Linear Noise Approximation of Intensity-Driven Signal Transduction Channels

Gregory R. Hessler[1], Andrew W. Eckford[2], and Peter J. Thomas[3]

[1,3] Dept. of Math., Appl. Math., and Stats., Case Western Reserve University, Cleveland, OH, USA

[2] Dept. of EECS, York University, Toronto, ON, Canada

Emails: [1] gregory.hessler@case.edu, [2] aeckford@yorku.ca, [3] pjthomas@case.edu

Abstract—Biochemical signal transduction, a form of molecular communication, can be modeled using graphical Markov channels with input-modulated transition rates. Such channel models are strongly non-Gaussian. In this paper we use a linear noise approximation to construct a novel class of Gaussian additive white noise channels that capture essential features of fully- and partially-observed intensity-driven signal transduction. When channel state transitions that are sensitive to the input signal are directly observable, high-frequency information is transduced more efficiently than low-frequency information, and the mutual information rate per bandwidth (spectral efficiency) is significantly greater than when sensitive transitions and observable transitions are disjoint. When both observable and hidden transitions are input-sensitive, we observe a superadditive increase in spectral efficiency.

I. Introduction

Communication within biological systems often involves chemical processes in which a receptor (protein, DNA or a living cell) transitions among several metastable states, with transitions rates modulated by the intensity of an input signal. Examples of such input signals include: the concentration of a ligand, transcription factor, or metabolic substrate; intensity of a light source; and potential difference across a cell membrane.

A simple example of such a channel is the BIND channel [1] in which a receptor protein has two states: bound to the ligand, or unbound; the receptor is sensitive to the signal (ligand concentration) in the unbound state, and insensitive in the bound state. More elaborate channel schemes abound in biology: examples include the Channelrhodopsin-2 receptor (ChR2) [2] and the acetylcholine receptor (ACh) [3], [4], with three and five states, respectively. These systems are not restricted to receptors in signal transduction: similar models have been proposed for transcription of RNA in the presence of a promoter [5].

Receptor proteins may be modelled as Markov processes on a finite graph, with vertices representing the state of the receptor, and a weighted directed edge $a \rightarrow b$, with weight representing the transition rate from state a to state b. In general, not all of these edges have rates that are sensitive to the input signal, and not all the state transitions are observable by the organism. We will describe these systems more formally

in the next section, but from this description the system output can also be modelled as a *hidden Markov chain*.

$$
\begin{array}{ccccccc}
X_1 & & X_2 & & X_3 & \cdots & \\
\downarrow & & \downarrow & & \downarrow & & \\
(Y_0) \longrightarrow & Y_1 & \longrightarrow & Y_2 & \longrightarrow & Y_3 & \longrightarrow \cdots \\
& \downarrow & & \downarrow & & \downarrow & \\
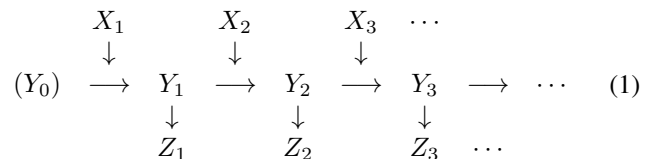& Z_1 & & Z_2 & & Z_3 & \cdots
\end{array}
\tag{1}
$$

Diagram (1) illustrates the structure of such a partially observed channel model, in a discrete time formulation: A sequence of input signals $(X_1, X_2, \ldots)$ modulates the transition rates governing the evolution of a channel state vector $(Y_1, Y_2, \ldots)$, a subset of which form the observable channel output $(Z_1, Z_2, \ldots)$. The *fully observed* channel refers to $X \rightarrow Y$; the *partially observed* channel to $X \rightarrow Z$.

There are many ways to evaluate the performance and capabilities of a signalling system, including ideal observer and signal detection analysis [6] and fidelity of stimulus reconstruction [7]. In this paper we consider the framework of information theory, particularly the mutual information (MI) between the input signal and the state of the receptor (either all states, Y, or the subset of observable states, Z). In our previous work, we calculated mutual information for individual receptors including ChR2 and ACh [8], and the BIND receptor for multiple receptors [9]; elsewhere, mutual information has been used as a figure of merit in biological systems such as the sensitivity of gene expression [10], [11].

Since the partially observed system consists of the features of signal transduction that are used by the organism, a complete analysis of the partially observed system is essential. While the fully observed system MI can be obtained in closed form for IID inputs [8], for the partially observed system the MI is generally intractable. Explicit expressions for the entropy rates of hidden Markov chains are often difficult to obtain, and new, approximate methods are required.

Our previous investigations revealed a qualitative relationship between the structure of *sensitivity* (which state transitions are modulated by the input) and *observability* (which components of Y are included in Z) and the MI of the fully and partially observed systems. In some cases the fully observed system has the same, or nearly the same MI as the partially observed system, so the latter can be studied using the closed form expression for the former. In other cases, such
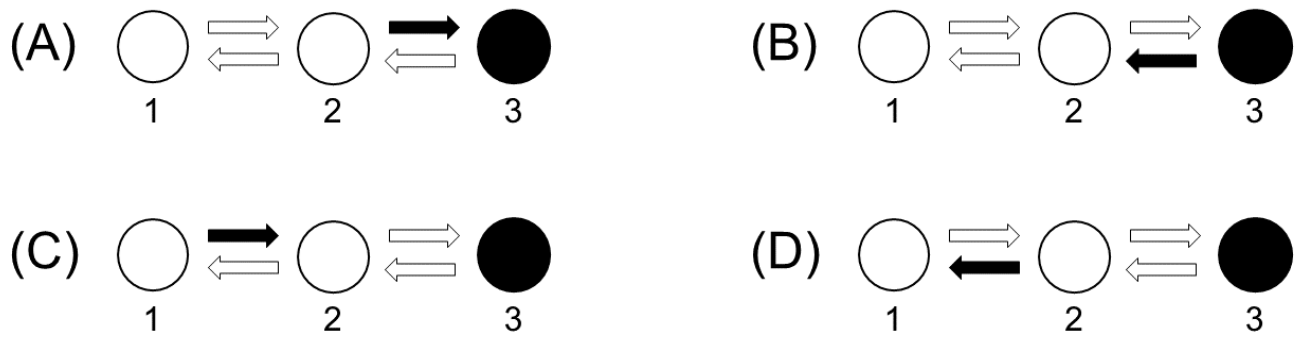
Fig. 1. Structure of a partially-observed intensity-modulated Markov channel with three states. The channel state is $Y = [Y_1, Y_2, Y_3]$; the channel output is $Z = [0, 0, 1] \cdot Y = Y_3$. Thus transitions along edge $2 \to 3$ and $3 \to 2$ are observable, while transitions $1 \to 2$ and $2 \to 1$ are hidden. In each version of the model, a single edge is sensitive to the input put X (filled arrows) and the remaining edges are insensitive (open arrows). In A and B the sensitive edge is observable; In C and D the sensitive edge is hidden. Qualitatively, we expect the mutual information $I(X; Z)$ to be larger for A and B than for C and D.

as the important neurotransmitter ACh, the fully observed and partially observed MI differ by an order of magnitude.

The key contribution of this paper is to develop a linear noise approximation [12], [13] applicable to large populations of senders and receivers, that gives a closed-form solution for the spectral efficiency (MI per bandwidth) of both the fully- and partially-observed signal transduction channel models. For our linear channel approximation, we show that high frequency information is transduced more efficiently when observable transitions are also sensitive; we also show that the spectral efficiency of receptors is dependent on whether sensitive transitions are observable. These results have biological significance, for example in predicting how an organism will observe high-frequency phenomena.

II. SYSTEM MODEL

A. Physical model

Consider a biochemical receptor that is sensitive to a given input process $X(t)$. We first describe a discrete-time channel model, $t \in \Delta t \cdot \mathbb{N}$, from which we will derive a continuous time model expressed as a stochastic differential equation (8) with $t \in \mathbb{R}$. As noted in the introduction, different types of receptors are sensitive to a variety of input stimuli.

For concreteness, here we will consider a simple three-state chain with one observable state (cf. Fig. 1). (However, our framework can be extended to general signal transduction systems, such as those considered in [8].) We see that:

- *Only some of the state transitions are sensitive to the input.* In Fig. 1, only the filled arrows are sensitive to the concentration of the ligand: their transition rate is directly proportional to concentration. The unfilled arrows always occur at the same rate, independent of concentration.
- *Only some of the state transitions are observable by the organism.* In Fig. 1, the unfilled bubbles (representing states 1 and 2) are indistinguishable to the organism. Thus, only transitions between states 2 and 3 are observable, while transitions between 1 and 2 occur internally to the receptor, hidden from the organism. Physically, this

may occur if the receptor has an ion channel, which is closed in states 1 and 2, and open in state 3.

We consider both "fully observed" systems (the channel with signal $X(t)$ as input and channel state $Y(t)$ as output) and "partially observed" systems (the channel with $X(t)$ as input and $Z(t) = C^{\mathsf{T}} Y(t)$ as output, where C is a rectangular matrix giving the observable output). Intuitively, we expect most of the information transduced about the signal to come from the edges that are both observable and sensitive. But even in networks for which all sensitive edges are hidden, there can still be a positive mutual information and capacity.

We consider an input signal generated by an ensemble of $M \gg 1$ independent binary sources that together produce a concentration or intensity $c(t) = c_{\mathsf{L}} + \Delta c \, V(t)/M$, where $\Delta c = (c_{\mathsf{H}} - c_{\mathsf{L}})$, $V(t) \sim \text{Binom}(M, p)$, and $0 \leq c_{\mathsf{L}} < c_{\mathsf{H}}$ are the low and high concentration signals that would obtain were all M sources simultaneously inactive, or active, respectively. As a simplification, we assume the input process is uncorrelated across discrete timesteps of size Δt (see Discussion). We let $\epsilon = 1/\sqrt{M}$ represent the relative size of fluctuations in the input concentration signal. By the central limit theorem, for sufficiently large M, we may approximate the input signal $c(t)$ with a discrete time Gaussian white noise input, and write the input process $X(t)$ as

$$X(t) = c(t) - \overline{c} = \epsilon \, \Delta c \, \sqrt{p(1-p)} \xi(t) \qquad (2)$$

where $\overline{c} = c_{\mathsf{L}} + p\Delta c$ is the mean concentration, the variance of the input is $(\Delta c)^2 \, p(1-p)/M$, ξ is Gaussian white noise with $\langle \xi(t) \rangle = 0$, and $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$. As a rule of thumb we restrict attention to parameter values for which the Gaussian approximation is appropriate, namely $10 < Mp < (M - 10)$. For the systems we consider, $M \gtrsim 1000$, so $.01 < p < .99$ is a suitable parameter range.

Next, consider a population of $N \gg 1$ independent receptors, each individually described by a channel state given by one of the standard unit vectors $Y(t) \in \mathcal{S} \equiv \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. We have shown previously that for a population of independent binary-state BIND receptors receiving common IID input, the

mutual information and capacity is exactly N times the single-receptor mutual information and capacity, respectively [9]. An analogous result holds for arbitrary finite-state receptors with IID input (the proof is straightforward but lengthy; we omit it for space). We assume that each receptor acts, independently of the others, as a conditional Markov process on $\mathcal{S}$, with transition rates depending on the input concentration $c = \bar{c} + X(t)$. Assuming standard first-order mass-action kinetics [14], the transition probability from state i to state j is affine-linear. Thus for each receptor we have

$$\Pr[Y(t+\Delta t) = j \mid Y(t) = i, X(t) = x] \qquad (3)$$
$$= (\alpha_{ij} + \beta_{ij}x)\,\Delta t + o(\Delta t), \text{ as } \Delta t \to 0.$$

Recall that only some of the state transitions are sensitive to the input concentration. If $\beta_{ij} = 0$, we define the $i \to j$ transition to be *insensitive*; otherwise it is *sensitive*. For the sensitive transitions, the mean input concentration affects the mean transition rate. Thus we have $\alpha_{ij}(p) = k_{ij} + c_{\mathsf{L}}\beta_{ij} + p\beta_{ij}\Delta c$, where k_{ij} is the intrinsic $i \to j$ transition rate. (Introducing a nonlinear dependence of the α_{ij} on p would not qualitatively change our results.) We will require that the channel state process is irreducible for all $0 < p < 1$.

Just as we assume the population of independent transmitters is large enough to justify a Gaussian approximation for the input concentration signal, we also assume a large population of $N \gg 1$ independent receptors exposed to the same input signal. For large N, the fluctuations in number of transitions per time among states, relative to the mean number of transitions per unit time, scale as $1/\sqrt{N}$ [15]. In general, we may consider a variety of scaling relations between M, the number of sources, and N, the number of receptors. For this paper we assume $M = N$ for simplicity. Our results would not change significantly with $N \propto M$ in some other fixed proportion. Thus we consider $N = M$ independent receptors all exposed to an identical input concentration $c(t) = \bar{c} + X(t)$. Kurtz's theorem guarantees that (over any given finite time horizon) the discrete process converges as $N \to \infty$ to a Gaussian process [16]. Thus for sufficiently large N, we may approximate the receptor state *via* a chemical Langevin equation [17], [18]. We write the three-state fractional population vector as $Y(t) = (Y_1, Y_2, Y_3)^\intercal$ with $Y_1 + Y_2 + Y_3 \equiv 1$ and $Y_i \geq 0$. To derive a tractable channel model based on the linear noise approximation, we begin with the exact (discrete) representation with N channels, in which case $Y_i \in \frac{1}{N}\{0, 1, 2, \ldots, N\}$. For $N \gg 1$, we may approximate the population state vector $Y(t)$ as a Gaussian random vector process in continuous time, based on several observations:

1) In the absence of input fluctuations ($c(t) = \bar{c}$) the mean channel state obeys a linear evolution equation with stable eigenvalues (except the Perron-Frobenius eigenvalue corresponding to the steady state channel distribution).
2) The input fluctuations are stationary, so Y has a stationary distribution, given by $\overline{Y}_1 = \alpha_{32}\alpha_{21}/\mathcal{Z}$, $\overline{Y}_2 =$

$\alpha_{32}\alpha_{12}/\mathcal{Z}$, $\overline{Y}_3 = \alpha_{12}\alpha_{23}/\mathcal{Z}$,, with $\mathcal{Z} = \alpha_{32}\alpha_{21} + \alpha_{32}\alpha_{12} + \alpha_{12}\alpha_{23}$. Generally both $\overline{Y}$ and the normalization $\mathcal{Z}$ depend on p.
3) For $M \gg 1$, the fluctuating part of the input concentration signal $\bar{c} + X(t)$ will be small (variance $O(1/M)$).
4) The variance of the channel state, $E[(Y_i - \overline{Y}_i)^2]$, will be small when both M and N are large.

To obtain a linear noise approximation [13], we expand the joint (X, Y) process around the mean, and neglect terms smaller than size $O(\epsilon)$. We subtract the mean with a linear change of coordinates $Y \to Y - \overline{Y}$. Thus, Y represents the deviations of the channel population vector from its mean value (therefore $E[Y] = 0$). We thereby obtain a first-order approximation with additive Gaussian noise represented by the increments of several independent Wiener processes (ΔW_{ij} for the intrinsic fluctuations in the $i \to j$ transition):

$$Y_1(t+\Delta t) - Y_1(t) = \qquad (4)$$
$$- \left((\overline{Y}_1\beta_{12}X + Y_1\alpha_{12})\,\Delta t + \epsilon\sqrt{\overline{Y}_1\alpha_{12}\Delta t}\,W_{12} \right)$$
$$+ \left((\overline{Y}_2\beta_{21}X + Y_2\alpha_{21})\,\Delta t + \epsilon\sqrt{\overline{Y}_2\alpha_{21}\Delta t}\,W_{21} \right)$$
$$Y_3(t+\Delta t) - Y_3(t) = \qquad (5)$$
$$- \left((\overline{Y}_3\beta_{32}X + Y_3\alpha_{32})\,\Delta t + \epsilon\sqrt{\overline{Y}_3\alpha_{32}\Delta t}\,W_{32} \right)$$
$$+ \left((\overline{Y}_2\beta_{23}X + Y_2\alpha_{23})\,\Delta t + \epsilon\sqrt{\overline{Y}_2\alpha_{23}\Delta t}\,W_{23} \right)$$
$$Y_2(t+\Delta t) = 1 - (Y_1(t+\Delta t) + Y_3(t+\Delta t)). \qquad (6)$$

Writing the system more compactly, equations (4)-(6) have the form

$$\mathbf{Y}(t+\Delta t) = (I + A\Delta t)\mathbf{Y}(t) + \mathbf{b}\,\Delta t\,X(t) + \epsilon\,\sqrt{\Delta t}\,\Gamma\,\mathbf{W}(t) \quad (7)$$

where the 3×3 matrix A, the 3×1 vector $\mathbf{b}$, and the 3×4 matrix Γ all depend on the parameter p, and $\mathbf{W}$ is a 4×1 vector of standard Gaussian random variables, with independent components and uncorrelated in time. Thus we arrive at a linear channel model, with structure arising from the channel state transition graph. The input signal is $X(t)$, a mean-zero Gaussian discrete time process with variance $\epsilon^2 p(1-p)(\Delta c)^2$, the channel noise is given by the Wiener processes comprising $\mathbf{W}$, and the channel memory arises from the mean transition rate matrix A.

In order to obtain analytical results we will consider a continuous time limit in which the channel model takes the form of a multidimensional Ornstein-Uhlenbeck process [17]

$$d\mathbf{Y} = A\mathbf{Y}\,dt + \epsilon\mathbf{b}\,\sqrt{p(1-p)}\,\Delta c\,dW_x + \epsilon\mathcal{B}\,d\mathbf{W}. \quad (8)$$

Here $\mathbf{Y}$ is a vector with components summing to zero, representing the fluctuations of the channel state around the mean occupancy (with the mean occupancy normalized to unity); A is a 3×3 matrix (depending on p) with a single null eigenvector corresponding to the stationary distribution, and remaining eigenvalues real and negative. The independent Wiener process increments $dW_x(t)$ represents the input. The

input fluctuations are "transduced" according to the vector $\mathbf{b} = \sum_{k \in \mathcal{E}^*} \beta_{ij(k)} \zeta_k$, where $\zeta_k = -\mathbf{e}_{i(k)} + \mathbf{e}_{j(k)}$ is the stoichiometry vector for directed edge $k : i(k) \to j(k)$ and $\mathcal{E}^*$ is the collection of sensitive edges in the graph. Note that the components of $\mathbf{b}$ sum to zero. The channel noise arises from the vector $d\mathbf{W}$ of independent Wiener process increments; note $d\mathbf{W}$ is also independent of dW_x. For the three-state model with four edges, $\mathcal{B}$ is a 3×4 matrix

$$\mathcal{B} = \left(\begin{array}{c|c|c|c} \sqrt{J_1}\zeta_1 & \sqrt{J_2}\zeta_2 & \sqrt{J_3}\zeta_3 & \sqrt{J_4}\zeta_4 \end{array} \right). \quad (9)$$

Here $J_i = \overline{Y}_i \alpha_{ij(k)}$ is the mean flux across the kth directed edge, with $ij(k)$ denoting the pair $i \to j$ corresponding to edge k. A similar construction results for arbitrary signal-transduction systems represented as a directed graph $(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}|$ vertices and $|\mathcal{E}|$ edges; in general $\mathcal{B}$ will be $|\mathcal{V}| \times |\mathcal{E}|$.

The mutual information rate between the Gaussian input and multidimensional Gaussian output processes $X(t)$ and $Y(t)$ in (8) is given by the integral of the spectral efficiency [19]–[21]

$$\text{SE}^{\text{FULL}}(\omega) = \frac{1}{2} \log \left(\mathcal{R}(\omega) \right) \quad (10)$$

$$\mathcal{R} = \frac{\det^* \left((A + i\omega)^{-1} (\mathcal{B}\mathcal{B}^T + p(1-p)\mathbf{b}\mathbf{b}^T)(A^T - i\omega)^{-1} \right)}{\det^* \left((A + i\omega)^{-1} \mathcal{B}\mathcal{B}^T (A^T - i\omega)^{-1} \right)}.$$

In this formula $\det^*$ denotes the pseudodeterminant, the product of the nonzero eigenvalues of a square matrix. For $|\omega| > 0$ the matrix $(A \pm i\omega I)$ is full rank; the matrix $\mathcal{B}\mathcal{B}^T$ has rank $|\mathcal{V}| - 1$. The rank-one matrix $\mathbf{b}\mathbf{b}^T$ is the outer product of a vector $\mathbf{b}$ that is a linear combination of the columns of $\mathcal{B}$, each of which is proportional to one of the stoichiometry vectors ζ_k. It follows that the matrix $\mathcal{B}\mathcal{B}^T + p(1-p)\mathbf{b}\mathbf{b}^T$ has the same rank as $\mathcal{B}\mathcal{B}^T$. The matrices in (10) are Hermitian and have the same number of nonzero (and real) eigenvalues, thus the ratio of pseudodeterminants is well defined. Note that the factor of ϵ multiplying $\mathbf{b}$ and $\mathcal{B}$ in (8) cancels in the ratio $\mathcal{R}$. Therefore, provided $\epsilon \ll 1$ is small enough (that is, $M, N \gg 1$ are large enough) to ensure validity of the underlying Gaussian approximations, the following results hold independent of the specific value of ϵ.

We wish to compare the spectral efficiency for the fully-observed system (10) with that of the partially-observed system. Let the observation vector $C = (0, 0, 1)^T$ (meaning transitions $2 \rightleftharpoons 3$ are directly observable, while transitions $1 \rightleftharpoons 2$ are not). Then the spectral efficiency appearing in the expression for the mutual information rate between the input $X(t)$ and the observed channel state $Z(t) = C^T Y(t)$ satisfies

$$\text{SE}^{\text{PART}}(\omega) = \frac{1}{2} \log \left(\mathcal{Q}(\omega) \right) \quad (11)$$

$$\mathcal{Q} = \frac{C^T (A + i\omega)^{-1} (\mathcal{B}\mathcal{B}^T + p(1-p)\mathbf{b}\mathbf{b}^T)(A^T - i\omega)^{-1} C}{C^T (A + i\omega)^{-1} \mathcal{B}\mathcal{B}^T (A^T - i\omega)^{-1} C}.$$

In general, the integrals $\int_{|\omega| < \omega_{\max}} \log \mathcal{R}(\omega) \, d\omega$ and $\int_{|\omega| < \omega_{\max}} \log \mathcal{Q}(\omega) \, d\omega$ diverge as $\omega_{\max}$ grows without bound. Rather than impose an arbitrary cutoff frequency, we study the integrands directly in their high- and low-frequency limits,

$$\text{SE}_0 \equiv \lim_{\omega \to 0} \text{SE}(\omega); \qquad \text{SE}_\infty \equiv \lim_{|\omega| \to \infty} \text{SE}(\omega). \quad (12)$$

While $\text{SE}^{\text{PART}}(\omega)$ varies over a finite range as ω varies $0 \leq |\omega| < \infty$, we find that SE^{FULL} is independent of ω and can be simplified to

$$\text{SE}^{\text{FULL}}(\omega) = \frac{1}{2} \log \left(\frac{\det^* \left(\mathcal{B}\mathcal{B}^T + p(1-p)\mathbf{b}\mathbf{b}^T \right)}{\det^* \left(\mathcal{B}\mathcal{B}^T \right)} \right). \quad (13)$$

III. RESULTS

Figs. 2-3 shows the spectral efficiency (SE) of the fully observed system (solid black lines), together with the spectral efficiency of the partially observed system in the high-frequency (dashed red lines) and low-frequency (dashed blue lines) limits, as the high-concentration input probability p ranges from 0.01 to 0.99. For all systems we use standard parameters $\alpha_{ij} = 1, \beta_{ij} = 0$ for insensitive edges and $\alpha_{ij} = 1, \beta_{ij} = 1$ for sensitive edges. Inset diagrams indicate which edges are sensitive (solid black arrows), resp. insensitive (open arrows), following the same scheme as Fig. 1.

In Fig. 2, a single transition is sensitive to the input. When this transition is observable, the high-frequency SE is indistinguishable from the fully-observed SE (Fig. 2A, 2B, black and red dashed lines), while the low-frequency SE is smaller (blue dashed lines). In contrast, when only the observable transitions are insensitive, the low-frequency SE is reduced, but the high-frequency SE is effectively zero (Fig. 2C, 2D). Thus, heuristically, only the slow time scale carries information about the hidden transition.

In each case the maximal value of SE occurs at an interior value of p, ranging roughly from 0.3 to 0.4. The maximal SE for the fully observed system is 0.025 for cases A and D (which are identical, by symmetry, for the parameters we use) and is 0.022 for cases B and C.

Fig. 3 contrasts a system with two sensitive transitions directed towards the third state (the transition to which is observable) (Fig. 3A), and a system with two sensitive transitions directed away from the third state (Fig. 3B). For the fully observed system, these two cases are identical and hence have identical spectral efficiency curves. However, for the partially observed system, we find that both the low and high frequency spectral efficiency limits are higher for the system with sensitive transitions moving towards state 3 than the system with sensitive transitions moving away from state 3. In contrast to the systems with a single sensitive edge, we find that the spectral efficiency at low frequencies exceeds that at high frequencies by a significant margin, when two sensitive edges operate in concert.

Panel 3C shows the sum of the two SE curves for the systems with a single edge directed towards the third state (sum of curves in 2A and 2C). Panel 3D shows the sum of the SE curves for the systems with a single edge directed away from the third state (sum of curves in 2B and 2D).
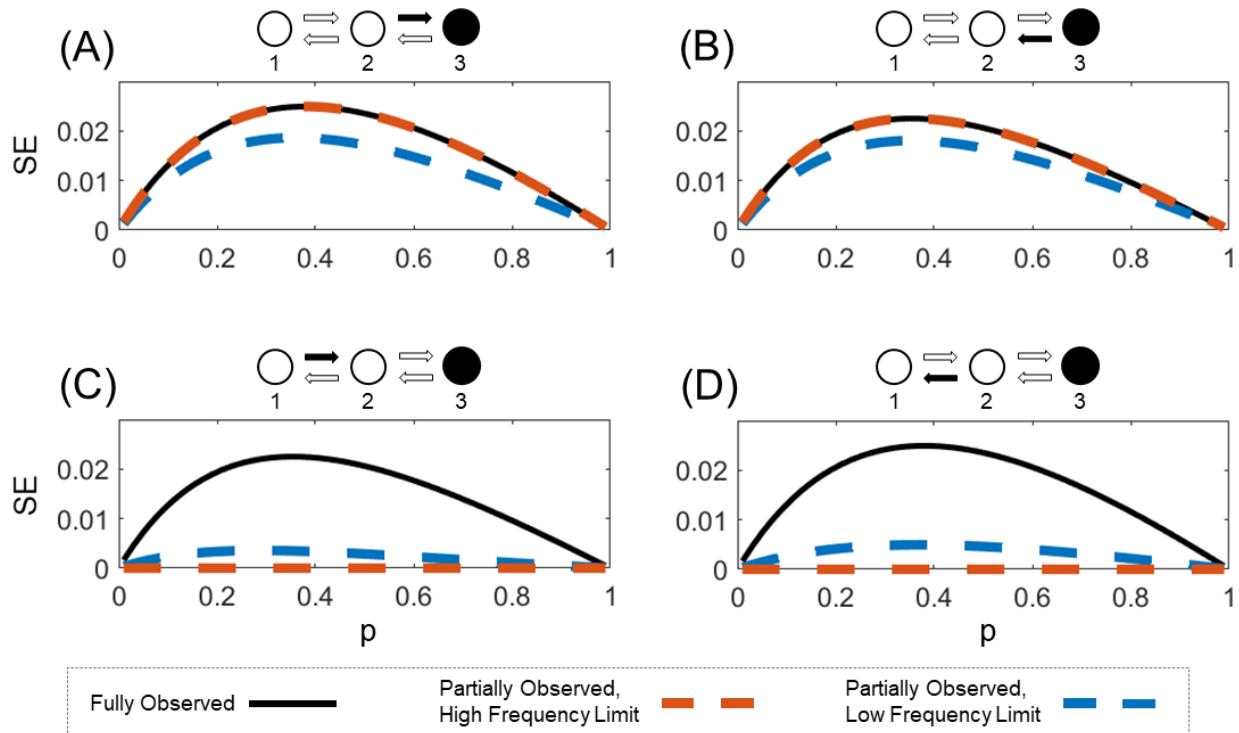
Fig. 2. Spectral efficiency (SE, (bits/sec)/Hz) as a function of input parameter p, when the system is fully observable (solid black line) or when only the transition to state 3 is observable (dashed lines). Under partial observability, SE is frequency-dependent. Red (resp. blue) dashing shows the high (resp. low) frequency SE limits. In each panel a single transition is sensitive (filled arrow) to the input, and three transitions are insensitive (open arrows). **A:** Observable transition $2 \to 3$ is sensitive. **B:** Observable transition $3 \to 2$ is sensitive. **C:** Hidden transition $1 \to 2$ is sensitive. **D:** Hidden transition $2 \to 1$ is sensitive.

The spectral efficiency of the fully observed system with edges $1 \to 2 \to 3$ sensitive is within 5% of the sum of the corresponding single-edge SE curves (compare Fig. 3A, 3C). The SE of the fully observed system with sensitive edges $3 \to 2 \to 1$ is within 12% of the sum of the $3 \to 2$ and $2 \to 1$ systems. The mechanism leading to higher SE for low- versus high-frequency inputs (Figs. 3A-B) remains to be explored.

IV. DISCUSSION

By considering a large population of independent sources signaling via concentration fluctuations, and a large population of independent protein molecules transducing the concentration signal through intensity-mediated state changes, we derived an additive Gaussian channel based on a linear noise approximation. Although models we have studied recently are more realistic for single receptors [1], [8], [9], [22], the model in this paper is analytically tractable for large populations with hidden states, and shows how signal transduction depends on the structure of observability and sensitivity in a channel. For example, Figs. 2A and 2C suggest that directly observable transitions carry information about the input on both fast and slow timescales, while hidden transitions carry information on slow timescales. This reflects the stochastic shielding phenomenon [18], [23]; cf. Fig. 4A of [15]; see also [24].

A shortcoming of the present treatment is the assumption of unlimited input signal power. Moreover, rapid signaling via molecular concentrations is limited by the physics of diffusion,

and a more realistic treatment would involve an input with a natural high frequency cutoff. The typical persistence time τ of a signaling molecule with diffusion constant D, in a geometry of size L, is roughly $\tau \sim L^2/2D$. For acetylcholine signaling in the neuromuscular junction (NMJ), typical values are $D = 4.0 \times 10^{-4} \mu m^2/\mu s$ and $L \lesssim 1\mu m$ [25], giving $\tau \lesssim 1$ msec. For comparison, nerve cells signaling through the NMJ transmit action potentials with maximal rates in the tens of Hz [26]. In [9] we considered the two-state BIND model driven both with IID inputs and with inputs having finite correlation time, and in [27] we considered thermodynamic constraints on diffusion-based signaling schemes. Further analysis along these lines would be a natural next step for the linear channel model explored here.

REFERENCES

[1] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Transactions on Information Theory*, vol. 62, pp. 7358–7382, Dec 2016.

[2] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, and E. Bamberg, "Channelrhodopsin-2, a directly light-gated cation-selective membrane channel," *PNAS*, vol. 100, no. 24, pp. 13940–13945, 2003.

[3] B. Sakmann, J. Patlak, and E. Neher, "Single acetylcholine-activated channels show burst-kinetics in presence of desensitizing concentrations of agonist," *Nature*, vol. 286, no. 5768, p. 71, 1980.

[4] D. Colquhoun and A. G. Hawkes, *Single-Channel Recording*, ch. The Principles of the Stochastic Interpretation of Ion-Channel Mechanisms. Plenum Press, New York, 1983.

[5] G. Rieckh and G. Tkačik, "Noise and information transmission in promoters with multiple internal states," *Biophys. J.*, vol. 106, pp. 1194–1204, 2014.
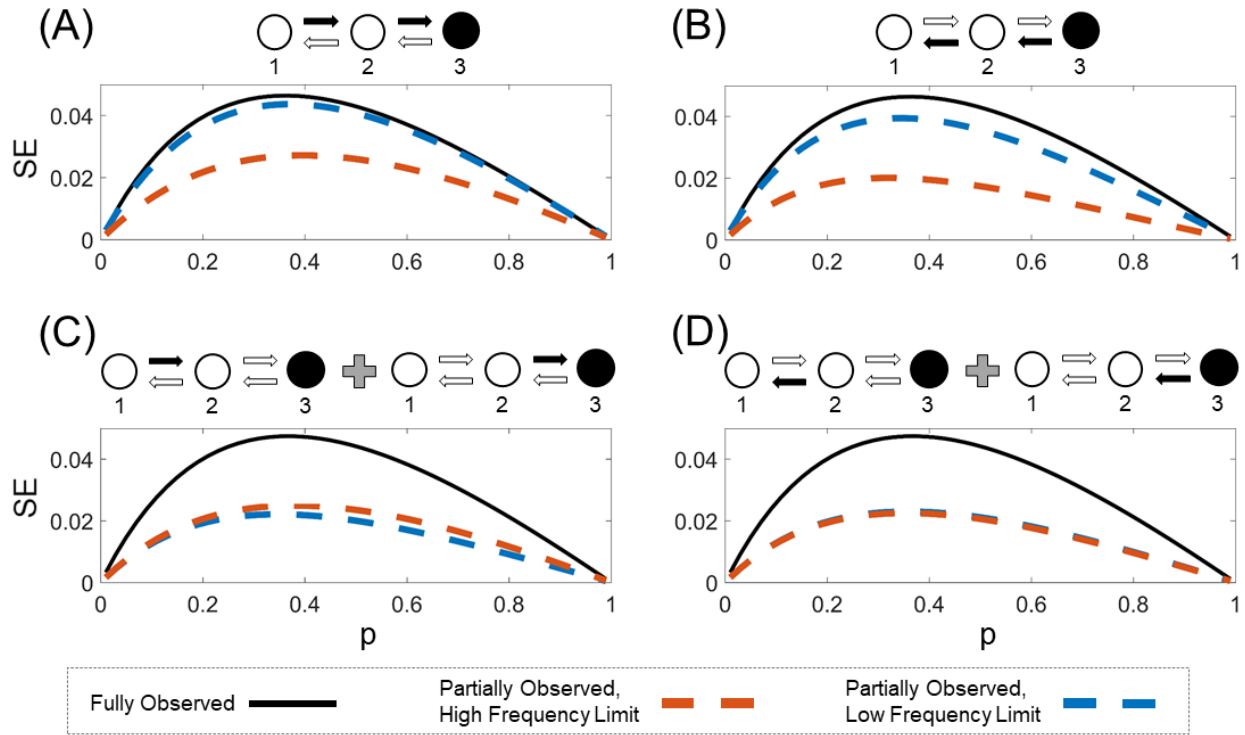
Fig. 3. Spectral efficiency (SE, (bits/sec)/Hz) as a function of input parameter p, when two transitions are sensitive and two insensitive. Conventions as in Fig. 2 (black line: fully observed; red dashed line: partially observed, high-frequency limit; blue dashed line: partially observed, low-frequency limit). **A:** Hidden transition $1 \rightarrow 2$ and observable transition $2 \rightarrow 3$ are sensitive. **B:** Hidden transition $2 \rightarrow 1$ and observable transition $3 \rightarrow 2$ are sensitive. **C:** Sum of SE from Figs. 2A ($2 \rightarrow 3$ sensitive) and 2C ($1 \rightarrow 2$ sensitive). **D:** Sum of SE from Figs. 2B ($3 \rightarrow 2$ sensitive) and 2D ($2 \rightarrow 1$ sensitive).

[6] J. A. Swets, *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Psychology Press, 2014.

[7] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Transactions on Information Theory*, vol. 44, pp. 2693–2723, 1998.

[8] A. W. Eckford and P. J. Thomas, "The channel capacity of channel-rhodopsin and other intensity-driven signal transduction receptors," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, 2019. In press (arXiv preprint arXiv:1804.04533).

[9] P. J. Thomas and A. W. Eckford, "Shannon capacity of signal transduction for multiple independent receptors," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1804–1808, July 2016.

[10] G. Tkačik, C. G. Callan, and W. Bialek, "Information flow and optimization in transcriptional regulation," *Proceedings of the National Academy of Sciences*, 2008.

[11] G. Tkačik and A. M. Walczak, "Information transmission in genetic regulatory networks: a review," *Journal of Physics: Condensed Matter*, vol. 23, no. 15, p. 153102, 2011.

[12] J. Elf and M. Ehrenberg, "Fast evaluation of fluctuations in biochemical networks with the linear noise approximation," *Genome research*, vol. 13, no. 11, pp. 2475–2484, 2003.

[13] E. Wallace, D. Gillespie, K. Sanft, and L. Petzold, "Linear noise approximation is valid over limited times for any chemical system that is sufficiently large," *IET systems biology*, vol. 6, no. 4, pp. 102–115, 2012.

[14] G. E. Briggs and J. B. S. Haldane, "A note on the kinetics of enzyme action," *Biochemical journal*, vol. 19, no. 2, p. 338, 1925.

[15] D. R. Schmidt, R. F. Galán, and P. J. Thomas, "Stochastic shielding and edge importance for markov chains with timescale separation," *PLoS computational biology*, vol. 14, no. 6, p. e1006206, 2018.

[16] T. G. Kurtz, "Limit theorems for sequences of jump markov processes approximating ordinary differential processes," *Journal of Applied Probability*, vol. 8, no. 2, pp. 344–356, 1971.

[17] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer Verlag, 2nd ed., 2004.

[18] 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.

[19] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 102–108, 1956.

[20] M. Pinsker, "Information and information stability of random variables and processes. Translated and edited by Amiel Feinstein, Holden-Day, Inc., San Francisco, Calif," *London-Amsterdam xii*, vol. 243, 1964. Cf. Equation 10.4.1.

[21] S. Ihara, *Information theory for continuous systems*, vol. 2. World Scientific, 1993.

[22] P. J. Thomas, D. J. Spencer, S. K. Hampton, P. Park, and J. Zurkus, "The diffusion-limited biochemical signal-relay channel," in *Advances in NIPS*, 2004.

[23] N. T. Schmandt and R. F. Galán, "Stochastic-shielding approximation of markov chains and its application to efficiently simulate random ion-channel gating," *Physical review letters*, vol. 109, no. 11, p. 118101, 2012.

[24] D. Bernardi and B. Lindner, "A frequency-resolved mutual information rate and its application to neural systems," *Journal of neurophysiology*, vol. 113, no. 5, pp. 1342–1357, 2014.

[25] K. Tai, S. D. Bond, H. R. MacMillan, N. A. Baker, M. J. Holst, and J. A. McCammon, "Finite element simulations of acetylcholine diffusion in neuromuscular junctions," *Biophysical journal*, vol. 84, no. 4, pp. 2234–2241, 2003.

[26] B. Bigland-Ritchie, R. Johansson, O. C. Lippold, S. Smith, and J. J. Woods, "Changes in motoneurone firing rates during sustained maximal voluntary contractions.," *The Journal of physiology*, vol. 340, no. 1, pp. 335–346, 1983.

[27] A. W. Eckford, B. Kuznets-Speck, M. Hinczewski, and P. J. Thomas, "Thermodynamic properties of molecular communication," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2545–2549, IEEE, 2018.