Capacity of a Simple Intercellular Signal Transduction Channel

Peter J. Thomas and Andrew W. Eckford

Abstract

We model the ligand-receptor molecular communication channel with a discrete-time Markov model, and show how to obtain the capacity of this channel. We show that the capacity-achieving input distribution is IID. Further, unusually for a channel with memory, we show that feedback does not increase the capacity of this channel. We show how the capacity of the discrete-time channel approaches the capacity of Kabanov’s Poisson channel, in the limit of short time steps and rapid ligand release.

I. INTRODUCTION

A. Overview

Research at the intersection of biology and information theory stretches back almost to Shannon’s founding papers, with notable work by Yockey [1], [2], Attneave [3], Barlow [4], and Berger [5]. This work had long remained on the margins of the wider information theory community, perhaps owing to the difficulty of this inherently interdisciplinary area. However,
the past several years have seen a surge of interest in biological applications among mainstream information theorists (e.g., [6]–[10]), exploring information-theoretic applications from molecular biology to neuroscience. At the same time, quantitative biologists have increasingly begun applying information theoretic ideas to the analysis of high throughput, individually resolved laboratory data [11]. This new interest has arisen in parallel with increasing interest in the mathematical and conceptual foundations of biology, as well as interest in the techniques that communications engineers can learn from biological communication, such as in nanoscale or other nonconventional communication scenarios [12], [13].

The current paper focuses on communication systems that employ chemical principles, broadly known as molecular communication [14]. Recent work in molecular communication can be divided into two categories. In the first category, work has focused on the engineering possibilities: to exploit molecular communication for specialized applications, such as nanoscale networking [14], [15]. In this direction, information-theoretic work has focused on the ultimate capacity of these channels, regardless of biological mechanisms (e.g., [9], [10]). In the second category, work has focused on analyzing the biological machinery of molecular communication (particularly ligand-receptor systems), both to describe the components of a possible communication system [16] and to describe their capacity [6], [17]–[20]. Our paper, which builds on work presented in [18], fits into the second category, and many tools in the information-theoretic literature can be used to solve problems of this type. Related work is also found in [19], where capacity-achieving input distributions were found for a simplified “ideal” receptor; that paper also discusses but does not solve the capacity for the channel model we use.

Our contribution in this paper is to prove several important properties of capacity for a two-state signal transduction channel, for instance as found in the Dictyostelium model organism [21], [22] as well as in models of neural communication systems taking into account refractoriness or synaptic dynamics [23]. Using a discrete-time Markov chain model, we show that the capacity-achieving input distribution is IID, with all the probability weight on the minimum and maximum possible ligand concentration. Further, we show that feedback does not increase the capacity of the channel. Finally, given an IID input distribution, we give a simple closed-form expression for the mutual information, which can be maximized to find capacity. In addition to the capacity results, we discuss the mutual information of the channel when the channel inputs are Markov distributed, and we link our capacity results to earlier known results on the capacity of Poisson
counting channels.

B. Biological Motivation

As some readers of the Transactions may be unfamiliar with the details of biological signal transduction, we devote the remainder of the introduction to an overview of such systems.

Living cells communicate with one another through a web of biochemical interactions referred to as *signal transduction networks* [24]–[26]. Just as neural networks underly the interactions of many multicellular organisms with their environments, these biochemical networks allow individual cells to perceive, evaluate and react to chemical stimuli [27], [28]. Examples include chemical signaling across the synaptic cleft connecting the axon of one nerve cell to the dendrite of another [29], calcium signaling within the postsynaptic spines of a dendrite [30], pathogen localization by migratory cells in the immune system [31], growth-cone guidance during neuronal development [32], phototransduction in the retina [33], gradient sensing by the social amoeba *Dictyostelium* [34], and many others.

Signal transduction at the cellular and subcellular level typically involves a complex macromolecular apparatus involving multiple proteins. For example, transmission of neural signals often depends on diffusion of neurotransmitter molecules across a narrow gap (the synaptic cleft) to receptor proteins on the postsynaptic membrane. These neurotransmitter receptors are connected to large protein “signaling machines” [35] that control the downstream effects of neurotransmitter signaling, including signaling mediated by the influx of extracellular calcium ions. In general, activation of a receptor will produce *second messengers* within the cell, which control its behaviour.

In this paper we are most interested in the process at the receiving end of signal transduction, where a signaling molecule (ligand molecule) binds to a receiver molecule (protein) at a destination cell. Despite the apparent complexity of this process, a key simplifying observation is that the receptor proteins are driven through a finite series of *states* by the presence of signaling molecules [36].

A two-state example, where the receptor can be either bound to the ligand (signaling) molecule or else unbound, is shown in Figure [1] if the receptor is unbound, an available ligand can bind with it, changing its state; the receptor must then go through an “unbinding” process, processing the ligand and reverting to the initial state, before it can bind with another ligand. This two-
state, bound-unbound receptor model is appropriate for the 3'-5'-cyclic adenosine monophosphate (cAMP) receptor in the Dictyostelium amoeba, which is used as a model organism for studies of signal transduction \[18\], \[37\], \[38\]. This is the simplest nontrivial example of a ligand binding to a receptor, and forms the basis for the results in this paper.

![Diagram of receptor binding and unbinding](image)

Fig. 1. An example of a two-state binding and unbinding process of the receptor. If the receptor is unbound (U), then a ligand can be absorbed by the receptor; this process occurs at a rate $\alpha(c)$, as a function of the ligand concentration $c$. If the receptor is bound (B), it reverts to the unbound state at a rate $\beta$, independent of ligand concentration.

The complexity of these systems can be much higher. In many instances, signal transduction molecules possess a number of sites at which ligand molecules can bind to the receiver protein. A protein with $k$ binding sites potentially has $2^k$ distinct binding states. The transition corresponding to a single binding event occurs with a rate proportional to the ligand concentration, so these transitions are sensitive to the input signal. The reverse transition also occurs, in which one occupied binding site releases or degrades the bound ligand, and returns to the unbound, receptive state. This transition happens at a constant rate, independent of the ligand concentration, so it is insensitive to the input signal and its occurrence conveys no information about the instantaneous signal level.

Basic mechanisms of signal transduction have been known for decades \[39\], \[40\]. However,
recent technological advances have dramatically increased the ability to manipulate and measure the signals entering and leaving signal transduction networks at the molecular level. These advances create an opportunity for quantitative understanding of molecular communication. For example, microfluidics combined with cell-by-cell single track measurements have been used to estimate the mutual information between a chemical gradient and the motile response of the *Dictyostelium* amoeba [41], [42]. Single molecule fluorescence methods have allowed visualizing the binding and unbinding of signaling molecules to single receptors in real time [43]. High throughput measurements have led to sufficiently precise capacity estimates, for a cancer-related signaling network, to extract information about the network topology [44]. Optogenetics methods have created a new paradigm for manipulating molecular communication devices using applied light sources [45]. Thus, our results come at an opportune time for biological researchers, both in terms of their analytical capabilities and their interest in exploiting information theory.

C. Capacity problem for a general point process channel

We now introduce a formal definition of the continuous time signal transduction channel with arbitrary (bounded) scalar input and discrete output. Finite state Markov processes conditional on an input process provide models of signal transduction and communication in a variety of biological systems, as detailed in the preceding section. Typically a single ion channel, or receptor, is in one of *n* states. The states form a finite directed graph \( \mathcal{Y} \) with *n* vertices, with edges connecting states that intercommunicate through a conformational or chemical change, or ligand-binding/unbinding event. The receptor performs a continuous time random walk on the graph, with one or more transition rates being influenced by the external input signal, \( X(t) \). The input signal can be the concentration of a diffusing signaling molecule for a ligand-gated receptor; it can be the transmembrane electrical potential for a voltage-gated receptor [46].

There is a rich literature on the use of so-called “master equations” for representing stochastic chemical reactions [47] and algorithms for generating sample trajectories [48]–[50]. In the master equation representation of a signal transduction channel, the instantaneous transition rate matrix \( Q = [q_{jk}] \) depends on the external input \( X(t) \). The probability, \( p_k \), that the channel is in state \( Y(t) = k \in \mathcal{Y} \) evolves according to

\[
\frac{dp_k}{dt} = \sum_{j=1}^{n} p_j(t) q_{jk}(X(t))
\]
where for \((j \neq k)\), \(q_{jk} \geq 0\) is the input-dependent rate at which the receptor transitions from state \(j\) to state \(k\), and \(q_{jj} = -\sum_{k, k \neq j} q_{jk}\). Taking \(\{X(t)\}^T_{t=0}\) as the input, and the receptor state \(Y(t) \in \mathcal{Y}\) as the output, gives a channel model, the capacity of which is of general interest.

The majority of biological signal transduction systems operate without regulation by a fast clock, i.e. they operate as continuous time stochastic systems. Nevertheless, discrete time channel models arise as approximations to continuous time systems by fixing a small time step. While most of our analysis falls in the discrete time framework, we discuss the relation to continuous time systems further in §IV.

In §II-A we specialize from (1) to the typical case of a single receptor that can be in one of two states \((n = 2)\). In this case, (1) reduces to

\[
\frac{dp}{dt} = k_+ c(t) (1 - p(t)) - k_- p(t),
\]

where \(c(t)\) is the time-varying ligand concentration, and \(p(t)\) is the probability that the receptor is in the bound-to-ligand state \((Y = B)\) as opposed to the unbound state \((Y = U)\). Thus \(\mathcal{Y} = \{U, B\}\). When the receptor is bound by a ligand molecule, the signal is said to be transduced; typically the receptor (a large protein molecule) undergoes a conformational shift upon binding the ligand. Subsequently, the presence of the ligand is communicated to the cell interior by a cascade of reactions catalyzed by the bound receptor. The forward (binding) reaction happens at rate \(k_+ c(t)\), where the forward rate constant \(k_+\) has units of \((\text{time} \times \text{concentration})^{-1}\). The reverse (unbinding) reaction happens at rate \(k_-\). This rate constant has units \((\text{time})^{-1}\).

A key feature distinguishing this channel is that the receptor is insensitive to the input when in the state \(Y = B\), and can only transduce information about the input, \(X(t) = k_+ c(t)\), when \(Y = U\). Thus, analysis of the ligand-binding channel is complicated by the receptor’s insensitivity to changes in concentration occurring while the receptor is in the occupied state.

A related classical Poisson channel was solved by Kabanov \[51\], \[52\]. In the rapid-unbinding limit, in which transition from the bound state back to the unbound state is instantaneous, the ligand-binding channel becomes a simple counting process, with the input encoded in the time varying intensity. This situation is exactly the one considered in Kabanov’s analysis of the

\[\text{In practice, signals are transduced in parallel by multiple receptor protein molecules. However, in many instances individual receptor proteins act to a good approximation as independent receivers of a common ligand concentration signal. In this case, analysis of the single molecule channel provides a useful reference point.}\]
capacity of a Poisson channel, under a max/min intensity constraint \([51], [52]\). For the Poisson channel, the capacity may be achieved by setting the input to be a two-valued random process fluctuating between the maximum and minimum intensities. If the intensity is restricted to lie in the interval \([1, 1 + c]\), the capacity is \([52]\)

\[
C_{Kab}(c) = \frac{(c + 1)^{1+1/c}}{e} - \left(1 + \frac{1}{c}\right) \ln(c + 1).
\]

Wyner showed that Kabanov’s formula may be obtained by restricting the input to a two-state discrete time Markov process with input \(X(t)\) taking the values \(X_{lo} = 1\) and \(X_{hi} = 1 + c\), with transitions \(X_{lo} \rightarrow X_{hi}\) happening with probability \(r\), and transitions \(X_{hi} \rightarrow X_{lo}\) with probability \(s\), per time step \([53]\). Maximizing the mutual information with respect to \(r\) and \(s\), and taking the limit of small time steps, yields (3). For completeness, we recapitulate this derivation in §A1.

In addition, Kabanov proved that the capacity of the Poisson channel cannot be increased by allowing feedback.

Kabanov’s approach, focusing on instantaneous unbinding and restricted intensity, is not directly applicable to molecular signal transduction. However, our long-term goal is to obtain expressions analogous to (3) for the continuous-time systems (1) and (2). As a first step, we restrict attention to a discrete time analog of the two-state system (2). As we show in §IV-B, our channel model can be seen as a natural generalization of Kabanov’s counting process channel model.

In the next section we develop an analysis of a two-state signal-transduction channel model in discrete time and solve its capacity.
II. CAPACITY OF THE DISCRETE INTERCELLULAR TRANSDUCTION CHANNEL

In this section, we prove our main capacity results. A roadmap for these results is given as follows:

1) In (38), we give a closed-form expression for the mutual information if the inputs are IID. This expression can be maximized to find capacity with input distribution constrained to be IID (the IID capacity, see (39)), but IID capacity is not available in closed form.

2) In Theorem 1, we show that the IID capacity is achieved when only the minimum and maximum possible ligand concentrations are used, and no intermediate concentration. (This theorem is given first as it simplifies the proof of the main result.)

3) In Theorem 2, we then show that capacity of this channel is achieved by the IID input distribution, with inputs only on minimum and maximum ligand concentration. We do so by showing that the feedback capacity of the channel is satisfied by an IID input distribution, relying on the important results on feedback capacity from [54], [55].

4) In Corollary 1, combining Theorems 1 and 2 with Equation (39), IID capacity is given by (74). An illustration is given in Figure 2.

5) In §II-E, we show how capacity behaves as a function of the system parameters, and give the parameters that maximize capacity.

6) Finally, in §II-F, we consider whether capacity and feedback capacity are equal for signal transduction channels with more than two receptor states. In general they are not, which we illustrate with a counterexample.

A. Discrete-input, discrete-time model

In this section we examine a discrete-time, two-state Markov channel representation of the signal reception process, an example of which may be found in the Dictyostelium cAMP receptor. The channel input, channel output, and input-output relationship are described as follows.

Channel input. The channel input is the local concentration of ligands at the receptor: at the interface between the receptor and the environment, the receptor is sensitive to the concentration of ligands, binding more frequently as concentration increases. We assume that the input concentration $c_j$ is one of $m$ discrete levels, and without loss of generality, we will assume $c_1 \leq c_2 \leq \ldots \leq c_m$. The lowest concentration $c_1$ and highest concentration $c_m$ are especially
important in our analysis; we will give them the special symbols \( c_L := c_1 \) and \( c_H := c_m \). Thus, the input (concentration) alphabet is

\[
X = \{c_L, c_2, c_3, \ldots, c_{m-1}, c_H\}.
\]  

(4)

Further, let \( X^n = [X_1, X_2, \ldots, X_n] \in X^n \) represent a sequence of inputs to the receptor.

**Channel output.** The channel output is the state of the receptor. As in Figure 1, the receptor may either be in an unbound state, in which the receptor is waiting for a molecule to bind; or in a bound state, in which the receptor has captured a molecule, and cannot capture another until the molecule is degraded or released. Thus, channel output is binary: let \( \mathcal{Y} = \{U, B\} \) represent the output alphabet, where \( U \) represents the unbound state and \( B \) represents the bound state. Further, let \( Y^n = [Y_1, Y_2, \ldots, Y_n] \in \mathcal{Y}^n \) represent a sequence of receptor states.

**Input-output relationship.** The state of the receptor is dependent on the previous input and the previous state, forming a Markov transition PMF \( p_{Y_i | X_i, Y_{i-1}}(y_i | x_i, y_{i-1}) \). Under our model, if \( Y_{i-1} = U \), i.e. the receptor was previously unbound, then the distribution of \( Y_i \) depends on the input concentration \( X_i \). However, if \( Y_{i-1} = B \), then \( Y_i \) is independent of \( X_i \).

Thus, the Markov transition PMF \( p_{Y_i | X_i, Y_{i-1}}(y_i | x_i, y_{i-1}) \) has \( m+1 \) parameters: the \( m \)-dimensional vector \( \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m] \) of binding rates, where

\[
\alpha_j := p_{Y_i | X_i, Y_{i-1}}(B | c_j, U);
\]  

(5)

and \( \beta \), the unbinding rate, independent of input signal concentration, where

\[
\beta := p_{Y_i | X_i, Y_{i-1}}(U | c_j, B)
\]  

(6)

which is constant for all \( c_j \in X \). This may also be written as a state transition probability matrix

\[
P_{Y_i | X = c_j} = \begin{bmatrix} 1 - \alpha_j & \alpha_j \\ \beta & 1 - \beta \end{bmatrix}.
\]  

(7)

Recalling the notation from (4), we write \( \alpha_L \) and \( \alpha_H \) for the lowest and highest binding rates, respectively. Thus, we can write \( \alpha = [\alpha_L, \alpha_2, \ldots, \alpha_{m-1}, \alpha_H] \).

From the above discussion, the sequence \( Y^n \), given \( X^n \) and initial state \( Y_0 \), forms a time-inhomogeneous Markov chain with PMF

\[
p_{Y^n | X^n, Y_0}(y^n | x^n, y_0) = \prod_{i=1}^{n} p_{Y_i | X_i, Y_{i-1}}(y_i | x_i, y_{i-1}).
\]  

(8)
The dependencies of the transition probabilities can be illustrated graphically as follows:

\[
\begin{array}{cccccc}
X_1 & X_2 & X_3 & X_4 & X_5 & \cdots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
Y_0 & \rightarrow & Y_1 & \rightarrow & Y_2 & \rightarrow \cdots
\end{array}
\]

We give the following expressions and definitions, which will be useful in the remainder of this section. For an IID input distribution \( p_X(x) \), since there are \( m \) possible values for \( x \), we will express \( p_X(x) \) as a vector \( p \), with elements

\[
p = [p_1, p_2, \ldots, p_m] \quad (9)
\]

\[
= [p_L, p_2, p_3, \ldots, p_{m-1}, p_H]. \quad (10)
\]

For the IID input distribution vector \( p \), let \( \bar{\alpha}_p \) represent the average binding probability, given by

\[
\bar{\alpha}_p = \sum_{j=1}^{m} \alpha_j p_j. \quad (11)
\]

Finally, we give a condition on the parameters that will be used in many of our results:

**Definition 1 (Strictly Ordered Parameters):** The parameters \( \alpha \) and \( \beta \) are said to be strictly ordered if they satisfy

\[
0 < \alpha_L < \alpha_2 < \alpha_3 < \ldots < \alpha_{m-1} < \alpha_H < 1 \quad (12)
\]

and

\[
0 < \beta < 1. \quad (13)
\]

**B. Mutual information and capacity under IID inputs**

Let \( C \) represent the Shannon capacity of the system; as this is a channel with memory, capacity is defined by

\[
C = \lim_{n \to \infty} \max_{p_X^n} \frac{1}{n} I(X^n; Y^n), \quad (14)
\]

Let \( C_{\text{IID}} \) represent the capacity from (14) where \( p_X^n(x^n) \) is constrained to be IID, i.e., we can write \( p_X^n(x^n) = \prod_{i=1}^{n} p_X(x_i) \).
If the input distribution is IID, then $Y^n$ forms a homogeneous Markov chain [55]. To see this, we start with

$$p_{Y^n, X^n|y_0}(y^n, x^n \mid y_0) = p_{Y^n|X^n, y_0}(y^n \mid x^n, y_0)p_X(x^n)$$  \hspace{1cm} (15)

$$= p_{Y^n|X^n, y_0}(y^n \mid x^n, y_0) \prod_{i=1}^{n} p_X(x_i)$$  \hspace{1cm} (16)

$$= \prod_{i=1}^{n} p_{Y_i|X_i, Y_{i-1}}(y_i \mid x_i, y_{i-1})p_X(x_i),$$  \hspace{1cm} (17)

where (17) follows from (8). Continue by letting

$$p_{Y_i|Y_{i-1}}(y_i \mid y_{i-1}) = \sum_{x_i} p_{Y_i|X_i, Y_{i-1}}(y_i \mid x_i, y_{i-1})p_X(x_i).$$  \hspace{1cm} (18)

Finally, marginalizing over $X^n$,

$$p_{Y^n|y_0}(y^n \mid y_0) = \sum_{x^n} p_{Y^n, X^n|y_0}(y^n, x^n \mid y_0)$$  \hspace{1cm} (19)

$$= \prod_{i=1}^{n} \sum_{x_i} p_{Y_i|X_i, Y_{i-1}}(y_i \mid x_i, y_{i-1})p_X(x_i)$$  \hspace{1cm} (20)

$$= \prod_{i=1}^{n} p_{Y_i|Y_{i-1}}(y_i \mid y_{i-1}),$$  \hspace{1cm} (21)

which is the distribution of a homogeneous Markov chain. If $y_{i-1} = U$,

$$p_{Y_i|Y_{i-1}}(B \mid U) = \sum_{x_i} p_{Y_i|X_i, Y_{i-1}}(B \mid x_i, U)p_X(x_i)$$  \hspace{1cm} (22)

$$= \sum_{j=1}^{m} \alpha_j p_j$$  \hspace{1cm} (23)

$$= \bar{\alpha} p,$$  \hspace{1cm} (24)

with $p_{Y_i|Y_{i-1}}(U \mid U) = 1 - \bar{\alpha} p$. If $y_{i-1} = B$,

$$p_{Y_i|Y_{i-1}}(U \mid B) = \sum_{x_i} p_{Y_i|X_i, Y_{i-1}}(U \mid x_i, B)p_X(x_i)$$  \hspace{1cm} (25)

$$= \sum_{j=1}^{m} \beta p_j$$  \hspace{1cm} (26)

$$= \beta,$$  \hspace{1cm} (27)
with \( p_{Y_i|Y_{i-1}}(B \mid B) = 1 - \beta \). The transition probability matrix for \( Y \) is given by
\[
P_Y = \begin{bmatrix}
1 - \bar{\alpha}_p & \bar{\alpha}_p \\
\beta & 1 - \beta
\end{bmatrix}.
\] (28)

Suppose the parameters are strictly ordered (Definition 1). Then \( Y \) has a stationary distribution, by inspection of (28). The stationary probability of state \( U \) is given by
\[
p_Y(U) = \frac{1}{1 + \bar{\alpha}_p/\beta},
\] (29)
and \( p_Y(B) = 1 - p_Y(U) \).

When \( Y^n \) is a stationary, homogeneous Markov chain, we can write the mutual information rate as
\[
\lim_{n \to \infty} \frac{1}{n}I(X; Y) = H(Y_n \mid Y_{n-1}) - H(Y_n \mid X_n, Y_{n-1}).
\] (30)

Let
\[
\mathcal{H}(p) = -p \log p - (1-p) \log(1-p)
\] (31)
represent the binary entropy function. Dealing with each term on the right hand side of (30) individually,
\[
H(Y_n \mid Y_{n-1}) = p_Y(U)H(Y_n \mid Y_{n-1} = U) + p_Y(B)H(Y_n \mid Y_{n-1} = B)
\] (32)
\[
= p_Y(U)\mathcal{H}(\bar{\alpha}_p) + p_Y(B)\mathcal{H}(\beta)
\] (33)
which follows from (24)-(27); and
\[
H(Y_n \mid X_n, Y_{n-1}) = \sum_{x_n} p_X(x_n)p_Y(U)H(Y_n \mid X_n = x_n, Y_{n-1} = U)
\]
\[
+ \sum_{x_n} p_X(x_n)p_Y(B)H(Y_n \mid X_n = x_n, Y_{n-1} = B)
\] (34)
\[
= p_Y(U)\sum_{j=1}^m p_j\mathcal{H}(\alpha_j) + p_Y(B)\mathcal{H}(\beta).
\] (35)

Then
\[
\lim_{n \to \infty} \frac{1}{n}I(X; Y) = H(Y_n \mid Y_{n-1}) - H(Y_n \mid X_n, Y_{n-1})
\] (36)
\[
= p_Y(U)\left(\mathcal{H}(\bar{\alpha}_p) - \sum_{j=1}^m p_j\mathcal{H}(\alpha_j)\right)
\] (37)
\[
= \frac{\mathcal{H}(\bar{\alpha}_p) - \sum_{j=1}^m p_j\mathcal{H}(\alpha_j)}{1 + \bar{\alpha}_p/\beta}.
\] (38)
Finally, $C_{\text{IID}}$ is given by

$$C_{\text{IID}} = \max_p \frac{\mathcal{H}(\bar{\alpha}_p) - \sum_{j=1}^m p_j \mathcal{H}(\alpha_j)}{1 + \bar{\alpha}_p / \beta}. \quad (39)$$

**C. $C_{\text{IID}}$ is achieved with all probability mass on $x = L$ and $x = H$**

Here we show that the $C_{\text{IID}}$-achieving input distribution $p_X(x)$ uses only the extreme values of concentration: $L$ and $H$. The result is stated as follows.

*Theorem 1*: Let $p^* = [p^*_1, p^*_2, \ldots, p^*_m]$ represent an IID distribution that maximizes \(39\). If the parameters are strictly ordered (see Definition 1), then $p^*_2 = p^*_3 = \ldots = p^*_m-1 = 0$.

*Proof*: The proof proceeds by contradiction. Assume the theorem is false: that $p^*_i > 0$ for at least one index $i$ in $\{2, 3, \ldots, m-1\}$. Let $u$ represent the smallest index in $\{2, 3, \ldots, m-1\}$ such that $p^*_u > 0$. From the initial assumption, $u$ must exist, and $\alpha_u$ is the corresponding binding probability.

Since $\alpha_1 < \alpha_u < \alpha_m$, there exist constants $\pi_1$ and $\pi_m$ such that

$$0 < \pi_1, \pi_m < 1 \quad (40)$$

$$\pi_1 + \pi_m = 1 \quad (41)$$

$$\pi_1 \alpha_1 + \pi_m \alpha_m = \alpha_u \quad (42)$$

Let $q = [q_1, \ldots, q_m]$ represent a distribution constructed as follows:

$$q_1 = p^*_1 + p^*_u \pi_1 \quad (43)$$

$$q_m = p^*_m + p^*_u \pi_m \quad (44)$$

$$q_u = 0 \quad (45)$$

$$q_j = p^*_j \quad \forall j \neq \{1, u, m\}. \quad (46)$$

Note that $q$ is constructed so that $\bar{\alpha}_q = \bar{\alpha}_{p^*}$ (see \(11\)).

From \(38\), the mutual information under distribution $p^*$ is

$$I_{p^*}(X; Y) = \frac{\mathcal{H}(\bar{\alpha}_{p^*}) - \sum_{i=1}^m p^*_i \mathcal{H}(\alpha_i)}{1 + \bar{\alpha}_{p^*} / \beta}, \quad (47)$$

and under distribution $q$ it is (recalling $\bar{\alpha}_q = \bar{\alpha}_{p^*}$)

$$I_q(X; Y) = \frac{\mathcal{H}(\bar{\alpha}_{p^*}) - \sum_{i=1}^m q_i \mathcal{H}(\alpha_i)}{1 + \bar{\alpha}_{p^*} / \beta}. \quad (48)$$
Equations (47)-(48) differ only in the term under summation. We can write
\[
\sum_{i=1}^{m} p_i^* \mathcal{H} (\alpha_i) - \sum_{i=1}^{m} q_i \mathcal{H} (\alpha_i) = \left( p_1^* \mathcal{H} (\alpha_1) + p_u^* \mathcal{H} (\alpha_u) + p_m^* \mathcal{H} (\alpha_m) \right) - \left( q_1 \mathcal{H} (\alpha_1) + q_u \mathcal{H} (\alpha_u) + q_m \mathcal{H} (\alpha_m) \right) \tag{49}
\]
\[
= p_u^* \left( \mathcal{H} (\alpha_u) - \left( \pi_1 \mathcal{H} (\alpha_1) + \pi_m \mathcal{H} (\alpha_m) \right) \right) \tag{50}
\]
\[
= p_u^* \left( \pi_1 \alpha_1 + \pi_m \alpha_m - \left( \pi_1 \mathcal{H} (\alpha_1) + \pi_m \mathcal{H} (\alpha_m) \right) \right), \tag{51}
\]
where (49) follows from (46), (50) follows from (43)-(45), and (51) follows from (42). \(\mathcal{H}\) is strictly concave, \(0 < \pi_1, \pi_m < 1\) (from (40)), and \(\alpha_1 < \alpha_m\) (by assumption), so (51) is always positive. This implies that
\[
I_{p^*}(X;Y) < I_q(X;Y). \tag{52}
\]
However, \(p^*\) is the maximizing IID input distribution (by definition), which is a contradiction. The theorem follows.

D. Capacity and feedback capacity are achieved by an IID input distribution

The directed information [56] between vectors \(X^n\) and \(Y^n\), written \(I(X^n \rightarrow Y^n)\), is given by
\[
I(X^n \rightarrow Y^n) = \sum_{i=1}^{n} I(X_i; Y_i | Y_{i-1}). \tag{53}
\]
The per-symbol directed information rate is given by
\[
\lim_{n \to \infty} \frac{1}{n} I(X^n \rightarrow Y^n). \tag{54}
\]
Let \(Y_0^n = [Y_0, Y_1, \ldots, Y_n]\) and let \(\mathcal{P}\) represent the set of causal-conditional feedback input distributions: \(p_{X^n|Y_0^n}(x^n | y_0^n) \in \mathcal{P}\) if and only if \(p_{X^n|Y_0^n}(x^n | y_0^n)\) can be written as
\[
p_{X^n|Y_0^n}(x^n | y_0^n) = \prod_{k=1}^{n} p_{X_k|X_{k-1},Y_{k-1}}(x_k | x_{k-1}, y_{k-1}), \tag{55}
\]
where \(p_{X_i|X^0,Y^0}(x_1 | x^0, y^0) = p_{X_i|Y_0}(x_1 | y_0)\). Feedback capacity, \(C_{FB}\), is then given by
\[
C_{FB} = \max_{p_{X^n|Y_0^n}(x^n | y_0^n) \in \mathcal{P}} \left( \lim_{n \to \infty} \frac{1}{n} I(X^n \rightarrow Y^n) \right). \tag{56}
\]
Our capacity result is stated as follows.
Theorem 2: If the parameters are strictly ordered (see Definition 1), then

\[ C_{FB} = C = C_{IID}. \]  \hspace{1cm} (57)

We give several lemmas prior to proving the main result. Let \( P^* \subseteq P \) (see (56)-(55)) represent the set of feedback input distributions that can be written

\[ p_{X^n|Y_0^n}(x^n \mid y_0^n) = \prod_{i=1}^{n} p_{X_i|Y_{i-1}}(x_i \mid y_{i-1}). \]  \hspace{1cm} (58)

(Note that distributions in \( P^* \) need not be stationary: \( p_{X_i|Y_{i-1}}(x \mid y) \) can depend on \( i \).) Then \( P^* \subset P \) for \( n > 2 \). The following result, found in the literature, says there is at least one feedback-capacity–achieving input distribution in \( P^* \).

Lemma 1: Taking the maximum in (56) over \( P^* \subset P \),

\[ \max_{p_{X^n|Y_0^n}(x^n \mid y_0^n) \in P^*} \left( \lim_{n \to \infty} \frac{1}{n} I(X^n \to Y^n) \right) = C_{FB}. \]  \hspace{1cm} (59)

**Proof:** The lemma follows from [54, Thm. 1]. \[\Box\]

If the feedback-capacity–achieving input distribution is in \( P^* \), then \( Y^n \) is a Markov chain (the reader may check; see also [54], [55]). That is,

\[ p_{Y_i|Y_{i-1}}(y_i \mid y_{i-1}) = p_{Y_i|Y_{i-1}}(y_i \mid y_{i-1}). \]  \hspace{1cm} (60)

Using the following shorthand notation:

\[ p_{j|i}^{(i)} := p_{X_i|Y_{i-1}}(e_j \mid B) \]  \hspace{1cm} (61)

\[ p_{j|i}^{(i)} := p_{X_i|Y_{i-1}}(e_j \mid U), \]  \hspace{1cm} (62)

where the superscripts represent the time index, the transition probability \( p_{Y_i|Y_{i-1}}(y_i \mid y_{i-1}) \) may be represented as a matrix \( P_Y^{(i)} \), where

\[ P_Y^{(i)} = \begin{bmatrix} 1 - \sum_{j=1}^{m} \alpha_j p_{j|i}^{(i)} \sum_{j=1}^{m} \alpha_j p_{j|i}^{(i)} & \beta \\ \beta & 1 - \beta \end{bmatrix} \]  \hspace{1cm} (63)

(cf. (28), where the input distribution is IID).

We now consider stationary distributions. Let \( P^{**} \subset P^* \) represent the feedback input distributions that can be written with stationary \( p_{X_i|Y_{i-1}}(x_i \mid y_{i-1}) \), i.e., with some time-independent distribution \( p_{X|Y} \) such that

\[ p_{X^n|Y_0^n}(x^n \mid y_0^n) = \prod_{k=1}^{n} p_{X|Y}(x_i \mid y_{i-1}). \]  \hspace{1cm} (64)
Then:

**Lemma 2**: Suppose the parameters are strictly ordered (Definition 1). Taking the maximum in \((66)\) over \(\mathcal{P}^{**} \subset \mathcal{P}^* \subset \mathcal{P}^*\),
\[
\max_{p_{X^n|Y^n_0} (x^n | y^n_0) \in \mathcal{P}^{**}} \left( \lim_{n \to \infty} \frac{1}{n} I(X^n \to Y^n) \right) = C_{HD}.
\]

**Proof**: We start by showing that \(I(X^i; Y_i | Y^{i-1})\) is independent of \(p^{(k)}_{L|B}\) for all \(k\). There is a feedback-capacity–achieving input distribution in \(\mathcal{P}^*\) (from Lemma 1). Using this input distribution,
\[
\begin{align*}
I(X^i; Y_i | Y^{i-1}) &= H(Y_i | Y^{i-1}) - \underbrace{H(Y_i | Y_{i-1}, X^i)}_{(67)} \quad (66) \\
&= H(Y_i | Y_{i-1}) - \underbrace{H(Y_i | Y_{i-1}, X_i)}_{(67)}.
\end{align*}
\]

where \((67)\) follows since (by definition) \(Y_i\) is conditionally independent of \(X^{i-1}\) given \(Y_{i-1}\), and since (from the parameters being strictly ordered) \(Y^i\) is a stationary, homogeneous first-order Markov chain. Expanding \((67)\),
\[
\begin{align*}
I(X^i; Y_i | Y^{i-1}) &= \sum_{y_{i-1}} p_{Y_{i-1}}(y_{i-1}) \sum_{x_i} p_{X_i|Y_{i-1}}(x_i | y_{i-1}) \\
&\quad \cdot \sum_{y_i} p_{Y_i|Y_{i-1}, X_i}(y_i | y_{i-1}, x_i) \log \frac{p_{Y_i|Y_{i-1}, X_i}(y_i | y_{i-1}, x_i)}{p_{Y_i|Y_{i-1}}(y_i | y_{i-1})}.
\end{align*}
\]

From \((63)\), \(p_{Y_{i-1}}(y_{i-1})\) is calculated from parameters in \(\mathcal{P}^{(j)}_Y\) and the initial state, so \(p_{Y_{i-1}}(y_{i-1})\) is independent of \(p^{(k)}_{j|B}\) for all \(j\) and \(k\). Further, everything under the last sum (over \(y_i\)) is independent of \(p^{(k)}_{j|B}\), from \((63)\) and the definition of \(p_{Y_i|Y_{i-1}, X_i}(y_i | y_{i-1}, x_i)\). There remains the term \(p_{X_i|Y_{i-1}}(x_i | y_{i-1})\), which is dependent on \(p^{(i-1)}_{j|B}\) when \(y_{i-1} = B\). However, if \(y_{i-1} = B\), then
\[
\begin{align*}
\sum_{y_i} p_{Y_i|Y_{i-1}, X_i}(y_i | B, x_i) \log \frac{p_{Y_i|Y_{i-1}, X_i}(y_i | B, x_i)}{p_{Y_i|Y_{i-1}}(y_i | B)} &= \sum_{y_i} p_{Y_i|Y_{i-1}}(y_i | B) \log \frac{p_{Y_i|Y_{i-1}}(y_i | B)}{p_{Y_i|Y_{i-1}}(y_i | B)} \\
&= \sum_{y_i} p_{Y_i|Y_{i-1}}(y_i | B) \log 1 = 0
\end{align*}
\]
where (69) follows since $y_i$ is independent of $x_i$ in state $B$. Thus, the entire expression is independent of $p^{(k)}_{L|B}$ for all $k$. Moreover, from (53), directed information is independent of $p^{(k)}_{j|B}$ for all $k$.

To prove (65), distributions in $\mathcal{P}^{**}$ have $p^{(1)}_{j|U} = p^{(2)}_{j|U} = \ldots$, and $p^{(1)}_{j|B} = p^{(2)}_{j|B} = \ldots$. Since $I(X^i; Y_i | Y^{i-1})$ is independent of $p^{(k)}_{j|B}$ for all $k$ (by the preceding argument), we may set $p^{(k)}_{j|B} = p^{(k)}_{j|U}$ for all $j$ and $k$, without changing $I(X^i; Y_i | Y^{i-1})$. Thus, inside $\mathcal{P}^{**}$, there exists a maximizing input distribution that is independent for each channel use. By the definition of $\mathcal{P}^{**}$, that maximizing input distribution is IID, and there cannot exist an IID input distribution outside of $\mathcal{P}^{**}$.

Finally, we must show that $C_{FB}$ is itself achieved by a distribution in $\mathcal{P}^{**}$. To do so, we rely on [55, Thm. 4], which shows that this is the case, as long as several technical conditions are satisfied. Stating the conditions and proving that they hold for this channel requires restatement of definitions from [55], so we give this result in Appendix B as Lemma 3.

We can now return to the proof of Theorem 2, where we relate these results to the Shannon capacity $C$.

**Proof:** From Lemma 1, $C_{FB}$ is satisfied by an input distribution in $\mathcal{P}^*$. From Lemma 2, if we restrict ourselves to the stationary input distributions $\mathcal{P}^{**}$ (where $\mathcal{P}^{**} \subset \mathcal{P}^*$), then the feedback capacity is $C_{IID}$. From Lemma 3, the conditions of [55, Thm. 4] are satisfied, which implies that there is a feedback-capacity–achieving input distribution in $\mathcal{P}^{**}$. Therefore,

$$C_{FB} = C_{IID}. \quad (72)$$

For general channels,

$$C_{FB} \geq C \geq C_{IID}. \quad (73)$$

because the set $\mathcal{P}$ includes the set of input distributions without feedback, and because the set of input distributions without feedback includes the IID input distributions. The theorem follows from (72) and (73).

From Theorems 1 and 2 and Equation (39), we have the following.

**Corollary 1:** Capacity $C$ of the discrete channel model given in (8) is given by

$$C = \max_{p_H} \frac{H(p_L \alpha_L + p_H \alpha_H) - p_L H(\alpha_L) - p_H H(\alpha_H)}{1 + (p_L \alpha_L + p_H \alpha_H)/\beta}, \quad (74)$$

where it is sufficient to maximize over $p_H$, since $p_L = 1 - p_H$. 

This result has an intuitively appealing form: the mutual information rate appearing in (39) and (74) is the product of the binary channel MI rate with transition probabilities \(\{\alpha_L, \alpha_H\}\), and the fraction of time the channel is in the sensitive (unbound) state.

In Figure 2, we illustrate the behaviour of the maximizing value of \(p_H\): in this figure, the mutual information in (38) is plotted for \(\alpha_L = 0.1, \beta = 0.9\), and various values of \(\alpha_H\); in the input distribution, all \(p_j\) are equal to zero except \(p_L\) and \(p_H\). The maximum values on each mutual information curve are illustrated. In Figure 3, we give a contour plot of capacity for values of \(\alpha_H\) and \(\alpha_L\), where \(\beta = 0.9\).

E. Capacity- and mutual-information–maximizing parameter values

Recall the definition of capacity in (74). Here we show the values of the parameters \(\alpha_L, \alpha_H, \text{ and } \beta\) that maximize the mutual information, and hence the capacity. (It will turn out that all these parameters are equal to either 0 or 1, thus violating the strict ordering assumption.)

Dropping the maximization from (74), mutual information is written

\[
I(X;Y) = \mathcal{H}(\alpha_H p_H + \alpha_L p_L) - p_H \mathcal{H}(\alpha_H) - p_L \mathcal{H}(\alpha_L) \frac{1}{1 + (p_H \alpha_H + p_L \alpha_L)/\beta}.
\]

(75)

We assume that the parameters are strictly ordered (see Definition 1). With this assumption, \(I(X;Y) > 0\); therefore, the same is true of the numerator in (75), since the denominator is positive. Also note that

\[
\frac{d}{dp} \mathcal{H}(p) = \log \frac{1 - p}{p}.
\]

(76)

We will use these properties below.

First consider \(\beta\). By inspection of (75), \(\beta\) only appears in the denominator, and the denominator decreases with increasing \(\beta\). Thus, \(I(X;Y)\) is increasing in \(\beta\), and \(\beta = 1\) is optimal.

Now consider \(\alpha_L\). By inspection of (75), the denominator is increasing in \(\alpha_L\). We can show that the numerator is decreasing in \(\alpha_L\): we can write

\[
\frac{d}{d\alpha_L} \left( \mathcal{H}(\alpha_H p_H + \alpha_L p_L) - p_H \mathcal{H}(\alpha_H) - p_L \mathcal{H}(\alpha_L) \right)
\]

\[
= p_L \log \frac{(1 - \alpha_H p_H - \alpha_L p_L) \alpha_L}{(\alpha_H p_H + \alpha_L p_L) (1 - \alpha_L)}
\]

(77)

\[
= p_L \log \frac{\alpha_L - \alpha_L (\alpha_H p_H + \alpha_L p_L)}{\alpha_H p_H + \alpha_L p_L - \alpha_L (\alpha_H p_H + \alpha_L p_L)}
\]

\[
\leq 0,
\]

(78)
where the final inequality follows since $\alpha_L \leq \alpha_H p_H + \alpha_L p_L$ (since $\alpha_L \leq \alpha_H$). Thus, $I(X; Y)$ is decreasing in $\alpha_L$, and $\alpha_L = 0$ is optimal.

Finally, consider $\alpha_H$: this case is slightly trickier than $\alpha_L$, since both the numerator and denominator of (75) are increasing. For simplicity, we start by substituting $\beta = 1$ and $\alpha_L = 0$: we have

$$I(X; Y) = \frac{\mathcal{H}(\alpha_H p_H) - p_H \mathcal{H}(\alpha_H)}{1 + p_H \alpha_H}. \tag{80}$$
Fig. 3. Contour plot of capacity with respect to $\alpha_L$ and $\alpha_H$, fixing $\beta = 0.9$. Note that $\alpha_L > \alpha_H$ in the upper left triangle, so capacity here is undefined.

To show that this quantity is increasing with $\alpha_H$, the first derivative with respect to $\alpha_H$ is

$$
\frac{d}{d\alpha_H} I(X; Y) = \frac{\left(1 + p_H \alpha_H\right) p_H \log \left(\frac{1 - \alpha_H p_H}{(1 - \alpha_H p_H)}\right) - p_H H(\alpha_H p_H) - p_H H(\alpha_H)}{(1 + \alpha_H p_H)^2}.
$$

The goal is to determine whether $dI(X; Y)/d\alpha_H$ is positive. It is useful to write

$$
H(\alpha_H p_H) - p_H H(\alpha_H) = \log \frac{(1 - \alpha_H)^p_H}{1 - \alpha_H p_H} + \alpha_H p_H \log \frac{(1 - \alpha_H p_H)}{(1 - \alpha_H)^p_H}.
$$
Thus, (81) becomes
\[
\frac{d}{d\alpha} I(X; Y) = \frac{p_H}{(1 + \alpha_H p_H)^2} (1 + p_H \alpha_H) \log \frac{(1 - \alpha_H p_H)}{(1 - \alpha_H) p_H}
\]
\[
- \frac{p_H}{(1 + \alpha_H p_H)^2} \left( \log \frac{(1 - \alpha_H)^{p_H}}{1 - \alpha_H p_H} - \alpha_H p_H \log \frac{(1 - \alpha_H p_H)}{(1 - \alpha_H) p_H} \right)
\]
\[
= \frac{p_H}{(1 + \alpha_H p_H)^2} \log \frac{(1 - \alpha_H p_H)^2}{(1 - \alpha_H)^{p_H+1} p_H}.
\] (83)

By inspection of (84), the derivative is positive when
\[
\frac{(1 - \alpha_H p_H)^2}{(1 - \alpha_H)^{p_H+1} p_H} \geq 1.
\] (84)

Inequality (85) is satisfied for \( \alpha_H = 0 \) (as \( 1/p_H \geq 1 \)); to show that it is satisfied for all strictly ordered \( \alpha_H \), we show that the left side of (85) is increasing for \( \alpha_H \geq 0 \). After some manipulation, we have
\[
\frac{d}{d\alpha} \frac{(1 - \alpha_H p_H)^2}{(1 - \alpha_H)^{p_H+1} p_H}
\]
\[
= \frac{p_H (1 - \alpha_H)^{p_H} (1 - \alpha_H p_H) (1 - p_H) (1 + \alpha_H p_H)}{((1 - \alpha_H)^{p_H+1} p_H)^2},
\] (86)

which is positive for all strictly ordered parameters. Thus, \( I(X; Y) \) is increasing in \( \alpha_H \), and \( \alpha_H = 1 \) is optimal.

The proceeding analysis is true for any valid setting of \( p_H \) and \( p_L \). Therefore, it applies to capacity as well as mutual information.

### F. Feedback Can Increase Capacity in a General Signal Transduction Network Channel

To close this section, we may wonder if it is true that \( C_{FB} = C \) in general signal transduction models. The answer is no, which we show using a (somewhat pathological) counterexample.

Consider a signal transduction channel with binary inputs (H and L) and three receptor states. Suppose the transition probability matrices for concentration H and concentration L are given, respectively, by
\[
T_H = \begin{bmatrix}
1 - \alpha_{12}^H & \alpha_{12}^H & 0 \\
\beta & 1 - (\alpha_{23}^H + \beta) & \alpha_{23}^H \\
0 & 0 & 1
\end{bmatrix}, \quad T_L = \begin{bmatrix}
1 - \alpha_{12}^L & \alpha_{12}^L & 0 \\
\beta & 1 - \beta & 0 \\
0 & 0 & 1
\end{bmatrix}
\] (87)

where \( 0 < \alpha_{ij}^L < \alpha_{ij}^H < 1 \) and \( 0 < \beta < 1 \). The reader will note that the bottom row of \( T_H \) and \( T_L \) represents an absorbing state: once the receptor enters that state it remains there forever, and
subsequent mutual information between the input and output is zero. The other two states are
direct analogs of the “unbound” and “bound” states considered in the two-state receptor binding
model. (Note that this example violates the assumptions of Lemma 3: the parameters are not
strictly ordered, and the matrices $T$ are not strongly irreducible.)

In the absence of feedback, it is easy to see that the mutual information rate is identically
zero. If the input remains in the low state (i.e. $\forall n, \Pr(X_n = L) = 1$), then $I(X^n; Y^n) \equiv 0$. If the
input leaves the low state at most a finite number of times, then $\lim_{n \to \infty} (1/n) I(X^n; Y^n) = 0$.
If the input leaves the low state repeatedly, then within a finite number of instances of $X_n = H$
the channel will enter the third state, after which subsequent information gain will be nil.

In the presence of feedback, on the other hand, the sender can opt to adopt condition 2
($X_n = H$) whenever $Y_{n-1} = 2$, thereby avoiding the possibility of becoming trapped in state
3. Under this strategy, the channel can transmit a nonzero amount of information whenever it
is in state $Y_{n-1} = U$, which occurs with nonzero probability. Hence the channel capacity with
feedback allowed is strictly positive, while in the absence of feedback, the capacity is zero.
III. MARKOV INPUTS

Although we found in the previous section that there exists a capacity-achieving input distribution that is IID, physical concentration does not behave like an IID random variable: concentrations, either low or high, can persist for long periods of time. In this section, we analyze the capacity of the discrete time channel when the channel inputs are Markov, though we restrict ourselves to binary Markov inputs (L and H) for simplicity.

**Change of indices:** When $X^n$ forms a Markov chain, the joint process $(X^n, Y^n)$ also forms a Markov chain. Modifying (17), we can write

$$ p_{Y^n, X^n | Y_0, X_0}(y^n, x^n | y_0, x_0) = \prod_{i=1}^{n-1} p_{X_{i+1} | X_i}(x_{i+1} | x_i)p_{Y_i | X_i, Y_{i-1}}(y_i | x_i, y_{i-1}) p_{X_1 | X_0}(x_1 | x_0) $$

$$ = \left( \prod_{i=1}^{n} p_{Y_i | X_i, Y_{i-1}}(y_i | x_i, y_{i-1}) \right) \left( \prod_{j=0}^{n-1} p_{X_{j+1} | X_j}(x_{j+1} | x_j) \right) . \quad (88) $$

The state of the joint Markov process is given by $(x_i, y_{i-1})$. The misaligned indices result in the complicated expression in (88). We used this indexing in Section II because this indexing is more common in the directed information literature. However, in this section, it will be more convenient to change the indexing to align the indices, as follows:

$$ y_{i-1} \rightarrow y_i $$

for all indices $i$. Using the new indexing in (89), the state of the joint Markov process is $z_i = (x_i, y_i)$, and (88) becomes

$$ p_{Y^n, X^n | Y_0, X_0}(y^n, x^n | y_0, x_0) = \prod_{i=1}^{n} p_{Y_i | X_{i-1}, Y_{i-1}}(y_i | x_{i-1}, y_{i-1}) p_{X_i | X_{i-1}}(x_i | x_{i-1}) . \quad (90) $$

The graph depicting the relationship among inputs and outputs is now given by

```
X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow \cdots
```

```
Y_0 \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_4 \rightarrow \cdots
```

This indexing will be used for the remainder of the paper, unless specified. Since we are careful to ensure that our Markov chains are stationary, the results in Section II are still valid under this change of indices.
A. Mathematical model with Markov inputs

Assume the sequence $X^n$ forms a Markov chain with two parameters, $r$ (the L-to-H transition probability) and $s$ (the H-to-L transition probability), giving a transition probability matrix of

$$
P_X = \begin{bmatrix}
1 - r & r \\
s & 1 - s
\end{bmatrix},
$$

with entries for L on the first row and column, and H on the second row and column.

The joint sequence $Z^n$ forms a four-state Markov chain with states $\{LU, LB, HU, HB\}$, with transition probability matrix (see Figure 4).

![Transition diagram](image)

Fig. 4. Transition diagram for a 2-state channel ($Y = U$, unbound receptor; $Y = B$, bound receptor) driven by a 2-state input Markov process ($X = L$, low concentration of signaling molecule; $X = H$, high concentration of signaling molecule). Probability per time step of $X_k = L$ to $X_{k+1} = H$ transition is $0 < r \leq 1$. Probability per time step of $X_k = H$ to $X_{k+1} = L$ transition is $0 < s \leq 1$. Probability per time step of $Y_k = U$ to $Y_{k+1} = B$ transition is $\alpha_L$ when $X_k = L$, and is $\alpha_H$ when $X_k = H$; these probabilities satisfy the strictly-ordered condition (Def. 1). Probability per time step of $Y_k = B$ to $Y_{k+1} = U$ is $0 < \beta < 1$, regardless of $X_k$. Compare Equation (92).

$$
P_Z = \begin{bmatrix}
(1 - \alpha_L)(1 - r) & \alpha_L(1 - r) & (1 - \alpha_L)r & \alpha_Lr \\
\beta(1 - r) & (1 - \beta)(1 - r) & \beta r & (1 - \beta)r \\
(1 - \alpha_H)s & \alpha_Hs & (1 - \alpha_H)(1 - s) & \alpha_H(1 - s) \\
\beta s & (1 - \beta)s & \beta(1 - s) & (1 - \beta)(1 - s)
\end{bmatrix}.
$$

(92)
The input $X^n$ has a unique stationary distribution if $0 < r, s < 1$. The chain $Z^n$ has a stationary distribution if $X^n$ has a stationary distribution, and the parameters are strictly ordered. (These conditions are sufficient, but not necessary.) The steady-state distribution on $X$ is given by

$$p_L = p_X(L) = \frac{s}{r+s}, \quad p_H = p_X(H) = \frac{r}{r+s}. \quad (93)$$

The stationary distribution of $Z$ is given by the (normalized) eigenvector of $P_Z$ with unit eigenvalue. This is given by $[p_{X,Y}(L, U), p_{X,Y}(L, B), p_{X,Y}(H, U), p_{X,Y}(H, B)]$, where

$$p_{X,Y}(L, U) = \frac{1}{K} (\beta s (-r - s + \alpha_H (r + s - 1) + \beta (r + s - 1))) \quad (94)$$
$$p_{X,Y}(L, B) = \frac{1}{K} (s (\alpha_H (\alpha_L (r + s - 1) - r) + \alpha_L (\beta (r + s - 1) - s))) \quad (95)$$
$$p_{X,Y}(H, U) = \frac{1}{K} (\beta r (-r - s + \alpha_L (r + s - 1) + \beta (r + s - 1))) \quad (96)$$
$$p_{X,Y}(H, B) = \frac{1}{K} (r (\alpha_H (-r + \alpha_L (r + s - 1) + \beta (r + s - 1)) - \alpha_L s)), \quad (97)$$

where $K$ is the normalization constant, to ensure the probabilities sum to 1. The expressions (94-97) may be simplified by introducing the notation

$$\bar{\alpha} = \frac{r \alpha_H + s \alpha_L}{r+s} \quad (98)$$
$$\lambda = 1 - r - s \quad (99)$$
$$\mu = \frac{\lambda}{1 - \lambda}. \quad (100)$$

The quantity $\bar{\alpha}$ is the mean value of $\alpha$ under the equilibrium distribution for $X$ (cf. (11) for IID inputs); $\lambda$ is the second eigenvalue of the matrix $P_X$; and $0 \leq \mu < \infty$ is a monotonically increasing function of $\lambda$. With this notation, the stationary distribution of the joint process satisfies

$$p_{X,Y}(L, U) = \frac{1}{K} \beta s (1 + (\alpha_H + \beta) \mu) \quad (101)$$
$$p_{X,Y}(L, B) = \frac{1}{K} s (\bar{\alpha} + \alpha_L (\alpha_H + \beta) \mu) \quad (102)$$
$$p_{X,Y}(H, U) = \frac{1}{K} \beta r (1 + (\alpha_L + \beta) \mu) \quad (103)$$
$$p_{X,Y}(H, B) = \frac{1}{K} r (\bar{\alpha} + \alpha_H (\alpha_L + \beta) \mu), \quad (104)$$

with normalization constant

$$K = \lambda (\alpha_L + \beta)(\alpha_H + \beta) + (r + s)(\bar{\alpha} + \beta). \quad (105)$$
From \( p_{X,Y}(X,Y) \) one may obtain the stationary marginal distribution \( p_Y(y) \). Define
\[
\Delta = \frac{s - r}{s + r}
\]
(106)
to represent the relative difference in probabilities between the low-to-high and high-to-low transitions. Then
\[
p_Y(U) = \frac{(1 + \mu(\bar{\alpha} + \beta + \Delta(\alpha_H - \alpha_L)))/K'}{(1 + \mu(\bar{\alpha} + \beta + \alpha_H\alpha_L))/K'},
\]
(107)
\[
p_Y(B) = \frac{(\bar{\alpha} + \mu(\bar{\alpha}\beta + \alpha_H\alpha_L))/K'}{K'}
\]
(108)
with normalization constant
\[
K' = 1 + \bar{\alpha} + \mu(\bar{\alpha} + \bar{\alpha}\beta + \beta + \alpha_H\alpha_L + \Delta(\alpha_H - \alpha_L)).
\]
(109)

B. Capacity estimates for Markov inputs

To estimate capacity for Markov inputs, we need the entropy rates for \( X, Y, \) and \( Z \). Since \( X \) and \( Z \) are stationary Markov processes, their entropy rates are available in closed form. The entropy rate of \( X \) is given as a function of \( r \) and \( s \) by
\[
\mathcal{H}(X) = \lim_{n \to \infty} \frac{1}{n} H(X^n)
\]
(110)
\[
= H(X_i \mid X_{i-1})
\]
(111)
\[
= p_L \left( r \log \frac{1}{r} + (1 - r) \log \frac{1}{1 - r} \right) + p_H \left( s \log \frac{1}{s} + (1 - s) \log \frac{1}{1 - s} \right).
\]
(112)
The joint entropy rate \( \mathcal{H}(Z) = \mathcal{H}(X,Y) \) can be calculated directly from (101)-(105). Let
\[
\pi_{xy} = p_{X,Y}(x, y)
\]
(113)
denote the stationary density of the Markov process, i.e. the four terms given in (101)-(104). Denote the joint transition probabilities from the matrix \( \mathbf{P}_Z \) as
\[
T_{xy \to x'y'} = p_{X_i,Y_i|X_{i-1},Y_{i-1}}(x', y' \mid x, y)
\]
(114)
\[
= p_{Y_i|X_{i-1},Y_{i-1}}(y' \mid x, y) p_{X_i|X_{i-1}}(x' \mid x).
\]
(115)
Further let
\[
\phi(p) = -p \log p.
\]
(116)
Then the joint entropy rate is
\[
\mathcal{H}(X, Y) = \sum_{x=L}^{H} \sum_{y=U}^{B} \pi_{xy} \left( \sum_{x'=L}^{H} \sum_{y'=U}^{B} \phi(T_{xy \rightarrow x'y'}) \right).
\] (117)

However, the output process \(Y\) is not a Markov process in general, and its entropy rate is difficult to express. To bound the entropy rate of \(Y\), we use the fact that this rate \(\mathcal{H}(Y)\) is bounded above and below by entropies conditioned on a finite number of previous channel states. From standard inequalities ([57], Theorem 4.4.1) we have, for each \(n\),
\[
H(Y_n|X_0, Y_0, \cdots, Y_{n-1}) \leq \mathcal{H}(Y) \leq H(Y_n|Y_0, \cdots, Y_{n-1}).
\] (118)
and
\[
\lim_{n \to \infty} H(Y_n|X_0, Y_0, \cdots, Y_{n-1}) = \mathcal{H}(Y) = \lim_{n \to \infty} H(Y_n|Y_0, \cdots, Y_{n-1}).
\] (119)

Using the inequalities (118), we have
\[
\mathcal{I}(X; Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y)
\] (120)
\[
\leq \mathcal{H}(X) - \mathcal{H}(X, Y) + H(Y_n|Y_0, \cdots, Y_{n-1}) =: \mathcal{I}^+_n(X; Y)
\] (121)
and
\[
\mathcal{I}(X; Y) \geq \mathcal{H}(X) - \mathcal{H}(X, Y) + H(Y_n|X_0, Y_0, \cdots, Y_{n-1}) =: \mathcal{I}^-_n(X; Y).
\] (122)
The required bounds on \(\mathcal{H}(Y)\) are derived below.

1) Upper Bounds on \(\mathcal{H}(Y)\) and \(\mathcal{I}(X; Y)\): First consider the one-step conditional entropy of the \(Y\) sequence,
\[
H(Y_1|Y_0) = \sum_{y_0} p_{Y}(y_0) \mathcal{H}(p_{Y_1|Y_0}(B|y_0))
\] (123)
\[
= \sum_{y_0} p_{Y}(y_0) \mathcal{H} \left( \sum_{x} \sum_{x'} T_{x_0y_0 \rightarrow x_1y_1B} \frac{\pi_{x_0y_0}}{\pi_{L}y_0 + \pi_{H}y_0} \right),
\] (124)
where \(\mathcal{H}\) is the binary entropy function, and \(\pi_{xy}\) and \(T_{x_0y_0 \rightarrow x_1y_1B}\) are the steady-state probability (resp. transition probability) of the \((X, Y)\) Markov chain, defined in (113) (resp. (115)). At the same time, the mutual information rate is bounded above by the entropy rate of the input (112).
Thus, from (112), (117), and (124), the first upper bound $I^+_1(X;Y)$ is given by

$$I^+_1(X;Y) = p_L \left( r \log \frac{1}{r} + (1 - r) \log \frac{1}{1 - r} \right) + p_H \left( s \log \frac{1}{s} + (1 - s) \log \frac{1}{1 - s} \right)$$

$$+ \left\{ - \sum_{x=L}^{H} \sum_{y=U}^{B} \pi_{xy} \left( \sum_{x'=L}^{H} \sum_{y'=U}^{B} \phi(T_{xy} \rightarrow x'y') \right) \right. \right.$$ 

$$+ \left. \sum_{y_0} p_Y(y_0) \mathcal{H} \left( \sum_{x_1} \sum_{x_0} T_{x_0, y_0 \rightarrow x_1} B \frac{\pi_{x_0, y_0}}{\pi_{L, y_0} + \pi_{H, y_0}} \right) \right\} \wedge 0,$$  \hspace{1cm} (125)

where $A \wedge B$ represents the lesser of $A$ and $B$. The bound $I^+_1(X;Y)$ is illustrated in Figure 5.

Fig. 5. The mutual information upper bound given by (125), for $\alpha_L = 0.1, \alpha_H = 0.9, \beta = 0.5$. Horizontal axis: $r$; vertical axis: $s$.

Next consider the two-step entropy, $H(Y_2|Y_0, Y_1)$. We calculate this entropy explicitly as follows:

$$H(Y_2|Y_0, Y_1) = \sum_{y_0, y_1} p_{Y_0, Y_1}(y_0, y_1) \mathcal{H} \left( p_{Y_2|Y_0, Y_1}(B \mid y_0, y_1) \right),$$  \hspace{1cm} (126)
where

\[ p_{Y_0,Y_1}(y_0, y_1) = \sum_{x_0,x_1} \pi_{x_0,y_0} T_{x_0,y_0 \rightarrow x_1,y_1} \]  

(127)

\[ p_{Y_2|Y_0,Y_1}(B \mid y_0, y_1) = \sum_{x_2} p_{X_2,Y_2|Y_0,Y_1}(x_2, B \mid y_0, y_1) \]  

(128)

and, writing \( X_0^n \) for \((X_0, \cdots, X_n)\),

\[ p_{X_2,Y_2,Y_0,Y_1}(x_2, B \mid y_0, y_1) = \sum_{x_0,x_1} p_{X_2^2,Y_2|Y_0,Y_1}(x_0^2, B \mid y_0, y_1) \]  

(129)

\[ = \sum_{x_0,x_1} \left( \frac{T_{x_1,y_1 \rightarrow x_2} B P_{X_0,Y_0,X_1,Y_1}(x_0, y_0, x_1, y_1)}{\sum_{x_0,x_1} P_{X_0,Y_0,X_1,Y_1}(x_0, y_0, x_1, y_1)} \right) \]  

(130)

\[ = \sum_{x_0,x_1} \left( \frac{T_{x_1,y_1 \rightarrow x_2} B T_{x_0,y_0 \rightarrow x_1,y_1} \pi_{x_0,y_0}}{\sum_{x_0,x_1} T_{x_0,y_0 \rightarrow x_1,y_1} \pi_{x_0,y_0}} \right). \]  

(131)

In general, the \( n \)th upper bound of this form is obtained from the \( n \)-step upper bound of the entropy rate of the channel state. Writing \( Y_0^{n-1} \) for \((Y_0, \cdots, Y_{n-1}) \in \{U,B\}^n\), the \( n \)-step upper bound is given by a sum involving \( 2^n \) terms

\[ \mathcal{H}_n^+ := H(Y_n \mid Y_0^{n-1}) \]  

(132)

\[ = \sum_{y_0^{n-1} \in \{U,B\}^n} p_{Y_0^{n-1}}(y_0^{n-1}) \mathcal{H}(p_{Y_n,Y_0^{n-1}}(B \mid y_0^{n-1})). \]  

(133)

In appendix \[ \square \] we briefly show how to use the sum-product algorithm to calculate the general \( n \)-step bound. Figure \[ \square \] illustrates the convergence of the sequence of upper bounds \( \mathcal{H}_n^+ \) with a similar sequence of lower bounds (next section) for \( n = 2, 3, 4, 5 \).

2) Lower Bounds on \( \mathcal{H}(Y) \) and \( \mathcal{I}(X;Y) \): In a similar fashion, we can formulate a lower bound on \( \mathcal{H}(Y) \) involving \( n \) prior states of \( Y \) and the initial state of \( X \), namely

\[ \mathcal{H}_n^- := H(Y_n \mid X_0, Y_0^{n-1}) \]  

(134)

\[ = \sum_{x_0 \in \{L,H\}} \sum_{y_0^{n-1} \in \{U,B\}^n} p_{X_0,Y_0^{n-1}}(x_0, y_0^{n-1}) \mathcal{H}(p_{Y_n,X_0,Y_0^{n-1}}(B \mid x_0, y_0^{n-1})). \]  

(135)

Moreover, we also have the trivial lower bound on the mutual information rate \( \mathcal{I}(X;Y) \geq 0 \).

Appendix \[ \square \] briefly shows how to perform this calculation using the sum-product algorithm. Figure \[ \square \] illustrates the convergence of \( \mathcal{H}_n^\pm \) in the interior of the region \( 0 < r, s < 1 \), for \( n = 2, 3, 4, 5 \). The upper and lower bounds obtained by conditioning \( Y \) to a depth of five steps
Fig. 6. Difference between upper and lower bounds for the mutual information rate $I(X;Y)$ of the 2-state Markov channel, as a function of switching parameters $0 < r < 1$ and $0 < s < 1$. The upper and lower bounds are given by (121) and (122), respectively. Each panel shows $\log_{10}(I^+_n - I^-_n)$ for $n = 2$ (panel A), $n = 3$ (panel B), $n = 4$ (panel C), $n = 5$ (panel D). Parameter values are $\alpha_L = 0.1, \beta = 0.5, \alpha_H = 0.9$. Each increase in the depth of conditioning decreases the gap between the upper and lower bounds by roughly an order of magnitude.

constrains the mutual information to within less than 1% for input switching rates satisfying $|r + s - 1| \lesssim 0.9$, or roughly all but 1% of the $(r,s)$ plane, for the parameters $(\alpha_L = 0.1, \beta = 0.5, \alpha_H = 0.9)$ illustrated in Fig. 6. We confirmed this result using Monte Carlo sampling to obtain empirical mutual information rates. Figure 7 shows the mutual information surface, as a function of $r$ and $s$, for the same parameters.
Fig. 7. Mutual information of the two state channel driven by a two-state Markov input process. The channel parameters are $\alpha_L = 0.1, \beta = 0.5, \alpha_H = 0.9$. The input switching rates are $r$ (low-to-high) and $s$ (high-to-low). **Top:** Three sets of curves show the upper bound $I^+_2(X;Y)$ (red), lower bound $I^-_2(X;Y)$ (black), and Monte Carlo estimate (blue). **Bottom:** Two sets of curves show the upper and lower bounds $\mathcal{I}^+_5(X;Y)$, which are indistinguishable over most of the $(r,s)$ plane. Horizontal axis: $r$; vertical axis: $s$. 
IV. CONTINUOUS TIME CHANNEL MODELS

In this section we consider the capacity in two limiting cases. In §IV-A we derive an expression for the capacity of the two-state channel in continuous time, by evaluating the limiting behavior of the discrete time mutual information rate in the limit of short time steps, and maximizing with respect to parameters. While this approach does not provide a rigorous proof of the capacity formula, it suggests an intuitively appealing form for the continuous time mutual information rate. Namely, the product of the mutual information rate of a counting process when the channel is in the receptive or unbound state, multiplied by the fraction of time it is in that state under stationary input conditions. In §IV-B we again consider the short time-step limit of the mutual information, but do so while fixing the per time-step release probability to be unity. In this case the continuous time channel without dead time gives the same mutual information rate and capacity expression as Kabanov’s Poisson channel [53].

A. Derivation of a capacity expression for the 2-state signal transduction channel

We start with the expression for the discrete time mutual information rate (74). Assuming the input distribution is IID, the mutual information per discrete time step is given by

\[ I(X;Y) = \mathcal{H}(\alpha_H p_H + \alpha_L p_L) - p_H \mathcal{H}(\alpha_H) - p_L \mathcal{H}(\alpha_L) \]

\[ + \frac{1 + (p_H \alpha_H + p_L \alpha_L)/\beta}{1 + (p_H \alpha_H + p_L \alpha_L)/\beta}, \]

(136)

where

\[ \mathcal{H}(p) = -p \log p - (1 - p) \log(1 - p). \]

(137)

The IID capacity is obtained by maximizing over the set \( p_H \in [0, 1] \) with \( p_L = 1 - p_H \). For convenience, we use \( x \) to represent \( p_H \) in the rest of this section.

The model for the channel assumes that the probability of transition per time step is \( \alpha_H, \alpha_L, \) or \( \beta \), depending on the state of the input and the state of the channel. The IID input approximation assumes the input can flicker back and forth arbitrarily fast, so that successive time steps are uncorrelated. For the following calculation, we will assume that the input can remain IID even as the time step goes to zero. To represent discretization with finer and finer time steps, we set

\[ \alpha_H^* = \epsilon \alpha_H \]

\[ \alpha_L^* = \epsilon \alpha_L \]

\[ \beta^* = \epsilon \beta \]
where $\epsilon > 0$ is the size of the time step. The case $\epsilon = 1$ corresponds to the discrete time model considered up to this point. The fixed constants $\alpha_{H/L}$ and $\beta$ now represent transition rates per unit time, rather than probabilities per time step.

Figure 8 shows the per-time-step mutual information, as a function of $0 \leq x \leq 1$, for $\alpha_{H}^{*}L = 0.1\epsilon$, $\alpha_{H}^{*} = 0.9\epsilon$, and $\beta^{*} = 0.5\epsilon$, for $\epsilon$ ranging from 1 to $10^{-4}$. The curves suggest that, as expected, the optimal value of $x$ lies in the interior of the interval $(0, 1)$. Moreover, the optimal value appears to converge to a given value $x_{opt}$ as $\epsilon \to 0$.

We define the mutual information rate for a given $\epsilon > 0$, $I_\epsilon$, to be $I/\epsilon$, and we study how this quantity scales as $\epsilon \to 0^+$. In the remainder of this section we do the following:

1) We study the information rate $I_\epsilon(x)$ for small $\epsilon$ and show that it has a unique maximum in the interior of the interval $0 \leq x \leq 1$, where $x = p_{H}$ is the probability that the input signal is in the “high” concentration state.

2) By taking the limit as $\epsilon \to 0^+$, and optimizing over $x$, we obtain an expression for the capacity of the continuous time channel, as a function of the binding and unbinding rates $\alpha_{H}$, $\alpha_{L}$, and $\beta$.

In the following section §IV-B, we further show that by taking the limit of the capacity for the continuous time channel, as the unbinding rate $\beta \to \infty$, we recover Kabanov’s expression...
for the capacity for the Poisson channel.

1) Critical point of the information rate $I$, for small $\epsilon > 0$: First we study the behavior of the optimal value of $x$ in the limit of small $\epsilon$. Assuming an interior maximum for $I$, we set the derivative of the RHS of Equation (75) equal to zero to obtain the necessary and sufficient condition

\[
0 = -\alpha_H \beta \epsilon \log \left( \frac{1}{-\alpha_H \epsilon x + \alpha_L (x-1)\epsilon + 1} \right) + \alpha_L \beta \epsilon \log \left( \frac{1}{-\alpha_H \epsilon x + \alpha_L (x-1)\epsilon + 1} \right) + \beta \epsilon (\alpha_H - \alpha_L) \log \left( \frac{1}{\alpha_H \epsilon x - \alpha_L \epsilon + \alpha_L \epsilon} \right) - \alpha_H \epsilon (\alpha_L + \beta) \log \left( \frac{1}{\alpha_H \epsilon} \right) + \alpha_L \epsilon (\alpha_H + \beta) \log \left( \frac{1}{\alpha_L \epsilon} \right) - \alpha_L \log \left( \frac{1}{-\alpha_H \epsilon + \alpha_L (x-1)\epsilon + 1} \right) + \alpha_H \alpha_L \epsilon \log \left( \frac{1}{1 - \alpha_H \epsilon} \right) + \alpha_H \beta \epsilon \log \left( \frac{1}{\alpha_H \epsilon} \right) + \beta \log \left( \frac{1}{\alpha_L \epsilon} \right) - \alpha_L \beta \epsilon \log \left( \frac{1}{\alpha_L \epsilon} \right).
\]

In Appendix D we show that this condition leads to an interior maximum at a unique value of $x$ as $\epsilon \to 0$.

2) Implicit expression for $x_{\text{opt}}$ in the limit $\epsilon \to 0$, and an expression for the capacity of the continuous time channel: Define the continuous time information rate, as a function of the fraction of time the input is in the higher state ($x = p_H$), as

\[
I = \lim_{\epsilon \to 0^+} \left( \frac{1}{\epsilon} \right) I = \lim_{\epsilon \to 0^+} \left( \frac{1}{\epsilon} \mathcal{H}(\epsilon \alpha_H x + \epsilon \alpha_L (1-x)) - x \mathcal{H}((\epsilon \alpha_H) - (1-x) \mathcal{H}(\epsilon \alpha_L)) \right).
\]

From the preceding section, we know this expression converges to a finite value, and moreover $I$ has a unique maximum in the range $0 \leq x \leq 1$. Let $x_{\text{opt}}$ denote the optimal value of the high-input probability. It is straightforward to show that

\[
\mathcal{I}(x) = -\left( \frac{\beta}{\beta + \bar{\alpha}(x)} \right) (\bar{\alpha} \log(\bar{\alpha}) - (x \alpha_H \log \alpha_H + (1-x) \alpha_L \log \alpha_L))
\]

where, as above, $\bar{\alpha}(x) = x \alpha_H + (1-x) \alpha_L$ is the average value of $\alpha$ given $x$. Thus the mutual information rate is given by the product of the fraction of time the channel is in the receptive state,

\[
f(x) \equiv \left( \frac{\beta}{\beta + \bar{\alpha}(x)} \right)
\]
and the mutual information rate conditional on the channel being in the receptive state,
\[ g(x) \equiv - (\bar{\alpha}(x) \log \bar{\alpha}(x) - x \alpha_H \log \alpha_H + (1 - x) \alpha_L \log \alpha_L)). \]

Although the optimal value of the high input probability \( x \) is not available explicitly, we can obtain a useful implicit expression for the capacity of the continuous time channel. Setting \( \mathcal{I}' = f'g + fg' = 0 \), and noting that \( f' < 0 \), we have \( g(x_{opt}) = f(x_{opt})g'(x_{opt})/f'(x_{opt}) \), from which
\[
\mathcal{I}(x_{opt}) = \frac{g'(x_{opt})}{f'(x_{opt})} f(x_{opt})^2 \quad (140)
\]

\[
= \left( \frac{\beta}{\alpha_H - \alpha_L} \right) (\alpha_H \log \alpha_H - \alpha_L \log \alpha_L - (\alpha_H - \alpha_L)(1 + \log \bar{\alpha}(x_{opt}))). \quad (141)
\]

In the continuous time setting, we have an ambiguity associated with the choice of the time unit. Provided the low binding rate is not identically zero, we can choose time units with respect to which the low binding rate \( \alpha_L \equiv 1 \). Let \( \alpha_H = 1 + c \) in the same units. Thus, from (141), the capacity is given by
\[
\mathcal{I}(x_{opt}) = \beta \left( \frac{1 + c}{c} \log(1 + c) - 1 - \log(1 + x_{opt}c) \right). \quad (142)
\]

Although \( x_{opt} \) is not known explicitly, it lies in the interior of the unit interval, so we have the upper and lower bounds
\[
\beta \left( \frac{1 + c}{c} \log(1 + c) - 1 - \log(1 + c) \right) \leq \mathcal{I}(x_{opt}) \leq \beta \left( \frac{1 + c}{c} \log(1 + c) - 1 \right). \quad (143)
\]

Kabanov obtained the capacity of the Poisson channel with signal intensity bounded above by a constant \( c \), and unit background intensity. The “high” and “low” rates of the incoming process (combining signal and noise) were \( 1 + c \) and 1, respectively. In the limit as the unbinding rate grows without bound, we expect that our channel should be equivalent to Kabanov’s Poisson channel. Note that because the optimal sending input distribution depends on the channel parameters, including \( \beta \), the expression (142) does not necessarily diverge as \( \beta \to \infty \). In the next section we recover Kabanov’s capacity formula in this limit.

**B. Reduction of the 2-State Signal-Transduction Channel to Kabanov's Poisson Channel**

It is of interest to consider the case of a “perfect” receiver that has an unbinding rate much faster than any other rate in the system. As a thought experiment, let us suppose that with a
discrete time step $\epsilon$, we may set the unbinding rate $\beta = 1/\epsilon$, so that the unbinding probability is $\beta^* = \epsilon \beta = 1$. In this case, the continuous time channel capacity is still bounded, provided the low sending rate $\alpha_L > 0$. Scaling $\beta$ in this way gives

$$I_{\epsilon}(x) = \frac{1}{\epsilon} \left( \frac{1}{1 + \epsilon \bar{\alpha}(x)} \right) (\mathcal{H}(\epsilon \bar{\alpha}) - x \mathcal{H}(\epsilon \alpha_H) - (1 - x) \mathcal{H}(\epsilon \alpha_L))$$

(144)

$$\lim_{\epsilon \to 0} I_{\epsilon}(x) \equiv I_0(x) = x \alpha_H \log \alpha_H + (1 - x) \alpha_L \log \alpha_L - \bar{\alpha} \log \bar{\alpha}$$

(145)

$$\frac{d}{dx} I_0(x) = \alpha_H \log \alpha_H - \alpha_L \log \alpha_L - (\alpha_H - \alpha_L)(1 + \log \bar{\alpha}(x))$$

(146)

$$= 0 \text{ when } \log \bar{\alpha} = \left( \frac{1}{\alpha_H - \alpha_L} \right) \left( \alpha_H \log \left( \frac{\alpha_H}{e} \right) - \alpha_L \log \left( \frac{\alpha_L}{e} \right) \right).$$

(147)

Recall that $x_{opt}$ denotes the optimal probability of the high-concentration signal. Since $\bar{\alpha} = \alpha_L + x_{opt}(\alpha_H - \alpha_L)$, we have

$$x_{opt} = \frac{\bar{\alpha} - \alpha_L}{\alpha_H - \alpha_L} \quad \text{and} \quad (1 - x_{opt}) = \frac{\alpha_H - \bar{\alpha}}{\alpha_H - \alpha_L}.$$

Consequently the capacity, $I_0^* = I_0(x_{opt})$, reduces to

$$I_0^* = \bar{\alpha}(x_{opt}) - \left( \frac{\alpha_H \alpha_L}{\alpha_H - \alpha_L} \right) \log \left( \frac{\alpha_H}{\alpha_L} \right)$$

(148)

$$= \exp \left[ \left( \frac{1}{\alpha_H - \alpha_L} \right) \left( \alpha_H \log \left( \frac{\alpha_H}{e} \right) - \alpha_L \log \left( \frac{\alpha_L}{e} \right) \right) \right] - \left( \frac{\alpha_H \alpha_L}{\alpha_H - \alpha_L} \right) \log \left( \frac{\alpha_H}{\alpha_L} \right)$$

(149)

Consider an optical detection channel for which the source of the signal is a Poisson process of modulated intensity with range from zero to an upper limit $c > 0$. Suppose the noise in the channel comes from a background Poisson source with constant rate $\lambda > 0$. The peak of the detection rate (false positive + true positive) is then $\lambda + c$. If the detector has a dead time following each detection that is exponentially distributed with parameter $\beta$, then the corresponding channel is formally identical to our ligand binding/unbinding channel with binding rates $\alpha_L = \lambda$ (low rate), $\alpha_H = \lambda + c$ (high rate), and unbinding rate $\beta$. Taking the $\beta \to \infty$ limit as above, we can
rewrite the capacity as

$$\mathcal{I}^*_0(c, \lambda) = \lambda + x_{\text{opt}} c - \left(\frac{\lambda(\lambda + c)}{c}\right) \log \left(\frac{\lambda + c}{\lambda}\right)$$  \hspace{1cm} (150)

$$= \exp \left[\frac{\lambda + c}{c} \log \left(\frac{\lambda + c}{e}\right) - \frac{\lambda}{c} \log \left(\frac{\lambda}{e}\right)\right] - \frac{\lambda}{c} (\lambda + c) \log \left(\frac{\lambda + c}{\lambda}\right)$$  \hspace{1cm} (151)

$$= \exp \left[\left(1 + \frac{\lambda}{c}\right) \log (\lambda + c) - \frac{\lambda}{c} \log \lambda - 1\right] - \lambda \left(1 + \frac{\lambda}{c}\right) \log \left(1 + \frac{c}{\lambda}\right)$$  \hspace{1cm} (152)

$$= \exp \left[\log \left(\frac{(\lambda + c)^{(1+\lambda/c)}}{\lambda^{(\lambda/c)} e}\right)\right] - \lambda \left(1 + \frac{\lambda}{c}\right) \log \left(1 + \frac{c}{\lambda}\right)$$  \hspace{1cm} (153)

$$= \lambda \left[\frac{1}{e} \left(1 + \frac{c}{\lambda}\right)^{(1+\lambda/c)} - \left(1 + \frac{\lambda}{c}\right) \log \left(1 + \frac{c}{\lambda}\right)\right].$$  \hspace{1cm} (154)

The capacity for the Poisson channel with background noise at rate $\lambda = 1$ and signal bounded by intensity $c$ was first obtained by Kabanov [52], and provides the foundation for information theoretic analysis of channels with Poisson signals (reviewed in [58]). Setting $\lambda = 1$ in (154), which is equivalent to choosing new units for time, immediately yields

$$C_{\text{Kabanov}}(c) = \frac{1}{e} \left((c + 1)^{1+c^{-1}} - \left(1 + \frac{1}{c}\right) \log (c + 1)\right).$$  \hspace{1cm} (155)

The analysis of the Kabanov/Poisson channel has been elaborated in numerous ways. In [51], Davis gives the following formula for the capacity of the Poisson channel with noise rate $\lambda$ and signal rate bounded by $c$, namely

$$C_{\text{Davis}}(c, \lambda) = \frac{1}{e} (\lambda + c) \left(1 + \frac{c}{\lambda}\right)^{\lambda/c} - \lambda \left(1 + \frac{\lambda}{c}\right) \log \left(1 + \frac{c}{\lambda}\right)$$

which is identical to (154), see Equation (4b) in [51] and also Equation (5) in [59].

We emphasize that Kabanov did much more than derive the formula. He proved in [52] that (155) is the capacity for the Poisson channel and also that the capacity cannot be increased via feedback. While our rigorous proofs are restricted to the discrete time case of the ligand binding/unbinding channel, the consistency of the limiting continuous time expressions with Kabanov’s formula suggests that the analogy is sound.
V. Discussion

In this paper, we calculated the capacity of a simple signal transduction channel, related to the cAMP receptor in Dictyostelium, and derived many useful properties of mutual information. Our contribution is one of a rapidly growing body of work applying information theory to biological communication problems. Indeed, a natural open problem suggested by our work is to extend Kabanov’s continuous time Poisson channel to a family of channels defined by continuous time Markov chains on finite graphs. Here we consider some features of this generalized problem.

One may consider the input signal to a generalized continuous-time Markov channel as any physical or biochemical process that varies the transition rate intensities between the nodes of the graph, with the output signal comprising either the transitions themselves or a related counting process on one or more of the nodes of the graph. Viewed in this way, the Kabanov-Poisson channel comprises a “graph” with a single node, with a single counting process instead of a multicomponent marked point process.

Analysis of the capacity for a general $n$-state signal-transduction channel, such as described by (1), remains an interesting open problem. In this paper, we considered the case $n = 2$, in a sense the simplest generalization of the Poisson channel. For our two-state signal-transduction channel, the mutual information rate in both the discrete time setting (74) and in the continuous time setting (139) decompose into the product of an information rate conditional on occupying a “sensitive” state, and the fraction of time the system occupies that state. However, as we already stated in the introduction, many higher-order Markov models are available for different kinds of receptors, so the generalized problem is of significant practical interest.

First, a simple extension of our results in Section III gives the mutual information of a general $n$-state receptor under IID inputs. For receptor states $i$ and $j$, $1 \leq i, j \leq n$, and input concentration $x$, taking discrete levels in $1 \leq x \leq m$, let $\alpha_{i,j,x}$ represent the transition probability from state $i$ to state $j$ under input concentration $x$. Let $p$ represent the $m$-dimensional vector containing the IID input distribution. Let $\bar{\alpha}_{i,j} = \sum_{x=1}^{m} \alpha_{i,j,x} p_x$ represent the average transition probability from $i$ to $j$. Under an IID input distribution, the sequence of receptor states $Y$ forms a regular Markov chain with transition probability matrix $P_Y = [\bar{\alpha}_{i,j}]$. If $p_Y$ is the stationary distribution on the receptor states, given by the normalized eigenvector of $P_Y$ with eigenvalue 1, and recalling $\phi(\cdot)$
from (116), then the mutual information under IID inputs is given by

$$I(X; Y) = \sum_{i=1}^{n} p_{Y,i} \sum_{j=1}^{n} \left( \phi(\bar{\alpha}_{i,j}) - \sum_{x=1}^{m} p_{x} \phi(\alpha_{i,j,x}) \right).$$ \hfill (156)

Given the above, and considering our other results on this paper, we may state a conjecture on the capacity of the corresponding continuous-time system. Let $K(a, b)$ be the Kabanov capacity of a Poisson channel with minimum intensity $a > 0$ and maximum intensity $b > a$. Let $Y(t)$ be the state of a continuous time channel with transition matrix rates $Q(x)$ depending monotonically on an input $a \leq x \leq b$. Let directed edge $k$ connect state $i(k)$ to state $j(k)$; denote the $k^{th}$ edge by $ij(k)$. Assume the transition rates have the property that all states intercommunicate for all values of $x$ in the allowed interval. That is, for $a \leq x \leq b$ the transition matrix $Q(x)$ satisfies the conditions of the Perron-Frobenius theorem. Let $\alpha_k(x)$ be the transition rate along edge $ij(k)$ as a continuous function of $x$, and let $\alpha_k^-$ and $\alpha_k^+$ be the extreme values of $\alpha_k$ over the interval $[a, b]$. Moreover we require $\alpha_k^- > 0$ for all $k$, and that all transitions be observable. If $X$ is a stationary stochastic process satisfying the constraint, then there is a corresponding joint process $(X, Y)$ that is also stationary. Let $\pi(x, y)$ be the steady state probability distribution for this process, with marginal distributions $\pi_x$ and $\pi_y$. Then we may conjecture that the information rate for the $X \rightarrow Y$ channel should given by

$$\sum_{k} \pi_y(i(k)) K(\alpha_k^-, \alpha_k^+),$$ \hfill (157)

and that this rate may be maximized by a stationary IID distribution on the input process $X$.

An analogous result has been shown for the capacity given full channel state information (equivalent to feedback in our formulation), for a channel model representing communication along a bacterial nanowire [60], [61]. In this model, the states of the channel comprise a graph in the form of a finite chain.

However, it is clear that for a Markov channel taking the form of an arbitrary network, it is not generally true that $C_{\text{IID}} = C_{\text{FB}}$, as the following example illustrates.

Consider a channel with three states arranged in a chain

$$
\begin{array}{c c c c}
(\alpha_L \text{ or } \alpha_H) & (\epsilon \text{ or } 1 - \beta) \\
1 & \rightleftharpoons & 2 & \rightleftharpoons & 3, \\
\beta & & \epsilon
\end{array}
$$ \hfill (158)
where $0 < \alpha_L < \alpha_H < 1$, $0 < \beta < 1$, and $0 < \epsilon < 1 - \beta$. The $1 \rightarrow 2$ and $2 \rightarrow 3$ transition probabilities depend on the input (assumed binary for this example) in the same manner as in §III. That is, we have $\alpha_{1,2,L} = \alpha_L$ and $\alpha_{1,2,H} = \alpha_H$, and $\alpha_{2,3,L} = \epsilon$ while $\alpha_{2,3,H} = 1 - \beta$. The other transitions are insensitive to the input value $x$, i.e. $\alpha_{2,1,x} = \beta$ and $\alpha_{3,2,x} = \epsilon$ independently of $x$. Hence the transitions out of state 3 do not carry information about the input. Given the input probabilities $p_H + p_L = 1$, the transition matrix of the channel state for IID input is

$$T_{\text{IID}} = \begin{pmatrix}
1 - \bar{\alpha} & \bar{\alpha} & 0 \\
\beta & 1 - \beta - (p_L \epsilon + p_H (1 - \beta)) & p_L \epsilon + p_H (1 - \beta) \\
0 & \epsilon & 1 - \epsilon
\end{pmatrix},$$

(159)

where $\bar{\alpha} = p_L \alpha_L + p_H \alpha_H$, and the stationary distribution is

$$p_{Y,1} = \frac{\beta}{Z_{\text{IID}}}, p_{Y,2} = \frac{\bar{\alpha}}{Z_{\text{IID}}}, p_{Y,3} = \frac{p_H (1 - \beta) \bar{\alpha}}{\epsilon Z_{\text{IID}}}, Z_{\text{IID}} = \bar{\alpha} + \beta + \frac{p_H (1 - \beta) \bar{\alpha}}{\epsilon}$$

(160)

If we denote the mutual information for the two-state channel with IID inputs by $I_2$, then the mutual information for the three-state channel with IID inputs is

$$I_{\text{IID}}^3 = p_{Y,1} I_2 + p_{Y,2} H \frac{\epsilon Z_{\text{IID}}}{\alpha}$$

(161)

The mutual information for a given input distribution is reduced, compared to that of the two-state channel, because the channel gets trapped in the long-lived, insensitive state 3, thus reducing the fraction of time spent in the sensitive state 1.

In case the sender is informed of the state of the channel, the sender may arrange to send input $x = L$ whenever the channel is in state 2, thus reducing the rate at which the channel enters the trap in state 3. In case the sender adopts this strategy, the transition matrix for the channel state becomes

$$T_{\text{FB}} = \begin{pmatrix}
1 - \bar{\alpha} & \bar{\alpha} & 0 \\
\beta & 1 - \beta - \epsilon & \epsilon \\
0 & \epsilon & 1 - \epsilon
\end{pmatrix},$$

(162)

and the stationary distribution is

$$p_{Y,1} = \frac{\beta}{Z_{\text{FB}}}, p_{Y,2} = \frac{\bar{\alpha}}{Z_{\text{FB}}}, p_{Y,3} = \frac{\bar{\alpha}}{Z_{\text{FB}}}, Z_{\text{FB}} = 2\bar{\alpha} + \beta.$$  

(163)

The capacity under this feedback scheme is

$$I_{\text{FB}}^3 = \frac{\beta I_2 + \bar{\alpha} H p_H}{Z_{\text{FB}}}.$$  

(164)
To compare the mutual information for any choice of input probabilities \(p_{L/H}\) and parameters \(\alpha_{L/H}, \beta, \epsilon\), consider the ratio of the mutual information under the IID inputs versus the feedback scheme:

\[
\frac{I^\text{IID}}{I^\text{FB}} = \frac{Z_{\text{FB}}}{Z_{\text{IID}}} = \frac{\epsilon(2\bar{\alpha} + \beta)}{\epsilon(\bar{\alpha} + \beta) + p_H(1 - \beta)\bar{\alpha}} \to 0, \quad \text{as } \epsilon \to 0^+.
\]  

That is, the ratio of the mutual information under the IID inputs versus inputs informed by the channel state can be made arbitrarily small, by taking the slow transition rates \(\epsilon\) sufficiently small. This suggests that the feedback capacity and the IID capacity cannot be equal for this simple example.

The question of the regular capacity for this channel, and channels with arbitrary state graphs, remains an interesting problem for future work.
A. An asymptotic derivation of the capacity for Kabanov’s Poisson channel in the case without feedback

As shown in §IV-B, the capacity of our two-state binding/unbinding channel in continuous time reduces to the capacity of the Poisson channel in the limit of rapid unbinding. Here we review Kabanov’s formula for the Poisson channel and show that it can be derived from a discrete time framework analogous to our discrete time channel model. See also [53].

1) Kabanov’s Capacity Formula: In [52], Kabanov derived an exact expression for the information capacity of a communications channel based on a variable rate Poisson process. The channel is a combination of two point processes. The first represents a constant nonzero background noise rate, represented as a uniform Poisson process with intensity \( \lambda \). The second Poisson process encodes the input signal through a time varying intensity \( \mu_t \geq 0 \), subject to a maximum intensity constraint \( \mu_t \leq c \). Exploiting Girsanov’s change-of-measure theorem in the point process setting, the convexity of the function \( x \ln x \), and Jensen’s inequality, Kabanov obtained a closed form expression for the capacity. As shown by Davis [51], this capacity can be written

\[
C_{Kab}(c, \lambda) = \lambda \left[ \frac{1}{e} \left( 1 + \frac{c}{\lambda} \right)^{1+\lambda/c} - \left( 1 + \frac{\lambda}{c} \right) \log \left( 1 + \frac{c}{\lambda} \right) \right].
\]

Moreover, Kabanov shows that allowing feedback from the output of the channel to influence the input signal cannot increase the capacity. In this section we consider only the channel without feedback.

2) Informal Derivation of Kabanov’s Formula: Here we derive Kabanov’s formula using elementary methods, to facilitate comparison with the binding/unbinding channel.

Let the input signal \( \mu_t \) be piecewise constant on intervals of length \( h \), taking on only the extreme values \( \mu_t \in \{0, c\} \). Hence the combined Poisson process fluctuates in intensity between \( \lambda \) and \( \lambda + c \). Furthermore, let the high (respectively, low) value of the intensity be chosen independently with probability \( p \) (respectively, \( 1 - p \)) at each time step. The probability of receiving \( n \) counts in an interval of length \( h \) is depends on whether the input is “high” (H) or
“low” (L). Specifically,

\[ P(n|H) = e^{-(c+\lambda)h((c + \lambda)h)^n / n!} \] (167)

\[ P(n|L) = e^{-\lambda h ((\lambda h)^n / n!}. \] (168)

Before a particular realization of the input signal has been determined, the \textit{a priori} probability of receiving \( n \) counts in any given interval is

\[ P(n) = \frac{pe^{-(c+\lambda)h((c + \lambda)h)^n + (1 - p)e^{-\lambda h (\lambda h)^n}}}{n!}. \] (169)

The entropy of the count during a single timestep, with the input signal treated as a parameter, is

\[ S_H = -\sum_{n=0}^{\infty} P(n|H) \log [P(n|H)] \] (170)

\[ S_L = -\sum_{n=0}^{\infty} P(n|L) \log [P(n|L)] . \] (171)

Because the input is presumed independent on successive time steps, the conditional entropy is \( pS_H + (1 - p)S_L \). Before the stimulus has been fixed, the \textit{a priori} entropy is determined by the prior distribution \( P(n) \). The mutual information, \( S_{\text{prior}} - S_{\text{post}} \), follows from

\[ S_{\text{prior}} = -\sum_{n=0}^{\infty} P(n) \log [P(n)] \] (172)

\[ S_{\text{post}} = pS_H + (1 - p)S_L \] (173)

\[ I = S_{\text{prior}} - S_{\text{post}}. \] (174)

The information rate, \( I/h \) provides a lower bound on the information capacity of the channel per unit time, \( C \).

The sums defining \( S_{\text{prior,post}} \) are intractable. However, numerical experiments (cf. [18]) suggest that optimal information transmission rates can occur in the limit of frequent noisy measurements (short time steps) rather than longer more accurate measurements. In the limit of short time steps, the probability of obtaining more than one count in interval \( h \) is \( o(h) \), as \( h \to 0 \). Consequently we consider only the information contained in those intervals in which either zero or one counts are received, and ignore those rare intervals in which \( n \geq 2 \) counts arrive. Since
this approximation only discards information in the signal, it should not compromise the lower bound on $C$. Truncating the sums for $S_H$, $S_L$ and $S_{\text{prior}}$ at $n = 1$ gives a tight lower bound $\hat{I}/h$:

$$\frac{\hat{I}}{h} = (1 - p)\lambda \log[\lambda] + p(c + \lambda) \log[c + \lambda] - (cp + \lambda) \log[cp + \lambda] + O(h), \quad \text{as } h \to 0. \quad (175)$$

Neglecting the $O(h)$ terms, we maximize $\hat{I}/h$ for given values of $c$ and $\lambda$. The critical value of $p$ is given by

$$p^*(c, \lambda) = \frac{1}{c} \exp\left[\left(1 + \frac{\lambda}{c}\right) \log[c + \lambda] - \left(1 + \frac{\lambda}{c} \log[\lambda]\right)\right] - \frac{\lambda}{c}. \quad (176)$$

Inserting $p^*$ into (175) yields an expression that after algebraic manipulation reduces exactly to (166). Hence in the limit of short time intervals an encoding scheme consisting of random selection of “high” and “low” sending rates with probabilities given by (176) achieves the Kabanov capacity for the Poisson channel.

**B. Stationary distributions achieve feedback capacity**

Here we use the same indexing of the input/output process $(X^n, Y^n)$ as in Section II.

We start with several definitions. Assuming that the input distribution is in $\mathcal{P}^*$ (i.e., $Y^n$ is a Markov chain), and recalling (7), let $\hat{\mathbf{P}} = [\hat{P}_{ij}]$ represent a $2 \times 2$ matrix, taking values in $\{0, 1\}$, with elements

$$\hat{P}_{ij} = \begin{cases} 1, & \min_{k \in \{1, 2, \ldots, m\}} P_{Y|X=k,ij} > 0 \\ 0, & \text{otherwise}, \end{cases} \quad (177)$$

and for positive integers $\ell$, let $\hat{P}_{ij}^{\ell}$ represent the $i, j$th element of $\hat{\mathbf{P}}^{\ell}$. Further, for the $i$th diagonal element of the $\ell$th matrix power $\hat{P}_{ii}^{\ell}$, let $\mathcal{D}_i$ contain the set of integers $\ell$ such that $\hat{P}_{ii}^{\ell} \neq 0$. Then:

- $Y^n$ is strongly irreducible if, for each pair $i, j$, there exists an integer $h > 0$ such that $\hat{P}_{ij}^{h} \neq 0$; and
- If $Y^n$ is strongly irreducible, it is also strongly aperiodic if, for all $i$, the greatest common divisor of $\mathcal{D}_i$ is 1.

These conditions are described in terms of graphs in [55], but our description is equivalent. Let

$$I(p, Q_i) = I(X_j; Y_j \mid Y_{j-1} = i) \quad (178)$$

If $i = U$, we have

$$I(p, Q_U) = \mathcal{H}(\bar{\alpha}_p) - \sum_{i=1}^{m} p_i \mathcal{H}(\alpha_i). \quad (179)$$
We will use the following corollary to Theorem 1.

**Corollary 2**: Let \( p' = [p'_1, p'_2, \ldots, p'_m] \) represent the distribution satisfying
\[
p' = \arg \max_p I(p, Q_U), \tag{180}
\]
using \( I(p, Q_U) \) from (178). Then \( p'_2 = p'_3 = \ldots = p'_{m-1} = 0 \).

**Proof**: The quantity in (179) is equal to the numerator of (39). To prove the theorem, we relied only on terms in the numerator, so the same argument applies to this corollary. \( \blacksquare \)

**Lemma 3**: If the parameters are strictly ordered (Definition 1), then the conditions of [55, Thm. 4] are satisfied, namely:

1) \( Y^n \) is strongly irreducible and strongly aperiodic.

2) For \( j \in \{U, B\} \), let \( Q_j \) be a \( 2 \times m \) matrix, defined as
\[
Q_j = \begin{bmatrix}
p_{Y|X,Y_{t-1}}(B|c_1,j) & p_{Y|X,Y_{t-1}}(B|c_2,j) & \ldots & p_{Y|X,Y_{t-1}}(B|c_m,j) \\
p_{Y|X,Y_{t-1}}(U|c_1,j) & p_{Y|X,Y_{t-1}}(U|c_2,j) & \ldots & p_{Y|X,Y_{t-1}}(U|c_m,j)
\end{bmatrix}. \tag{181}
\]
Then (reiterating [55, Defn. 6]) for the set of possible input distributions in \( \mathcal{P}^* \), and for all \( j \in \{U, B\} \), there exists a subset \( \tilde{\mathcal{P}}^* \) satisfying

a) \( \{Q_jp : p \in \mathcal{P}^*\} = \{Q_jp : p \in \tilde{\mathcal{P}}^*\} \).

b) For any \( q \in \{Q_jp : p \in \mathcal{P}^*\} \),
\[
\left\{ \max_{p \in \mathcal{P}^*, Q_jp = q} I(p, Q_j) \right\} \cap \left\{ \max_{p \in \mathcal{P}^*, Q_jp = q} I(p, Q_j) \right\} \neq \emptyset \tag{182}
\]
c) There exists a positive constant \( \lambda \) such that
\[
\frac{\partial I(p, Q_j)}{\partial \ell} - \frac{\partial I(q, Q_j)}{\partial \ell} \leq -\lambda ||p - q|| \tag{183}
\]
for any nonidentical \( p, q \in \tilde{\mathcal{P}}^* \), where \( \ell \) is in the direction from \( q \) to \( p \), and the norm is the Euclidean vector norm.

**Proof**: To prove the first part of the lemma, if the parameters are strictly ordered, then \( \hat{\mathcal{P}} \) is an all-one matrix, so \( Y^n \) is strongly irreducible (with \( h = 1 \)); further, since the positive powers of an all-one matrix can never have zero elements, \( D_i \) contains all positive integers from 1 to \( n \), whose greatest common divisor is 1, so \( Y^n \) is strongly aperiodic.
To prove the second part of the lemma, we first show that the definition is satisfied for $Q_B$, given by

$$Q_B = \begin{bmatrix} 1 - \beta & 1 - \beta & \ldots & 1 - \beta \\ \beta & \beta & \ldots & \beta \end{bmatrix}.$$  \hfill (184)

We choose the subset $\tilde{P}^*$ to consist of a single point $p \in \mathcal{P}^*$ (it can be any point, as all points give the same result). The columns of $Q_B$ are identical, since the output is not dependent on the input in state $B$. Then for every $p \in \mathcal{P}^*$,

$$Q_B p = \begin{bmatrix} 1 - \beta & \ldots & 1 - \beta \\ \beta & \ldots & \beta \end{bmatrix} \begin{bmatrix} p_1|B \\ p_2|B \\ \vdots \\ p_m|B \end{bmatrix} = \begin{bmatrix} (1 - \beta) \sum_{j=1}^{m} p_j|B \\ \beta \sum_{j=1}^{m} p_j|B \end{bmatrix} = \begin{bmatrix} (1 - \beta) \bar{\alpha} p \\ \bar{\alpha} p \end{bmatrix}.$$  \hfill (185)

This is also true of the single point in $\tilde{P}^*$, so condition (a) is satisfied. Similarly, by inspection of (184), when $Y_0 = B$, the output $Y_1$ is not dependent on the input $X_1$, so $I(p, Q_B) = 0$ for all $p \in \mathcal{P}$. Since all $p \in \mathcal{P}^*$ “maximize” $I(p, Q_B)$ and have identical values of $Q_p$ (including the single point in $\tilde{P}^*$), then the single point $p \in \tilde{P}^*$ is always in both sets, and the intersection (182) is nonempty; so condition (b) is satisfied. There is only one point in $\tilde{P}^*$, so there is no pair of nonidentical points, and condition (c) is satisfied trivially.

Now we show that the conditions are satisfied for $Q_U$, given by

$$Q_U = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_m \\ 1 - \alpha_1 & 1 - \alpha_2 & \ldots & 1 - \alpha_m \end{bmatrix}. \hfill (188)$$

Since the parameters are strictly ordered, $\text{rank}(Q_U) = 2$. (The lemma is satisfied if $\text{rank}(Q_U) = 1$, by the same argument we gave above, though in this case $\alpha_1 = \ldots = \alpha_m$ and the capacity is zero.) Now we have

$$Q_U p = \begin{bmatrix} \sum_{j=1}^{m} p_j \alpha_i \\ 1 - \sum_{j=1}^{m} p_j \alpha_i \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_p \\ 1 - \bar{\alpha}_p \end{bmatrix}. \hfill (189)$$
Since the parameters are strictly ordered, $\bar{\alpha}_p$ can take any value on the interval $[\alpha_1, \alpha_m]$.

Let $\tilde{P}^*$ represent the set of input distributions $p$ from Corollary 2 with $p_2 = p_3 = \ldots = p_{m-1} = 0$. For $p \in \tilde{P}^*$,
\[
Q_U p = \begin{bmatrix}
p_1 \alpha_1 + p_m \alpha_m \\
1 - p_1 \alpha_1 - p_m \alpha_m
\end{bmatrix},
\]
and $p_1 \alpha_1 + p_m \alpha_m$ can take any value on the interval $[\alpha_1, \alpha_m]$. Therefore, condition (a) is satisfied.

From Corollary 2 all distributions $p$ maximizing $I(p, Q_i)$ have $p_2 = p_3 = \ldots = p_{m-1} = 0$. Thus, all maximizing distributions in $P^*$ are also in $\tilde{P}^*$, and condition (b) is satisfied.

Finally, by the definition of the directional derivative, condition (c) is equivalent to
\[
(p - q) \cdot \left( \nabla_p I(p, Q_U) - \nabla_q I(q, Q_U) \right) < 0,
\]
where $\cdot$ represents vector dot product. Inequality (191) reduces to
\[
\frac{1}{\log_e 2} (\bar{\alpha}_p - \bar{\alpha}_q) \log \frac{\bar{\alpha}_q - \bar{\alpha}_p \bar{\alpha}_q}{\bar{\alpha}_p - \bar{\alpha}_p \bar{\alpha}_q} < 0
\]
By inspection, this inequality is satisfied as long as $\bar{\alpha}_p \neq \bar{\alpha}_q$. To check when this is satisfied in the subset $\tilde{P}^*$, we can write
\[
\bar{\alpha}_q - \bar{\alpha}_p = \alpha_1 q_1 + \alpha_m q_m - \alpha_1 p_1 - \alpha_m p_m
= \alpha_1 q_1 + \alpha_m (1 - q_1) - \alpha_1 p_1 - \alpha_m (1 - p_1)
= (q_1 - p_1)(\alpha_1 - \alpha_m).
\]
where (193) follows from the definition of $\tilde{P}^*$. By assumption, $\alpha_m \neq \alpha_1$. Thus, $\bar{\alpha}_q \neq \bar{\alpha}_p$ so long as $q_1 \neq p_1$, i.e., for any distinct points in $\tilde{P}^*$. Thus, condition (c) is satisfied, and the lemma follows.

Closely related results were given in the (unfortunately unpublished) [62], as well as stronger results for all possible binary-input, binary-output, unit-memory Markov channels.

C. Entropy Rates via the Sum-Product Algorithm

In this appendix, we briefly explain how to use the sum-product algorithm [63], both to calculate bounds on mutual information and to perform the Monte Carlo simulations that were discussed in Section III. (This appendix uses the same indexing of the input/output process as in Section III.)
The channel is specified by the conditional probabilities $p_{Y_{i+1}|X_i,Y_i}(y_{i+1}|x_i, y_i)$, with a Markov input process governed by transition probabilities $p_{X_{i+1}|X_i}(x_{i+1}|x_i)$. In order to approximately calculate the mutual information rate,

$$
\mathcal{I}(X, Y) = \lim_{n \to \infty} \left( H(Y_n|Y_0^{n-1}) - H(Y_n|X_0^n, Y_0^{n-1}) \right)
$$

(196)

the second term reduces (in the case of our Markov channel) to $H(Y_n|X_0^n, Y_0^{n-1}) = H(Y_n|X_n, Y_{n-1})$, which is available in closed form. Thus, we need a way to estimate the first term, $H(Y_n|Y_0^{n-1})$, which requires the calculation of two quantities: $p_{Y_k|Y_0^{k-1}}(y_k|y_0^{k-1})$ and $p_{Y_k|X_0,Y_0^{k-1}}(y_k|x_0, y_0^{k-1})$, for various values of $k$.

Calculation of $p_{Y_k|Y_0^{k-1}}(y_k|y_0^{k-1})$ can be accomplished efficiently using the sum-product algorithm. By defining a sequence of functions $\varphi_i(x_i)$, which act as “messages” propagating along the factor graph, one obtains a recursive algorithm:

$$
\varphi_0(x_0, y_0) = p_{X_0,Y_0}(x_0, y_0)
$$

(197)

$$
\varphi_i(x_i, y_i) = \sum_{x_{i-1}} p_{Y_i|X_{i-1},Y_{i-1}}(y_i|x_{i-1}, y_{i-1}) p_{X_i|X_{i-1}}(x_i|x_{i-1}) \varphi_{i-1}(x_{i-1}, y_{i-1}), \quad \text{for } 1 < i \leq k
$$

(198)

$$
p_{Y_0^k}(y_0^k) = \sum_{x_0^k} p_{Y_k|X_{k-1},Y_{k-1}}(y_k|x_{k-1}, y_{k-1}) \varphi_{k-1}(x_{k-1}, y_{k-1})
$$

(199)

where the probability on the right side of (197) is the steady-state probability for the Markov process $(x_i, y_i)$. This well-known algorithm arises from the decomposition of the probability $p_{Y_0^k}(y_0^k)$ into a sum of products:

$$
p_{Y_0^k}(y_0^k) = \sum_{x_0^k} p_{X_0,Y_0}(x_0, y_0) p_{Y_k|X_0,Y_0}(y_k|x_0, y_0) \prod_{i=1}^{k} p_{Y_i|X_{i-1},Y_{i-1}}(y_i|x_{i-1}, y_{i-1}) p_{X_i|X_{i-1}}(x_i|x_{i-1})
$$

valid for our channel driven by a Markov input source. Finally, we obtain $p_{Y_k|Y_0^{k-1}}(y_k|y_0^{k-1})$ by

$$
p_{Y_k|Y_0^{k-1}}(y_k|y_0^{k-1}) = \frac{p_{Y_0^k}(y_0^k)}{\sum_{y_k} p_{Y_0^k}(y_0^k)}
$$

(200)

Calculation of $p_{Y_k|X_0,Y_0^{k-1}}(y_k|x_0, y_0^{k-1})$ proceeds similarly, except for $k = 1$, (198) is replaced by

$$
\varphi_1(x_1, y_1) = p_{Y_1|X_0,Y_0}(y_1|x_0, y_0) p_{X_1|X_0}(x_1|x_0) \varphi(x_0, y_0),
$$

(201)
i.e., we do not sum over $x_0$. The final result in (199) is then the joint probability $p_{Y_k^k, X_0}(y_k^k, x_0)$. Finally, we obtain

$$p_{Y_k^k|X_0, Y_k^{k-1}}(y_k| x_0, y_0^{k-1}) = \frac{p_{Y_k^k}(y_k, x_0)}{\sum_{y_k} p_{Y_k^k}(y_k, x_0)}.$$  \tag{202}$$

The upper and lower bounds from Section III are obtained by substituting (200) into (133), and (202) into (135), respectively.

To calculate a Monte Carlo estimate of the information rate, we obtain an estimate of

$$H \left( Y_k | Y_k^{k-1} \right) = \mathbb{E} \left[ \log \left( \frac{1}{p_{Y_k|Y_k^{k-1}}(y_k|y_0^{k-1})} \right) \right]$$  \tag{203}$$

for sufficiently large $k$. Here we generate sample sequences $y_k^k$ with the correct distribution, calculate $p_{Y_k|Y_k^{k-1}}(y_k|y_0^{k-1})$ using (200), and take the sample mean to obtain the term under the expectation in (203).

D. Critical point of the continuous time information rate

In §IV-A1 we consider the information rate as the time step $\epsilon$ goes to zero. We assume the mutual information rate has an interior maximum as a function of the high-state probability $x \equiv p_H$. Here we show that this maximum is unique. Setting the derivative of the mutual information rate (75) equal to zero gives a necessary and sufficient condition for the maximum, Equation (138). We may simplify (138) by introducing $\bar{\alpha} = x \alpha_H + (1 - x) \alpha_L$, which gives

\begin{align*}
0 &= -\alpha_H \beta \epsilon \log \left( \frac{1}{1 - \epsilon \bar{\alpha}} \right) + \alpha_L \beta \epsilon \log \left( \frac{1}{1 - \epsilon \bar{a}} \right) \\
&\quad + \beta \epsilon (\alpha_H - \alpha_L) \log \left( \frac{1}{\epsilon \bar{\alpha}} \right) - \alpha_H \epsilon (\alpha_L + \beta) \log \left( \frac{1}{\alpha_H \epsilon} \right) \\
&\quad + \alpha_L \epsilon (\alpha_H + \beta) \log \left( \frac{1}{\alpha_L \epsilon} \right) - \alpha_H \log \left( \frac{1}{1 - \epsilon \bar{\alpha}} \right) \\
&\quad + \alpha_L \log \left( \frac{1}{1 - \epsilon \bar{\alpha}} \right) - \alpha_L \log \left( \frac{1}{1 - \alpha_H \epsilon} \right) + \alpha_H \alpha_L \epsilon \log \left( \frac{1}{1 - \alpha_H \epsilon} \right) \\
&\quad + \alpha_L \log \left( \frac{1}{1 - \alpha_L \epsilon} \right) - \alpha_H \alpha_L \epsilon \log \left( \frac{1}{1 - \alpha_L \epsilon} \right) - \beta \log \left( \frac{1}{1 - \alpha_H \epsilon} \right) \\
&\quad + \alpha_H \beta \epsilon \log \left( \frac{1}{1 - \alpha_H \epsilon} \right) + \beta \log \left( \frac{1}{1 - \alpha_L \epsilon} \right) - \alpha_L \beta \epsilon \log \left( \frac{1}{1 - \alpha_L \epsilon} \right) .
\end{align*}
Inverting,

\[ 0 = \alpha_H \beta \epsilon \log (1 - \epsilon \bar{a}) - \alpha_L \beta \epsilon \log (1 - \epsilon \bar{a}) \]
\[ - \beta \epsilon (\alpha_H - \alpha_L) \log (\epsilon \bar{a}) + \alpha_H \epsilon (\alpha_L + \beta) \log (\alpha_H \epsilon) \]
\[ - \alpha_L \epsilon (\alpha_H + \beta) \log (\alpha_L \epsilon) + \alpha_H \log (1 - \epsilon \bar{a}) \]
\[ - \alpha_L \log (1 - \epsilon \bar{a}) + \alpha_L \log (1 - \alpha_H \epsilon) - \alpha_H \alpha_L \epsilon \log (1 - \alpha_H \epsilon) \]
\[ - \alpha_H \log (1 - \alpha_L \epsilon) + \alpha_H \alpha_L \epsilon \log (1 - \alpha_L \epsilon) + \beta \log (1 - \alpha_H \epsilon) \]
\[ - \alpha_H \beta \epsilon \log (1 - \alpha_H \epsilon) - \beta \log (1 - \alpha_L \epsilon) + \alpha_L \beta \epsilon \log (1 - \alpha_L \epsilon). \]

Gathering like terms,

\[ 0 = \alpha_H \beta \epsilon \log (1 - \epsilon \bar{a}) - \alpha_L \beta \epsilon \log (1 - \epsilon \bar{a}) + \alpha_H \log (1 - \epsilon \bar{a}) - \alpha_L \log (1 - \epsilon \bar{a}) \]
\[ - \beta \epsilon (\alpha_H - \alpha_L) \log (\epsilon \bar{a}) + \alpha_H \epsilon (\alpha_L + \beta) \log (\alpha_H \epsilon) \]
\[ - \alpha_L \epsilon (\alpha_H + \beta) \log (\alpha_L \epsilon) + \alpha_L \log (1 - \alpha_H \epsilon) - \alpha_H \alpha_L \epsilon \log (1 - \alpha_H \epsilon) \]
\[ - \alpha_H \log (1 - \alpha_L \epsilon) + \alpha_H \alpha_L \epsilon \log (1 - \alpha_L \epsilon) - \beta \log (1 - \alpha_L \epsilon) + \alpha_L \beta \epsilon \log (1 - \alpha_L \epsilon) \]
\[ = (\alpha_H \beta \epsilon \log - \alpha_L \beta \epsilon + \alpha_H - \alpha_L) \log (1 - \epsilon \bar{a}) - \beta \epsilon (\alpha_H - \alpha_L) \log (\epsilon \bar{a}) \]
\[ + \alpha_H \epsilon (\alpha_L + \beta) \log (\alpha_H \epsilon) - \alpha_L \epsilon (\alpha_H + \beta) \log (\alpha_L \epsilon) \]
\[ + (\alpha_L - \alpha_H \alpha_L \epsilon + \beta - \alpha_H \beta \epsilon) \log (1 - \alpha_H \epsilon) \]
\[ + (- \alpha_H + \alpha_H \alpha_L \epsilon - \beta + \alpha_L \beta \epsilon) \log (1 - \alpha_L \epsilon). \]

Only the terms involving \( \bar{a} \) depend on \( x \). In order for equality to hold as \( \epsilon \to 0^+ \), we require the value \( x(\epsilon) \) for which we have

\[ f(x, \epsilon) = (\alpha_H \beta \epsilon - \alpha_L \beta \epsilon + \alpha_H - \alpha_L) \log (1 - \epsilon \bar{a}) - \beta \epsilon (\alpha_H - \alpha_L) \log (\epsilon \bar{a}) \]
\[ = - (\alpha_H \epsilon (\alpha_L + \beta) \log (\alpha_H \epsilon) - \alpha_L \epsilon (\alpha_L + \beta) \log (\alpha_L \epsilon) \]
\[ + (\alpha_L - \alpha_H \alpha_L \epsilon + \beta - \alpha_H \beta \epsilon) \log (1 - \alpha_H \epsilon) + (- \alpha_H + \alpha_H \alpha_L \epsilon - \beta + \alpha_L \beta \epsilon) \log (1 - \alpha_L \epsilon)) \]
\[ = g(\epsilon). \]
Expanding both sides in orders of $\epsilon$, and using the expansion $\log(1 + u) = u - u^2/2 + O(u^3)$, as $\epsilon \to 0^+$, we have:

$$f(x, \epsilon) = \{(\alpha_H - \alpha_L)\beta \epsilon + (\alpha_H - \alpha_L)\} \left\{ -\epsilon A - \frac{\epsilon^2 A^2}{2} + O(\epsilon^3) \right\} - \beta \epsilon (\alpha_H - \alpha_L) \{\log \epsilon + \log \bar{A}\}$$

That is to say, we have the regular perturbation expansion

$$f(x, \epsilon) = \epsilon \log(\epsilon) f_0(x) + \epsilon f_1(x) + \epsilon^2 f_2(x) + O(\epsilon^3), \text{ as } \epsilon \to 0^+, \text{ with}$$

$$f_0(x) = -\beta (\alpha_H - \alpha_L)$$
$$f_1(x) = -(\alpha_H - \alpha_L) \bar{A}(x) - \beta (\alpha_H - \alpha_L) \log(\bar{A}(x))$$
$$f_2(x) = -\bar{A}(x)(\alpha_H - \alpha_L) \beta - \frac{\bar{A}(x)^2}{2} (\alpha_H - \alpha_L).$$

Note that $f_0$ does not, in fact, depend on $x$. For the right hand side we have:

$$g(\epsilon) = -\alpha_H (\alpha_L + \beta) \log(\alpha_H \epsilon) + \alpha_L (\alpha_H + \beta) \log(\alpha_L \epsilon)$$
$$- (\alpha_L - \alpha_H \alpha_L \epsilon + \beta - \alpha_H \beta \epsilon) \log(1 - \alpha_H \epsilon) - (-\alpha_H + \alpha_H \alpha_L \epsilon - \beta + \alpha_L \beta \epsilon \log(1 - \alpha_L \epsilon))$$
$$= -\alpha_H (\alpha_L + \beta) \log(\alpha_H \epsilon) + \alpha_L (\alpha_H + \beta) \log(\alpha_L \epsilon)$$
$$+ (-\alpha_L + \alpha_H \alpha_L \epsilon - \beta + \alpha_H \beta \epsilon) \log(1 - \alpha_H \epsilon) + (\alpha_H - \alpha_H \alpha_L \epsilon + \beta - \alpha_L \beta \epsilon) \log(1 - \alpha_L \epsilon)$$
$$= -\alpha_H (\alpha_L + \beta) \log(\alpha_H \epsilon)$$
$$+ \alpha_L (\alpha_H + \beta) \log(\alpha_L \epsilon)$$
$$+ (-\alpha_L + \alpha_H \alpha_L \epsilon - \beta + \alpha_H \beta \epsilon) \log(1 - \alpha_H \epsilon)$$
$$+ (\alpha_H - \alpha_H \alpha_L \epsilon + \beta - \alpha_L \beta \epsilon) \log(1 - \alpha_L \epsilon)$$
$$= -\alpha_H (\alpha_L + \beta) \log(\alpha_H) \epsilon - \alpha_H (\alpha_L + \beta) \epsilon \log(\epsilon)$$
$$+ \alpha_L (\alpha_H + \beta) \log(\alpha_L) \epsilon + \alpha_L (\alpha_H + \beta) \epsilon \log(\epsilon)$$
$$+ (-\alpha_L - \beta + (\alpha_H \alpha_L + \alpha_H \beta) \epsilon) \left\{-\alpha_H \epsilon - \frac{\alpha_H^2 \epsilon^2}{2} + O(\epsilon^3) \right\}$$
$$+ (\alpha_H + \beta - (\alpha_H \alpha_L + \alpha_L \beta) \epsilon) \left\{-\alpha_L \epsilon - \frac{\alpha_L^2 \epsilon^2}{2} + O(\epsilon^3) \right\}, \text{ as } \epsilon \to 0^+. $$
Therefore, as $\epsilon \to 0^+$, we have
\[
g(\epsilon) = \epsilon \log(\epsilon)g_0 + \epsilon g_1 + \epsilon^2 g_2 + O(\epsilon^3), \quad \text{as } \epsilon \to 0^+, \text{ with}
\]
\[
g_0 = -\alpha_H(\alpha_L + \beta) + \alpha_L(\alpha_H + \beta)
\]
\[
g_1 = -\alpha_H(\alpha_L + \beta) \log(\alpha_H) + \alpha_L(\alpha_H + \beta) \log(\alpha_L) - \alpha_H(-\alpha_L - \beta) - \alpha_L(\alpha_H + \beta)
\]
\[
g_2 = -\alpha_H(\alpha_L+\beta \alpha_L) + \alpha_L(\alpha_H\alpha_L + \alpha_L \beta) + (-\alpha_L - \beta) \left(-\frac{\alpha_H^2}{2}\right) + (\alpha_H + \beta) \left(-\frac{\alpha_L^2}{2}\right).
\]
Comparing the terms of order $\epsilon \log \epsilon$, we see that $f_0 = -\beta(\alpha_H - \alpha_L) = g_0$ holds independently of $x$.

Moving to the $O(\epsilon)$ terms, we require $x \in (0, 1)$ for which $f_1(x) = g_1$. That is, we require that
\[
f_1(x) = -(\alpha_H - \alpha_L)\bar{\alpha}(x) - \beta(\alpha_H - \alpha_L) \log(\bar{\alpha}(x))
\]
\[
= -\alpha_H(\alpha_L + \beta) \log(\alpha_H) + \alpha_L(\alpha_H + \beta) \log(\alpha_L) - \alpha_H(-\alpha_L - \beta) - \alpha_L(\alpha_H + \beta)
\]
\[
= g_1.
\]
If we introduce the function $\psi(x) = \bar{\alpha}(x) + \beta(1 + \log(\bar{\alpha}(x)))$, and a constant
\[
\mathcal{G} = \frac{\alpha_H(\alpha_L + \beta) \log(\alpha_H) - \alpha_L(\alpha_H + \beta) \log(\alpha_L)}{\alpha_H - \alpha_L}, \tag{206}
\]
then we have the equivalent requirement on $x$:
\[
\psi(x) = \mathcal{G}. \tag{207}
\]
Since $\bar{\alpha}(x) = \alpha_H x + \alpha_L(1 - x)$, we have
\[
\frac{d\psi(x)}{dx} = (\alpha_H - \alpha_L) \left(1 + \frac{\beta}{\bar{\alpha}}\right) > 0,
\]
and so $\psi$ is monotonically increasing on $(0, 1)$, and has a smooth inverse $\psi^{-1}$. The range of $\psi$ is $\psi(0) = \alpha_L + \beta(1 + \log \alpha_L) < \psi(x) < \psi(1) = \alpha_H + \beta(1 + \log \alpha_H)$, so Equation (207) has a unique solution, provided $\mathcal{G}$ lies in this range. To check, we need to verify that
\[
\alpha_L + \beta(1 + \log \alpha_L) < \frac{\alpha_H(\alpha_L + \beta) \log(\alpha_H) - \alpha_L(\alpha_H + \beta) \log(\alpha_L)}{\alpha_H - \alpha_L} < \alpha_H + \beta(1 + \log \alpha_H).
\]
Upon assuming that $\beta > 0$ and $0 \leq \alpha_L < \alpha_H \leq 1$, and setting $y = (\alpha_H - \alpha_L)/\alpha_H$, these inequalities reduce to showing that
\[
\frac{y}{1+y} < \log(1+y) < y
\]
for $0 < y \leq 1$, which are readily verified.

This calculation shows that, for small $\epsilon > 0$, there can only be one maximum for $\tilde{I}_\epsilon(x)$ in the interior of the unit interval. Moreover, $\tilde{I}_\epsilon(x) \to 0$ for $x \to 0$ and $x \to 1$, and $\tilde{I}_\epsilon(x) > 0$ for $0 < x < 1$. Therefore, $\tilde{I}_\epsilon$ has to have at least one maximum, but it can have at most one critical point (by the preceding argument) so it has a unique maximum.

The corresponding value of $x$ will be the asymptotically optimal value $x_{\text{opt}}$, as $\epsilon \to 0^+$.

REFERENCES

[1] H. P. Yockey, R. P. Platzman, and H. Quastler, Eds., Symposium on Information Theory in Biology. New York, London: Pergamon Press, 1958.

[2] H. P. Yockey, “A study of aging, thermal killing and radiation damage by information theory,” in Symposium on Information Theory in Biology, H. P. Yockey, R. P. Platzman, and H. Quastler, Eds. New York, London: Pergamon Press, 1958, pp. 297–316.

[3] F. Attneave, “Some informational aspects of visual perception,” Psychol. Rev., vol. 61, pp. 183–93, 1954.

[4] H. Barlow, Sensory Communications. MIT Press, 1961, ch. Possible principles underlying the transformation of sensory messages.

[5] T. Berger, Rate distortion theory: A mathematical basis for data compression. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[6] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, “Capacity of discrete molecular diffusion channels,” in IEEE Intl. Symp. on Inform. Theory, 2011.

[7] C. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, “Estimating the directed information to infer causal relationships in ensemble neural spike train recordings,” Journal of Computational Neuroscience, vol. 30, no. 1, pp. 17–44, Jan. 2011.

[8] C. Quinn, N. Kiyavash, and T. P. Coleman, “Directed information graphs,” Information Theory, IEEE Transactions on, submitted.

[9] R. Song, C. Rose, Y.-L. Tsai, and I. S. Mian, “Wireless signalling with identical quanta,” in IEEE Wireless Commun. and Networking Conf., 2012, to appear.

[10] A. W. Eckford, “Nanoscale communication with Brownian motion,” in Proc. Conference on Information Sciences and Systems, Baltimore, MD, 2007, pp. 160–165.

[11] A. Levchenko and I. Nemenman, “Cellular noise and information transmission,” Curr Opin Biotechnol, vol. 28, pp. 156–64, Aug 2014.

[12] A. Einolghozati, M. Sardari, and F. Fekri, “Collective sensing-capacity of bacteria populations,” in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on. IEEE, 2012, pp. 2959–2963.

[13] N. Farsad, W. Guo, and A. W. Eckford, “Tabletop molecular communication: text messages through chemical signals,” PLoS One, vol. 8, no. 12, p. e82935, 2013.

[14] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, “Molecular communication,” in Proc. 2005 NSTI Nanotechnology Conference, 2005, pp. 391–394.

[15] L. Parcerisa and I. F. Akyildiz, “Molecular communication options for long range nano networks,” Computer Networks, vol. 53, no. 16, pp. 2753–2766, Nov. 2009.
[16] T. Nakano, T. Suda, T. Kojuin, T. Haraguchi, and Y. Hiraoka, “Molecular communication through gap junction channels: System design, experiments and modeling,” in Proc. 2nd International Conference on Bio-Inspired Models of Network, Information, and Computing Systems, Budapest, Hungary, 2007.

[17] B. Atakan and O. Akan, “An information theoretical approach for molecular communication,” in Proc. 2nd Intl. Conf. on Bio-Inspired Models of Network, Information, and Computing Systems, 2007.

[18] D. J. Spencer, S. K. Hampton, P. Park, J. P. Zurkus, and P. J. Thomas, “The diffusion-limited biochemical signal-relay channel,” in Advances in Neural Information Processing Systems 16, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2004.

[19] A. Einolghozati, M. Sardari, and F. Fekri, “Capacity of diffusion-based molecular communication with ligand receptors,” in IEEE Inform. Theory Workshop, 2011.

[20] H. Li, S. M. Moser, and D. Guo, “Capacity of the memoryless additive inverse gaussian noise channel,” 2013, to appear.

[21] K. Wang, W.-J. Rappel, R. Kerr, and H. Levine, “Quantifying noise levels of intercellular signals.” Phys Rev E Stat Nonlin Soft Matter Phys, vol. 75, no. 6 Pt 1, p. 061905, 2007 Jun.

[22] M. Ueda and T. Shibata, “Stochastic signal processing and transduction in chemotactic response of eukaryotic cells.” Biophys J, vol. 93, no. 1, pp. 11–20, Jul 2007.

[23] M. Deger, M. Helias, S. Cardanobile, F. M. Atay, and S. Rotter, “Nonequilibrium dynamics of stochastic point processes with refractoriness,” Phys Rev E Stat Nonlin Soft Matter Phys, vol. 82, no. 2 Pt 1, p. 021129, Aug 2010.

[24] H. Cai and P. N. Devreotes, “Moving in the right direction: how eukaryotic cells migrate along chemical gradients,” Semin Cell Dev Biol, vol. 22, no. 8, pp. 834–41, Oct 2011.

[25] N. R. Gough, “Focus issue: signaling across membranes,” Sci STKE, vol. 2005, no. 313, p. eg11, Dec 2005.

[26] D. J. Robbins, D. L. Fei, and N. A. Riobo, “The Hedgehog signal transduction network,” Sci Signal, vol. 5, no. 246, p. re6, Oct 2012.

[27] N. Fedoroff and W. Fontana, “Genetic networks. small numbers of big molecules,” Science, vol. 297, no. 5584, pp. 1129–31, Aug 2002.

[28] W.-J. Rappel, P. J. Thomas, H. Levine, and W. F. Loomis, “Establishing direction during chemotaxis in eukaryotic cells,” Biophysical Journal, vol. 83, no. 3, pp. 1361–7, Sep 2002.

[29] P. Jonas, “AMPA-type glutamate receptors–nonselective cation channels mediating fast excitatory transmission in the CNS,” EXS, vol. 66, pp. 61–76, 1993.

[30] D. X. Keller, K. M. Franks, T. M. Bartol, Jr, and T. J. Sejnowski, “Calmodulin activation by calcium transients in the postsynaptic density of dendritic spines,” PLoS One, vol. 3, no. 4, p. e2045, 2008.

[31] O. A. Khair, R. J. Davies, and J. L. Devalia, “Bacterial-induced release of inflammatory mediators by bronchial epithelial cells,” Eur Respir J, vol. 9, no. 9, pp. 1913–22, Sep 1996.

[32] G. J. Goodhill, “A theoretical model of axon guidance by the robo code,” Neural Comput, vol. 15, no. 3, pp. 549–64, Mar 2003.

[33] K.-W. Yau, “Phototransduction mechanism in retinal rods and cones. The Friedenwald Lecture.” Investigative ophthalmology & visual science, vol. 35, no. 1, pp. 9–32, 1994.

[34] P. A. Iglesias and P. N. Devreotes, “Navigating through models of chemotaxis,” Current Opinion in Cell Biology, vol. 20, no. 1, pp. 35 – 40, 2008, cell structure and dynamics. [Online]. Available: http://www.sciencedirect.com/science/article/B6VRW-4RM895Y-1/2/26bb6aaebeda87fd2f10f3d166b03

[35] M. B. Kennedy, “Signal-processing machines at the postsynaptic density,” Science, vol. 290, pp. 750–754, 2000.
[36] J. M. Shifman, M. H. Choi, S. Mihalas, S. L. Mayo, and M. B. Kennedy, “Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) is activated by calmodulin with two bound calciums,” Proc Natl Acad Sci USA, vol. 103, no. 38, pp. 13968–73, 2006.

[37] P. N. Devreotes, “G protein-linked signaling pathways control the developmental program of Dictyostelium,” Neuron, vol. 12, pp. 235–241, February 1994.

[38] P. S. Klein, T. J. Sun, C. L. Saxe, 3rd, A. R. Kimmel, R. L. Johnson, and P. N. Devreotes, “A chemoattractant receptor controls development in Dictyostelium discoideum,” Science, vol. 241, no. 4872, pp. 1467–72, Sep 1988.

[39] M. S. Springer, M. F. Goy, and J. Adler, “Protein methylation in behavioural control mechanisms and in signal transduction.” Nature, vol. 280, no. 5720, pp. 279–284, 1979.

[40] T. Hunter and J. A. Cooper, “Protein-tyrosine kinases,” Annual review of biochemistry, vol. 54, no. 1, pp. 897–930, 1985.

[41] D. Fuller, W. Chen, M. Adler, A. Groisman, H. Levine, W.-J. Rappel, and W. F. Loomis, “External and internal constraints on eukaryotic chemotaxis,” Proc Natl Acad Sci U S A, vol. 107, no. 21, pp. 9656–9, May 2010.

[42] B. Hu, W. Chen, H. Levine, and W.-J. Rappel, “Quantifying information transmission in eukaryotic gradient sensing and chemotactic response,” J Stat Phys, vol. 142, no. 6, pp. 1167–1186, Apr 2011.

[43] M. Ueda, Y. Sako, T. Tanaka, P. Devreotes, and T. Yanagida, “Single-molecule analysis of chemotactic signaling in Dictyostelium cells,” Science, vol. 294, no. 5543, pp. 864–7, Oct 2001.

[44] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, “Information transduction capacity of noisy biochemical signaling networks,” Science, Sep 2011.

[45] K. Deisseroth, “Optogenetics,” Nature methods, vol. 8, no. 1, pp. 26–29, 2010.

[46] D. R. Schmidt and P. J. Thomas, “Measuring edge importance: a quantitative analysis of the stochastic shielding approximation for random processes on graphs,” The Journal of Mathematical Neuroscience, vol. 4, no. 1, p. 6, 2014.

[47] D. J. Higham, “Modeling and simulating chemical reactions,” SIAM review, vol. 50, no. 2, pp. 347–368, 2008.

[48] D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” J. Phys. Chem., vol. 81, pp. 2340–2361, 1977.

[49] D. F. Anderson and T. G. Kurtz, Design and Analysis of Biomolecular Circuits. Springer, 2011, ch. 1. Continuous Time Markov Chain Models for Chemical Reaction Networks.

[50] D. F. Anderson, B. Ermentrout, and P. J. Thomas, “Stochastic representations of ion channel kinetics and exact stochastic simulation of neuronal dynamics,” 11 Feb 2014, http://arxiv.org/abs/1402.2584.

[51] M. Davis, “Capacity and cutoff rate for Poisson-type channels,” IEEE Transactions on Information Theory, vol. 26, no. 6, pp. 710 – 715, 1980.

[52] Y. M. Kabanov, “The capacity of a channel of the Poisson type,” Theory of Probability and Applications, vol. 23, pp. 143–147, 1978, in Russian. English translation by M. Silverman.

[53] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel.” Information Theory, IEEE Transactions on, vol. 34, no. 6, pp. 1462–1471, 1988.

[54] Y. Ying and T. Berger, “Characterizing optimum (input, output) processes for finite-state channels with feedback,” in Proc. IEEE Intl. Symp. on Info. Theory, 2003, p. 117.

[55] J. Chen and T. Berger, “The capacity of finite-state Markov channels with feedback,” IEEE Trans. Info. Theory, vol. 51, no. 3, pp. 780–798, Mar. 2005.

[56] J. L. Massey, “Causality, feedback and directed information,” in Proc. 1990 Intl. Symp. on Info. Th. and its Applications, 1991.
[57] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. John Wiley and Sons, Inc., New York, 1991.

[58] S. Verdú, “Poisson communication theory,” March 1999, Invited talk for the International Technion Communication Day in honor of Israel Bar-David. http://www.princeton.edu/~verdu/reprints/VerduPoisson1999.pdf.

[59] M. Frey, “Information capacity of the Poisson channel,” *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 244–256, 1991.

[60] N. Michelusi and U. Mitra, “Capacity of electron-based communication over bacterial cables: the full-CSI case,” 2014, to appear in IEEE Journal on Selected Areas in Communications.

[61] N. Michelusi, S. Pirbadian, M. Y. El-Naggar, and U. Mitra, “A stochastic model for electron transfer in bacterial cables,” 2014, submitted to IEEE Journal on Selected Areas in Communications.

[62] Y. Ying and T. Berger, “Feedback capacity of binary channels with unit output memory,” submitted to IEEE Trans. Info. Theory.

[63] F. Kschischang, B. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498–519, 2001.