An information-theoretic model of engineered cell–cell communication

Akshay Maheshwari
akshaym@stanford.edu

Anton Jackson-Smith
acjs@stanford.edu

Matthew Feldman
mattfel@stanford.edu

Abstract

Engineered cell–cell communication systems will be critical for performing tasks such as computation in living systems. Information theory provides a means to quantify and analyze the capacity of cellular communication systems for the purpose of comparison and design. We integrate information-theoretic models of several component channels of a complete cell–cell system in order to determine the limits on capacity and infer optimal strategies for communication.

1 Introduction

The engineering of cell–cell communication is limited by a lack of theoretical foundations that enable precise, reproducible design. Information theory, which is extensively used in traditional communication to interpret noisy communication over a channel, provides a natural framework to analyze the precision and information capacity of cellular communication. Cell sensing and associated intracellular dynamics are inherently noisy processes that integrate information from the environment (input), process it (channel), and actuate a cellular response (output). Specifically, cells can encode signals by controlling the rate of production of molecules and by releasing different types of molecules, and decode signals through receptors that actuate intracellular responses.

Many mechanistic details of biological processes are not fully understood. Recent developments in systems biology seek to abstract and model complete cell systems despite access to limited information. Experimentally, it is more tractable to capture input and output signals, rather than measuring internal cell state. Consequently, information theory can be applied to both theoretically model these cell systems as a black box channel, and as a tool to interpret experimental data.

Here we explore the theoretical information communication limits from an engineered signal-transmitting bacterium to other engineered signal-receiving bacteria. We integrate information-theoretic models for different components of the communication chain to produce a complete end-to-end measure of channel capacity. We conclude by roughly estimating the computation possible by communicating bacteria in a 2D layer culture.

2 Background

Cells often communicate by producing small molecules, which reach receiver cells through diffusion in the cells' growth medium. First, using a simple 2-dimensional diffusion model, we establish the rate at which information can be sent from a transmitter to a receiver cell. We model transmission in the diffusion channel as a Markov Chain with memory that can switch between low and high transmission using the work of Einolghozati, et al. [1]. Unlike many classical communication channels, the channel's memory affects the state transition times. The channel capacity is determined by these diffusion dynamics.

Cells respond to external signals through the use of receptors: proteins which span the cellular membrane, bind to external signaling molecules, and trigger downstream effects within the cell. Secondly, we model signal transduction by receptors on the signal-receiving bacterium. Previous

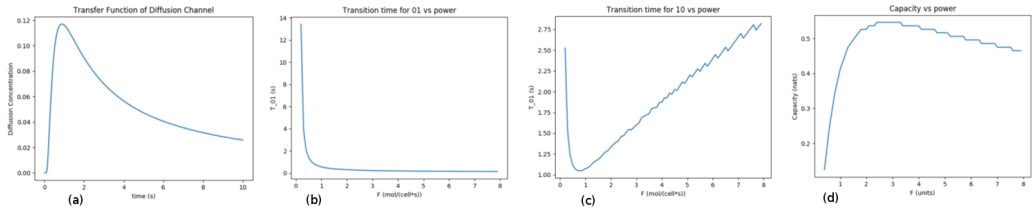

Figure 1: Dynamics of diffusion channel (a), transition delays (b,c), and capacity vs power (d)

work by Cheong et al. [2] have shown that the information transducible by Tumor Necrosis Factor (TNF) signaling is limited to 1.27 bits. Although multiple parallel intracellular pathways respond to TNF, the ability of the signal-receiving cell to accurately respond to an input signal is bottlenecked by the number of receptors that directly detect TNF. This form of relationship between receptors and their resulting intracellular cascades likely extends to other signaling systems as well. Consequently, in designing a swarm of communicating bacteria, we expect that additional cellular receptors would increase the capacity of information transfer, as shown by Thomas et al. [3]. We model the limits of transducible information with multiple receptors in response to time-varying signal. The overall communication system architecture is shown below in Figure 1.

Finally, we hypothesize that multiple cells detecting a given signal will increase the total information transfer rate. We use a bush model, commonly used to model one-to-many signal transfer, to evaluate this growth in signaling capacity. We also show using a tree network structure, as in Cheong et al. [2], that multiple copies of genes involved with intracellular signaling in a single cell gives diminishing capacity gains because of the receptor bottleneck.

3 Analysis

3.1 Diffusion channel

The diffusion dynamics given by Fick's second law provide the transfer function of the channel (Equation 1), so that the concentration given an input signal and a channel can be computed via convolution with this function (Equation 2). Figure 1 (a) shows the general shape of the transfer function

$$g_d(x,t) = \frac{1}{4\pi Dt} e^{\frac{-|x|^2}{4Dt}} \tag{1}$$

$$c(x,t) = g_d(x,t) r(x,t) \tag{2}$$

Unlike a classical discrete memoryless channel, the model is determined by a transition's start and end state, rather than just the end state. The amount of time a transmitter must allocate for transitioning from one state to another state are labeled as T_{00}, T_{01}, T_{10}, and T_{11}. If a receiver cell can differentiate between concentration levels that differ by S, we can compute the amount of time it must allocate for transitions from $0 \rightarrow 1$ and $1 \rightarrow 0$ by solving for the times in Equation 3. These transition times vary with power, as shown in Figure 1 (b) and (c).

$$\begin{aligned} c(x, T_{01}) &= 2S \\ c(x, T_{01} + T_{10}) &= S \end{aligned} \tag{3}$$

Finally, given these transition times, [1] derives the capacity over this channel as $C = logW$ Where W can be found by solving:

$$W^{T_{01}+T_{10}} - 2W^{T_{01}+T_{10}-1} + W^{T_{01}+T_{10}-2} = 1 \tag{4}$$

The power of the signal that the transmitter must send, in units of mol/s, can be swept to find the value which maximizes capacity for a given distance and channel parameters, shown in Figure 1. After

fixing this power for a single cell-cell channel, more receiving cells can be dropped in at varying distances to form a network. The capacity for each receiver is maximized at the distance that was used to compute the fixed power, and is smaller for all other cells in the network.

3.2 Multiple receptors

A cell with n receptors, can be in any of n+1 states, where state k refers to the state with any k receptors in the bound by a signaling molecule. For a given discrete timestep [tau], the state transition in response to the input signal can be modeled as a Markov chain. If the individual receptors do not interact with each other, the system is described by the transition diagram below:

$$0 \underset{\tau\beta}{\overset{n\tau\alpha_{H/L}}{\rightleftharpoons}} 1 \underset{2\tau\beta}{\overset{(n-1)\tau\alpha_{H/L}}{\rightleftharpoons}} 2 \cdots k \underset{(k+1)\tau\beta}{\overset{(n-k)\tau\alpha_{H/L}}{\rightleftharpoons}} k+1 \cdots n-1 \underset{n\tau\beta}{\overset{\tau\alpha_{H/L}}{\rightleftharpoons}} n$$

In the system used in our work, the input Xi is a sequence of signaling molecule concentrations, and output (and channel state) Yi is the number of bound receptors. This definition is a unit output memory channel, and was found to achieve capacity (as $\tau \to 0$) with an IID input source. The parameters α and β are derived from biochemical constants of the system under study; these kinetic constants are known for a number of specific cell receptors.

Overall, Thomas and Eckford proved that the IID capacity of this channel scales linearly with the number of independent cell-surface receptors [3]. Capacity of a single receptor is determined by its binding and unbinding parameters. This work also models the system if receptors interact (ie, binding of a single receptor affects the binding kinetics or one or more others), showing that in this regime feedback input can produce a higher capacity than IID. Given the challenges of implementing feedback input in a biological system, we focus on the independent and IID input model.

3.3 Multiple receivers

Multiple receivers that receive information across a noisy channel and are able to share their received signal with each other are better able to reconstruct the true original signal. We used this principle to construct seven-cell units where six cells receive information from a central transmitter and communicate with each other to achieve greater total capacity (figure 2(a)). We used the capacity of a single cell calculated in our analysis in previous sections to calculate the SNR that a single cell can achieve and then calculated the total capacity of 6 cells.

Multiple gene copies encoding for intracellular response to receptor binding are limited by the number of receptors on the cell. As in Cheong et al., we show this relationship using a tree network model that takes into account the SNR of the transmitted signal to the receptor as calculated above and the SNR of transducing an intracellular response after receptor binding . We used the SNR of a simplified TNF-alpha response cascade empirically derived by Cheong et al. to model intracellular response. In engineering a cell, this model can be used to determine the optimal balance between number of gene copies encoding intracellular response and the corresponding number of receptors needed to not bottleneck information.

4 Results

The models that we collected from previous work collectively describe the entire communication system from a transmitting cell which can emit signaling to a collection of receiving cells that act on this information. From the diffusion capacity model, we learn the trade-off between the power of the transmitted signal, the time steps required to accurately change the concentration at the receiving location, and the capacity over this channel. From the modeling of receptor binding, we learn how the cell probabilistically interacts with the concentration of molecules in its surrounding environment and uses this information to dynamically change its internal state. A larger number of receptors increases the output state space, and so increases the capacity of the channel. Finally, the network configuration of collections of cells, whether in a bush or tree topology affects the communication capacity from population to population. A receiver population of cells has more total capacity than a single receiver cell.

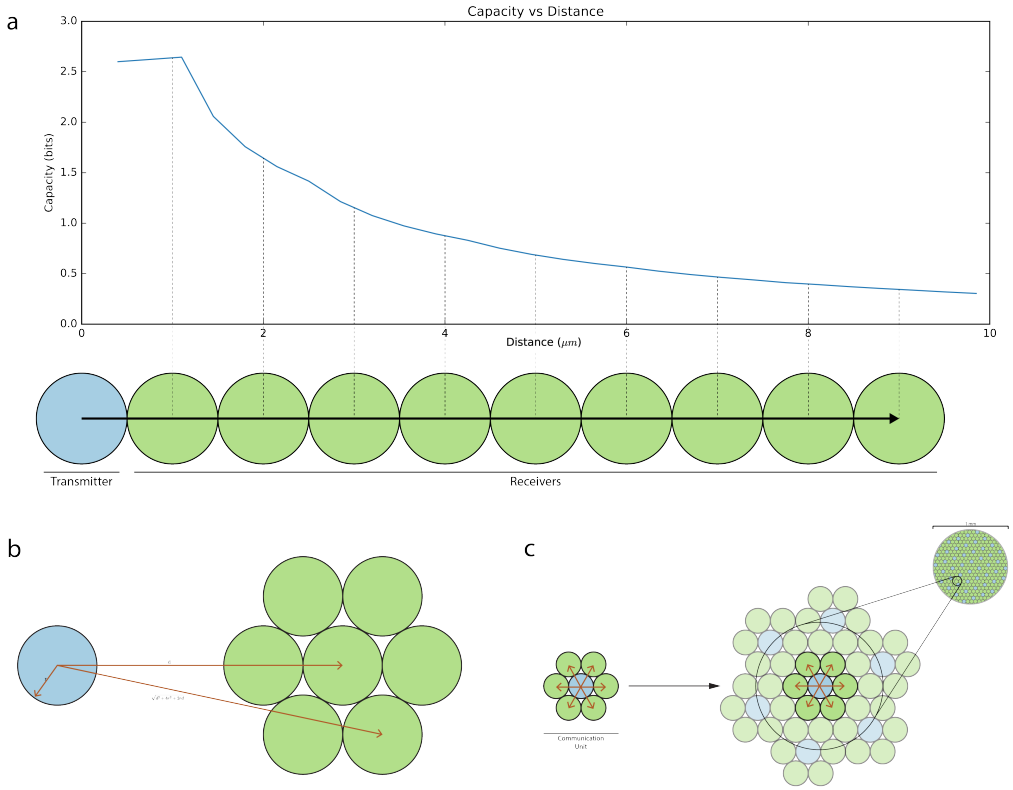

Figure 2: Integrated communication model for cell–cell signaling. **a** Channel capacity versus distance from transmitter to receiver. **b** Schematic of communication from a transmitter cell to a distant cluster of cells. **c** Schematic of communication across a macroscale colony: six-cell units are tessellated to form the colony.

By optimizing each step of the diffusion—receptor—signaling chain independently, we find that cell–cell communication is most effective where the transmitter and receiver are in close proximity, that in biologically-relevant scenarios receptor transduction of a signal is not limiting, and that multiple receiver-cells of a given signal provide more capacity than one alone. This suggests a unit design model of seven cells, where a center cell transmits to the six cells packed densely around it. In a two dimensional space, this communication unit can be tessellated to model a continuous 'lawn' of cells, as might exist in a macro-scale cell colony.

With a cell diameter of 1μm, we calculate the information capacity from a transmitter cell to its surrounding cells as 3.9 bits in total. We approximate the maximum possible information rate of a cell with a biologically relevant number of receptors (10,000 receptors for E. coli) to be 25,778 bits—-this capacity is not limiting. Finally, we calculate the theoretical maximal information flux across a 1 mm colony of 150,000 E. coli, where every cell transmits to and receives from its 6 neighbors, to be 271,103 bits.

The equations derived from these models can be used to globally optimize for communication capacity in a biological system, such as number of receptors per cell, transmitting power of a cell, and architecture of the population. There are several future directions for this work including performing such a global optimization, rather than optimizing parameters independently; integrating additional models, such as gene expression in response to a signal [4]; and investigating alternative signals, such as virus-encapsulated DNA messages [5]. Finally, the model could be expanded to a three-dimensional system to explore the communication and computational capacity of batch cell culture.

References

[1] Einolghozati, Arash, et al. "Capacity of discrete molecular diffusion channels." *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on.* IEEE, 2011.

[2] Cheong, Raymond, et al. "Information transduction capacity of noisy biochemical signaling networks." *Science* 334.6054 (2011): 354-358.

[3] Thomas, Peter J., and Andrew W. Eckford. "Capacity of a simple intercellular signal transduction channel." *IEEE Transactions on Information Theory* 62.12 (2016): 7358-7382.

[4] Tabbaa, Omar P., and C. Jayaprakash. "Mutual information and the fidelity of response of gene regulatory models." *Physical biology* 11.4 (2014): 046004.

[5] Ortiz, Monica E., and Drew Endy. "Engineered cell-cell communication via DNA messaging." *J Biol Eng* 6.1 (2012): 16-16.