 3T3 cells were the dominant species, they were

allowed to attach the glass bottom of the microfluidics devices first. MDA-MB-231 cells already loaded with CellTracker were then injected into the devices to reach the desired cell density. After incubating the microfluidics devices containing cell monolayers overnight, fluorescent calcium indicator was applied (Fluo4, Life Technologies) making the samples ready for imaging.

2.3 Results

To study the sensory responses of a multicellular network, we use single-channel microfluidic devices and deliver ATP solutions (Fig. 2.1B) to monolayers of fibroblast (NIH 3T3) cells. The ATP concentrations vary from 0 to 200μM, and the calcium dynamics of individual cells is obtained with fluorescent calcium indicator at 4 frames per second. We ensure the ATP arrival is sufficiently rapid to stimulate all cells at the same time (Fig. 2.1D).

We modulate the degree of communication in two ways. First, we vary the cell density. Smaller cell densities correspond to larger cell-to-cell distances, which we expect to reduce the probability of forming gap junctions. Second, we coculture the fibroblasts with breast cancer (MDA-MB-231) cells in the flow channel as shown in Fig. 2.1C. As we later show, MDA-MB-231 cells have reduced communication properties and therefore, act as defects in the multicellular network. To distinguish the two cell types, MDA-MB-231 cells are prelabeled with red fluorescent dye (Cell Tracker Red CMTPX; Life Technologies). Varying cell density and the fraction of cancer cells allow us to control the architecture of the multicellular network over a wide range.

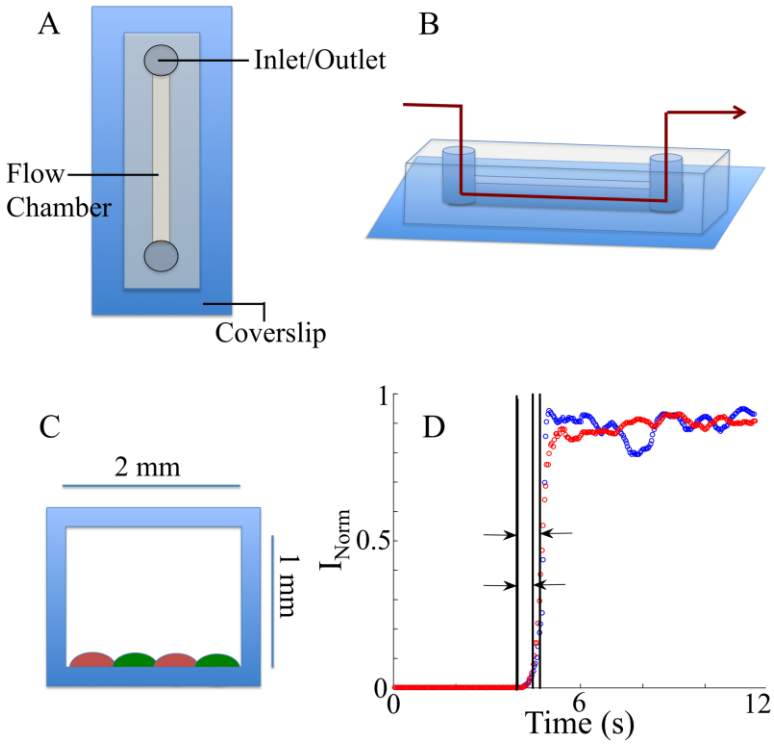

Figure 2.1 Design of the microfluidic device and stimuli arrival time (A) Top-view of microfluidic device showing glass slide, PDMS device, inlet/outlet, and connecting microfluidic chamber. (B) Side-view of microfluidic device indicating flow direction in red. (C) Cross-sectional view of device providing dimensions of flow layer and attached fibroblast (Green) and breast cancer (Red) cell monolayer. (D) Temporal profiles of chemical stimuli in the flow chamber evaluated using fluorescein. Two devices are tested (red and blue). Vertical lines correspond to arrival time and the times when half-maximum intensity reached.

Fig. 2.3A shows the composite image of a high-density cell monolayer with cocultured fibroblast and cancer cells. In this example, MDA-MB-231 cells make up a fraction F_C = 15% of the total population, which has a total cell density of ρ_T = 2,500 cells per 1 mm^2. At this density, each cell has an average of six nearest neighbors, from which extensive gap junction communication is expected. After identifying cell centers from the composite image (Fig. 2.2), we compute the time dependent average

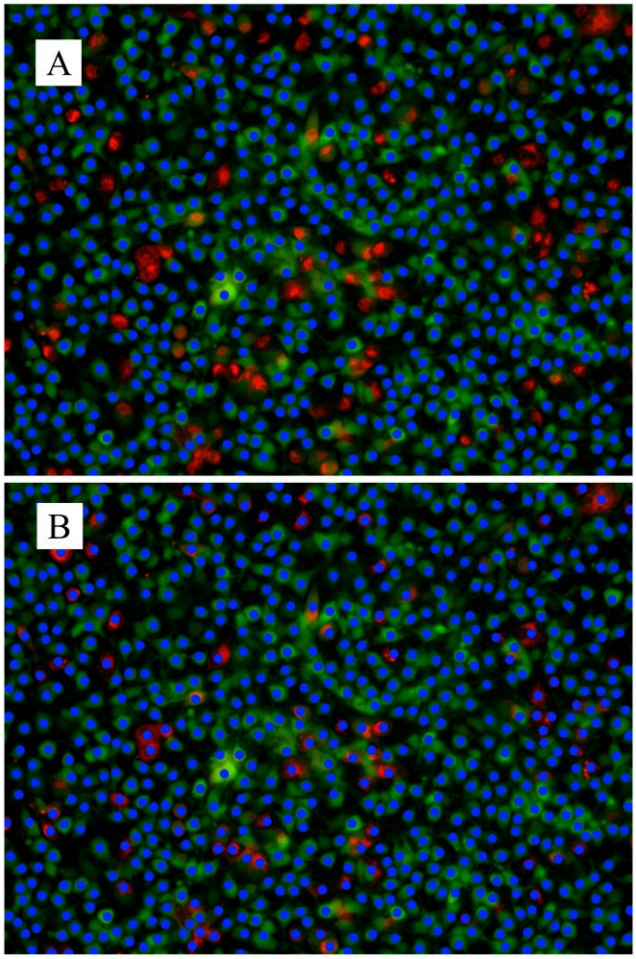

Figure 2.2 Determination of single cell centroids. (A) Composite image showing all cells (NIH 3T3 and MDA-MB-231, green), MDA-MB-231 (red), and manually determined centroids of NIH 3T3 cells (blue). (B) Composite image showing all cells (NIH 3T3 and MDA-MB-231, green), MDA-MB-231 (red), and manually determined centroids of all cells (NIH 3T3 and MDA-MB-231, blue).

fluorescent intensity near the center of each cell, which represents the instantaneous intracellular calcium concentration at the single-cell level.

2.3.1 Collective Response to ATP Stimuli

Typical responses of cells to excitation by ATP are shown in Fig. 2.3B. We see that, on average, higher concentrations of ATP trigger larger increases in calcium

levels. Cell-to-cell variations are significant; for example, response times as well as subsequent calcium dynamics of individual cells vary dramatically. In many cells, the initial calcium increase is followed by transient calcium oscillations. We quantify the oscillation propensity by computing the fraction of non-oscillating cells F_N using a peak-finding algorithm. We see in Fig. 2.3C that higher concentrations of ATP cause a larger percentage of cells to oscillate and thus, a smaller F_N.

The period of the oscillation is characterized by the interspike interval (ISI), which has been proposed to dynamically encode information about the stimuli[51,52]. To investigate the characteristics of ISI in the context of collective chemosensing, we study the statistics of the ISI from 30,000 cells (See Appendix C for further ISI information). Fig. 2.3D shows the histogram (event counts) of ISI values normalized by the number of cells of a typical experiment where the ATP concentration is 50μM. We see that the distribution is broad, which underscores the high degree of cell-to-cell variability in the responses. Fig. 2.3E summarizes the distribution at each ATP concentration using a box and whisker plot. We see that there is no significant dependence of the ISI on the ATP concentration. This observation is at odds with a familiar property of calcium oscillations, termed frequency encoding, in which the oscillation frequency (or ISI) depends on the strength of the stimulus[4,51,53,54]. However, we will see in the next section that the lack of a dependence here is likely caused by the high degree of cell-to-cell variability.

Finally, we characterize the spatial correlations of the ISI within the monolayer by computing the cross-correlation function C_{ISI} as a function of topological distance d between cells [defined by Delaunay triangulation[46], see Appendix C]. For each

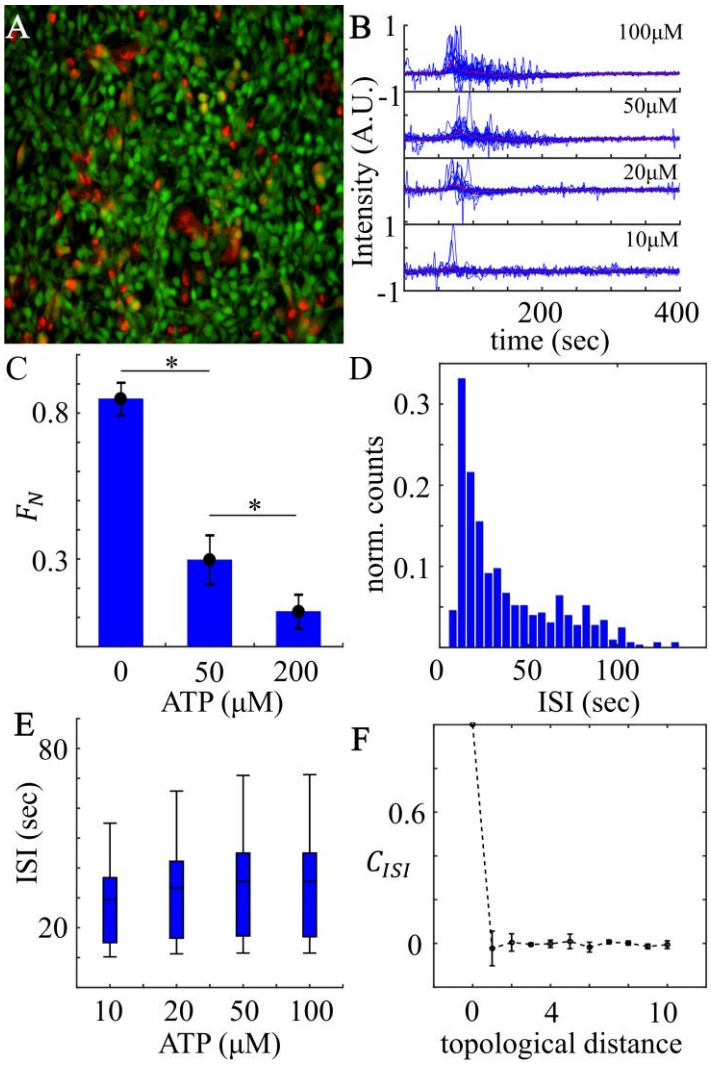

Figure 2.3 Calcium dynamics of cell monolayer in response to extracellular ATP. (A) Composite image showing the multicellular network of cocultured fibroblast (NIH 3T3) and breast cancer cells (MDA-MB-231). Green, fluorescent calcium signal for both cell types; red, MDA-MB-231. (B) Normalized fluorescence intensity profiles of one typical experiment for each ATP concentration tested. Blue, randomly selected single-cell calcium responses; red, average intensity profiles of all cells in each experiment. All time series begin ~50s before arrival of ATP stimuli. Intensity profiles of individual cells have been rescaled to [−1,1]. (C) Fraction of non-oscillating cells F_N as a function of ATP concentration at fixed cell density. Error bars: SEMs for n≥4. *P < 0.05. (D) ISI event counts normalized by number of cells for only NIH 3T3 cells. (E) Average experimentally measured ISI values of NIH 3T3 cells at varying ATP concentrations at fixed cell density. In B, C, and E, cell density ρ_T = 1,200±200 cells per 1 mm^2, and cancer cell fraction F_C = 15±6%. (F) ISI cross-correlation as a function of topological distance. Data from experiments with 50μM ATP at fixed cell density (ρ_T = 1,400±400 cells per 1 mm^2) and cancer fraction (F_C = 20±5%). Error bars show SDs from five experiments.

experiment, we compute the average ISI T_i for each oscillatory cell i. We then define $\delta T_i = T_i - \langle T_i \rangle$ and $C_{ISI}(d) = \langle \delta T_i \delta T_j \rangle_{D_{ij}=d} / \langle \delta T_i^2 \rangle$, where D_{ij} is the topological distance between cells i and j. Fig. 2.3F shows that C_{ISI} falls off immediately for $d > 0$. This observation is surprising, because one might hypothesize that communication between cells would result in ISI values for nearby cells being correlated. However, as described next, evidence from mathematical modeling suggests that this correlation is removed by the cell-to-cell variability.

2.3.2 Stochastic Modeling of the Collective Response

To obtain a mechanistic understanding of the experimental observations, we turn to mathematical modeling. We develop a stochastic model of collective calcium signaling based on the works of Tang and Othmer[22,51]. Their model captures the ATP-induced release of IP_3, the IP_3-triggered opening of calcium channels, and the nonlinear dependence of the opening probability on the calcium concentration as illustrated in Fig. 2.4A. The model neglects more complex features of calcium signaling observed in some cell types, such as upstream IP_3 oscillations[55,56] and spatial correlations among channels[3,57]. The model predicts that, at a critical ATP concentration, the calcium dynamics transitions from non-oscillating to oscillating. However, it was previously only analyzed deterministically for a single cell[22,51]. Therefore, we extend it to include both intrinsic noise and cell–cell communication via calcium exchange. We also explicitly include the dynamics of IP_3, which has a constant degradation rate and a production rate α that we take as proportional to the ATP concentration. We simulate

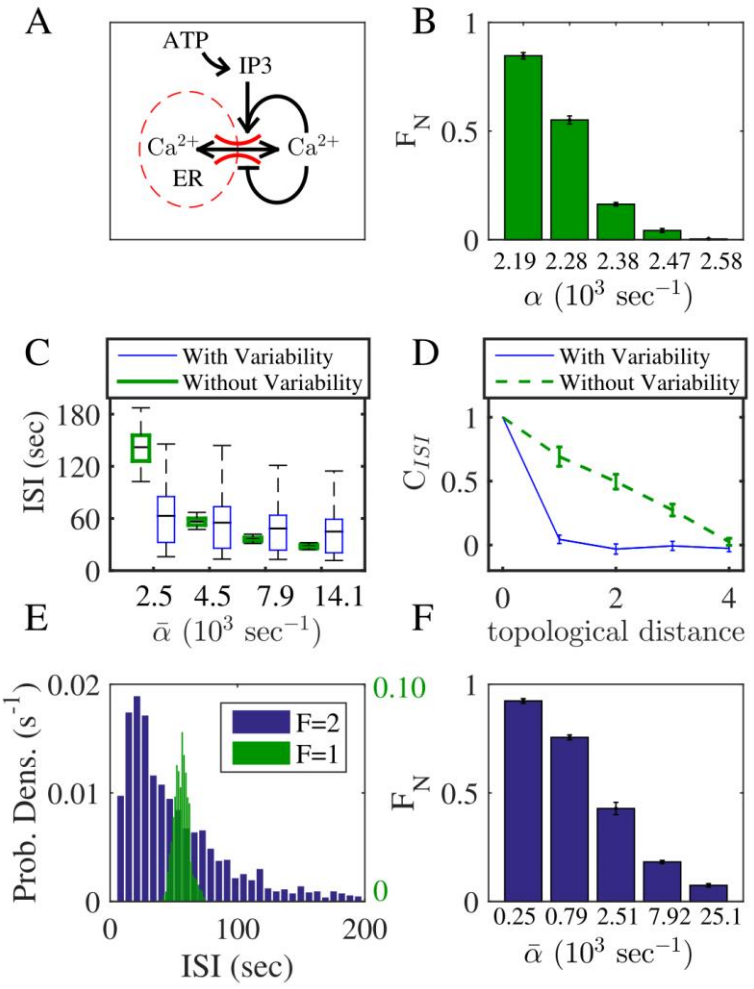

Figure 2.4 Model development and validation. (A) Schematic of the model. ATP stimulates IP_3 release at rate α, and IP_3 acts jointly with Ca^{2+} to open the endoplasmic reticulum (ER) calcium channels (positive feedback), whereas additional Ca^{2+} binding closes channels (negative feedback). Communication is modeled via diffusion of Ca^{2+} between adjacent cells. (B) Fraction of non-oscillating cells F_N as a function of ATP-induced IP_3 production rate α. (C) ISI decreases with $\bar{\alpha}$ (green). The decrease is severely weakened by cell-to-cell variability (blue). (D) ISI cross-correlation as a function of topological distance d (green). Cell-to-cell variability removes correlations for $d > 0$ (blue). (E) Distribution of ISI values (green). Cell-to-cell variability significantly broadens distribution (blue). (F) F_N vs. $\bar{\alpha}$ with cell-to-cell variability. Cells are simulated on a (B, C, E, and F) 3×3 or (D) 7×7 grid with density (B, D, and F) ρ_T =2.5×10^3 or (C and E) 1.4×10^3 mm^{-2}. In B and F, error bars are SEMs for n=5 subsamples.

the dynamics using the Gillespie algorithm[28], and we vary the density ρ_T of cells on a square grid, which modulates the degree of communication.

Fig. 2.4B shows the dependence of F_N on α, the model analog of the ATP concentration. Consistent with the experimental findings in Fig. 2.3C, we see that F_N decreases with α. In the model, the decrease is caused by the fact that intrinsic noise broadens the transition from the non-oscillating to the oscillating regime. Thus, instead of a sharp transition from F_N =1 to F_N =0 as predicted deterministically, the transition occurs gradually over the range of α shown in Fig. 2.4B. Fig. 2.4C shows the dependence of the ISI on α in the model (Fig. 2.4C, green box plots). We see that the ISI decreases with α, which is expected, because frequency encoding is a component of the Tang-Othmer model[22,51]. However, this property is not consistent with the experimental observation in Fig. 2.3E, where the ISI shows no clear dependence on ATP concentration. Furthermore, Fig. 2.4D shows the dependence of the correlation function C_{ISI} on the topological distance d in the model (green dashed curve in Fig. 2.4D). We see that C_{ISI} decreases gradually with d, indicating nonzero spatial correlations in the ISI, again inconsistent with the experimental findings (Fig. 2.3F).

Motivated by the high level of cell-to-cell variability evident in Fig. 2.3B and D, we hypothesize that cell-to-cell variability is responsible for these discrepancies between the model and the experiments. Indeed, inspecting the ISI histogram from the model reveals a very narrow distribution of ISI values, as seen in Fig. 2.4E, green bars, which is in contrast to the broad distribution observed experimentally in Fig. 2.3D. To incorporate cell-to-cell variability, we allow the model parameters to vary from cell to cell. Lacking information about the susceptibility of particular parameters to variation,

we allow all model parameters to vary by the same maximum fold change F. F is found by equating the variance of the resulting ISI distribution with that from the experiments, which yields F=2. As seen in Fig. 2.4E, blue bars, the resulting ISI distribution is consistent with that observed in Fig. 2.3D in both width and shape.

We see in Fig. 2.4C, blue box plots, that including cell-to-cell variability in the model severely weakens the decrease of the ISI with $\bar{\alpha}$, therefore agreeing with the experimental results shown in Fig. 2.3E (with variability, $\bar{\alpha}$ is defined as the mean of the α values sampled for each cell). We also see in Fig. 2.4D, blue curve, that variability removes the correlation C_{ISI} for $d > 0$, which is consistent with the immediate fall off observed experimentally in Fig. 2.3F. Importantly, even with variability, the decrease of F_N with α seen in Fig. 2.4B persists, as shown in Fig. 2.4F. This decrease remains consistent with the experimental observation in Fig. 2.3C. Indeed, variability significantly broadens the range of $\bar{\alpha}$ values over which the transition occurs, as expected (compare Fig. 2.4B and F), which is consistent with the broad range over which the transition occurs experimentally (Fig. 2.3C).

2.3.3 Effects of Communication on the Sensory Response

Having validated the model, we now use it to make predictions about the effect of cell–cell communication on collective calcium dynamics. Communication in the model is controlled by cell density, with higher density leading to more cell-to-cell contacts and thus, a higher degree of communication. Therefore, we first investigate the dependence of the oscillation propensity on the cell density. Fig. 2.5A shows F_N as a function of both cell density ρ_T and the ATP-induced IP_3 production rate $\bar{\alpha}$. We see

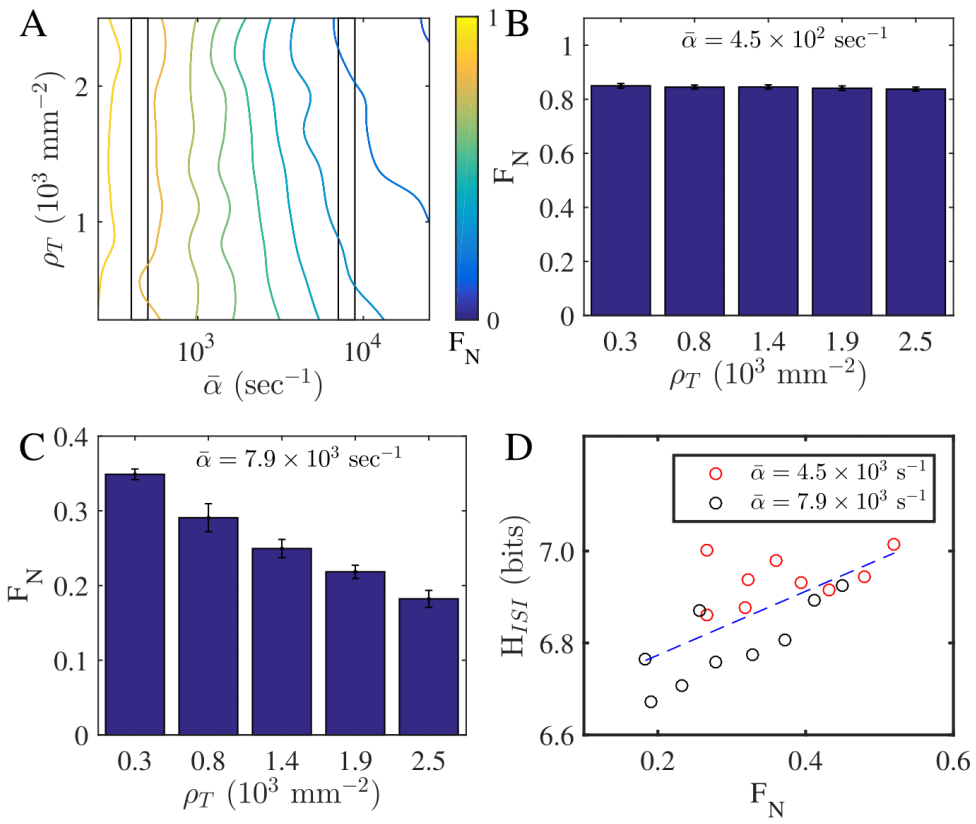

Figure 2.5 Model predictions. (A) Fraction of non-oscillating cells F_N as a function of cell density ρ_T and ATP-induced IP_3 production rate $\bar{\alpha}$. Left and right boxes correspond to B and C, respectively. (B) At small $\bar{\alpha}$, F_N is large and density-independent. (C) At intermediate $\bar{\alpha}$, F_N decreases with density. In B and C, error bars are SEMs for n=5 subsamples. (D) Entropy of ISI distribution H_{ISI} increases with F_N.

that the fraction of non-oscillating cells transitions from F_N =1 to F_N =0 as a function of $\bar{\alpha}$ and that there is also a dependence of F_N on ρ_T. At low $\bar{\alpha}$, F_N is everywhere large and independent of ρ_T (Fig. 2.5 A, Left and B). However, at intermediate $\bar{\alpha}$, F_N is a decreasing function of ρ_T (Fig. 2.5 A, Right and C). In this regime, increasing the cell density causes more cells to exhibit oscillatory calcium dynamics (thus decreasing F_N), even with a fixed sensory stimulus $\bar{\alpha}$. At large $\bar{\alpha}$ (beyond the range shown in Fig.

2.5A), we have checked that the non-oscillating fraction is driven to low values as expected, and the density dependence of F_N is weakened.

The prediction in Fig. 2.5C is striking, because it implies that cell–cell communication causes more cells to oscillate, even while cell-to-cell variability causes their ISI values to be spatially uncorrelated (Fig. 2.4D). Therefore, we wondered whether communication would have an effect on the width of the ISI distribution in this regime. The width or more generally, the amount of uncertainty in the ISI distribution is characterized by the entropy. For a continuous variable x, the entropy becomes the differential entropy defined as $H_{ISI} = \int \rho(x) log \rho(x) dx$, where $\rho(x)$ is the probability density. As seen in Fig. 2.5D, the entropy of the ISI distribution increases with F_N. This result indicates that, as communication decreases F_N, it also narrows the distribution of ISI values.

We now test these predictions in our experimental system. To test our predictions about how the non-oscillating fraction of cells should depend on cell density, we measure F_N as a function of ρ_T for various ATP concentrations. We see in Fig. 2.6A that, with no ATP, F_N is large at both low and high densities, and there is no statistically significant correlation between F_N and ρ_T . Then, we see in Fig. 2.6B that, at intermediate ATP concentrations (10–100μM), F_N significantly decreases with ρ_T . Finally, we see in Fig. 2.6C that, at large ATP concentration (200μM), F_N is small at both low and high densities, and again, there is no statistically significant correlation between F_N and ρ_T . These results confirm the predictions in Fig. 2.3.

To test the prediction that the entropy of the ISI distribution increases with the non-oscillating fraction of cells, we measure H_{ISI} as function of F_N. As seen in

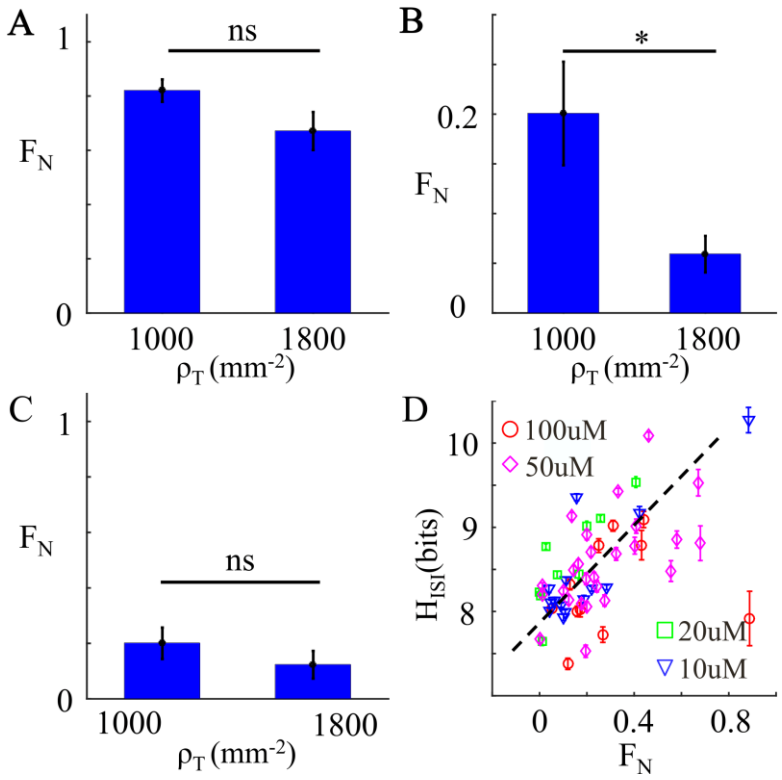

Figure 2.6 Experimental tests of model predictions. (A) Fraction of non-oscillating NIH 3T3 cells F_N as a function of cell density ρ_T when stimulated by 0μM ATP. Error bars: SEMs for $n > 4$. (B) As in A but with intermediate concentrations 10–100μM ATP. *P < 0.05. (C) As in A but with 200μM ATP. In A–C, the cancer cell fraction is fixed at F_C = 15±6%. (D) F_N is positively correlated with the differential entropy of ISIs H_{ISI} (ρ_T between 600 and 2,500 cells per 1 mm^2; F_C between 5% and 80%). Error bars represent SDs of 1,000 bootstrap resampled results. ns, not significant.

Fig. 2.6D, H_{ISI} increases with F_N, consistent with the prediction in Fig. 2.3D. This result implies that increasing the degree of communication narrows the distribution of ISI, making the ISI values less variable across the population. We have also checked that the entropy of the distribution of cross-correlation values for nearest neighbors' entire calcium trajectories C_{NN}[19,46] decreases as a function of cell density. Together, these results imply that cell– cell communication has a significant effect on the collective sensory response. This finding is especially striking given the strong effects of cell-to-

cell variability (Fig. 2.3 E and F). We conclude that the effects of communication observed here persist, despite extensive variability.

2.3.4 Effect of Cancer Cell Defects

We have seen that increasing cell density increases the propensity of cells to oscillate in response to an ATP stimulus. This behavior is consistent with our model, which predicts that the mechanism is through increased cell–cell communication. However, it could be in the experiments that increasing the cell density introduces other effects beyond increased gap junction communication, such as mechanical coupling between cells or coupling to the substrate[58]. To modulate the communication directly, we vary the fraction F_C of cancer cells with which the fibroblasts are cocultured, while keeping the density of all cells fixed. Because cancer cells are known to have reduced gap junction communication[48–50], we expect the fraction of non-oscillating cells F_N to have the opposite dependence on F_C that it does on cell density (Fig. 2.6B).

We first investigate whether MDA-MB-231 cells indeed have reduced communication in our system. Figure 2.7A shows several examples of single-cell calcium dynamics for NIH 3T3 and MDA-MB-231 cells in a typical experiment. We see that both cell types exhibit immediate increases in cytosolic calcium levels at the arrival of ATP, but cancer cells typically show long relaxation times, whereas fibroblast cells tend to more often exhibit oscillations after stimulation. These qualitative features are maintained across all ATP concentrations. Figure 2.7B shows a comparison of the

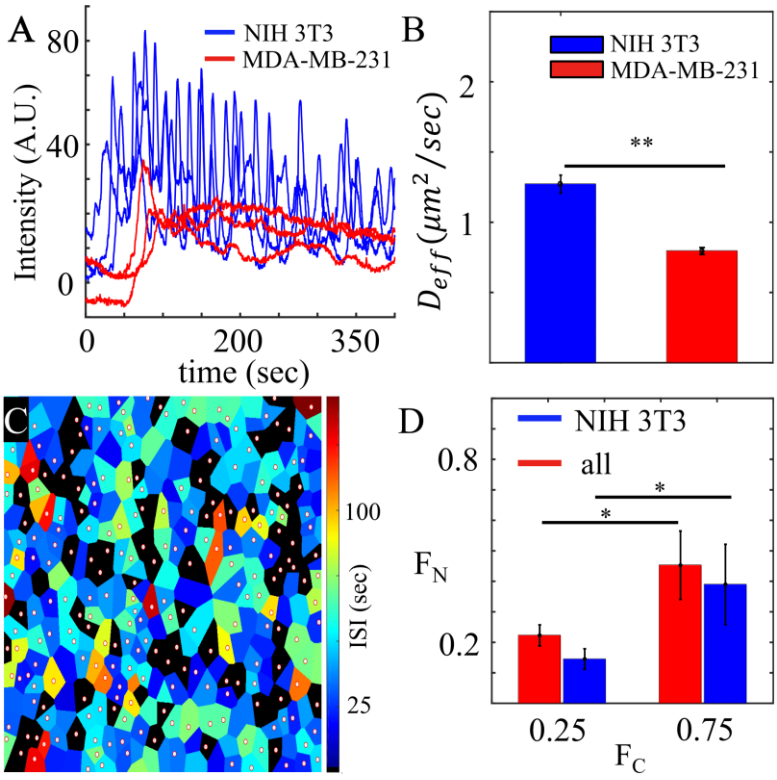

Figure 2.7 Effects of cancer cell defects on collective response. (A) Typical fluorescence intensity profiles showing the calcium dynamics on the single cell level, where basal-level intensity has been subtracted. For each cell, basal-level intensity is estimated by averaging 100s of its fluorescent intensity before ATP arrival (ATP concentration = 50μM; ρ_T = 2,400 cells per 1 mm^2; F_C = 12%). (B) Fluorescence recovery after photobleaching experiments confirm that MDA-MB-231 cells have weaker gap junction communication compared with NIH 3T3 cells (error bars: SEMs for $n > 100$). **P < 0.01. (C) Spatial map of average ISI of each individual cell. ATP concentration is 50μM. Black, non-oscillating cell; circle, MDA-MB-231 cell. (D) When stimulated by an intermediate range of ATP concentrations (10–100μM), the fraction of non-oscillating cells F_N increases with increased cancer fraction F_C at fixed total cell density (ρ_T = 1,200±200 cells per 1 mm^2). *P < 0.05. Blue, fraction of non-oscillating NIH 3T3 cells; red, fraction of non-oscillating cells including both cell types.

intercellular diffusion coefficients in the two cell types obtained from a fluorescence recovery after photobleaching analysis[59]. We see in Fig. 2.7B that gap junction-mediated diffusion between MDA-MB-231 cells is significantly weaker than that between NIH 3T3 cells, consistent with previous reports[48–50]. Therefore, it is evident

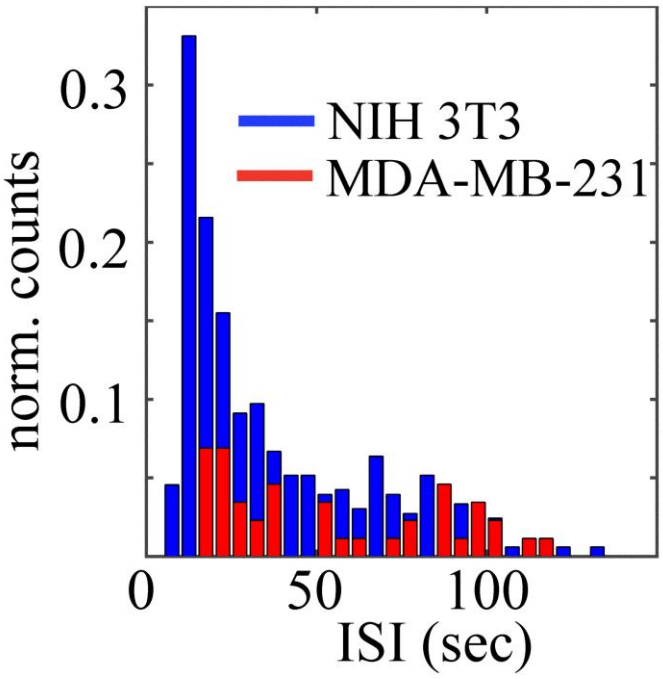

Figure 2.8 ISI characteristics of a typical experiment. ISI event counts normalized by number of cells. Blue: statistics of only NIH 3T3 cells, Red: statistics of only MDA-MB-231 cells.

that MDA-MB231 cells can be treated as communication defects in the cocultured multicellular network. Indeed, Fig. 2.7C shows the spatial distribution of these defects in the monolayer. In Fig. 2.7C, the mean ISI for each cell is shown in color, with non-oscillating cells in black. We see that cancer cells, labeled by white circles, are more likely to be non-oscillating, which is consistent with the qualitative characteristics shown in Fig. 2.7A. We have further quantified the distinction between the two cell types in Fig. 2.8 where we show using the distributions of ISI values that oscillatory events are at least five times less likely to occur for the MDA-MB-231 cells.

Having established that the presence of cancer cells reduces the degree of cell–cell communication in the monolayer, we now vary the fraction of cancer cells and measure the oscillation propensity of the remaining fibroblasts. Figure 2.7D shows the

non-oscillating fraction of fibroblasts F_N (blue bars) as a function of the cancer cell fraction F_C for a typical experiment at fixed cell density (ρ_T = 1,200±200 cells per 1 mm^2). We see that F_N significantly increases with F_C. We also see that F_N for all cells (both fibroblasts and cancer cells) (red bars in Fig. 2.7D) significantly increases with F_C and that, as expected, F_N is larger for all cells than for just fibroblasts. These findings imply that reduced cell–cell communication decreases the propensity for calcium oscillations, which is consistent with the effects of varying cell density (Fig. 2.6B). Finally, we also investigate the effect of cancer cells on the entropy of the ISI distribution. H_{ISI} is higher for cells that are surrounded by a large number of cancer cells and lower for cells with pure fibroblast neighbors. In the latter case, H_{ISI} also increases as the number of nearest neighbors decreases. These findings imply that reduced cell–cell communication increases the entropy of the ISI values, even at the local level of a cell's microenvironment, which is consistent with the effects seen in Fig. 2.6D. Taken together, we conclude that the calcium dynamics of individual cells is strongly regulated by the degree of gap junction communication inside the cell monolayer.

2.4 Discussion

We have characterized the collective calcium dynamics of multicellular networks with varying degrees of cell–cell communication when they respond to extracellular ATP. We have found that increasing the ATP stimulus increases the propensity for cells to exhibit calcium oscillations, which is expected at the single-cell level. However, we have also found that increasing the cell density alone, while keeping

the stimulus fixed, has a similar effect, revealing a purely collective component to the sensory response. Modeling suggests that this effect is caused by an increased degree of molecular communication between cells. In line with this prediction, we have found that increasing the fraction of cancer cells in the monolayer reduces the oscillation propensity, because cancer cells act as defects in the communication network. Based on these results, we conclude that the collective sensory response, in which nonlinear signaling dynamics is coupled with strong intrinsic and extrinsic noise, encodes both stimulus strength and degree of communication.

Our results suggest that the calcium response to extracellular ATP encodes multiplexed information under physiological conditions. Typical plasma and pericellular concentrations of ATP in animals and human have been reported to range from submicromolar to tens of micromolar[60–62], whereas hundreds of micromolar have been associated with tumor because of the hypoxia microenvironment[63]. The concentration range of ATP in Fig. 2.4B is associated with several physiological phenomena, including immunomodulation[64–66], traumatic shock[66], and platelet activation[67]. Within this range, our results show that calcium dynamics encodes both stimuli strength in the magnitude of intracellular calcium concentration (Fig. 2.3B) and cell density in the propensity of calcium oscillation (Fig. 2.6B). Such multiplexing has been shown to be possible with simple biochemical networks[68], and it is thought to underlie the ability of single networks to respond with specificity to multiple inputs, such as neuronal growth factor and EGF in the rat PC-12 system[69]. A possible reason for multiplexing is that it is beneficial for the responses to each input to be dependent on each other[70], which in our case, suggests a benefit for a collective component to the

ATP sensory response. The ways in which dynamic information is stored in and extracted from cellular signals are a topic of ongoing research[32,36].

Recent experiments have put our results in the context of a unique paradigm of cell signaling: cells may decode information from the dynamics and not just the magnitude of signaling molecules[71]. For instance, UV and γ-radiation differentially trigger non-oscillatory and oscillatory p53 dynamics[72]. Similarly, when endothelial cells are stimulated by VEGF, non-oscillatory and oscillatory calcium dynamics leads to migration and proliferation, respectively[73]. In light of these developments, our results suggest that cell density, via gap junctional communication and nonlinear signaling dynamics, can impact cellular function, similar to so-called dynamical quorum sensing[74–76].

Our results suggest that the dependence of the calcium response on both sensory and collective parameters persists, despite significant cell-to-cell variability. Certain measures are robust to variability, such as the oscillation propensity and the entropy of the ISI distribution, whereas others are not, such as spatial correlations in the ISI and its dependence on the ATP input (frequency encoding). This result implies that our main finding of communication-dependent sensing is generic, because it persists despite large variability, but that traditional measures of information processing, such as frequency encoding, may have to be rethought in contexts where cell-to-cell variability is pronounced. It is becoming increasingly understood that variability is common in cell populations, and recent examples suggest that it may even be beneficial. For example, recent studies in a related system (NF-κB oscillations in fibroblast

populations) also found a large degree of cell-to-cell variability[77] and showed that this variability allows entrainment of the population to a wider range of inputs[78].

In our model, the transition from the non-oscillatory to the oscillatory regime occurs because of a saddle-node bifurcation, a critical point in parameter space where the number of dynamical fixed points changes. This transition is broadened by intrinsic noise and cell-to-cell variability into a critical "region," and cell–cell communication causes the oscillation propensity to depend on cell density within this region (Fig. 2.3A). Our finding that this region is broad and our suggestion that it may be of some functional use for the system resonate with recent studies that have argued that biological systems are poised near critical points in their parameter space[79–81]. The connection between dynamical criticality, as in our model, and criticality in many-body statistical systems remains to be fully explored.

Gap junctional communications exist among many types of cells. Therefore, our results may have far-reaching implications for other biological model systems, such as neuronal networks or cardiovascular systems. Because gap junctions mediate fast, nearest neighbor communication, we expect our conclusions to also hold for 3D cell aggregates, such as tissue organoids. It will be interesting to explore whether distinctions in the calcium dynamics in these systems originate from differences in their degrees of cell–cell communication.

CHAPTER 3

Dynamic Sampling and Information Encoding in Biochemical Networks

Garrett Potter, Tommy Byrd, Andrew Mugler, and Bo Sun

Biophysical Journal, Volume 112, Issue 4, 795 – 804 (2017)

Cells use biochemical networks to translate environmental information into intracellular responses. These responses can be highly dynamic, but how the information is encoded in these dynamics remains poorly understood. Here, we investigate the dynamic encoding of information in the ATP-induced calcium responses of fibroblast cells, using a vectorial, or multi-time-point, measure from information theory. We find that the amount of extracted information depends on physiological constraints such as the sampling rate and memory capacity of the downstream network, and it is affected differentially by intrinsic versus extrinsic noise. By comparing to a minimal physical model, we find, surprisingly, that the information is often insensitive to the detailed structure of the underlying dynamics, and instead the decoding mechanism acts as a simple low-pass filter. These results demonstrate the mechanisms and limitations of dynamic information storage in cells.

3.1 Introduction

Cells utilize cascades of biochemical pathways in order to translate environmental cues into intracellular responses[82,83]. Due to extensive feedbacks and cross-talk among these signaling pathways[84–87], messenger molecules exhibit rich dynamic modes, such as waves, oscillations, and pulses. Recent work in cell biology has suggested a new perspective in cell signaling: the dynamics, or temporal profiles, of messenger molecules allow cells to encode and decode even more rich and complex information than static profiles do[88,89]. For instance, during inflammation response, exposure to tumor necrosis factor-α (TNFα) causes the transcription factor NF-κB to oscillate between the nucleus and cytoplasm of a cell[90], whereas bacterial lipopolysaccharide (LPS) triggers a single wave of NF-κB within the cell[91]. Therefore, the dynamics of NF-κB encode the identity of external stimuli. In another example, stimulation of pheochromocytoma cells (PC12) cells by epidermal growth factor (EGF) leads to transient mitogen-activated protein kinase (MAPK) activation and cell proliferation, whereas stimulation by nerve growth factor (NGF) leads to sustained MAPK activation and cell differentiation[69]. These and other examples raise the question of how one quantifies the information carried by signaling dynamics.

Information theory provides a useful framework to address such questions[30,92,93]. In the simplest case, one calculates the scalar mutual information between states of extracellular stimuli (typically well-controlled discretized values) and states of the cell (typically protein concentrations measured at a certain time). Mutual information characterizes the correlation between environmental cues and cell responses, and conveniently expresses such correlations in units of bits. For example,

if the mutual information between an environmental stimulus and a cell response is measured to be $\log_2 2 = 1$ bit, it means that effectively only two stimulus levels can be resolved by the response variable; any further resolution is not possible given the shape of the stimulus-response curve and the noise in the system[93]. Similarly, if the mutual information is $\log_2 3 = 1.6$ bits, then three stimulus levels can be resolved, and so on. This framework has been successfully employed to quantitatively understand the amount of information that can be transmitted through a biochemical pathway (channel capacity)[36], mechanisms of mitigating errors[94], and design principles of signaling network architectures[95].

Recently, inspired by the fact that cells utilize dynamic signaling to encode and decode information, a multivariate, or vectorial, mutual information has been proposed[32]. In this new framework, cellular responses are described by vectors of dimension n, which consist of cellular states sampled at multiple time points. The vectorial mutual information is generally higher than the scalar mutual information. Indeed, the scalar mutual information can be thought of as the limit of the vectorial mutual information when the length of the vector is one, and therefore the vectorial information cannot be lower than the scalar information. This suggests that signaling dynamics allow richer content to be transmitted than static information processing alone. It has also been shown that sampling cellular states at multiple time points eliminates extrinsic noise—noise that degrades information due to cell-to-cell variability[32].

In light of these results, we ask what is the optimal strategy for cells to utilize the power of vectorial mutual information? How should a cell sample its own temporal

profiles? Can cells use vectorial mutual information to distinguish different dynamic states of the underlying signaling pathways? To address these questions, we combine experimental measurements of ATP-induced Ca^{2+} responses with theoretical analysis, to systematically study scalar and vectorial mutual information in a dynamic signaling system. We find that given different physiological constraints, the optimal sampling depends on the starting time, sampling rate and memory capacity. We characterize how vectorial information is affected by intrinsic and extrinsic noise, in both the experimental system and a simple physical model. Surprisingly, we find that vectorial mutual information is often insensitive to the detailed structure of the underlying dynamics, failing to distinguish between, for example, oscillatory and relaxation dynamics. We explain this observation by deriving the connections between vectorial and scalar information, which reveals that in a particular regime vectorial encoding acts as a simple low-pass filter.

3.2 Materials and Methods

3.2.1 Cell Culture and Sample Preparation

The samples were prepared following previously reported protocol[46]. Briefly, NIH 3T3 cells were cultured in standard growth mediums (Dulbecco's modified Eagle medium (DMEM) supplemented with 10% bovine calf serum and1% penicillin. To prepare samples, cells were detached from culture dishes using TrypLE Select (Life Technologies) and suspended in growth mediums before pipetted into the microfluidics devices and allowed to form monolayers. After incubating the flow chamber devices

containing cell monolayers overnight, fluorescent calcium indicator was applied (Fluo4, Life Technologies) making the samples ready for imaging.

3.2.2 Fluorescence Imaging and Data Analysis

Fluorescence was detected using an inverted microscope (Leica DMI 6000B) coupled with a Hamamatsu Flash 2.8 camera. Movies were taken at a frame rate of 1 frame/sec with a 20x oil immersion objective. Image analysis and data processing were performed in MATLAB.

3.3 Results

To investigate properties of dynamic encoding, we focus on the calcium dynamics of fibroblast cells in response to extracellular adenosine triphosphate (ATP), a common signaling molecule involved in a range of physiological processes such as platelet aggregation[96] and vascular tone[97]. ATP is detected by P2 receptors on the cell membrane, and triggers the release of second messenger inositol trisphosphate (IP_3). IP_3 activates the ion channels on the endoplasmic reticulum (ER) which allows free calcium ions to flux into the cytosol. The nonlinear interactions of Ca^{2+}, IP_3, ion channels and ion pumps generate various types of calcium dynamics which may lead to distinct cellular functions[57,71].

3.3.1 Quantifying Information in Experimental Dynamics

In order to measure the calcium dynamics of fibroblast cells in response to external ATP stimuli, we employ microfluidic devices for cell culture and solution

delivery as described previously[46,98]. In brief, NIH 3T3 cells (ATCC) are cultured in the PDMS (Polydimethylsiloxane) bounded flow channels as shown in Fig. 3.1A. After attaching the glass bottom for 24 hours, the cells are loaded with fluorescent calcium indicators (FLUO-4, Thermo Fisher Scientific) according to the manufacture's protocol. ATP solutions diluted by DMEM (Dulbecco's Modified Eagle Medium) into 10, 20, 50, and 100μM concentrations are sucked into the flow channel with a two-way syringe pump (New Era Pump Systems Inc.) at a rate of 90μL/min. At the same time, we record fluorescent images of the cell monolayer at 1 Hz for a total of 10 minutes (Hamamatsu Flash 2.8).

In all experimental recordings, ATP arrives at approximately t = 10 sec, and stays at a constant concentration. Since most responses happen within 2 minutes, we use the first 160 seconds of recording for subsequent analysis. The time-lapse images are postprocessed to obtain the fluorescent intensity $I_i(t)$ of each cell i at a given time t. We define the calcium response as $R_i(t) = [I_i(t) - I_i^r(t)]/I_i^r$, where I_i^r is a reference obtained by averaging the fluorescent intensity of cell i before ATP arrives (Fig. 3.1B).

In order to quantify the information encoded in the calcium dynamics of fibroblast cells in response to ATP, we have analyzed a total of more than 10,000 cells over 4 different ATP concentrations (10, 20, 50, 100μM) as inputs. With the underlying assumption that each input appears at probability of 1⁄4, the same number of cells are selected for each ATP concentration. Since the mutual information is bounded from above by the minimum of the entropies of the two variables, the maximum possible

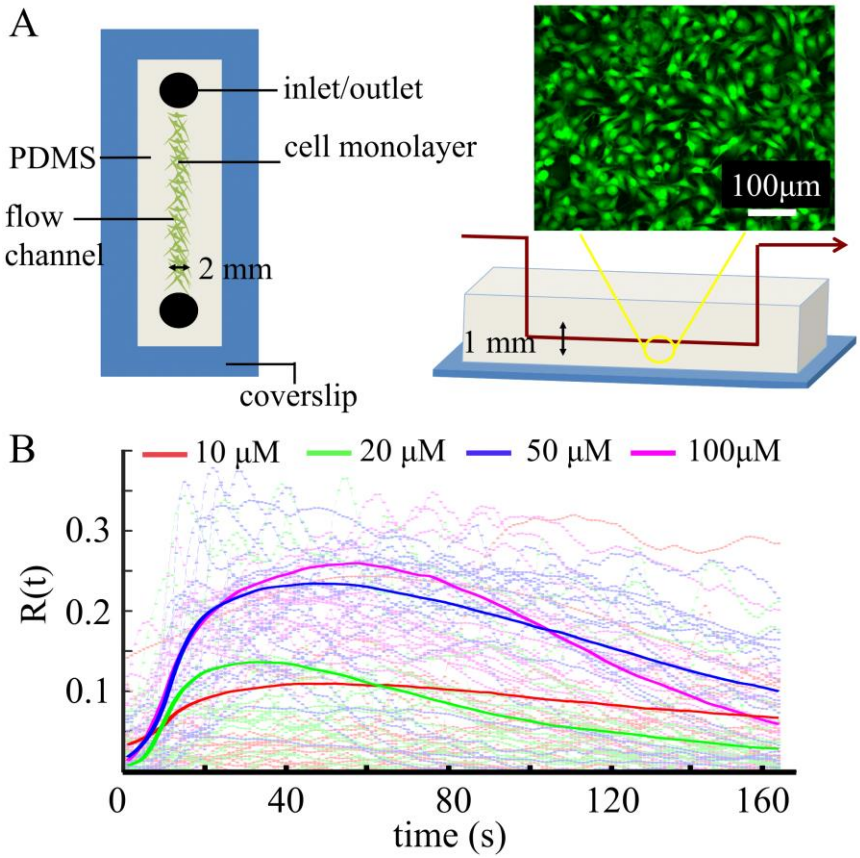

Figure 3.1: Schematics of experimental setup. (A) The top and cross view of the microfluidics device to deliver ATP solution to cultured fibroblast (NIH 3T3) cells. Inset: fluorescent calcium imaging of a typical experiment. (B) Relative fluorescent intensities indicating the calcium dynamics $R_i(t)$ of individual cells (dashed lines) and their average (solid lines) when stimulated by external ATP at four different concentrations.

mutual information between the input and output is the entropy of the input, or $\log_2 4 = 2$ bits.

The mutual information can be written in terms of the joint probability distribution between input and output variables, or equivalently as a difference between entropies[31]. In our case, we will find the latter more convenient. Denoting the dynamic calcium response as $R_i^{\alpha}(t)$, where α = 1,2,3,4 for each ATP concentration, and $i = 1,2,\ldots,N$ for each cell ($N \sim 2{,}500$), the scalar mutual information is defined as

$$MI_s(t) = H[\{R(t)\}] - \frac{1}{4}\sum_{\alpha} H[\{R^{\alpha}(t)\}], \tag{3.1}$$

where H represents the differential entropy. The first term is the unconditioned entropy calculated from cellular responses at time t of all four ATP concentrations. The second term is the average of differential entropy conditioned at each ATP concentration. Intuitively, the scalar mutual information $MI_s(t)$ measures how much the entropy (uncertainty) in the output (cellular responses) is reduced by knowledge of the input (ATP concentration), and therefore how much information one variable contains about the other. It is a function of the time t at which we take a snapshot of the system and evaluate the differential entropy across the ensemble of cells.

The vectorial mutual information is defined as

$$MI_v(t_s) = H\left[\{R(\vec{t})\}\right] - \frac{1}{4}\sum_{\alpha} H\left[\{R^{\alpha}(\vec{t})\}\right], \tag{3.2}$$

where $\vec{t} = (t_s, t_s + r^{-1}, t_s + 2r^{-1}, \dots, t_s + T_d)$. When generalizing to the vectorial mutual information MI_v, one has to specify not only the sampling start time t_s (equivalent to the time t in the case of MI_s), but also the sampling duration T_d and the sampling rate r, which opens the possibility of complex sampling strategies. In the time between t_s and $t_s + T_d$, a fibroblast cell sampling its calcium concentration at a rate r accumulates a vectorial representation of its calcium dynamics with vector dimension $n = 1 + rT_d$. Since the cell has to store the vector for further processing, n also represents its memory capacity.

In order to calculate the scalar and vectorial mutual information, we employ the k-nearest neighbor (kNN) method to estimate the differential entropies[32–34], as we have

done in our previous work[23]. This method does not require binning of data, and it has been shown to converge quickly even with a small number of data points[33]. The kNN method makes the approximation that the probability density $\rho(x)$ around a data point x is uniform within the n-dimensional sphere that encloses exactly k other data points. As a result, the performance of kNN method degrades particularly for distributions that have long tails or large spatial gradients. Despite these limitations, the kNN method is widely employed in quantitative biology for its ease of implementation and superior performance. That being said, some of the errors will cancel when calculating the mutual information[32,33], and we have taken $k = 50$ (square root of the sample size) based on the suggestions of [34].

3.3.2 Dynamic Encoding Increases Information

We first consider the situation where the sampling duration T_d is fixed. Fig. 3.2 shows the mutual information of both scalar (MI_s) and vectorial encoding (MI_v) from fibroblast calcium dynamics for T_d = 30 sec (A, C, E) and T_d = 60 sec (B, D, F). As seen in Fig. 3.2A and B, MI_s first rises, then falls, as a function of time. This is due to the separation, then convergence, of the four ATP-conditioned responses as a function of time, as seen in Fig. 3.1B: better-separated responses contain more information about the ATP level. This shape is also reflected in MI_v, with additional smoothing due to the repeated sampling.

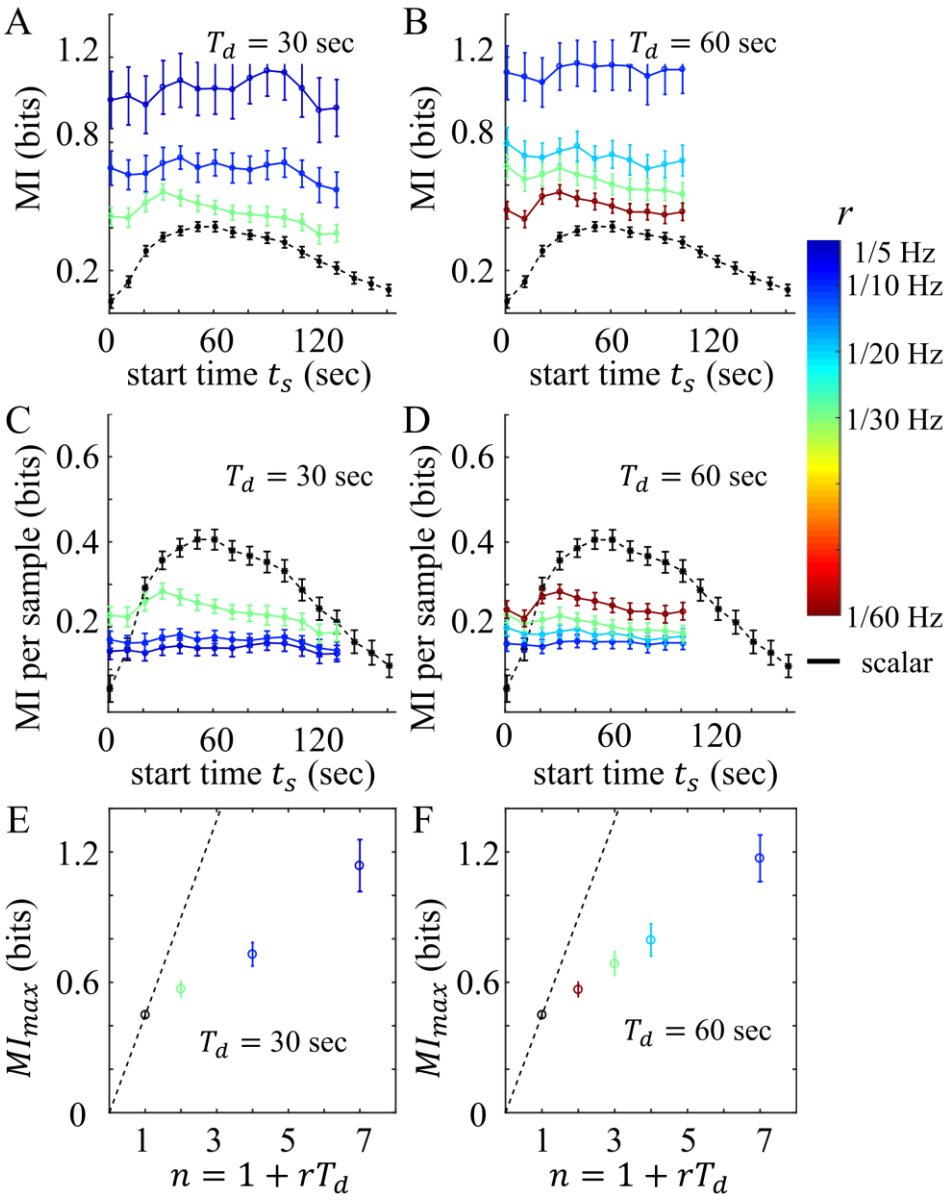

Figure 3.2: Information carried by calcium dynamics of fibroblast cells in response to ATP, for fixed sampling duration T_d. (A, B) Vectorial mutual information MI_v as a function of sampling start time t_s at different sampling rates r (color bar), for (A) T_d = 30 sec and (B) T_d = 60 sec. Black curve is scalar mutual information MI_s at each time point. (C, D) Mutual information per sample for the same conditions as A, B. (E, F) Maximum MI_v over all t_s values, as a function of the memory capacity n, for (E) T_d = 30 sec and (F) T_d = 60 sec. Maximum MI_s is plotted at n = 1. Error bars in A-F represent the means and standard deviations of 100 bootstraps.

Fig. 3.2A and B also show that MI_v increases with sampling rate r. This is intuitive, since a larger sampling rate produces a larger number of samples $n = 1 + rT_d$, which increases the amount of information extracted from the dynamics. While the results in Fig. 3.2A and B are intuitively expected, it is also important to know the efficiency for dynamic encoding. To this end, we have calculated the mutual information per sample, defined as MI_v/n, as shown in Fig. 3.2C and D. It is evident that higher coding efficiency is achieved at smaller sampling rate. This is because when the sampling rate is large, samples are spaced closely in time, and therefore contain increasingly redundant information, which lowers the coding efficiency. The results shown in Fig. 3.2C and D suggest that although dynamic encoding mitigates intrinsic noise, it is not enough to allow MI_v to grow faster than linearly with n. Indeed, scalar encoding generally offers better efficiency than vectorial encoding: as shown in both Fig. 3.2C and D, $MI_s(t_s) > MI_v(t_s)/n$, except at very early times when the cellular response has just started.

The results of Fig. 3.2A-D are summarized in Fig. 3.2E and F, which plot MI_{max}, the maximum mutual information over all possible sampling start times t_s. As seen in Fig. 3.2E and F, MI_{max}, monotonically increases with n, which shows that dynamic encoding improves the information capture. However, the increase is sublinear, i.e. below the dashed line defined by the scalar mutual information, which shows that the efficiency of dynamic encoding decreases with vector length n. Considering scalar encoding as the limiting case of $r \to 0$, we conclude that as the sampling rate increases, mutual information increases but the coding efficiency per measurement decreases.

3.3.3 Dynamics Determine Optimal Sampling Rate

Cells have limited ability to process dynamically encoded information. It is conceivable that a biochemical signaling network processing a vectorial code of high dimension will be complex and expensive, because it requires a high memory capacity n for storage and transfer. Therefore, a relevant question is, what sampling strategy can a cell apply when the memory capacity is fixed? Fig. 3.3A and B show the mutual

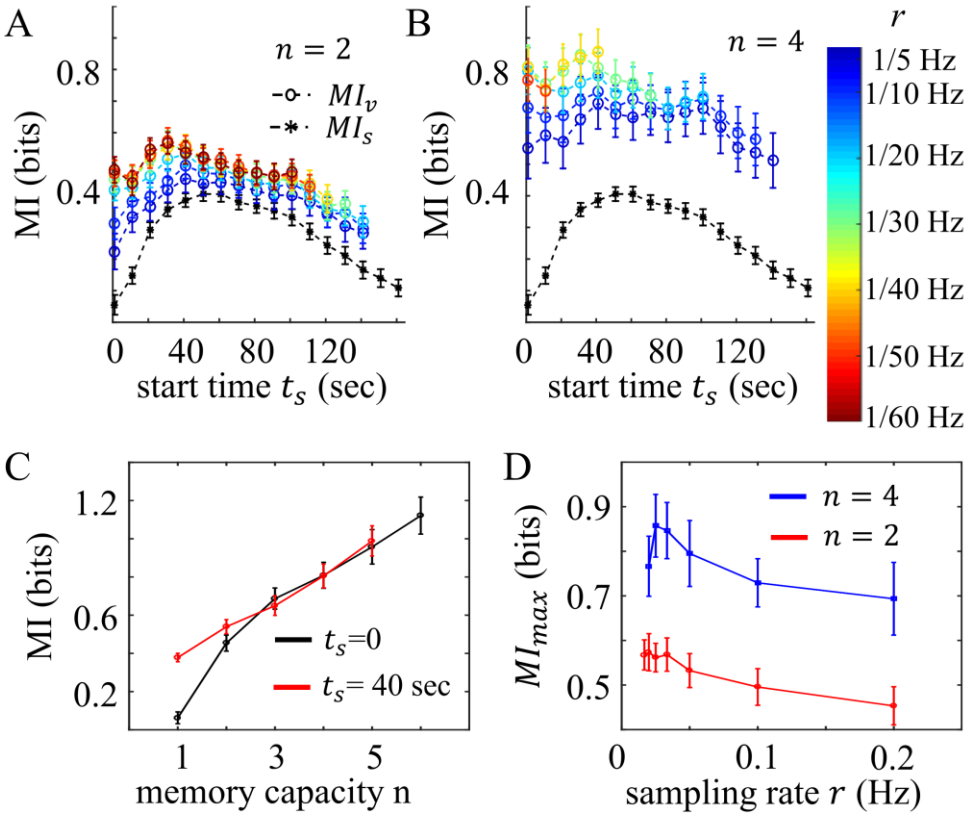

Figure 3.3 Information carried by calcium dynamics of fibroblast cells in response to ATP, for a given memory capacity n. (A, B) Vectorial mutual information MI_v as a function of sampling start time t_s at different sampling rates r (color bar) for (A) n = 2 and (B) n = 4. Black curve is scalar mutual information MI_s each time point. (C) MI_v as a function of n at fixed sampling rate r = 1/30 Hz and sampling start time (black curve t_s = 0, red curve t_s = 40 sec). n =1 corresponds to MI_s. (D) Maximum MI_v over all t_s values, as a function of r, for fixed memory capacity (red n = 2, blue n = 4). Error bars in A-D represent the means and standard deviations of 100 bootstraps.

information as a function of sampling start time t_s when the memory capacity n is fixed, while the sampling rate r, and therefore the duration $T_d = (n-1)/r$, are allowed to vary. Comparing Fig. 3.3A to B, we see that larger memory capacity n generally allows more information to be transmitted, as was the case in Fig. 3.2. This trend is quantified in Fig. 3.3C, which plots the mutual information as a function of n, for fixed sampling rate and at two particular starting times t_s. We see that the amount of information significantly depends on t_s for small n, while the difference diminishes at larger n. Therefore we see that larger memory capacity not only encodes higher information, but also helps cells to obtain more uniform readouts. We suspect that the convergence is due to the fact that information is upper-bounded (at 2 bits in our case), which requires that all curves, regardless of t_s, ultimately saturate with increasing n. Although our current sample size is not large enough to calculate MI_v at larger n, the saturation of MI_v has been shown in[32] to occur at around n = 12 for ATP-induced calcium dynamics.

We also see in Fig. 3.3A and B that for a given n, there is an optimal sampling rate r that maximizes the information. This is made more evident by considering, as before, the maximum mutual information MI_{max}, over all possible start time t_s, which is plotted as a function of r in Fig. 3.3D. Particularly, for n = 4 (blue curve), we see that MI_{max}, is maximal at a particular sampling rate. This is because, for a fixed number of samples n, sampling too frequently results in redundant information, as discussed above; while sampling too infrequently places samples at late times, when the dynamic responses have already relaxed (see Fig. 3.1B). Therefore it is generally beneficial to sample at a lower rate except when the sampled points are too far apart,

which places samples outside the 'high yield' temporal region. The tradeoff between these two effects leads to the optimal sampling rate, where the information gathered is the largest.

3.3.4 Vectorial Information is Insensitive to Detailed Dynamic Structure

Is vectorial encoding sensitive to the underlying details of the dynamic response? In order to answer this question, and to provide a mechanistic understanding of dynamic information transmission in biochemical networks, we construct a minimal stochastic model with the aim of recapitulating the key features of the fibroblast response. As a minimal model we consider a damped harmonic oscillator in a thermal bath, driven out of equilibrium by a time-dependent forcing $F(t)$. The magnitude of the external forcing is proportional to a scalar input, which is analogous to the ATP concentration. The displacement of the particle $x(t)$, like the calcium dynamics, can then be analyzed to infer the information that the oscillator encodes about the input.

The equation of motion for the oscillator is given by the Langevin equation[99]

$$m\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + kx = g_\alpha F(t) + \psi(t),$$

$$\langle \psi(t)\psi(t')\rangle = 2k_B T\gamma\delta(t-t'),$$

$$F(t) = \begin{cases} 0 & t < t_1 \\ 1 & t_1 \le t \le t_2 \\ e^{-\mathrm{B}(t-t_2)} & t > t_2 \end{cases} \qquad (3.3)$$

Here m is the mass, γ is the drag coefficient, and k is the spring constant. $\psi(t)$ is the random forcing arising from thermal fluctuations with energy $k_B T$; it is Gaussian and white, and represents intrinsic noise. The form of the external forcing $F(t)$, illustrated

in Fig. 3.4A for four magnitudes $g_{1,2,3,4}$, is chosen to reflect the fact that following initial elevation, cells relax to their resting level of cytosolic calcium concentration at the end of experimental recording (see Fig. 3.1B). To account for the extrinsic noise observed in our cellular system[23], we have allowed the spring constant for each oscillator trajectory to vary uniformly around $\langle k \rangle$ with a standard deviation δk. Fig. 3.4B shows sample trajectories $x(t)$ for two cases: when the oscillations are overdamped ($m < m_c$) or underdamped ($m > m_c$), where $m_c \equiv \gamma^2/4k$ is the mass at critical damping.

Fig. 3.4C and D show, for the overdamped and underdamped cases, the scalar and vectorial mutual information between input g_α and output $x(t)$, as a function of sampling start time t_s, for various sampling rates r and fixed memory capacity n = 2. Additionally, Fig. 3.4E shows the mutual information as a function of n at fixed r for the overdamped case, while Fig. 3.4F shows the maximum mutual information as a function of r at fixed n for both cases. Comparing Fig. 3.4 to Fig. 3.3, we see that our minimal model is sufficient to capture the key features of the experiments. Specifically, comparing Fig. 3.4C and D to Fig. 3.3A, we see that the model captures the non-monotonic shape of the mutual information as a function of start time t_s, as well as the improvement of vectorial encoding (colors) over scalar encoding (black). Comparing Fig. 3.4E to Fig. 3.3C, we see that the model captures the increase of mutual information with memory capacity n, as well as the large-n convergence of curves with different t_s (although it is evident that the model appears to saturate at a lower n value than the experimental results). Finally, comparing Fig. 3.4F to Fig. 3.3D, we see that the model captures the presence of an optimal sampling rate r that negotiates the

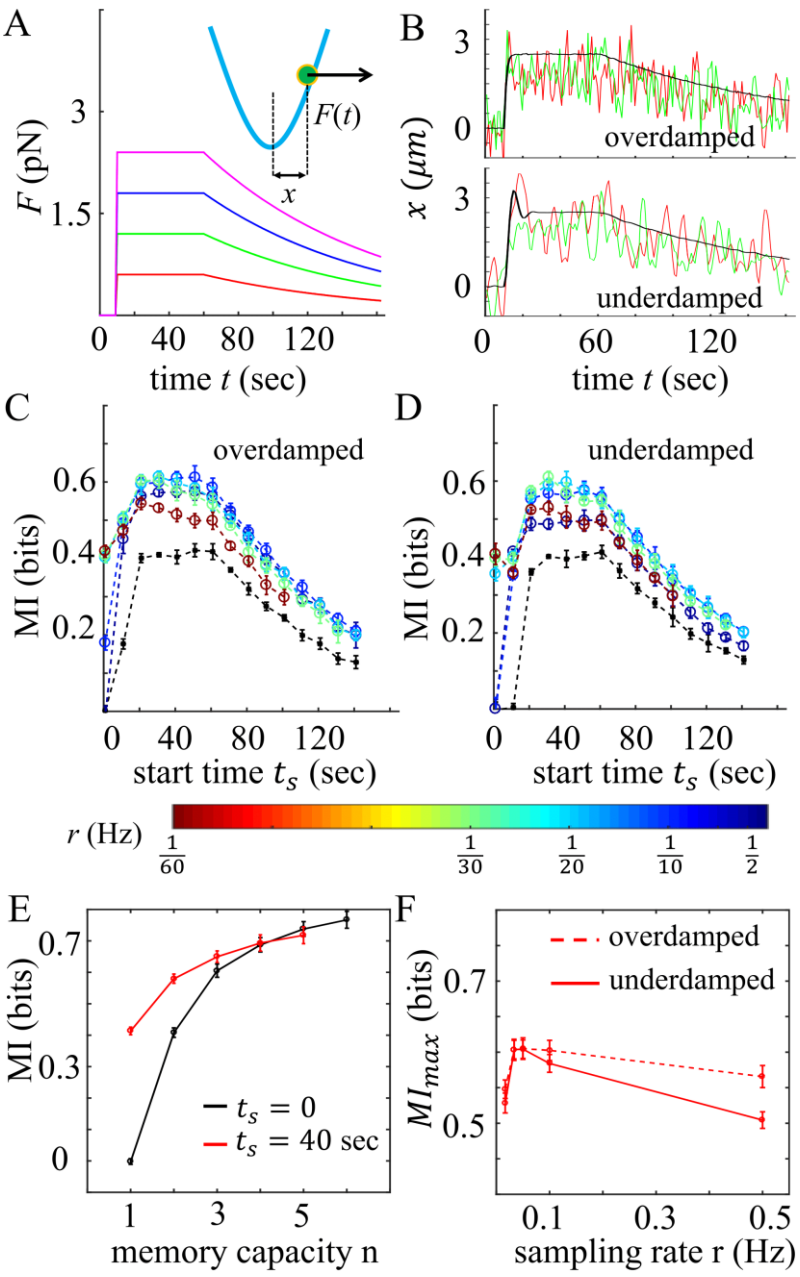

Figure 3.4 Information encoding in the noisy harmonic oscillator model. (A) Oscillator at position $x(t)$ is subjected to random thermal forcing as well as deterministic forcing $F(t)$ (Eq. 3). Four force magnitudes $g_{1,2,3,4}$ = {0.6 pN, 1.2 pN, 1.8 pN, 2.4 pN} serve as input, while $x(t)$ is output. Other parameters are t_1 = 10 sec, t_2= 60 sec, and β= 0.01 sec^{-1}. (B) Two sample trajectories (red and green curves) and the average of 5,000 trajectories (black curves) corresponding to g_4 = 2.4 pN. Upper: overdamped oscillators with m = 0.4m_c, where $m_c = \gamma^2/4k$ = 0.25 mg is the mass at critical damping, γ= 1 pN·sec/μm, and k = 1 pN/μm. Lower: underdamped oscillators with m = 9m_c. Other parameters are k_BT= 0.5 pN·μm and δk = 0.2 pN/μm. (C, D) Vectorial mutual

information MI_v as a function of sampling start time t_s at different sampling rates r (color bar) and memory capacity n = 2, for (C) overdamped and (D) underdamped oscillators. Black curve is scalar mutual information MI_s each time point. (E) MI_v as a function of n at fixed sampling rate r = 1⁄30 Hz and sampling start time (black curve t_s = 0, red curve t_s = 40 sec). n =1 corresponds to MI_s. (F) Maximum MI_v over all t_s values, as a function of r, for fixed memory capacity n = 2. Error bars in C-F represent the means and standard deviations of 20 independent trials each.

tradeoff between samples that are well-separated, yet confined to the high-yield region ($t_1 \leq t \leq t_2$ in the model). These correspondences validate the model, and allow us to use the model to ask how vectorial encoding depends on the structure of the underlying dynamic responses.

The noisy oscillator model allows us to explore two qualitatively different regimes of dynamic structure. In the overdamped regime, the thermal noise overpowers the oscillations, and the dynamics are dominated by fluctuations (Fig. 3.4B, upper). In contrast, in the underdamped regime, the oscillations overpower the thermal noise, and the dynamics are dominated by the underlying oscillatory structure (Fig. 3.4B, lower). Since vectorial mutual information corresponds to sampling the dynamics at regular intervals, it is natural to hypothesize that the amount of information extracted from underdamped dynamics will be higher than that extracted from overdamped dynamics, because underdamped dynamics have a more ordered structure. Fig. 3.4C and D compare the mutual information in the overdamped and underdamped cases. Surprisingly, we see that the amounts of information are roughly equivalent in the two cases. It is evident from Fig. 3.4C and D that the equivalence holds at varying sampling rates r and start times t_s (including the start time at which the information is maximal, Fig. 3.4F). In particular, the equivalence holds when the sampling rate r equals the

oscillation frequency of the underdamped oscillator, $\nu = \sqrt{(k/m)(1 - m_c/m)}/2\pi \approx$ 10 Hz in Fig. 3.4. This is true despite the fact that $r \approx \nu$ or nearby frequencies is the regime where one might have expected the vectorial information to benefit most from sampling the periodic dynamics instead of noisy dynamics. We have also checked that the equivalence holds for a large range of intrinsic and extrinsic noise levels. The previously demonstrated correspondence between the model and the experiments suggests that in the fibroblasts as well, ordered dynamics would not provide more information than noisy dynamics, at least as quantified by the vectorial mutual information. We expand upon this conclusion in the Discussion.

3.3.5 Differential Effects of Intrinsic and Extrinsic Noise

Vectorial mutual information MI_v is larger than scalar mutual information MI_s in part because repeated sampling helps to mitigate intrinsic noise[32]. Yet, in the case of the fibroblast cells, the gain of MI_v over MI_s is often small. For example, as seen in Fig. 3.3A, at n = 2 and r = 1/10 Hz, whereas MI_s can be as large as ~0.4 bits, the further increase of MI_v over MI_s is less than ~0.1 bits. We make this observation quantitative by defining the information gain $MI_{v,max} = MI_{s,max}$, where each is maximized over the start time t_s. Fig. 3.5 shows the information gain vs. $MI_{s,max}$ for the fibroblasts at n = 2 and r = 1/10 Hz (pink circle). The fact that the gain is small (0.1 bits) suggests that additional factors, apart from intrinsic noise alone, reduce the efficacy of vectorial encoding.

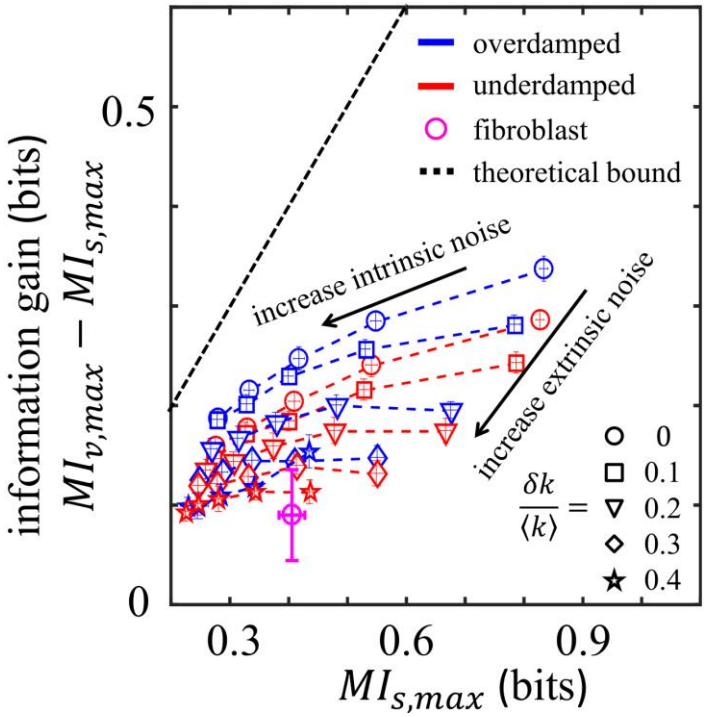

Figure 3.5 Gain of vectorial over scalar mutual information, where each is maximized over start times t_s. For vectorial information, memory capacity is n = 2 and sampling rate is r = 1/10 Hz. Fibroblast data are compared against over and underdamped oscillator model. In oscillator model, parameters are as in Fig. 3.4, and intrinsic noise is governed byk_BT, which varies from 0.2 to 1 pN·μm, while extrinsic noise is governed by $\delta k/\langle k \rangle$, which varies from 0 to 0.4. Error bars represent means and standard deviations of 100 bootstraps (fibroblast data), 20 independent trials of 5,000 trajectories each (oscillator model).

To explore this hypothesis in a systematic way, we again turn to our minimal oscillator model. For both the overdamped and underdamped oscillator, we compute $MI_{s,max}$ and the information gain. In the model, the intrinsic noise is governed by the thermal energy k_BT. The model also provides an opportunity to investigate the effects of extrinsic noise, which is governed by $\delta k/\langle k \rangle$, the relative width of the distribution of spring constants. As shown in Fig. 3.5, when the intrinsic noise increases while the extrinsic noise is fixed, both the scalar information and the information gain decrease,

as expected (dashed lines). The decrease in scalar information is more pronounced than the decrease in the gain, which is consistent with the fact that vector information is beneficial for mitigating intrinsic noise. On the other hand, when the extrinsic noise increases while the intrinsic noise is fixed, the gain decreases more rapidly, while the scalar information decreases less rapidly (the different symbols). This implies that the gain is more sensitive to extrinsic noise than intrinsic noise.

In the context of the fibroblast population, these results suggest that extrinsic noise (cell-to-cell variability), not intrinsic noise, is primarily responsible for degrading the performance of vectorial encoding and producing small information gains.

3.3.6 Redundant Information and Low-Pass Filtering

The vectorial mutual information MI_v can never be larger than the sum of the scalar mutual information values $MI_s(t_i)$ taken individually at each time point t_i. The reason is that there will always be some nonnegative amount of redundant information between the output at one time and the output at another time[100]. Denoting the redundant information as MI_{red}, we formalize this statement as

$$MI_{red} = \left[\sum_{i=1}^{n} MI_s(t_i)\right] - MI_v = n\langle MI_s\rangle - MI_v \geq 0, \tag{3.4}$$

whereas before $t_i = t_s + ir^{-1}$, and in the second step we rewrite the sum in terms of the temporal average $\langle MI_s\rangle = n^{-1}\sum_{i=1}^{n} MI_s(t_i)$. In the limit that the dynamics are approximately stationary, such as in the high-yield regions of Figs. 1B and 4B, MI_s is

approximately independent of time, and $\langle MI_s \rangle = MI_s$. For n = 2, as in Fig. 3.5, Eq. 4 then becomes

$$MI_v - MI_s \leq MI_s, \tag{3.5}$$

Eq. 5 expresses the intuitive fact that the gain upon making an additional measurement can never be more than the information from the original measurement, for a stationary process. Eq. 5 is plotted in Fig. 3.5 (dash-dotted line), and we see that it indeed bounds all data from above, as predicted.

The redundant information in Eq. 4 can be directly measured in the experiments. Fig. 3.6 shows the redundant information in the fibroblast calcium dynamics as a function of the memory capacity n, computed from the scalar and vectorial mutual information according to Eq. 4. Here t_s = 70 sec, and for each curve the sampling rate r is fixed, such that the duration T_d increases with n. We see that the redundant information depends on r and appears to be bounded from above by a roughly linear function of n. Can we understand this dependence theoretically? To address this question, we return to Eq. 4. We rearrange Eq. 4 as $MI_{red} = (n-1)\langle MI_s \rangle - \Delta$, where we define $\Delta = MI_v - \langle MI_s \rangle$. Since the vectorial information is not smaller than the scalar information corresponding to any of its time points, it is also not smaller than the average scalar information. Therefore, $\Delta \geq 0$, and we have

$$MI_{red} \leq (n-1)\langle MI_s \rangle, \tag{3.6}$$

Eq. 6 is a linear function of n, weakly modified by the fact that $\langle MI_s \rangle$ itself depends on n since it is computed for varying numbers of time points. Eq. 6 is compared with the data in Fig. 3.6, and we see that it indeed predicts the bound well. Eq. 6 makes another

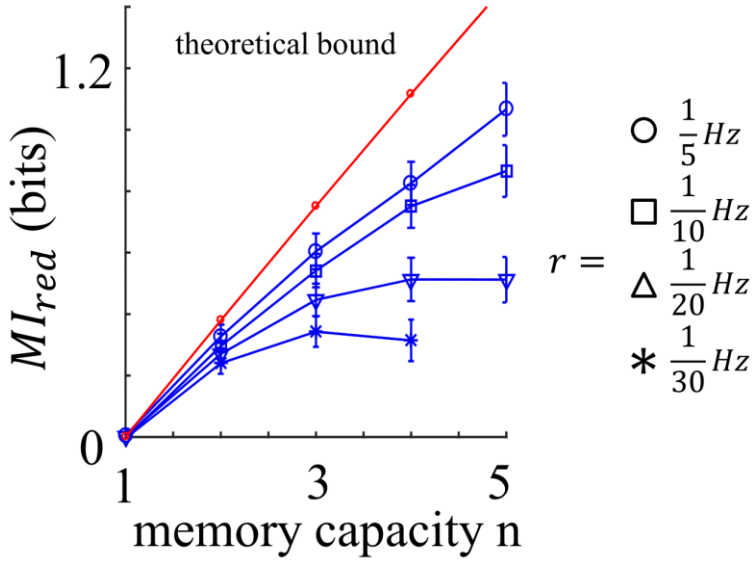

Figure 3.6 Redundant information of dynamic encoding for fibroblast cells. Redundant information MI_{red} (Eq. 4, t_s = 70 sec) is plotted as a function of memory capacity n for varying sampling rates r and compared with the theoretical bound (Eq. 6).

prediction, namely that the bound is reached for a stationary process when $\Delta = MI_v - \langle MI_s \rangle \to 0$, i.e. when the vector information provides vanishing improvement over the average scalar information. We expect this situation to occur in the limit of large sampling frequency r, when samples occur in close succession and offer little additional information beyond a single, scalar measurement. Indeed, we see from the data in Fig. 3.6 that consistent with this prediction, the bound is approached in the limit of increasing r.

Clearly, the benefit of vectorial encoding is largest when the redundant information is small (the lowest data points in Fig. 3.6). In the ideal case, there is no redundant information at all, and Eq. 4 becomes

$$MI_v = \sum_{i=1}^{n} MI_s(t_i)\,, \tag{3.7}$$

Here, we see that the vectorial information is simply the sum of the scalar information at each time point. In this sense, Eq. 7 describes a low-pass filter: vectorial encoding captures the temporal accumulation of scalar information, as long as the sampling is sufficiently slow to remove the redundancy. Therefore, in this limit the vectorial information records only the slow (low-frequency) variations in the dynamics. This feature may help explain the previous counterintuitive result that the vectorial information is insensitive to the detailed dynamic structure, as we expand upon in the discussion.

3.4 Discussion

The dynamic waveforms of signaling molecules have offered a new perspective to understand cellular information encoding. Indeed, dynamic encoding, as quantified by the vectorial mutual information MI_v, has larger channel capacity than the static encoding, as quantified by the scalar mutual information MI_s[32]. From both experimental data and a minimal model we presented here, we find that dynamic encoding has several key advantages over static encoding. First, the maximal vectorial information is larger than the maximal scalar information, suggesting that dynamic encoding provides a more reliable readout of environmental inputs than static encoding does. Second, while the scalar information can vary significantly with sampling time, the vectorial information is more uniform across sampling start times, even with small vector dimensions (Fig. 3.3C and Fig. 3.4E).

However, the benefit of dynamic encoding comes with the cost of increasing the memory capacity n of cells. For a fixed memory capacity, we have shown that the

best strategy for cells to adopt is to sample as slowly as possible while keeping their samples within a "high-yield" region, where the mean dynamics depend significantly on the input. Nonetheless, we find that within this region, the benefit of dynamic encoding can depend very little on the detailed structure of the dynamics (persistent oscillation vs. monotonic relaxation). Moreover, the gain of dynamic encoding over static encoding can be small, largely due to the presence of extrinsic, as opposed to intrinsic, noise.

The finding that vectorial information is largely insensitive to the detailed dynamics is surprising, and is likely a reflection of the type of dynamics we investigate here, as well as the vectorial measure itself. To accurately model the experimental dynamics, we have considered noisy dynamics arising from a driven oscillator. Although this has allowed us to probe both noise-dominated and oscillation-dominated regimes, these dynamics remain mean-reverting and confined to a stationary or cyclo-stationary state. It is likely that other classes of dynamics, such as temporal ramps, would emerge as having uniquely higher vectorial information than stationary dynamics. Furthermore, the vectorial information itself, as defined here, reports correlations between a categorial input variable and a regularly sampled output trajectory. It is likely that more sophisticated information-theoretic measures would be more sensitive to dynamic details, such as the mutual information between input and output trajectories, which has been argued to play a biological role in cell motility[101,102].

Our results suggest that dynamic and static encoding mechanisms are deeply connected. By invoking the redundant information MI_{red}[100], we have made this connection rigorous. Specifically, combining Eqs. 4 and 6 yields $\langle MI_s \rangle \leq MI_v \leq$

$n\langle MI_s\rangle$, which shows explicitly that the vectorial information is bounded from both above and below by quantities determined by the window-averaged scalar information $\langle MI_s\rangle$. Taking a window average of the scalar information is equivalent to the downstream network acting as a low-pass filter, accumulating temporal measurements at sufficiently low frequencies. We find that such low-pass filtering effects are evident from both the experimental and modeling results.

In this study, we have taken the approach that an understanding of both the static and dynamic encoding behaviors of the fibroblast cells can be obtained from a model based on noisy harmonic oscillators. Despite the simplicity of the model, we find that it reproduces the experimental results very well. The agreement between the experiment and this simple model highlights our central conclusion: the vectorial mutual information is intrinsically connected with the scalar mutual information and therefore has limited capability to distinguish underlying dynamics. Because the model is minimal, we anticipate that it can be extended to answer more general questions about information encoding on a large, multicellular scale. This is particularly desirable as understanding collective information processing is a new frontier in systems biophysics[23,46,98,103–105]. On the other hand, many interesting questions, such as the precise functional form of $MI_v(n)$, the dependence of mutual information on nonlinear effects such as feedback, bifurcations, and coupling of multiple time scales may require more realistic models beyond noisy harmonic oscillator.

CHAPTER 4

Conclusion

The unifying theme of this work is the communication of cells in multicellular networks in response to chemical stimuli. Cells routinely experience a wide variety of stimuli from their environments and – whether this is chemical, mechanical, radiative, or other – they need a reliable way of communicating information about these stimuli to ensure survival and proper functioning of both the individual cell and the higher order network of cells. We specifically focus on the chemosensing abilities of cells in multicellular networks and their ability to communicate information via biochemical signals generated in response to the applied chemical stimuli. Using the well-known Ca^{2+} signaling dynamics of cells, we look at how the Ca^{2+} signaling dynamics may be used to encode information about the sensed stimuli and network properties.

In chapter 2 we explore the collective Ca^{2+} dynamics of multicellular networks with varying degrees of cell-cell communication in response to differing concentrations of ATP. We do so by making use of a simple microfluidic device with one inlet and outlet that allow for the creation of the monolayer as well as the flowing of the ATP over the monolayer. Pre-labeling the monolayer with a Ca^{2+} binding dye coupled with fluorescence microscopy and image processing techniques allowed for the determination of the single cell Ca^{2+} dynamics. Our results show that the collective Ca^{2+} dynamics exhibit the ability to encode both stimulus strength as well as cell density. This was readily shown in Fig. 2.3B and Fig. 2.6B with the magnitude of the Ca^{2+} response increasing with ATP concentration and the propensity for oscillations increasing with cell density. Modeling using a modified Tang and Othmer Ca^{2+} dynamics model showed agreement with these experimental results further confirming the Ca^{2+} response multiplexing ability of the cell network. Recent work has shown that

this ability for multiplexing in biochemical signaling is advantageous in that it allows for a greater level of response specificity to an event.

In chapter 3 we look at two information encoding strategies of fibroblast cells in their Ca^{2+} dynamics: vector encoding of information and static encoding. Static encoding is the simplest case of encoding in which the cell samples the Ca^{2+} dynamics at a single timepoint whereas vector encoding relies on multiple samples over time. Vector encoding has been proposed as a method for cells to encode and decode more information in their dynamics than that of a static case. Our results show this to indeed be the case as we have found that vectoring encoding, in general, has a higher mutual information (a measure of the correlation between inputs and outputs) associated with it than static encoding. Further, the readouts from vector encoding are less sensitive to the underlying dynamics than scalar encoding. The trade-off with vector encoding is that it requires a larger memory capacity and we find that by fixing the memory capacity, the best strategy for a cell to employ is to sample infrequently over a larger period of time. Finally, our results show that the vector information is bounded above and below by the window averaged scalar information, showing that the downstream network acts as a low-pass filter.

The future of this work aims to focus more directly on the gap junction, the mode of communication that is at play in the exchange of Ca^{2+} between neighboring cells. To do so, we are stepping away from the multicellular network and instead working with a one-dimensional wire of cells that is created using a new two-layer microfluidic device. Direct injection of a cell suspension into narrow trapping channels allows for the cells to align in a wire such that communication is forced between nearest

neighbor cells via gap junctions. Use of pressure pumps in conjunction with ATP will creates strict control over the spatiotemporal profile of the stimulus to leading cells in the chain and will allow us to probe questions about intercellular wave and information propagation, as well as recent work suggesting that diffusion through gap junctions may be non-linear in nature as opposed to the often assumed linear diffusion.

BIBLIOGRAPHY

1. Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **1,** 11–21 (2000).
2. Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* **4,** 517–29 (2003).
3. Swillens, S., Dupont, G., Combettes, L. & Champeil, P. From calcium blips to calcium puffs: Theoretical analysis of the requirements for interchannel communication. *Proc. Natl. Acad. Sci.* **96,** 13750–13755 (1999).
4. Dupont, G., Combettes, L., Bird, G. S. & Putney, J. W. Calcium oscillations. *Cold Spring Harbor Perspectives in Biology* **3,** 1–18 (2011).
5. Tran, Q. K., Ohashi, K. & Watanabe, H. Calcium signalling in endothelial cells. *Cardiovasc. Res.* **48,** 13–22 (2000).
6. Bootman, M. D. & Lipp, P. Calcium signalling: Ringing changes to the 'bell-shaped curve'. *Curr. Biol.* **9,** R876–R878 (1999).
7. Mak, D.-O. D., McBride, S. & Foskett, J. K. Inositol 1,4,5-tris-phosphate activation of inositol tris-phosphate receptor Ca2+ channel by ligand tuning of Ca2+ inhibition. *Proc. Natl. Acad. Sci.* **95,** 15821–15825 (1998).
8. Reece, J. & Urry, L. *Campbell biology*. *Campbell Biology* (2011). doi:10.1039/c3ra44507k
9. Kadamur, G. & Ross, E. M. Mammalian phospholipase C. *Annu Rev Physiol* **75,** 127–154 (2013).
10. Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361,** 315–325 (1993).
11. Clapham, D. E. Calcium Signaling. *Cell* **131,** 1047–1058 (2007).
12. Iino, M. & Tsukioka, M. Feedback control of inositol trisphosphate signalling bycalcium. *Mol. Cell. Endocrinol.* **98,** 141–146 (1994).
13. Fill, M. & Copello, J. a. Ryanodine receptor calcium release channels. *Physiol Rev Rev* **82,** 893–922 (2002).

14. Wu, M. M., Buchanan, J., Luik, R. M. & Lewis, R. S. Ca2+ store depletion causes STIM1 to accumulate in ER regions closely associated with the plasma membrane. *J. Cell Biol.* **174,** 803–813 (2006).
15. Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441,** 179–185 (2006).
16. Leybaert, L. & Sanderson, M. J. Intercellular Ca(2+) waves: mechanisms and function. *Physiol. Rev.* **92,** 1359–1392 (2012).
17. Kumar, N. M. & Gilula, N. B. The gap junction communication channel. *Cell* **84,** 381–388 (1996).
18. Löwenstein, W. Junctional intercellular communication: the cell-to-cell membrane channel. *Physiol.Rev.* **61,** 829–913 (1981).
19. Sun, B., Duclos, G. & Stone, H. A. Network characteristics of collective chemosensing. *Phys. Rev. Lett.* **110,** 1–5 (2013).
20. Evans, W. H. & Martin, P. E. M. Gap junctions: structure and function (Review). *Mol. Membr. Biol.* **19,** 121–136 (2002).
21. Goldberg, M., de Pittà, M., Volman, V., Berry, H. & Ben-Jacob, E. Nonlinear gap junctions enable long-distance propagation of pulsating calcium waves in astrocyte networks. *PLoS Comput. Biol.* **6,** (2010).
22. Othmer, H. G. & Tang, Y. in *Experiemntal and theoretical advances in biological pattern formation* 277–299 (1993).
23. Potter, G. D., Byrd, T. A., Mugler, A. & Sun, B. Communication shapes sensory response in multicellular networks. *Proc. Natl. Acad. Sci.* **113,** 10334–10339 (2016).
24. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115,** 1716–1733 (2001).
25. Rathinam, M., Petzold, L. R., Cao, Y. & Gillespie, D. T. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J. Chem. Phys.* **119,** 12784–12794 (2003).
26. Cao, Y., Gillespie, D. T. & Petzold, L. R. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.* **124,** 44109 (2006).

27. Cao, Y., Gillespie, D. T. & Petzold, L. R. Adaptive explicit-implicit tau-leaping method with automatic tau selection. *J. Chem. Phys.* **126,** (2007).
28. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81,** 2340–2361 (1977).
29. Rhee, A., Cheong, R. & Levchenko, A. The application of information theory to biochemical signaling systems. *Phys. Biol.* **9,** 45011 (2012).
30. Bialek, W. *Biophysics: Searching for Principles*. (Princeton University Press, 2012).
31. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379–423 (1948).
32. Selimkhanov, J. *et al.* Accurate information transmission through dynamic biochemical signaling networks. *Science* **346,** 1370–3 (2014).
33. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **69,** (2004).
34. Loftsgaarden, D. O. & Quesenberry, C. P. A Nonparametric Estimate of a Multivariate Density Function. 1049–1051 (1965). doi:10.1214/aoms/1177700079
35. Dubuis, J. O., Tkacik, G., Wieschaus, E. F., Gregor, T. & Bialek, W. Positional information, in bits. *Proc. Natl. Acad. Sci.* **110,** 16301–16308 (2013).
36. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science (80).* **334,** 354–358 (2011).
37. Lauffenburger, D. A. Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5031–5033 (2000).
38. Lodish, H. *et al. Molecular Cell Biology, 4th Edition*. (W.H. Freeman, 2000).
39. Barritt, G. J. *Communication Within Animal Cells*. (Oxford University Press, 1992).
40. Miller, M. B. & Bassler, B. L. Quorum Sensing in Bacteria. *Annu. Rev. Microbiol.* **55,** 165–199 (2001).
41. Jones, W. D., Cayirlioglu, P., Grunwald Kadow, I. & Vosshall, L. B. Two

chemosensory receptors together mediate carbon dioxide detection in Drosophila. *Nature* **445,** 86–90 (2007).

42. Smear, M., Shusterman, R., O'Connor, R., Bozza, T. & Rinberg, D. Perception of sniff phase in mouse olfaction. *Nature* **479,** 397–400 (2011).
43. Benninger, R. K. P., Zhang, M., Head, W. S., Satin, L. S. & Piston, D. W. Gap Junction Coupling and Calcium Waves in the Pancreatic Islet. *Biophys. J.* **95,** 5048–5061 (2008).
44. Greschner, M. *et al.* Correlated firing among major ganglion cell types in primate retina. *J. Physiol.* **589,** 75–86 (2011).
45. Schneidman, E., Berry, M. J., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440,** 1007–1012 (2006).
46. Sun, B., Lembong, J., Normand, V., Rogers, M. & Stone, H. A. Spatial-temporal dynamics of collective chemosensing. *Proc. Natl. Acad. Sci.* **109,** 7753–7758 (2012).
47. Berridge, M. J. & Irvine, R. F. Inositol phosphates and cell signalling. *Nature* **341,** 197–205 (1989).
48. Lowenstein, W. R. & Kanno, Y. Intercellular Communication and the Control of Tissue Growth: Lack of Communication between Cancer Cells. *Nature* **209,** 1248–1249 (1966).
49. McLachlan, E., Shao, Q., Wang, H. L., Langlois, S. & Laird, D. W. Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. *Cancer Res.* **66,** 9886–9894 (2006).
50. Zhou, J. Z. & Jiang, J. X. Gap junction and hemichannel-independent actions of connexins on cell and tissue functions - An update. in *FEBS Letters* **588,** 1186–1192 (2014).
51. Tang, Y. & Othmer, H. G. Frequency encoding in excitable systems with applications to calcium oscillations. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7869–7873 (1995).

52. Thurley, K. *et al.* Reliable encoding of stimulus intensities within random sequences of intracellular Ca2+ spikes. *Science Signaling* **7,** ra59 (2014).

53. Woods, N. M., Cuthbertson, K. S. R. & Cobbold, P. H. Repetitive transient rises in cytoplasmic free calcium in hormone-stimulated hepatocytes. *Nature* **319,** 600–602 (1986).

54. Meyer, T. & Stryer, L. Calcium spiking. *Annu. Rev. Biophys. Biophys. Chem.* **20,** 153–174 (1991).

55. Meyer, T. & Stryer, L. Molecular model for receptor-stimulated calcium spiking. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5051–5055 (1988).

56. Politi, A., Gaspers, L. D., Thomas, A. P. & Höfer, T. Models of IP3 and Ca2+ oscillations: frequency encoding and identification of underlying feedbacks. *Biophys. J.* **90,** 3120–3133 (2006).

57. Falcke, M. Reading the patterns in living cells —the physics of ca2+ signaling. *Adv. Phys.* **53,** 255–440 (2004).

58. Janmey, P. A. & Miller, R. T. Mechanisms of mechanical signaling in development and disease. *J. Cell Sci.* **124,** 9–18 (2011).

59. Abbaci, M., Barberi-Heyob, M., Blondel, W., Guillemin, F. & Didelon, J. Advantages and limitations of commonly used methods to assay the molecular permeability of gap junctional intercellular communication. *BioTechniques* **45,** 33–62 (2008).

60. Gordon, J. L. Extracellular ATP: effects, sources and fate. *Biochem. J.* **233,** 309–319 (1986).

61. Falzoni, S., Donvito, G. & Di Virgilio, F. Detecting adenosine triphosphate in the pericellular space. *Interface Focus* **3,** 20120101 (2013).

62. Trabanelli, S. *et al.* Extracellular ATP exerts opposite effects on activated and regulatory CD4+ T cells via purinergic P2 receptor activation. *J Immunol* **189,** 1303–1310 (2012).

63. Pellegatti, P. *et al.* Increased level of extracellular ATP at tumor sites: In vivo imaging with plasma membrane luciferase. *PLoS One* **3,** (2008).

64. Cameron, D. J. Inhibition of macrophage mediated cytotoxicity by exogenous

adenosine 5’-triphosphate. *J. Clin. Lab. Immunol.* **15,** 215—218 (1984).

65. Schmidt, A., Ortaldo, J. R. & Herberman, R. B. Inhibition of human natural killer cell reactivity by exogenous adenosine 5'-triphosphate. *J. Immunol.* **132,** 146 LP-150 (1984).
66. Green, H. N. & Stoner, H. B. *Biological actions of the adenine nucleotides,*. (H.K. Lewis, 1950).
67. Ingerman, C. M., Smith, J. B. & Silver, M. J. Direct measurement of platelet secretion in whole blood. *Thromb. Res.* **16,** 335–344 (2017).
68. de Ronde, W., Tostevin, F. & ten Wolde, P. R. Multiplexing Biochemical Signals. *Phys. Rev. Lett.* **107,** 48101 (2011).
69. Marshall, C. J. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80,** 179–185 (1995).
70. Schwartz, M. a & Madhani, H. D. Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* **38,** 725–748 (2004).
71. Purvis, J. E. & Lahav, G. Encoding and Decoding Cellular Information through Signaling Dynamics. *Cell* **152,** 945–956 (2017).
72. Batchelor, E., Loewer, A., Mock, C. & Lahav, G. Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* **7,** 488 (2011).
73. Noren, D. P. *et al.* Endothelial cells decode VEGF-mediated Ca2+ signaling patterns to produce distinct functional responses. *Sci. Signal.* **9,** ra20-ra20 (2016).
74. Mehta, P. & Gregor, T. Approaching the molecular origins of collective dynamics in oscillating cell populations. *Current Opinion in Genetics and Development* **20,** 574–580 (2010).
75. De Monte, S., d’Ovidio, F., Danø, S. & Sørensen, P. G. Dynamical quorum sensing: Population density encoded in cellular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18377–81 (2007).
76. Taylor, A. F., Tinsley, M. R., Wang, F., Huang, Z. & Showalter, K. Dynamical quorum sensing and synchronization in large populations of chemical

oscillators. *Science* **323,** 614–617 (2009).

77. Tay, S. *et al.* Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature* **466,** 267–71 (2010).
78. Kellogg, R. A. & Tay, S. Noise facilitates transcriptional control under dynamic inputs. *Cell* **160,** 381–392 (2015).
79. Mora, T. & Bialek, W. Are Biological Systems Poised at Criticality? *J. Stat. Phys.* **144,** 268–302 (2011).
80. Krotov, D., Dubuis, J. O., Gregor, T. & Bialek, W. Morphogenesis at criticality. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 3683–3688 (2014).
81. Hidalgo, J. *et al.* Information-based fitness and the emergence of criticality in living systems. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 10095–100 (2014).
82. Pires-daSilva, A. & Sommer, R. J. The Evolution of Signaling Pathways in Animal Development. *Nat. Rev. Genet.* **4,** 39–50 (2003).
83. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. a. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75,** 655–680 (2006).
84. Papin, J., Hunter, T., Palsson, B. & Subramaniam, S. Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* **6,** 99–111 (2005).
85. Lim, W., Mayer, B. & Pawson, T. *Cell Signaling: Principles and Mechanisms*. (Garland Science, 2014).
86. Mestre, J. R. *et al.* Redundancy in the Signaling Pathways and Promoter Elements Regulating Cyclooxygenase-2 Gene Expression in Endotoxin-treated Macrophage/Monocytic Cells. *J. Biol. Chem.* **276,** 3977–3982 (2001).
87. Logue, J. S. & Morrison, D. K. Complexity in the signaling network: Insights from the use of targeted inhibitors in cancer therapy. *Genes and Development* **26,** 641–650 (2012).
88. Kholodenko, B. N. Cell-signalling dynamics in time and space. *Nat. Rev. Mol. cell Biol.* **7,** 165–176 (2006).
89. Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through

signaling dynamics. *Cell* **152,** 945–956 (2013).

90. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science* **298,** 1241–1245 (2002).
91. Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving stability of lipopolysaccharide-induced NF-κB activation. *Science* **309,** 1854–1857 (2005).
92. Adami, C. Information theory in molecular biology. *Physics of Life Reviews* **1,** 3–22 (2004).
93. Levchenko, A. & Nemenman, I. Cellular noise and information transmission. *Current Opinion in Biotechnology* **28,** 156–164 (2014).
94. Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341,** 558–61 (2013).
95. Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A. & Bowsher, C. G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E326-33 (2014).
96. Léon, C. *et al.* Defective platelet aggregation and increased resistance to thrombosis in purinergic P2Y1 receptor-null mice. *J. Clin. Invest.* **104,** 1731–1737 (1999).
97. Yitzhaki, S. *et al.* Uridine-5'-triphosphate (UTP) reduces infarct size and improves rat heart function after myocardial infarct. *Biochem. Pharmacol.* **72,** 949–955 (2006).
98. Sun, B., Duclos, G. & Stone, H. A. Network characteristics of collective chemosensing. *Phys. Rev. Lett.* **110,** (2013).
99. Van Kampen, N. G. Stochastic Processes in Physics and Chemistry. *Stochastic Processes in Physics and Chemistry* 96–133 (2007). doi:10.1016/B978-044452965-7/50016-7
100. Schneidman, E., Bialek, W. & Berry, M. J. Synergy, redundancy, and independence in population codes, revisited. *J. Neurosci.* **25,** 5195–206 (2005).
101. Tostevin, F. & Ten Wolde, P. R. Mutual information between input and output

trajectories of biochemical networks. *Phys. Rev. Lett.* **102,** 1–4 (2009).

102. Tostevin, F. & Ten Wolde, P. R. Mutual information in time-varying biochemical systems. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **81,** (2010).
103. Mugler, A., Levchenko, A. & Nemenman, I. Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proc. Natl. Acad. Sci.* **113,** E689–E695 (2016).
104. Ellison, D. *et al.* Cell-cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E679–E688 (2016).
105. Camley, B. A., Zimmermann, J., Levine, H. & Rappel, W. J. Emergent Collective Chemotaxis without Single-Cell Gradient Sensing. *Phys. Rev. Lett.* **116,** (2016).
106. Varennes, J., Han, B. & Mugler, A. Collective Chemotaxis through Noisy Multicellular Gradient Sensing. *Biophys. J.* **111,** 640–649 (2016).
107. Garcia, D. Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.* **54,** 1167–1178 (2010).
108. Meister, M. Multineuronal codes in retinal signaling. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 609–614 (1996).
109. Schnitzer, M. J. & Meister, M. Multineuronal firing patterns in the signal from eye to brain. *Neuron* **37,** 499–511 (2003).

APPENDICES

Appendix A: Photolithography

The ability of photolithography to reliably generate high-resolution features ranging from hundreds of microns to submicron in scale makes it an ideal candidate for the creation of microfluidic devices capable of studying cells and cell signaling behavior. Limitations of photolithographic techniques for these purposes are generally due to equipment limits and/or improper development procedure though many creative solutions have been found to create devices spanning orders of complexity based on the needs of the investigator. The general process involves the deposition of a photoresist onto a substrate that, in conjunction with a pre-generated photomask, prints a desired pattern into the resist that is then used for device creation. One type of photolithography utilizing negative photoresist is the basis for our device fabrication as the final features of the master are permanent, allowing for near limitless uses in the creation of microfluidic devices for our studies.

Creation of the master can be broken down into a series of steps: photomask generation, spin-coating, soft bake, exposure, post-exposure bake, and development.

A1 Photomask Generation

Photomasks are comprised of a square, fused quartz substrate coated on one side with photoresist and chromium. Mask generation begins by loading a design file generated from CAD software such as AutoCAD into the mask writing software. The mask writing software converts the file into a series of instructions that, while the laser of the mask writer raster scans the photomask, causes the exposure of the photoresist in the pattern described in the design file. A series of chemical baths after the mask

writing results in the removal of photoresist and chromium in the regions exposed to laser light creating the design in the fused quartz that allows the passage of UV light (See Figure A1).

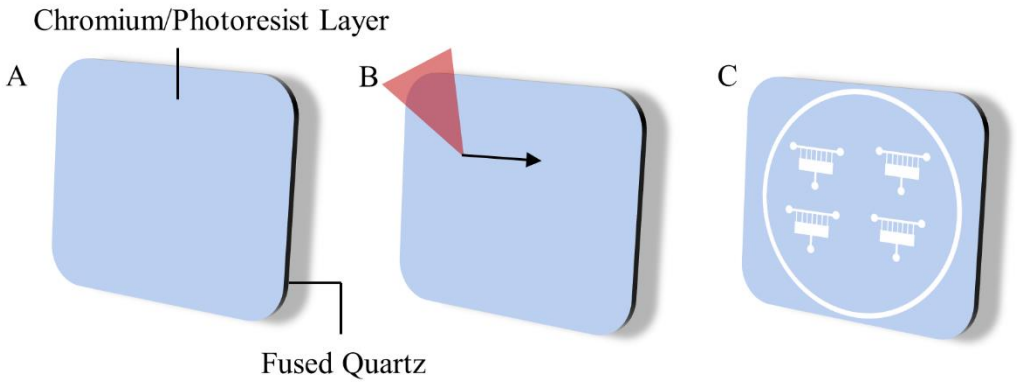

Figure A1 Photomask generation. (A) Mask blanks are pre-purchased consisting of fused quartz with a thin layer of photoresist and chromium deposited on one side of the photomask. (B) Laser light is raster scanned across the mask blank, exposing the chromium/photoresist layer according to the pattern created in the design file. (C) Development of mask in series of chemical baths removes the exposed chromium and photoresist allowing for passage of UV light in desired pattern.

A2 Spin Coating

A uniform photoresist layer is deposited on a silicon wafer by a series of controlled ramp rates by the spin-coater and spin time duration. Manufacturers of negative photoresist have detailed spin speed versus thickness documentation and suggested ramp rate and duration times. Note that it is possible that a non-uniform photoresist thickness is generated rendering the wafer unusable. This will generally look like pooling of the photoresist on the substrate and is due to insufficient wetting possibly due to an unclean surface or a high humidity environment. If this is the case, a photoresist primer can be spun onto the wafer and baked before the application of the

photoresist to increase wetting or the wafer can be plasma treated for several minutes to achieve the same effect (See Appendix A8 for more details).

A3 Soft Bake

After a uniform photoresist is achieved, the wafer is transferred to a leveled hot plate for a soft bake. The soft bake takes place at two temperatures (typically 65°C and 95°C) to reduce thermal stress and lasts for several minutes depending on the thickness of the photoresist. If the photoresist does not pre-bake for the correct amount of time it is likely to result in poor adhesion quality of the features to the substrate after development and may result in removal during rinsing of the wafer after development.

A4 Exposure

Following the soft bake is the exposure of the photoresist to the UV source via the photomask shown in Fig. A2. During the exposure step a strong acid is formed that, in conjunction with the post-exposure bake (PEB) catalyzes the epoxy cross-linking. The exposure step is vital to the proper formation of the microfeatures and is one of the main steps that is tweaked to generate the desired outcome. Exposure time is directly related to the desired film thickness and intensity of the UV source; it is best practice to rely on the photoresist manufacturers recommendations and to measure the intensity of the source being worked with before exposure to ensure replicability of device fabrication. Improper contact between the photomask and the coated wafer or too long/short of exposure time can cause poor feature resolution. Too short of exposure

time has the added disadvantage of not generating enough of the acid to form all the features resulting in poor substrate adhesion.

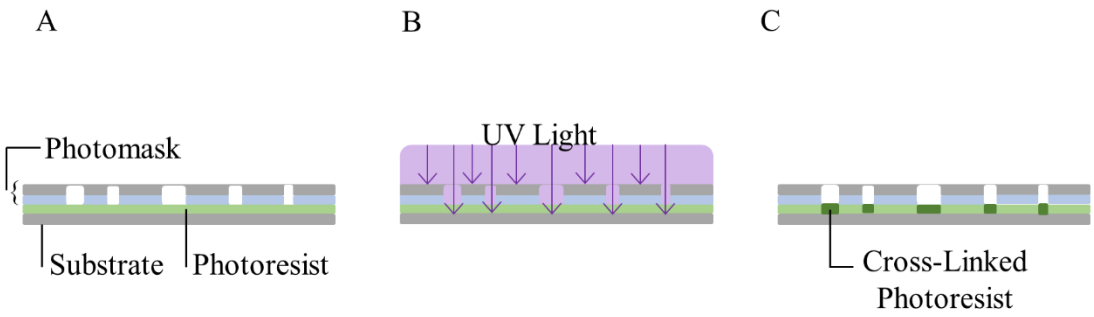

Figure A2 Cross-section of photoresist exposure. (A) Photomask is brought into contact with substrate coated with negative photoresist after spin-coating and soft bake steps. (B) Regions of photomask where chromium and photoresist were removed in development allow UV light to pass through the fused quartz to begin acid formation of the negative photoresist. All other UV light is blocked to prevent exposure in undesired regions. (C) Acid formed during exposure of the negative photoresist begins the cross-linking process resulting in rigid features in desired pattern.

A6 Post-Exposure Bake (PEB)

The acid formed during the exposure step is thermally driven with the PEB to catalyze the cross-linking step creating the insoluble epoxy features. The PEB, like the soft bake, takes place at 65°C and 95°C and has bake time dependent on the desired feature height, with larger features requiring longer bake times. Improper PEB time creates insufficient epoxy formation resulting in poor adhesion and feature resolution. A trace of the features should appear within the first few minutes of the PEB as an indicator that the epoxy is forming and serve as a check that things are progressing as desired.

A7 Development

Having removed the wafer from the PEB hotplate and allowed it to cool to room temperature, the wafer is submerged in developer to remove any un-crosslinked

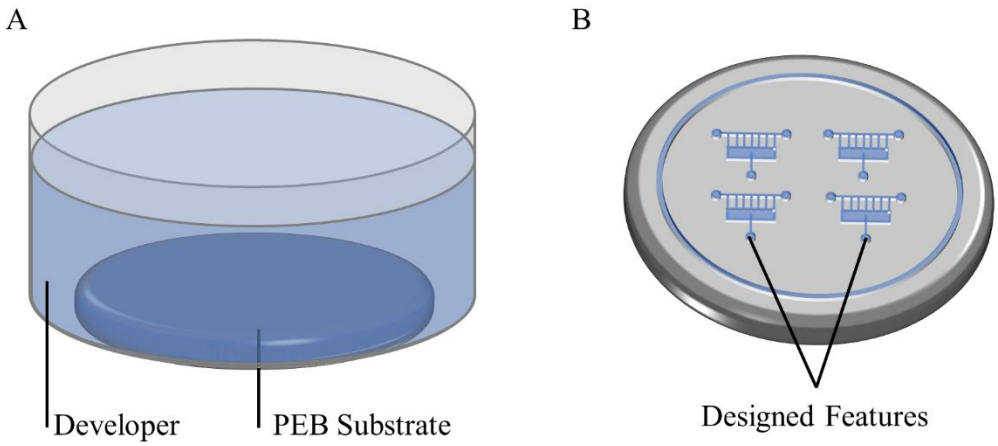

Figure A3 Wafer development process. (A) Photomask is brought into contact with substrate coated with negative photoresist after spin-coating and soft bake steps. (B) Regions of photomask where chromium and photoresist were removed in development allow UV light to pass through the fused quartz to begin acid formation of the negative photoresist. All other UV light is blocked to prevent exposure in undesired regions.

photoresist (Fig. A3). Total time in the developer is dependent on feature height but is expedited with gentle agitation. For particularly tight features, holes, or large vertical walls, a more vigorous agitation is generally required though caution must be taken as too much time in developer or too vigorous agitation may cause the erosion of cross-linked photoresist resulting in rounding or removal of features. When finished, the wafer is removed from the developer and rinsed with isopropyl alcohol followed by de-ionized water and blown dry with nitrogen. A white film appearing during the rinsing of the wafer is an indicator that further time in the developer is needed. If, during the

rinsing or drying process, the epoxy features warp or are removed from the substrate, adjusting of the baking times – particularly during the soft bake – may need to be made.

A8 Protocols

The protocol for generating a single layer and multi-layer SU-8 master mold on a standard 3-inch silicon wafer is found below. These protocols assume cleanroom conditions and access to a spin-coater, two hotplates, plasma cleaner, and contact aligner, all of which can be found in Owen Hall on the Oregon State University campus under the supervision of Dr. Chris Tasker and Rick Presley. Also assumed is that the proper training and certification has been obtained.

Necessary materials: Silicon wafer (at least one highly polished side), negative photoresist (SU-8 2000 series, MicroChem), SU-8 developer (MicroChem), lithography mask, acetone, isopropyl alcohol (IPA), de-ionized water (DIH_2O), tongs, syringes (at least 5ml), glassware for developer, developer waste container, wipes.

A8.1 Single layer protocol

(1) Select silicon wafer from carrier, visually inspect for defects.

(2) Over acetone, IPA, and DIH_2O waste container and holding wafer with tongs, rinse wafer with acetone, IPA, and DIH_2O in that order. Blow wafer dry with compressed nitrogen.

(3) (Optional) Plasma treat wafer at 50% power for 4-5 minutes. Note: this tool requires additional training to be certified on.

(4) Place wafer on spin-coater chuck, test to make sure wafer is centered properly.

(5) Using syringe, deposit negative photoresist (approximately 1 ml/inch of wafer)

(6) Begin pre-programmed spin-coating sequence. Sequence depends on desired feature height as well as the resist being used. Refer to resist manufacturers protocol for best practice. Sample instructions for ~2μm feature height using SU-8 2002 given below:

    a. Ramp spin speed to 500 rpm/s at rate of 100 rpm/s
    b. Hold at 500 rpm/s for 5s
    c. Ramp spin speed to 3000 rpm/s at rate of 300 rpm/s
    d. Hold at 3000 rpm/s for 30s
    e. End sequence

(7) Transfer wafer to 65°C hotplate and let stand for 1 minutes.

(8) Transfer wafer to 95°C hotplate and let stand for 3 minutes.

(9) Allow wafer to cool to room temperature on benchtop surface or similar. For thinner films, it may be beneficial to bring temperature down at controlled, slow rate to prevent thermal stress.

(10) Transfer wafer to contact aligner and ensure proper contact and separation of wafer from photolithography mask is achieved to create best resolution possible in SU-8 features.

(11) When in contact mode, expose wafer to UV light for ~8s

(12) Transfer wafer to 65°C hotplate for 3 minutes. A trace of the desired design should be apparent on the wafer after ~1 minute.

(13) Transfer wafer to 95°C hotplate for 5 minutes.

(14) Remove wafer from hotplates and allow to cool to room temperature.

(15) Place wafer in glassware containing fresh SU-8 developer and gently agitate wafer/developer for ~2 minutes. Enough developer to cover the silicon wafer in 1-2 inches of liquid is generally all that is needed.

(16) Remove wafer and rinse with IPA followed by DIH_2O. If white residue forms after the application of the IPA this is an indicator that further development time is needed. The wafer can be returned to the developer for additional time such that the residue disappears. Rinse again with IPA and DIH_2O.

(17) Blow dry with compressed nitrogen and view under microscope to verify design meets requirements.

(18) (Optional) Hard bake the final product on a 120-150°C hotplate for 20-30 minutes to create more permanent features.

A8.2 Multi-Layer Protocol

The protocol for the multi-layer case follows the same steps as above but is repeated to stack differing feature heights to generate the desired design. As such, each set of bake times, exposure times, and development times will vary per design needs. Additionally, it is critical that alignment marks are created for each photolithography mask so that all features align properly. Below is a truncated protocol for a two-layer design with ~2μm features connected to ~40μm features.

(1) Complete steps 1-18 above to generate the 2μm features and necessary alignment markers.

(2) Pre-treatment of wafer with plasma cleaner for second layer is generally unnecessary and can prove damaging to the SU-8 features already in place.

(3) Place wafer on the spin-coater and adjust ramp rates as necessary for feature height. For 40μm features using SU-8 2025, the same ramp rates and from the previous description are suitable. Ensure wafer is centered.

(4) Load syringe with 3ml of SU-8 2025 and carefully deposit the resist *around* the alignment markers making sure not to cover alignment markers.

> Note: Covering of alignment markers will result in the inability to align to the markers when using the contact aligner due to the similar indices of refraction. Other techniques using a positive resist for alignment markers have been suggested to solve this problem in the literature due to its deep red color, however, we haven't achieved success using this method.

(5) Pre-bake the wafer at 65°C and 95°C for 3 minutes and 5 minutes respectively.

(6) Ensuring proper alignment in the contact aligner utilizing second layer mask, expose wafer for ~9.5s

(7) Post-exposure bake wafer at 65°C and 95°C for 5 minutes and 8 minutes respectively.

(8) Develop wafer in SU-8 developer for ~3.5 minutes.

(9) Rinse with IPA, DIH_2O, and blow dry with compressed nitrogen as usual.

(10) Hard bake finished wafer at 120-150°C for 20-30 minutes.

A9 Troubleshooting

Due to the uniqueness of each designs needs and variability in the photolithography equipment and even daily weather conditions, these protocols serve only as a starting point. In general, the most sensitive parameters to the successful generation of the SU-8 master are the pre-bake, exposure, and post-exposure bake times and are good places to start when adjusting a protocol. A table of common problems and causes in master generation and potential fixes is given below.

Problem	Possible Cause	Fix
Poor feature resolution	Under/overexposure of SU-8 features in contact aligner	Adjust exposure time in small increments around initial time
	Overdeveloped in SU-8 developer	Lower development time
	Poor contact of wafer to photolithography mask	Ensure tighter contact in contact aligner
		Use edge-bead removal techniques to remove edge-bead
Poor adhesion of features to wafer	Unclean wafer surface before spin-coating	Clean surface thoroughly with acetone, IPA, and DIH_2O

		Use plasma cleaner or increase plasma cleaning time before SU-8 deposit
	Pre-bake times insufficient	Increase pre-bake time at both temperatures by 1-2 minutes
Rounding of features on wafer	Overdevelopment in SU-8 developer	Reduce development time by small increments, avoiding characteristic white residue of underdevelopment
Uneven coating of SU-8 on wafer after spin-coating	Poor wetting of SU-8	Plasma clean or increase plasma cleaning time of wafer
		Purchase a wafer pre-treatment solution and apply to wafer before SU-8 to increase wetting
Cracked features	Thermal stress	Adjust baking times or add intermediate baking temperatures to slow the temperature ramp
Previous exposure time now insufficient	Bulb failure in contact aligner	Use power meter to test UV power to ensure it is up to specifications. Replace if necessary

Appendix B: Cell Wire and Chemical Gradient Generation

It has been demonstrated previously[46] that cell colonies exposed to pulses of a chemical stimulus (ATP) highlighted the existence of pacemaker cells and that the temporal order exhibited in the Ca^{2+} response of nearest neighbor cells to the pacemakers suggested communication via gap junctions. When the gap junctions were chemically blocked, the temporal order between pacemaker and nearest neighbors vanished providing evidence for the vital role gap junctions play in cell-cell communication. To further explore this role and better understand the spatiotemporal dynamics involved in Ca^{2+} signaling, we created a microfluidic device capable of aligning cells in a 1D wire as shown in Fig. B1.

The device is a two-layer microfluidic device that is everywhere $\sim 40 \mu m$ tall except for a series of $\sim 2 \mu m$ tall pads connecting the trapping channels to a flow channel where the ATP stimuli is applied to the leading cells in the wire. Trapping channel widths have been empirically determined to be $\sim 15 \mu m$ to allow for sufficient room for cell attachment and viability without compromising the one-dimensionality of the wire. A series of inlets and outlets allow for the direct injection of the cell suspension necessary for wire generation, the application of growth media and Ca^{2+} binding dye, and the application of the ATP stimuli. A pressure pump system (AF1/OB1, Elvesys) allows for the tight control over the spatiotemporal profile of the stimulus. Previous single layer designs proved incapable of generating the cell wires due to the limitations of the photolithography process and deformability of cells as they slipped past the junctions between trapping channel and flow channel.

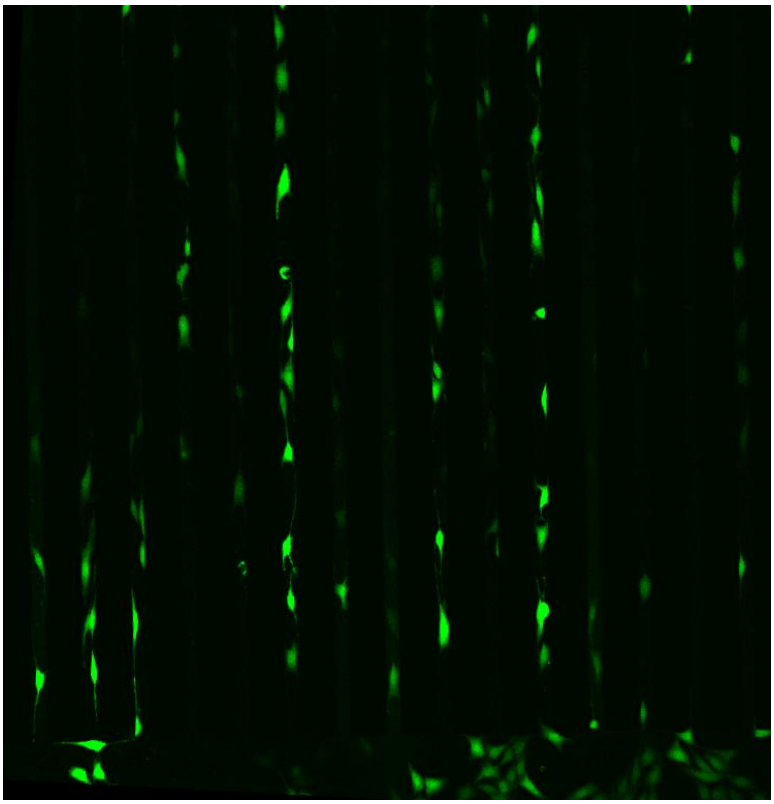

Figure B1 Generation of cell wire in two-layer device. Insertion of adherent cells via a syringe loaded with a cell suspension allows for alignment and attachment of cells on glass bottom. Side channels allow for application of the Ca^{2+} binding dye Fluo-4 (Life Technologies) while subsequent fluorescent imaging allows for monitoring of calcium dynamics in response to stimuli.

Currently, we have demonstrated the ability to reliably insert cells into the two-layer device, keep them viable for experimentation, apply and remove the Ca^{2+} binding dye, and apply a wide range of stimuli profiles via the pressure pump and a recently added flow switching device (Fluigent). Preliminary evidence shows the ability to stimulate the release of Ca^{2+} from the leading cells in the wire and some evidence showing propagation of a intercellular calcium wave. Further testing is required to confirm the existence of intercellular waves and adjust the experimental protocol to reliably generate such waves. Once this is achieved we can begin to explore the

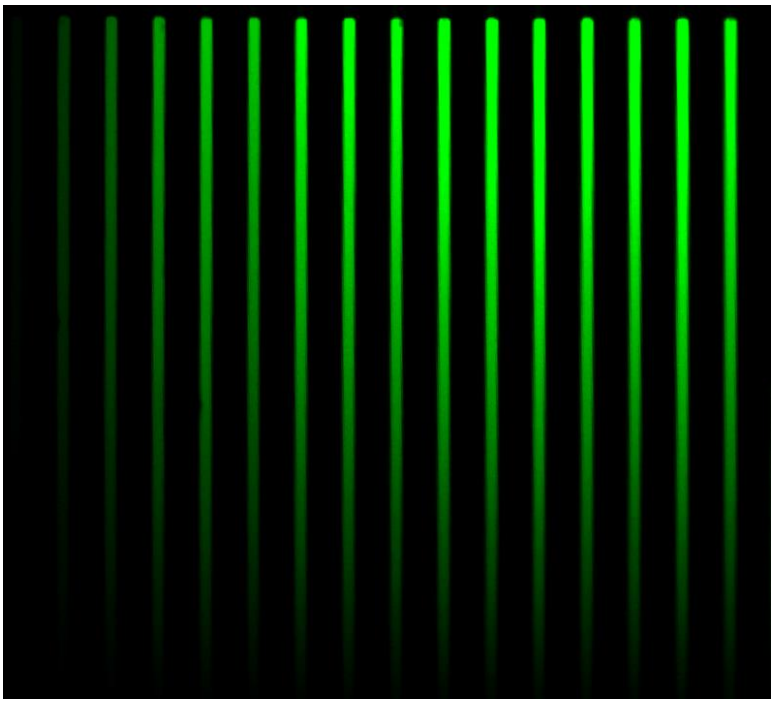

Figure B2 Gradient formation in two-layer microfluidic device. Constant flow provided by two syringe pumps in conjunction with fluorescein (Sigma Aldrich) and growth media (Thermo Fisher) allow for demonstration of a stable chemical gradient.

spatiotemporal dynamics of Ca^{2+} and the role of gap junction communication outlined at the beginning, answering questions such as: what is the natural propagation length of a Ca^{2+} wave? Is consistent stimulation and propagation achievable and, if so, what are these conditions? and is diffusion through gap junctions linear or non-linear?

Discussions with our collaborator at Purdue, Dr. Mugler, has led to another potentials use for the two-layer device and that is the generation of chemical gradients. Recent work by Dr. Mugler has predicted that large clusters of cells in a chemical gradient are able to sense the gradient with greater precision but that these larger clusters reduce mobility in collective motion[106]. Though unable to experimentally verify this with the current two-layer design, we are able to explore the gradient sensing ability of individual cells and how gradient strength impacts mobility, potentially

linking Ca^{2+} to such behavior. We have thus far been able to demonstrate in several instances the ability to create stable gradients (Fig. B2) in bare two-layer devices. Further work is necessary to develop consistent control of the gradient characteristics and the ability to implement such a system with cells in device. Such experiments would require high levels of stability due to the long imaging times required for studying cell migration.

Appendix C: Signal Analysis Techniques

C1 Inter-Spike Interval

As calcium spike trains are unsuitable for Fourier analysis due to their varying amplitudes and irregular periods, the inter-spike interval (ISI) has been used as a suitable substitute for determining how the period of calcium oscillation affects information encoding within and amongst cells in the network. To determine the ISI's from a single cell's calcium dynamics $R(t)$, we first smooth the time series $R(t)$ to remove outlying data points[107], to obtain $R_s(t)$. We then find all local peaks $\{t_i\}$ in $R_s(t)$ using the following criteria: a local peak must be higher than its neighboring valleys (local minimums) on both directions by a certain value, which we refer to as the sensitivity parameter; the first and last frames are excluded as peaks or valleys. Third, we obtain ISI's $\{\delta t_i\}$ from the distances of peaks $\{\delta t_i = |t_{i+1} - t_i|\}$. Finally, to obtain the true ISI's corresponding to calcium oscillations, we only keep $\{\delta t_i\}$ that fall in the range of 5 seconds and 150 seconds. This final step excludes about 5% of ISI's.

Figure C1A and C1B provide two examples of the peak finding algorithm's ability to determine peak locations for a fibroblast cell calcium profile (A) and cancer cell calcium profile (B). The sensitivity parameter used for all experiments and in Figure C1A and C1B is ¼ the standard deviation of the normalized intensity σ_I as determined empirically. Figure C1C and C1D show the algorithms diminished ability to determine peaks and to detect spurious peaks when the sensitivity parameter is set to σ_I and $\frac{\sigma_I}{8}$ respectively.

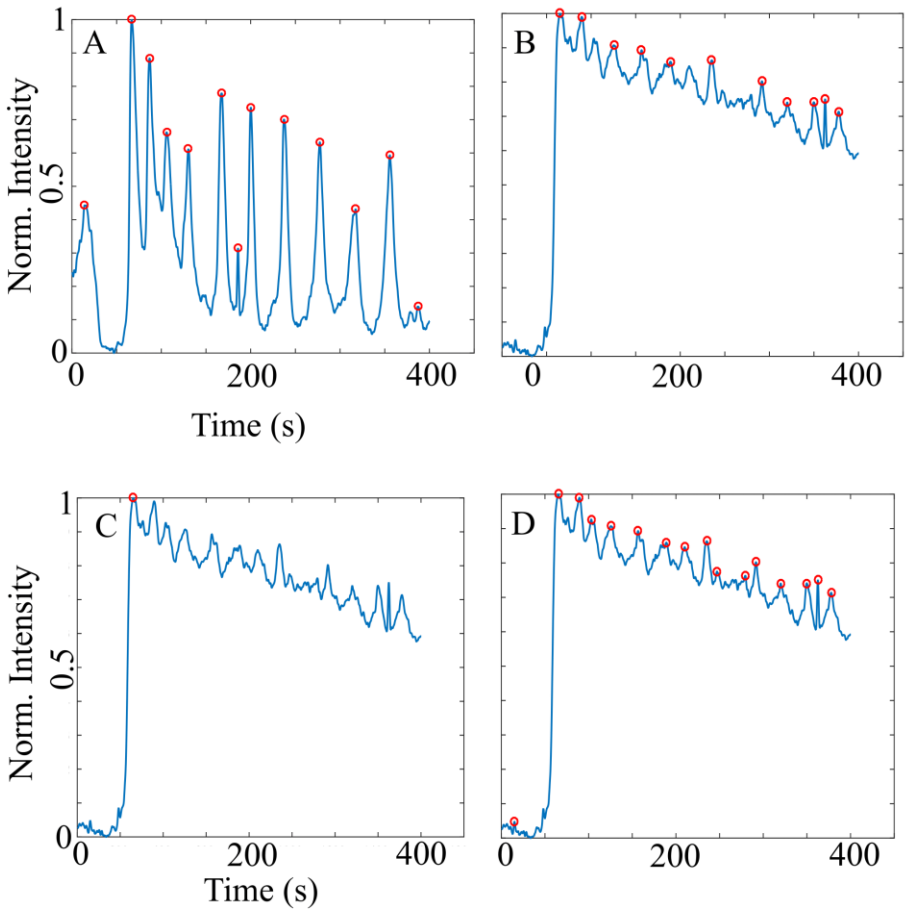

Figure C1 Intensity peak determination for single NIH 3T3 and MDA-MB-231 cell profiles. (A) Labeled peaks for a NIH 3T3 cell. Sensitivity parameter set to ¼ of the standard deviation of the time series. (B) Labeled peaks for a MDA-MB-231 cell. Sensitivity parameter set to ¼ of the standard deviation of the time series. (C) and (D) Same MDA-MB-231 intensity profile in (B) but with sensitivity parameter set to be the standard deviation and $\frac{1}{8}$ of the standard deviation of the time series respectively.

C2 Differential Intensity and Nearest Neighbor Cross Correlation

It has previously been demonstrated[46] the effectiveness of using nearest neighbor cross correlation functions (C_{NN}) to characterize the calcium dynamics of collective chemosensing. C_{NN} is not only a statistical characterization of the collective cellular dynamics, but also may provide robust channel for cellular information encoding[45,108,109]. As a result, we can study how network architecture may affect the

spectrum of C_{NN}. For any cell i, we represent its calcium dynamics (fluorescent intensity profile) as $R_i(t)$, and define C_{NN} as:

$$\dot{R}_i(t) = \frac{dR_i}{dt}(t) - \langle \frac{dR_i}{dt}(t) \rangle$$

$$C(\tau)_{ij} = \frac{1}{\sigma_i \sigma_j} \langle \dot{R}_i(t) \dot{R}_j(t+\tau) \rangle$$

$$C_{NN_{i,j}} = C(\tau = 0)_{ij} \,|D_{ij} = 1$$

To evaluate $\frac{dR_i}{dt}(t)$, we numerically differentiate the response curve $R_i(t)$ using the five-point stencil method; $\langle \frac{dR_i}{dt}(t) \rangle$ is the time average. Note that σ_i^2 is the variance of $R_i(t)$, which normalizes $C(\tau)_{ij}$ to be dimensionless. Based on Delaunay triangulation of the multicellular network, topological distance D can be defined for each cell pair where $D = 1$ for nearest neighbors. The mean nearest neighbor cross-correlation function $\bar{C}_{NN}$ is obtained by averaging $C_{NN_{i,j}}$ over all nearest neighbor pairs i, j.

Appendix D: Standard Lab Protocols

D1 Subculture Protocol

(1) Ensure culture flask is ≥70% confluent before subculturing cells. Percentages lower than this risk low yield and poor pellet formation.

(2) Pipette 7ml/15ml of 1x TrypLE Select (Thermo Fisher Scientific) into a $25cm^2/75cm^2$ flask respectively. In general, enough TrypLE Select should be used to completely cover cells in a few millimeters of liquid.

(3) Place flask back in incubator, wait ~10 minutes for cells to detach off the bottom of flask. This should be visible by holding the flask up to a light source and seeing floating cells. If uncertain, use microscope to check detachment.

(4) Transfer cell suspension to 15ml tube, labelling appropriately.

(5) Insert cell suspension into centrifuge being careful to counterbalance centrifuge to ensure proper functioning and prevent damage to the instrument.

(6) Using settings outlined by cell provider, set the rpm and time on centrifuge. Typical settings for centrifuge in Sun lab is 1000-1200 rpms for 5 minutes.

(7) Pour off the supernatant and re-suspend cell pellet in fresh growth medium. The quantity of fresh growth medium is dependent on application specifics and size of pellet.

(8) Transfer fraction of cell suspension to new flask and add appropriate growth media until desired final volume. Fraction of cells transferred depends on density of suspension from (7) and desired time to next subculturing – more

cells decreases time to subculture, fewer results in longer times until next subculture.

D2 Liquid Nitrogen Storage

(1) Turn on -80°C freezer and let reach -80°C (typically takes ~1 hour).

(2) Complete subculture protocol listed above until (7). Pour off supernatant and re-suspend cells in 1.5ml of Recovery Cell Culture Freezing Medium (Thermo Fisher Scientific).

(3) Transfer freezing medium cell suspension into 2ml cryovial for liquid nitrogen storage. Label cryovial with cell type, date, and initials.

(4) Place vial in Mr. Frosty (Nalgene) and place in -20°C freezer for ~2-4 hours.

(5) Transfer Mr. Frosty to -80°C freezer and store there for ~48 hours.

(6) Transfer to liquid nitrogen tank for long term storage.

D3 Liquid Nitrogen Retrieval

(1) Prepare a shallow dish with ~2cm of lukewarm water.

(2) Remove desired vial from liquid nitrogen tank, place in dish with lukewarm water for thawing.

(3) With ~50% of vial thawed, add 500μl of growth medium to contents of the vial and gently pipette to create cell suspension.

(4) Continue with subculturing procedure from step (4) as normal.