IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 2, 2019, accepted July 23, 2019, date of publication July 26, 2019, date of current version August 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931367

On the IID Capacity-Achieving Input for Binding Channels With Multiple Ligand Receptors

JIANFENG SUN AND HUI LI

Key Laboratory of Wireless-Optical Communications, Chinese Academy of Sciences, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

Corresponding author: Hui Li (mythlee@ustc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471335.

ABSTRACT This paper studies the molecular communication system where the transmitter has limited productivity and the receiver employs ligand-binding receptors. By simplifying the release and propagation process of the information particles, the ligand-binding process is regarded as a binding channel with peak and average constraints and modeled by a finite-state Markov chain. It is proved that the capacity of the constrained independent and identically distributed (IID) binding channel, defined as the IID capacity, is achieved by a discrete input distribution. The sufficient and necessary conditions of an IID input distribution being capacity-achieving is presented. Moreover, an algorithm called modified steepest ascent cutting-plane algorithm is proposed to efficiently compute the IID capacity-achieving distributions. The numerical results show that the IID capacity is a tight lower bound of the capacity for the binding channel.

INDEX TERMS Capacity-achieving distribution, channel capacity, channel models, iterative algorithms, molecular communication, ligand-binding receptors.

I. INTRODUCTION

Molecular communication (MC) is a promising research area which focuses on the communication paradigm for the nanoscale information exchange. In MC systems nano-machines exchange information by releasing molecules which act as the information carriers. These information particles propagate in an aqueous or a gaseous medium and are sensed or detected at the receiver.

Generally, the molecular receivers can be categorized into two classes, i.e., passive receiver and active receiver [1]. A passive receiver is usually assumed to be able to detect or sense the information particles without affecting their movement. On the other hand, active receivers employ receptor proteins, called ligand receptors, to detect information particles, namely, ligand. As an essential feature of the active reception, these receptor proteins participate in a reversible reaction, i.e., binding and unbinding, with the arrived ligands and form degradable ligand-receptor complexes. The ligand-binding receiver can be used in the designing of synthetic molecular receiver [2], [3]. In [4], the analytical signal models at the reversibly reactive receivers for diffusion-based molecular communication systems were investigated. The received signal model for the communication scenario in spherical bounded fluid environments was studied in [5]. The noise model and detection algorithms for molecular communication systems with ligand-binding receiver were also developed [3], [6], [7].

To evaluate the performance of molecular communication systems using ligand-binding reception mechanism, channel modeling and capacity analysis are required. In [8], the ligand-receptor binding process is modeled as a memoryless channel and Jeffery's prior is proved to be the capacity-achieving distribution. In [9], multiple ligand-binding receptors were modeled as independent Markov channels and the capacity-achieving distribution was also investigated. For the molecular communication system with a single ligand-binding receptor, it was proved in [10] that the channel capacity is achieved by an IID input distribution. When the input ligand concentration is binary and the transmission interval approaches 0, the capacity-achieving distribution of the binding channel with multiple independent ligand-binding receptors was further proved to be IID in [11]. More recently, by considering the ligand-receptor binding process as a signal transduction channel, an applied framework was presented in [12].

However, these previous works on the channel capacity analysis of the ligand-binding process or the binding channels

The associate editor coordinating the review of this manuscript and approving it for publication was Weisi Guo.

only considered a peak constraint on the channel input, which implies the ligand concentration can be chosen arbitrarily in a closed range. Whereas, due to the finite reaction rate and limited storage space, most ligand generators, including living cells and synthetic nano-machines,have limited productivity and hence can not continuously transmit ligands in high concentrations [13]. Therefore, an average constraint on the input should be considered, which will lead to some novel characteristics on the capacity-achieving distribution.

Besides, it is still an open problem to obtain the closed form of the capacity-achieving distribution of the binding channels, the numerically solution to the channel capacity and the capacity-achieving distribution is hence necessary, which is also studied in this work. By regarding the ligand-binding process as a special channel, called binding channel, it was proved in [9] that the IID channel capacity is achieved by a discrete distribution. A classical method to compute the capacity-achieving distribution for discrete channels is the Blahut-Arimoto algorithm (BAA) [14], [15]. In [16], a generalized BAA is proposed for the numerical optimization of finite-state channels. A modified BAA is developed in [17] to calculate the capacity of channels with feedback. All the above algorithms require that the input alphabet is finite or countably infinite and the capacity can be expressed in a min-max form. The dynamic assignment BAA proposed for the binomial channels in [18] can cope with uncountably infinite input alphabets, but still requires the min-max form for the capacity. The channel input of the binding channels is uncountably infinite and the capacity cannot expressed in a min-max form, which make the BAA-based algorithm unavailable in this case. It is necessary to develop new methods to numerically study the capacity of the binding channel.

In this paper, we concentrate on the point-to-point molecular communication systems where the transmitters have limited productivity and the receivers employ ligand-binding receptors. By simplifying the release and propagation process of the information particles, the ligand-binding process is regarded as a binding channel and modeled by a finite-state Markov chain.

It is worthy to note that the binding channel is not a fictional model, but instead is an abstract from the realistic, biological systems and can be used to describe the signal transduction processes in the biologically inspired molecular communication systems. For example, in the interneuronal communication, a neuron connects to another one across the synaptic cleft via the chemical messengers, called neurotransmitters. After being released by the presynaptic axon terminal, these neurotransmitters act by binding to the receptors on the membrane of the postsynaptic dendrite [19]. This reception process converts the molecular signal into the electrical signal and can be modeled as a binding channel. Besides this, the biological background of the investigation of the binding channels is formed by the signal transduction processes in the scenarios, including phototransduction in the retina [20], antigen presenting process by the dendritic cells in the immune system [21], chemotaxis of the tumor cells in cancer [22] and so on. The abstracted binding channel model can be used in many areas, e.g., the designing of the bio-electro interface in the healthcare delivery systems [23] and the analysis of the target cell localization in the targeted drug delivery systems [24].

Furthermore, the recent advances in the biotechnology provide an insight into the microscale reactions and facilitate the theoretical analysis on the ligand-receptor binding process. For example, the binding and unbinding processes can be visualized by the fluorescence method on the single-molecule level [25], and in neuroscience, the intracellular signal transduction process can be precisely modulated by the stimulation of light, which is known as the optogenetics [26]. These advances motivate us to develop an analytical model on the ligand-receptor binding process, which will not only provide biological researchers a perspective from information theory, but also benefit the further research on the designing of the artificial molecular communication systems.

Comparing to the binding channel model proposed in the previous works, in this paper the channel transition probability is modeled as a function of the communication duration, which underlies a more general and realistic model. By considering the limited productivity as an additional average constraint on the ligand concentration, the main contribution of this work is concluded as follows.

1) The capacity of the constrained IID binding channel is proved to be achieved by a discrete input distribution by using the Dubins' theorem.
2) The sufficient and necessary conditions on the capacity-achieving input distribution are proposed for the IID binding channel with peak and average constraints.
3) A numerical algorithm is proposed to efficiently compute the IID capacity-achieving distribution, and the computed results is proved to be convergent.
4) The effects of the average constraint on the IID capacity-achieving distribution are verified using numerical method.
5) It is verified by the numerical method that the IID capacity is a tightly lower bound of the capacity for the binding channel.

The rest of this paper is organized as follows: the system model is formulated in Section II, and the capacities of the binding channel are defined in Section III. The discreteness of the capacity-achieving distribution under peak and average constraints for the IID binding channels is proved in Section IV and the sufficient and necessary conditions of the IID input distribution being capacity-achieving is also studied in this section. In Section V, an algorithm called modified steepest ascent cutting-plane algorithm is developed to compute the capacity-achieving distribution, and the corresponding numerical results are shown in Section VI. Finally, conclusions are presented in Section VII.

II. MODEL FORMULATION

A. NOTATION

Let upper case letters denote random variables, e.g., X, Y, and their realizations are denoted by the corresponding lower case letters, e.g., x, y.

For $j \geq i$, a sequence of random variables $(X_i, \ldots, X_j)$ is denoted by X_i^j $(j \geq i)$, and their realizations can be denoted by x_i^j. For $j < i$, X_i^j and x_i^j denote empty sequence.

Let ν denote a probability measure defined on the Borel algebra on the nonnegative real numbers. The probability mass function (PMF) of a discrete random variable is denoted by P. Different probability measures or PMFs are discriminated by their subscripts. For example, ν_X denotes the probability measure of the random variable X, $P_{X,Y}$ denotes the joint distribution of discrete random variables X and Y, and $P_{Y|X}$ denotes the conditional distribution of Y given X. For simplicity, the subscripts is omitted when we are talking about an particular probability, e.g., $\nu(x)$, $P(x, y)$, and $P(y|x)$ are short for $\nu_X(x)$, $P_{X,Y}(x, y)$, and $P_{Y|X}(y \mid x)$, respectively.

Sets are denoted by the capital script letters, e.g., $\mathcal{X}$. Matrices are denoted by bold, capital letters, e.g., $\mathbf{A}$. Specially, the identity matrix is denoted by $\mathbf{I}$. Vectors are denoted by bold, small letters, e.g., $\mathbf{a}$. And the vectors are row vectors by default.

B. SYSTEM MODEL

In this work, we focus on the ligand-binding receiver and the channel model is sketched in Fig. 1.

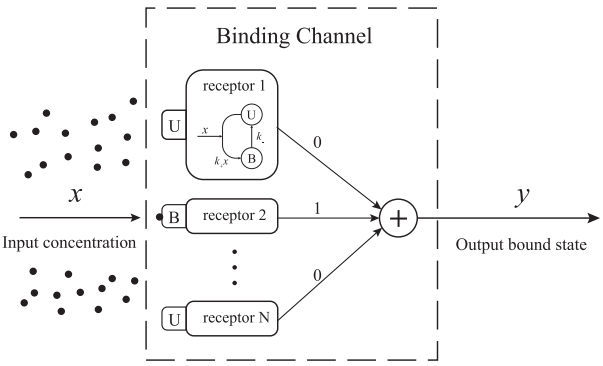

FIGURE 1. Binding channel with input concentration x and output bound state y.

Suppose there are N ligand-binding receptors at receiver. Each receptor has two states, i.e., bound and unbound. The input of the binding channel is the concentration of information particles and the channel output is the total number of the receptors in bound state.

Assume that the channel is divided equally into slots in time, denoted by T. The input sequence is X_1^k where X_i is the concentration of information molecules in i-th time slot for $i = 1, \ldots, k$. Following [9], the concentrations of information molecules are assumed to change continuously in $\mathcal{X} = [c_{\mathrm{L}}, c_{\mathrm{H}}]$ and can be held in a certain level in each time slot.

In addition, considering that the molecule generator in the source has a limited productivity, the concentrations is hence assumed to satisfy an average constraint $\bar{c} \in [c_{\mathrm{L}}, c_{\mathrm{H}}]$. Note that this assumption has been adopted in the previous works [13] and [27] to make the model more realistic.

The set of distributions which satisfy peak constraint $[c_{\mathrm{L}}, c_{\mathrm{H}}]$ and average constraint $\bar{c}$ is denoted by

$$\mathcal{V}^k(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c}) = \left\{ \nu_{X_1^k} \middle| \nu(\mathcal{X}^k) = 1, \mathbb{E}_{\nu_{X_i}}(X_i) \leq \bar{c} \right\}. \tag{1}$$

For simplicity, we shall use the shorthand $\mathcal{V}^k = \mathcal{V}^k(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ and the superscript will be omitted when $k = 1$.

The output sequence is denoted by Y_1^k, where Y_i denote the number of bound receptors at the end of the i-th time slot. Let Y_0 denote the initial state of the receptors and $\mathcal{N} = \{0, 1, \ldots, N\}$ be the state space, then $Y_i \in \mathcal{N}$, $i = 0, \ldots, k$.

Then the binding channel can be formulated by a finite-state Markov process based on the chemical master equation (CME). Let k_+ and k_- denote the rate coefficients corresponding to the bound reaction and unbound reaction, respectively. In the i-th transmission, the ligand-receptor binding occurs with a rate k_+X_i while the bound receptors unbind with a rate k_-. Let $Y_i(t)$ denote the number of bound receptors at time t in the i-th transmission, where $t \in (0, T]$.

The CME for the binding process can be described as the following equation.

$$N - Y_i(t) \underset{k_-}{\overset{k_+X_i}{\rightleftharpoons}} Y_i(t), \tag{2}$$

Let $p_n(t)$ denote the probability of $Y_i(t) = n$. A differential equation about $p_n(t)$ can be derived from the CME [6] as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_n(t) = {} & k_+X_i(N-n+1)p_{n-1}(t) + k_-(n+1)p_{n+1}(t) \\ & - [k_-n + k_+X_i(N-n)]\, p_n(t).\end{aligned} \tag{3}$$

Equation (3) can be interpreted as a homogeneous Markov process illustrated in Fig. 2. For $a, b \in \mathcal{N}$ and $X_i = x_i$, the transition intensity of the Markov process $\{Y_i(t), t \in (0, T]\}$ changing from state a to state b is denoted by $q_{a,b}(x_i)$.

$$q_{a,b}(x_i) = \begin{cases} a\mu, & a = b+1 \\ -[a\mu + (N-a)\lambda_i], & a = b \\ (N-a)\lambda_i, & a = b-1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\lambda_i = k_+x_i$ is proportional to the input concentration, $\mu = k_-$ is a constant.

The intensity matrix of the Markov Process is denoted by $\mathbf{Q}(x_i) = \left[q_{a,b}(x_i)\right]$. Let $\mathbf{d}(x_i)$ be the vector comprising the eigenvalues of $\mathbf{Q}(x_i)$ in descending order,

$$\mathbf{d}(x_i) = \Big(0, -(\lambda_i + \mu), \ldots, -N(\lambda_i + \mu)\Big). \tag{5}$$

Denote $\mathbf{P}(t; x_i) = \left[p_{a,b}(t; x_i)\right]$ as the transition matrix of the Markov process $\{Y_i(t), t \in (0, T]\}$, where $p_{a,b}(t; x_i)$ implies the probability of $Y_i(t)$ transitioning from state a to state b in time t, given the input concentration $X_i = x_i$.

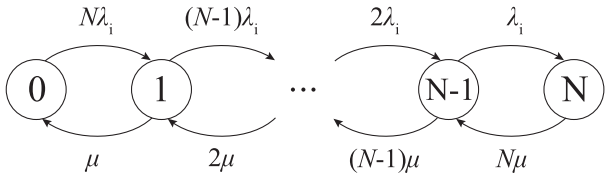

FIGURE 2. State transition diagram of the Markov chain, where $\lambda_i = k_+ X_i$ and $\mu = k_-$.

Using the Kolmogorov equation [28], we have

$$\mathbf{P}(t; x_i) = \mathbf{U}(x_i)\,\mathrm{diag}(\exp(\mathbf{d}(x_i)t))\,\mathbf{V}(x_i), \tag{6}$$

where $\mathbf{U}(x_i) = [u_{a,b}(x_i)]$ is the matrix consisting of the eigenvectors of $\mathbf{Q}(x_i)$ and satisfies $\mathbf{QU} = \mathbf{U}\,\mathrm{diag}(\mathbf{d})$; $\mathbf{V}(x_i) = [v_{a,b}(x_i)]$ is the inverse of $\mathbf{U}(x_i)$.

For $a, b \in \mathcal{N}$, let $\gamma = \min(a, b)$. The explicit expressions of $u_{a,b}(x_i)$ and $v_{a,b}(x_i)$ are

$$u_{a,b}(x_i) = \frac{\sum_{j=0}^{\gamma} \binom{a}{j}\binom{N-a}{b-j}\mu^j(-\lambda_i)^{b-j}}{\binom{N}{b}\mu^b}, \tag{7}$$

$$v_{a,b}(x_i) = \frac{\binom{N}{b}\mu^{N-b}\sum_{j=0}^{\gamma}\binom{N-b}{a-j}(-1)^{a-j}\lambda_i^{b-j}\mu^j}{(\lambda_i+\mu)^N}. \tag{8}$$

Therefore, the channel state transition probability in the i-th transmission is

$$\begin{aligned} P(y_i \mid y_{i-1}, x_i) &= p_{y_{i-1},y_i}(T; x_i) \\ &= \sum_{n=0}^{N} u_{y_{i-1},n} v_{n,y_i} \mathrm{e}^{-n(\lambda_i+\mu)T}. \end{aligned} \tag{9}$$

Remark 1: In [10], [11], the channel state transition probability is an approximate results as $T \to 0$. Different from these previous works, the channel state transition probability given in (9) is an accurate expression, which is a function of T, the length of time slot. For example, in the inter-neuronal communication, the receiver, i.e., the post-synaptic neuron, mainly contains two types of receptors, namely α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor and N-Methyl-D-Aspartate (NMDA) receptor. Since the signal transmission mediated by NMDA receptors is much slower than which mediated by AMPA receptors [29], it is hence reasonable and necessary to consider the cases where each communication is persisting hundreds millisecond. The channel model derived from (9) will be more realistic to describe the scenario with NMDA receptors.

According the biological nature of the binding process, the i-th output Y_i is supposed to depend only on the i-th input X_i and the adjacent previous output state Y_{i-1}, which follows the assumption used in the previous works [9]–[12] and reveals the Markov property of the ligand-binding process [30]. Hence, the channel model can be represented as

$$P(y_1^k \mid y_0, x_1^k) = \prod_{i=1}^{k} P(y_i \mid y_{i-1}, x_i), \tag{10}$$

where $Y_0 = y_0$ is the initial state of receptors and is supposed to be independent with X_1^k.

The PMF of the output distribution is

$$p(y_1^k) = \sum_{y_0} \int p(y_0) p(y_1^k \mid y_0, x_1^k) \nu(\mathrm{d}x_1^k). \tag{11}$$

III. CHANNEL CAPACITY

According to the channel characteristic in (10), the input-output mutual information can be expressed as

$$I(X_1^k; Y_1^k \mid Y_0) = \sum_{i=1}^{k} \Big[H(Y_i \mid Y_0^{i-1}) - H(Y_i \mid Y_{i-1}, X_i)\Big]. \tag{12}$$

The channel capacity C equals to the maximized mutual information rate per unit time:

$$C = \frac{1}{T} \lim_{k\to\infty} \max_{\nu_{X_1^k} \in \mathcal{V}^k} \frac{1}{k} I(X_1^k; Y_1^k \mid Y_0) \tag{13}$$

If the inputs are identically and independently distributed (IID), the outputs form a first-order Markov chain. For $k = 1, 2, \ldots,$

$$\begin{aligned} P(y_k | y_0^{k-1}) &= \int P(y_k \mid y_0^{k-1}, x_k) \nu(\mathrm{d}x_k \mid y_0^{k-1}) \\ &\overset{(a)}{=} \int P(y_k \mid y_0^{k-1}, x_k) \nu(\mathrm{d}x_k) \\ &= P(y_k \mid y_{k-1}), \end{aligned} \tag{14}$$

where (a) is established since X_k is independent of previous outputs.

Thus, the mutual information shown in (12) can be simplified as

$$\begin{aligned} I(X_1^k; Y_1^k \mid Y_0) &= \sum_{i=1}^{k} [H(Y_i \mid Y_{i-1}) - H(Y_i \mid Y_{i-1}, X_i)] \\ &= \sum_{i=1}^{k} I(X_i; Y_i \mid Y_{i-1}). \end{aligned} \tag{15}$$

Then the IID channel capacity C_{IID} is defined as

$$C_{\mathrm{IID}} = \frac{1}{T} \max_{\nu_X \in \mathcal{V}} \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} I(X_i; Y_i \mid Y_{i-1}), \tag{16}$$

where $\mathcal{V}$ is short for $\mathcal{V}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$.

Define $I_\infty(\nu_X)$ as

$$I_\infty(\nu_X) \triangleq \lim_{i\to\infty} I(X_i; Y_i \mid Y_{i-1}), \tag{17}$$

which can be expanded as

$$\begin{aligned} I_\infty(\nu_X) = \sum_{y_{i-1}=0}^{N} \pi(y_{i-1}) \int \sum_{y_i=0}^{N} P(y_i \mid y_{i-1}, x_i) \\ \cdot \log\left[\frac{P(y_i \mid y_{i-1}, x_i)}{P(y_i \mid y_{i-1})}\right] \cdot \nu_X(\mathrm{d}x_i), \quad \forall i > 0. \end{aligned} \tag{18}$$

where π is the unique stationary distribution for the output of the aperiodic and irreducible finite-state Markov chain Y_1^k according to (9) and (11).

Then the IID channel capacity C_{IID} can be further expressed as

$$C_{\text{IID}} = \frac{1}{T} \max_{\nu_X \in \mathcal{V}} I_\infty(\nu_X), \tag{19}$$

where $I_\infty(\nu_X) = \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} I(X_i; Y_i \mid Y_{i-1})$ based on the result of Cesáro mean in Theorem 4.2.3 in [31].

IV. CAPACITY-ACHIEVING INPUT

In this section, we focus on the characterization of the capacity-achieving distributions for the IID binding channels, which is defined as the input distribution maximizing the channel capacity defined in (19).

A. THE IID CAPACITY-ACHIEVING DISTRIBUTION IS DISCRETE

Since the binding channels have discrete outputs, it is intuitive to conjecture the capacity-achieving distribution is discrete.

Theorem 1: For the binding channel with IID inputs, the channel capacity is achieved by a discrete distribution in the set $\mathcal{V}(c_{\text{L}}, c_{\text{H}}, \bar{c})$.

Remark 2: In [9], it is proved that the IID channel capacity is achieved by a discrete input distribution, where the channel output is a vector consisting of the bound states of all receptors. However, the proof of the discreteness in [9] is based on the special expressions of the channel state transition probabilities, where the binding rate for a single receptor can be regarded as a special case of our channel model in (9) for $N = 1$ and $T \to \infty$.

In order to prove the discreteness of the capacity-achieving input of the binding channel for more general situations, we first introduce Dubins' theorem [32] and the definition of a linearly closed and linearly bounded set.

Definition 1: A convex set $\mathcal{S}$ is said to be *linearly closed and linearly bounded* if every straight line intersects $\mathcal{S}$ in a closed line segment.

Theorem 2 (Dubins' Theorem): Assume that $\mathcal{S}$ is a linearly closed and linearly bounded convex set. Let $\mathcal{I}$ be the intersection of $\mathcal{S}$ with N hyperplanes. Then every extreme point of $\mathcal{I}$ is a convex combination of $N+1$ or fewer extreme points of $\mathcal{S}$.

Then the proof of Theorem 1 is as follows.

Proof 1 (Proof of Theorem 1): For different distributions $\nu_0, \nu_1 \in \mathcal{V}(c_{\text{L}}, c_{\text{H}}, \bar{c})$ satisfying the peak and average constraints, define a straight line L as

$$L: \ \nu = (1-\theta)\nu_0 + \theta\nu_1, \ -\infty < \theta < \infty. \tag{20}$$

It can be verified that the line L intersects $\mathcal{V}(c_{\text{L}}, c_{\text{H}}, \bar{c})$ in a closed line segment. Hence $\mathcal{V}(c_{\text{L}}, c_{\text{H}}, \bar{c})$ is a linearly closed and linearly bounded convex set. In the following proof, we use $\mathcal{V}$ for short.

According to the Weierstrass extreme value theorem [33], $I_\infty(\nu_X)$ attains its maximum on $\mathcal{V}$. Denote the capacity-achieving distribution by ν_X^*, and the corresponding transmission probability distribution of the output Markov chain is denoted by $P^*_{Y_i|Y_{i-1}}$.

For $y_{i-1} \in \mathcal{N}$ and $y_i \in \{1, \ldots, N\}$, let H_{y_{i-1},y_i} denote the hyperplane defined as

$$H_{y_{i-1},y_i}: \int P(y_i|y_{i-1}, x_i)\nu(\mathrm{d}x_i) = P^*(y_i|y_{i-1}). \tag{21}$$

Let $\mathcal{A}$ denote the intersection of the set $\mathcal{V}$ and the above $N(N+1)$ hyperplanes. Since the stationary distribution π depends only on $P_{Y_i|Y_{i-1}}$, $I_\infty(\nu_X)$ defined in (18) is linear in ν_X for any $\nu_X \in \mathcal{A}$.

Thus, $I_\infty(\nu_X)$ is maximized by an extreme point of the set $\mathcal{A}$, namely, ν_X^* is an extreme point of $\mathcal{A}$.

By using the Dubins' theorem, it is proved that ν_X^* is a convex combination of at most $N(N+1)+1$ extreme points of $\mathcal{V}$.

The term "extreme point" of a convex set implies the point that cannot be expressed as a convex combination of other points in the set.

Let $\delta(\cdot)$ be the Dirac delta function. For $c \in [c_{\text{L}}, \bar{c})$, the point measure defined as $\delta_c(x) = \delta(x - c)$ must be a extreme point of $\mathcal{V}$. Besides, let B_{c_1,c_2} be the Bernoulli distribution defined as

$$B_{c_1,c_2} = \frac{c_2 - \bar{c}}{c_2 - c_1}\delta_{c_1} + \frac{\bar{c} - c_1}{c_2 - c_1}\delta_{c_2}, \tag{22}$$

where $c_1 \in [c_{\text{L}}, \bar{c}]$, $c_2 \in (\bar{c}, c_{\text{H}}]$.

It can be verified that $\mathbb{E}_{B_{c_1,c_2}}(X) = \bar{c}$ and B_{c_1,c_2} is a extreme point of $\mathcal{V}$. All other distributions in $\mathcal{V}$ can be expressed as a combination of distributions in $\{\delta_c, B_{c_1,c_2} \mid c \in [c_{\text{L}}, \bar{c}), c_1 \in [c_{\text{L}}, \bar{c}], c_2 \in (\bar{c}, c_{\text{H}}], \mathbb{E}_{B_{c_1,c_2}}(X) = \bar{c}\}$.

Therefore, the capacity-achieving distribution is discrete and take values at most $2(N^2 + N + 1)$ points.

Remark 3: Actually, the support set of the capacity-achieving distribution for IID binding channels, denoted by $\mathcal{S}^*$, contains much fewer points than the upper bound, i.e., $2(N^2+N+1)$ stated in Theorem 1. It is proved in [9] that the capacity-achieving distribution for IID binding channels takes values at most $(N+4)/2$ points for the specified channel model, where the channel output is the vector containing the states of all the receptors. According to the numerical results shown in Section VI, $(N+4)/2$ seems to be a much tight upper bound on the cardinality of $\mathcal{S}^*$. However, the proof of this conjecture is currently not available.

B. SUFFICIENT AND NECESSARY CONDITIONS OF BEING CAPACITY-ACHIEVING

Next, the sufficient and necessary conditions on the IID channel input to be capacity-achieving are investigated.

For a binding channel with a discrete IID input distribution P_X, let $\mathcal{S}$ be the corresponding support set and $P_{Y_i|Y_{i-1}}$ be the state transition matrix of the output Markov chain.

Let $\mathbf{R}\big(P_{Y_i|Y_{i-1}}, \mathcal{S}\big) = [r(y_{i-1}, x_i)]$ be a $(N+1)\times|\mathcal{S}|$ matrix, where the entries are defined as follows for any $y_{i-1} \in \mathcal{N}$ and $x_i \in \mathcal{S}$.

$$r(y_{i-1}, x_i) = \sum_{y_i=0}^{N} P(y_i|y_{i-1}, x_i) \log \frac{P(y_i|y_{i-1}, x_i)}{P(y_i|y_{i-1})} \tag{23}$$

The notation $\mathbf{p}_X$ is used to denote the probability vector corresponding to the PMF P_X. Similarly, $\boldsymbol{\pi}$ is the probability vector corresponding to the stationary distribution π. The notation $\mathbf{P}_{Y_i|Y_{i-1}}$ is used to denote the probability matrix corresponding to the conditional PMF $P_{Y_i|Y_{i-1}}$.

A PMF can be regarded as the combination of its probability vector and its support set, e.g., $P_X = (\mathbf{p}_X, \mathcal{S})$. The limitations of the mutual information defined in (17) can be rewritten in matrix form.

$$I(P_X) = \boldsymbol{\pi}\mathbf{R}\big(P_{Y_i|Y_{i-1}}, \mathcal{S}\big)\,\mathbf{p}_X^{\mathsf{T}}, \tag{24}$$

where $\mathbf{p}_X^{\mathsf{T}}$ is the transpose of $\mathbf{p}_X$, $P_{Y_i|Y_{i-1}}$ is a function of P_X and can be derived from (14).

$$P_{Y_i|Y_{i-1}} = \sum_{x_i \in \mathcal{S}} P_X(x_i) P_{Y_i|Y_{i-1}, X_i = x_i}. \tag{25}$$

Let $\mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ be the set of all the discrete PMFs satisfying the peak constraint $[c_{\mathrm{L}}, c_{\mathrm{H}}]$ and average constraint $\bar{c}$, and $\mathcal{P}$ be the shorthand of it. It is easy to verify that $\mathcal{P}$ is a convex subset of $\mathcal{V}$.

We use some decorations to make distinctions between different input distributions, e.g., P_X, P'_X indicate different PMFs, respectively.

Moreover, the variables induced by the input PMF are decorated with the same symbol, e.g., $\mathcal{S}'$ and $\mathbf{p}'_X$ are the support set and the probability vector corresponding to the distribution P'_X, respectively. Similar to define $\mathbf{P}'_{Y_i|Y_{i-1}}$ and $\boldsymbol{\pi}'$.

Definition 2: For any $P_X, P'_X \in \mathcal{P}$, the *weak derivative* of I_∞ at P_X is defined as [34]

$$I'_\infty(P'_X; P_X) = \lim_{\theta \downarrow 0} \frac{I_\infty\left(P_X^\theta\right) - I_\infty\left(P_X\right)}{\theta}, \tag{26}$$

where $P_X^\theta \triangleq (1-\theta)P_X + \theta P'_X$, and the down arrow notation is used to emphasize that θ is approaching 0 from above.

The weak derivative function defined in (26) can be expressed as follows (details in Appendix A).

$$\begin{aligned} I'_\infty(P'_X; P_X) = \boldsymbol{\pi}\left(\mathbf{P}'_{Y_i|Y_{i-1}} - \mathbf{W}\right)\boldsymbol{\Psi}^{-1}\mathbf{R}(P_{Y_i|Y_{i-1}}, \mathcal{S})\mathbf{p}_X^{\mathsf{T}} \\ + \boldsymbol{\pi}\,\mathbf{R}(P_{Y_i|Y_{i-1}}, \mathcal{S}')\left(\mathbf{p}'_X\right)^{\mathsf{T}}, \end{aligned} \tag{27}$$

where $\boldsymbol{\Psi} \triangleq \mathbf{W} - \mathbf{P}_{Y_i|Y_{i-1}}$; $\mathbf{W}$ is a $(N+1)\times(N+1)$ matrix defined in (59).

The sufficient and necessary conditions on the capacity-achieving distribution for the IID binding channels are stated in the following theorem.

Theorem 3: Suppose that $P_X^* \in \mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ is an IID input distribution of a binding channel under the peak and average constraints. Then P_X^* is capacity-achieving if and only if there exists a $\vartheta \geq 0$ such that for every $c \in [c_{\mathrm{L}}, c_{\mathrm{H}}]$,

$$\begin{aligned} \boldsymbol{\pi}^*(\mathbf{Q}_c - \mathbf{W})\left(\boldsymbol{\Psi}^*\right)^{-1}\mathbf{R}(P^*_{Y_i|Y_{i-1}}, \mathcal{S}^*)\mathbf{p}_X^{*\mathsf{T}} \\ + \boldsymbol{\pi}^*\mathbf{R}(P^*_{Y_i|Y_{i-1}}, c) - \vartheta(c - \bar{c}) \leq 0, \end{aligned} \tag{28}$$

where $\boldsymbol{\Psi}^* \triangleq \mathbf{W} - \mathbf{P}^*_{Y_i|Y_{i-1}}$, $\mathbf{Q}_c \triangleq \mathbf{P}(T; c)$. Furthermore, (28) hold with equality for $c \in \mathcal{S}^*$, respectively. Specially, if $\bar{c} = c_{\mathrm{H}}$, the sufficient and necessary conditions are degraded as

$$\begin{aligned} \boldsymbol{\pi}^*(\mathbf{Q}_c - \mathbf{W})\left(\boldsymbol{\Psi}^*\right)^{-1}\mathbf{R}(P^*_{Y_i|Y_{i-1}}, \mathcal{S}^*)\mathbf{p}_X^{*\mathsf{T}} \\ + \boldsymbol{\pi}^*\mathbf{R}(P^*_{Y_i|Y_{i-1}}, c) \leq 0 \end{aligned} \tag{29}$$

with equality if and only if $c \in \mathcal{S}^*$.

Proof 2: See Appendix B.

Let us define the effective average constraint as follows.

Definition 3: For an IID binding channel, an average constraint is said to be effective if the capacity-achieving distribution is changed after considering this constraint.

Based on Theorem 3, a property of the optimal support, which is defined as the support set of the capacity-achieving distribution, is investigated.

Corollary 1: For an IID binding channel with a peak constraint $[c_{\mathrm{L}}, c_{\mathrm{H}}]$ and an effective average constraint $\mathbb{E}_{P_X}(X) \leq \bar{c}$, the lowest concentration c_{L} must be included in the optimal support.

Proof 3: Let $P_X^* \in \mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ be the capacity-achieving distribution supported on $\mathcal{S}^* = \left\{s_1^*, \ldots, s_M^*\right\}$.

Without loss of generality, suppose that the elements in $\mathcal{S}^*$ are strictly ordered satisfying $c_{\mathrm{L}} \leq s_1^* < \cdots < s_M^* \leq c_H$.

By contradiction, suppose that $c_{\mathrm{L}} < s_1^*$. For any $\epsilon \geq 0$, the sufficient and necessary conditions says

$$L'\left(\delta_{s_1^*-\epsilon}; P_X^*\right) \leq 0, \tag{30}$$

with equality if and only if $\epsilon = 0$, where $L'(\cdot;\cdot)$ is defined in (64) in Appendix.

Since P_X^* is supposed to be optimal, $\mathbb{E}_{P_X^*}(X^*) = \bar{c}$. According to (64), we have

$$\begin{aligned} L'\left(\delta_{s_1^*-\epsilon}; P_X^*\right) &= I'_\infty(\delta_{s_1^*-\epsilon}; P_X^*) - \vartheta\left(\mathbb{E}_{\delta_{s_1^*-\epsilon}}(X) - \bar{c}\right) \\ &= I'_\infty(\delta_{s_1^*-\epsilon}; P_X^*) - \vartheta\left(s_1^* - \epsilon - \bar{c}\right) \leq 0. \end{aligned} \tag{31}$$

For a binding channel with an effective average constraint, the Lagrangian multiplier $\vartheta > 0$. Hence,

$$I'_\infty(\delta_{s_1^*-\epsilon}; P_X^*) \leq \vartheta\left(s_1^* - \epsilon - \bar{c}\right) < 0. \tag{32}$$

Based on the definition in (26), there must be a distribution $P_X^\theta = (1-\theta)P_X^* + \theta\delta_{s_1^*-\epsilon} \in \mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ such that

$$I_\infty(P_X^\theta) > I_\infty(P_X^*), \tag{33}$$

which contradicts the optimal-input assumption. Hence, $c_{\mathrm{L}} \in \mathcal{S}^*$.

For a binding channel with only a peak constraint, it is proved that the capacity-achieving distribution must take values of $X = c_{\mathrm{L}}$ and $X = c_{\mathrm{H}}$ [9], [10]. However, with an additionally effective average constraint, only c_{L} is certainly included in the optimal support, and the highest concentration in the optimal support may be remarkably less than c_{H}. The numerical results in Section VI affirm this conclusion.

V. CALCULATION OF CAPACITY-ACHIEVING DISTRIBUTION

It is proved in Section IV that the capacity-achieving distribution for IID binding channels is discrete. Note the fact that the input concentration can be chosen in a continuous range, i.e., $[c_{\mathrm{L}}, c_{\mathrm{H}}]$. It leads to the difficulty in the computation of the capacity-achieving input distribution. The classical BAA [14]–[16], which requires the finiteness of the input alphabet, is hence inapplicable for our model.

In [18], a dynamic assignment BAA is proposed to compute the capacity-achieving input distributions for the binomial channels with uncountably infinite input alphabets. However, the dynamic assignment BAA requires expressing the channel capacity in the min-max form, which is really a challenge for the capacity defined in (19).

A numerical algorithm, known as the steepest ascent cutting-plane algorithm (SACPA), can efficiently solve the problem. It is a piecewise-linear approximation method firstly proposed in [35]. In this section, a modified SACPA is proposed to compute the capacity-achieving input distributions of binding channels.

Before giving the exact description of the modified SACPA, a lemma is introduced at first.

Lemma 1: For any $P_X \in \mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$, the following equation holds.

$$I_\infty(P_X) = \min_{P'_X \in \mathcal{P}} \boldsymbol{\pi}\mathbf{R}\left(P'_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}. \tag{34}$$

where $P'_{Y_i|Y_{i-1}}$ is determined by P'_X.

Proof 4: According to (24) and (23), we have

$$\begin{aligned}&\boldsymbol{\pi}\mathbf{R}\left(P'_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}\\ &\quad= \boldsymbol{\pi}\mathbf{R}\left(P_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}} + \boldsymbol{\pi}\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P'_{Y_i|Y_{i-1}}\right)\\ &\quad\geq \boldsymbol{\pi}\mathbf{R}\left(P_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}},\end{aligned} \tag{35}$$

where $\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P'_{Y_i|Y_{i-1}}\right)$ is a column vector defined in (51). Hence $\boldsymbol{\pi}\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P'_{Y_i|Y_{i-1}}\right)$ must be non-negative.

Based on Lemma 1, the IID capacity defined in (19) can be expressed as

$$C_{\mathrm{IID}} = \frac{1}{T}\max_{P_X\in\mathcal{P}}\min_{P'_X\in\mathcal{P}} \boldsymbol{\pi}\mathbf{R}\left(P'_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}. \tag{36}$$

The original SACPA is an iterative algorithm, of which the main idea is listed as follows.

1) At each iteration, find the piecewise approximation of the capacity-achieving input distribution on a fixed input alphabet.
2) Update the input alphabet with the most likely point, where the likelihood can be measured using Theorem 3.

Let $P_X^{(n)}$ denote the input distribution found at the n-th iteration. For a discrete set $\mathcal{S} \subset [c_{\mathrm{L}}, c_{\mathrm{H}}]$, define that

$$\mathcal{P}(\mathcal{S}) = \left\{P_X \in \mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c}) \middle| \sum_{x\in\mathcal{S}} P_X(x) = 1\right\}. \tag{37}$$

The details of the modified SACPA is described in Algorithm 1.

Algorithm 1 Modified Steepest Ascent Cutting-Plane Algorithm

Initialize algorithm with an arbitrary input distribution $P_X^{(0)} = \left(\mathbf{p}_X^{(0)}, \mathcal{S}^{(0)}\right)$.

repeat

1. Generate the piecewise approximation function in the n-th iteration.

$$I^{(n)}(P_X) = \min_{0<i<n-1} \boldsymbol{\pi}\mathbf{R}\left(P^{(i)}_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}, \tag{38}$$

where $P^{(i)}_{Y_i|Y_{i-1}}$ is the transition matrix of the output states determined by $P_X^{(i)}$.

2. Generate the n-th distribution

$$P_X^{(n)} = \arg\max_{P_X\in\mathcal{P}(\mathcal{S}^{(n-1)})} I^{(n)}(P_X). \tag{39}$$

3. Update the support. Let

$$x^{(n)} = \arg\max_{x\in[c_{\mathrm{L}}, c_{\mathrm{H}}]}\left[I'_\infty\left(\delta_x; P_X^{(n)}\right) - \nu^{(n)}(x-\bar{c})\right], \tag{40}$$

where $\nu^{(n)}$ is an associated Lagrange multiplier calculated in (39).

The new support $S^{(n)}$ is obtained by

$$S^{(n)} = S^{(n-1)} \bigcup \left\{x^{(n)}\right\} \tag{41}$$

until $I^{(n)}\left(P_X^{(n)}\right) - I_\infty\left(P_X^{(n)}\right) < \epsilon$.

It can be proved that, after infinite iteration, the distribution found by the modified SACPA will converge to the capacity-achieving IID input distribution.

Theorem 4: Let P_X^* be the capacity-achieving IID input distribution. Then $P_X^{(n)} \to P_X^*$ as $n \to \infty$.

Proof 5: Note that $\mathcal{P}(c_{\mathrm{L}}, c_{\mathrm{H}}, \bar{c})$ is a compact space. The infinite sequence $\left\{P_X^{(n)}\right\}_{n=0}^{\infty}$ is hence have an infinite convergent subsequence, denoted by $\left\{P_X^{(n_k)}\right\}_{k=0}^{\infty}$.

By using (34), (38) and (39), the following inequality is established for all $P_X \in \mathcal{P}$.

$$I_\infty(P_X) \leq I^{(n_k)}(P_X) \leq I^{(n_k)}\left(P_X^{(n_k)}\right) \tag{42}$$

Another inequality obtained from (38) is

$$I^{(n_k)}\left(P_X^{(n_k)}\right) \leq \boldsymbol{\pi}^{(n_k)}\mathbf{R}\left(P^{(n_k-1)}_{Y_i|Y_{i-1}}, \mathcal{S}^{(n_k)}\right)\left(\mathbf{p}_X^{(n_k)}\right)^{\mathsf{T}}. \tag{43}$$

As $k \to \infty$, the right-hand side item in (43) is convergent to $I_\infty\left(P_X^{(\infty)}\right)$, which implies that, for $\forall P_X \in \mathcal{P}$,

$$I_\infty\left(P_X^{(\infty)}\right) \geq I_\infty\left(P_X\right). \tag{44}$$

Thus, $P_X^{(\infty)}$ is the capacity-achieving input distribution.

VI. NUMERICAL RESULTS

In this section, the numerical results are presented based on the environment parameters in Table 1, where the rate coefficients corresponding to the bound reaction and unbound reaction are obtained from [36] and μM refers to the unit "micromolar" where $1\,\mu\text{M} = 10^{-6}\text{mol}\,\text{L}^{-1}$.

TABLE 1. Parameters in numerical experiments.

Notations	Parameters	Values
k_+	Binding rate coefficient	$4.6\,\mu\text{M}^{-1}\text{s}^{-1}$
k_-	Unbinding rate coefficient	$30\,\text{s}^{-1}$
c_L	Lowest input concentrations	$0\,\mu\text{M}$
c_L	Highest input concentrations	$50\,\mu\text{M}$

A. WITH ONLY A PEAK CONSTRAINT

For the IID channels with only a peak constraint $[c_\text{L}, c_\text{H}]$, The capacity-achieving input distributions calculated by the modified SACPA are shown in Fig. 3, where the probabilities of the mass points are represented by the area of the circle. The scatter plot numerically verifies the discreteness stated in Theorem 1. As increasing the number of receptors form 1 to 10, the cardinality of the support set increases slowly. Comparing to the upper bound stated in Theorem 1, the support set contains much fewer points.

The orange asterisks against the left y-axis in Fig. 3 represent the expectations of the capacity-achieving inputs. As we can see, the expectation against the right y-axis attains its maximum when $N = 2$ and then decreases with increasing N.

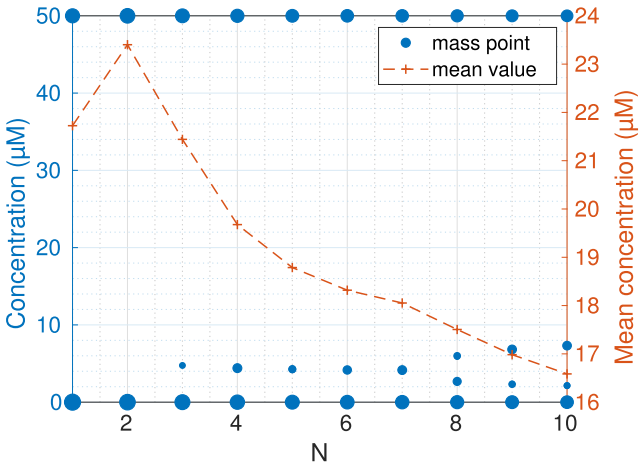

FIGURE 3. Capacity-achieving distributions with only a peak constraint, fixing $T = 100\,\text{ms}$.

The validity of the capacity-achieving input distributions obtained by the modified SACPA are evaluated by the criteria proposed in Theorem 3. Fig. 4 shows the criterion functions when $N \in \{2, 5, 8\}$. As is shown, the criterion function is non-positive and equal to 0 only when it takes values in the optimal support set.

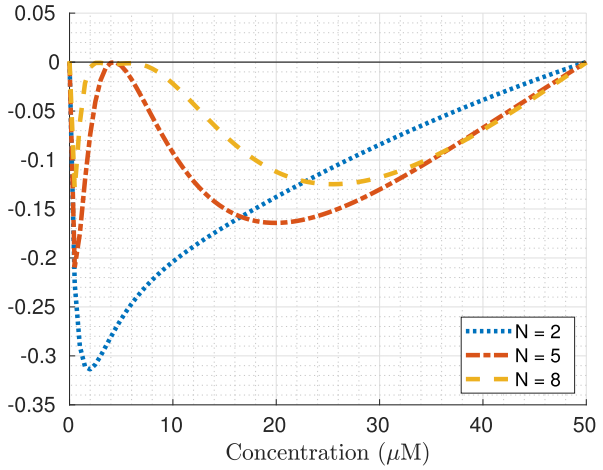

FIGURE 4. The validity of the calculated capacity-achieving input distributions, fixing $T = 100\,\text{ms}$.

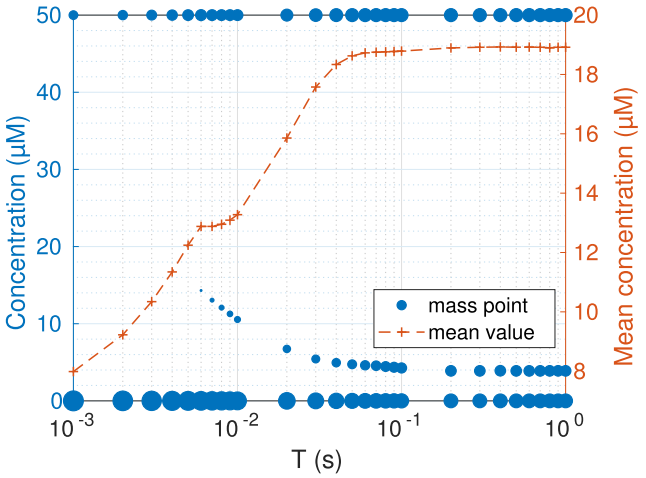

FIGURE 5. Effect of T on the capacity-achieving distribution with $N = 5$.

The effect of the length of a time slot, i.e., T, on the capacity-achieving distribution is also investigated. As is shown in Fig. 5, the IID channel capacity is achieved by a Bernoulli distribution as $T \to 0$ and converges to a special discrete distribution as $T \to 1\,\text{s}$. The expectation of the capacity-achieving input, represented by the orange asterisk against the right y-axis, increases with increasing T. By regarding as a function of T, the turn point of the expectation implies the appearance of a new support point.

B. WITH PEAK AND AVERAGE CONSTRAINTS

For the IID channels with the peak and average constraints, the capacity-achieving input distributions are shown in Fig. 6. The probability of a mass points is represented by the area of the circle. Different average constraints are distinguished by the colors of the circles. It can be seen that c_H may not be included in the optimal support set. The average constraint mainly affects the location of the maximum support point.

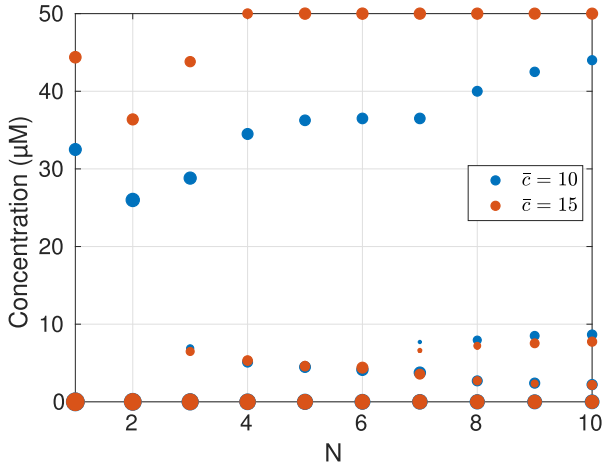

FIGURE 6. Capacity achieving input distribution with the peak and average constraints, where $\bar{c} = 10\,\mu\text{M}$ and $T = 100\,\text{ms}$.

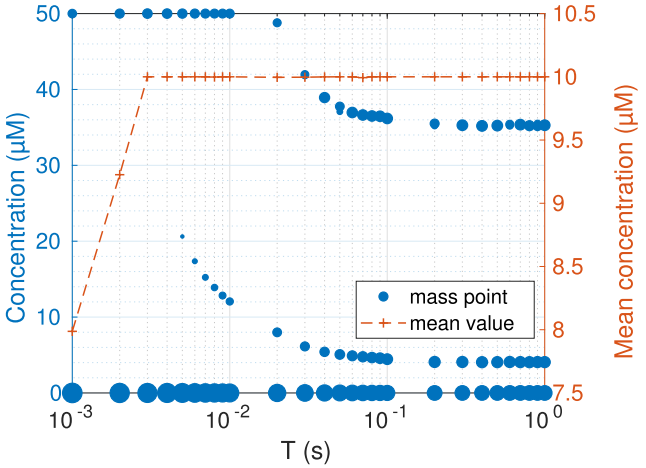

FIGURE 8. Effect of T on the capacity-achieving distribution under peak and average constraints, where $\bar{c} = 10\,\mu\text{M}$ and $N = 5$.

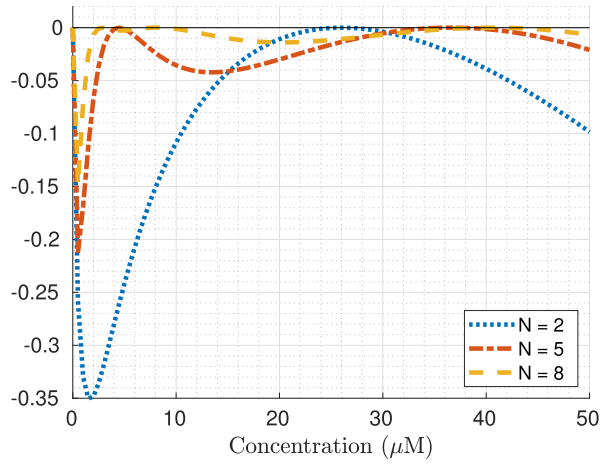

FIGURE 7. The validity of the calculated capacity-achieving input distributions, where $\bar{c} = 10\,\mu\text{M}$ and $T = 100\,\text{ms}$.

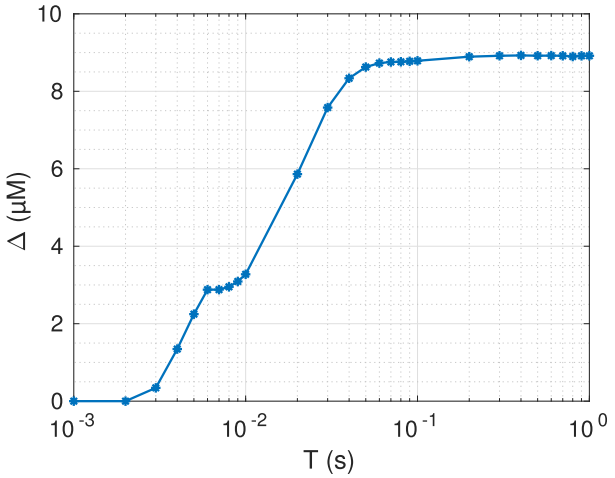

FIGURE 9. Effect of T on Δ, where $\bar{c} = 10\,\mu\text{M}$ and $N = 5$.

Let P_X^* denote the capacity-achieving input distributions of the IID binding channel without average constraint and X^* be the random variable following P_X^*. The effect of the average constraint depends on the gap between $\bar{c}$ and the expectation $\mathbb{E}_{P_X^*}(X^*)$, i.e., $\mathbb{E}_{P_X^*}(X^*) - \bar{c}$. As shown in Fig. 3, $\mathbb{E}_{P_X^*}(X^*)$ (orange asterisks) is greater than 15 μM for $N \in \{1, 2, \ldots, 10\}$. Hence the gap $\mathbb{E}_{P_X^*}(X^*) - \bar{c}$ has the same changing trend as $\mathbb{E}_{P_X^*}(X^*)$ for $N \in \{1, 2, \ldots, 10\}$ when $\bar{c} = 10$ or 15 μM. Combining with Fig. 6, it can be concluded that the value of the maximum support point decreases with increasing the gap $\mathbb{E}_{P_X^*}(X^*) - \bar{c}$.

The validity of the capacity-achieving input distributions shown in Fig. 6 are also evaluated by the criteria proposed in Theorem 3. For the average constraint $\bar{c} = 10\,\mu\text{M}$ and $N = 2, 5, 8$, the criterion functions are depicted in Fig. 7. Note that the criterion function is negative at the ending point c_H, which coincides the results shown in Fig. 6.

For the IID binding channel with peak and average constraints, the effect of T on the capacity-achieving distribution is shown in Fig. 8. When $T \to 0$, the capacity-achieving distribution is a Bernoulli distribution. When $T \geq 3 \times 10^{-3}$ s, the average constraint $\bar{c} = 10\,\mu\text{M}$ is less than $\mathbb{E}_{P_X^*}(X^*)$ shown by the orange asterisks in Fig. 5. Let $\hat{P}_X$ denote the capacity-achieving input distribution for the IID binding channel with the peak constraint $c_\text{L} = 0\,\mu\text{M}$, $c_\text{H} = 50\,\mu\text{M}$ and the average constraint $\bar{c} = 10\,\mu\text{M}$. Hence we have

$$\mathbb{E}_{\hat{P}_X}(\hat{X}) = \begin{cases} \mathbb{E}_{P_X^*}(X^*) & T \leq 2 \times 10^{-3}\,\text{s} \\ \bar{c} & T \geq 3 \times 10^{-3}\,\text{s}, \end{cases} \tag{45}$$

where $\hat{X}$ is the random variable following the distribution $\hat{P}_X$.

Let $\Delta = \mathbb{E}_{P_X^*}(X^*) - \mathbb{E}_{\hat{P}_X}(\hat{X})$. As T increases, the value of Δ becomes larger, which is plotted in Fig. 9 and leads to a remarkable decreasing on the maximum support point in Fig. 8.

C. ACHIEVING RATES

Plots of achieving rates of the IID binding channels are shown in Fig. 10 with fixing $T = 100\,\text{ms}$. It can be seen that

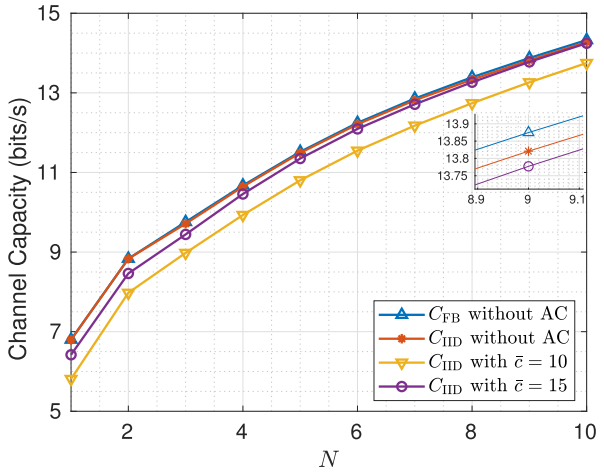

FIGURE 10. The achieving rates of the binding channels for different N, where $T = 100\,\mathrm{ms}$.

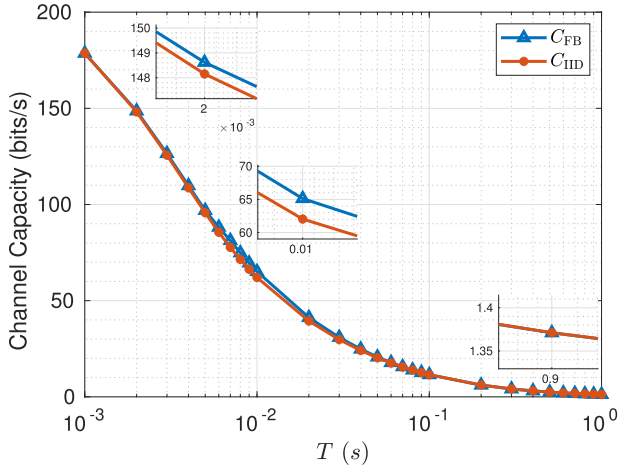

FIGURE 11. The achieving rates of the binding channels for different T, where $N = 5$.

the IID capacity increases with increasing the number of receptors. The effect of the average constraint on the IID channel capacity is related to the gap $\mathbb{E}_{P_X^*}(X^*) - \bar{c}$. The larger the gap, the lower the IID channel capacity.

Moreover, in Fig. 10, the numerical capacity of the binding channel with feedback (feedback capacity), denoted by C_{FB}, is also plotted, which is defined as

$$C_{\mathrm{FB}} = \frac{1}{T} \lim_{k\to\infty} \max_{P_{X_1^k \| Y_1^k}} I(X_1^k \to Y_0^k), \tag{46}$$

where $P_{X_1^k \| Y_0^k} \triangleq \prod_{i=1}^{k} P_{X_i|X_1^{i-1},Y_0^{i-1}}$ is the feedback input distribution, called causal conditioning distribution, and $I(X_1^k \to Y_0^k) \triangleq \sum_{i=1}^{k} I(X_1^k; Y_k \mid Y_0^{k-1})$.

The feedback capacity C_{FB} can be computed using the modified SACPA, where the optimization vector is replaced by $P_{X_1^k \| Y_0^k}$ and the updating strategy is based on the sequential sufficient and necessary conditions proposed in [38].

As an upper bound of the channel capacity C defined in (13), C_{FB} can be used to evaluate the tightness of C_{IID} approximating to the channel capacity C. It can be concluded that C_{IID} provides a tight lower bound of C.

For the binding channel with $N = 5$ receptors, the feedback capacities and IID capacities related to different lengths of each time slot are plotted in Fig. 11. As increasing T, the achieving rate decreases. However, the binding process may consume hundred milliseconds in some special cases, e.g., the aforementioned synaptic transmission mediated by NMDA receptors [29]. When T is long enough (in the experiment is $T \to 1\,\mathrm{s}$), C_{FB} and C_{IID} are almost equal, which implies that the capacity of the binding channel is achieved by an IID distribution in the limiting case.

VII. CONCLUSION

In this work, we investigate the binding channel derived from the ligand-receptor binding process in receiver. The channel model is established by using a finite-state Markov process.

For a binding channel with peak and average constraints, the discreteness of the capacity-achieving distribution for IID binding channels is proved. Based on the characteristics of the channel model, a criterion function is then proposed to verify the optimality of the IID input distributions. Furthermore, a numerical algorithm, called modified steepest ascent cutting-plane algorithm, is proposed to efficiently calculated the capacity-achieving input distribution. The numerical results show the special effect of the average constraint on the IID capacity-achieving input distributions and verifies the tightness of IID capacity to the capacity of binding channels.

Comparing to the previous works, our channel model is more realistic, derived from which the mathematical and numerical results can hence be used in the actual scenarios, e.g., modeling on the molecular-electro signal transduction process in inter-neuronal communication or designing of the bio-cyber interface in the healthcare delivery systems, in molecular communication.

APPENDIX A
CALCULATION OF THE WEAK DERIVATIVE

As is defined, $P_X^\theta = (1-\theta)P_X + \theta P'_X$, which means

$$\mathcal{S}^\theta = \mathcal{S} \bigcup \mathcal{S}', \tag{47}$$

$$\mathbf{p}_X^\theta = (1-\theta)\hat{\mathbf{p}}_X + \theta \hat{\mathbf{p}}'_X, \tag{48}$$

where $\hat{\mathbf{p}}_X$ and $\hat{\mathbf{p}}'_X$ are $|\mathcal{S}^\theta|$ dimensional row vectors expanded with zero entries from the probability vectors $\mathbf{p}_X$ and $\mathbf{p}'_X$, respectively, and $|\mathcal{S}^\theta|$ is the cardinality of the set $\mathcal{S}^\theta$.

According to (25), $P_{Y_i|Y_{i-1}}$ is a linear function of P_X, namely,

$$P^\theta_{Y_i|Y_{i-1}} = (1-\theta)P_{Y_i|Y_{i-1}} + \theta P'_{Y_i|Y_{i-1}}. \tag{49}$$

Thus we have the following expansion.

$$\begin{aligned}
&\mathbf{R}\left(P^\theta_{Y_i|Y_{i-1}}, \mathcal{S}^\theta\right)\left(\mathbf{p}_X^\theta\right)^{\mathsf{T}} \\
&\overset{(a)}{=} (1-\theta)\mathbf{R}\left(P^\theta_{Y_i|Y_{i-1}}, \mathcal{S}^\theta\right)\hat{\mathbf{p}}_X^{\mathsf{T}} + \theta\mathbf{R}\left(P^\theta_{Y_i|Y_{i-1}}, \mathcal{S}^\theta\right)\left(\hat{\mathbf{p}}'_X\right)^{\mathsf{T}} \\
&\overset{(b)}{=} (1-\theta)\mathbf{R}\left(P^\theta_{Y_i|Y_{i-1}}, \mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}} + \theta\mathbf{R}\left(P^\theta_{Y_i|Y_{i-1}}, \mathcal{S}'\right)\left(\mathbf{p}'_X\right)^{\mathsf{T}}
\end{aligned}$$

$$\overset{(c)}{=}(1-\theta)\left[\mathbf{R}\left(P_{Y_i|Y_{i-1}},\mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}+\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P^{\theta}_{Y_i|Y_{i-1}}\right)\right]+\theta\mathbf{R}\left(P^{\theta}_{Y_i|Y_{i-1}},\mathcal{S}'\right)\left(\mathbf{p}'_X\right)^{\mathsf{T}}, \tag{50}$$

where (a) is based on (48); (b) is established since P_X is only positive on its support set $\mathcal{S}$, so is P'_X; In (c), $\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P^{\theta}_{Y_i|Y_{i-1}}\right)\triangleq\left(d_0(\theta),\ldots,d_N(\theta)\right)^{\mathsf{T}}$ is a column vector consisting of relative entropies.

$$d_{y_{i-1}}(\theta)=\sum_{y_i=0}^{N}P(y_i|y_{i-1})\log\frac{P(y_i|y_{i-1})}{P^{\theta}(y_i|y_{i-1})},\quad y_{i-1}\in\mathcal{N}. \tag{51}$$

Lemma 2: According to the definition of $d_{y_{i-1}}(\theta)$, the following limitations are established.

1) $\lim_{\theta\downarrow 0}\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P^{\theta}_{Y_i|Y_{i-1}}\right)=\mathbf{0}$;
2) $\lim_{\theta\downarrow 0}\frac{1}{\theta}\boldsymbol{\pi}\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P^{\theta}_{Y_i|Y_{i-1}}\right)=0$.

Proof 6:

1) Let $\theta=0$, then

$$d_{y_{i-1}}(\theta)=\sum_{y_i=0}^{N}P(y_i|y_{i-1})\log\frac{P(y_i|y_{i-1})}{P(y_i|y_{i-1})}=0 \tag{52}$$

2) The expression in the limitation can be expanded as

$$\begin{aligned}&\boldsymbol{\pi}\mathbf{D}\left(P_{Y_i|Y_{i-1}}\middle\|P^{\theta}_{Y_i|Y_{i-1}}\right)\\&=\sum_{y_{i-1}=0}^{N}\pi(y_{i-1})\sum_{y_i=0}^{N}P(y_i|y_{i-1})\log\frac{P(y_i|y_{i-1})}{P^{\theta}(y_i|y_{i-1})}\\&=\sum_{y_{i-1}=0}^{N}\pi(y_{i-1})\sum_{y_i=0}^{N}P(y_i|y_{i-1})\\&\quad\cdot\log\left(1+\frac{\theta\left(P(y_i|y_{i-1})-P'(y_i|y_{i-1})\right)}{P^{\theta}(y_i|y_{i-1})}\right)\\&=\sum_{y_{i-1}=0}^{N}\pi(y_{i-1})\sum_{y_i=0}^{N}P(y_i|y_{i-1})\\&\quad\cdot\left(\frac{\theta\left(P(y_i|y_{i-1})-P'(y_i|y_{i-1})\right)}{P^{\theta}(y_i|y_{i-1})}+o(\theta)\right)\\&=\theta\,\mathbb{E}_{P_{Y_{i-1},Y_i}}\left(\frac{P(y_i|y_{i-1})-P'(y_i|y_{i-1})}{P^{\theta}(y_i|y_{i-1})}\right)+o(\theta),\end{aligned} \tag{53}$$

where $o(\theta)$ is the higher-order infinitesimal referring to θ as $\theta\to 0$.

The expectation in the first item in (53) is zero when $\theta=0$, which leads to the second equation in Lemma (2).

Since π (π^{θ}) is the unique stationary distribution corresponding to $P_{Y_i|Y_{i-1}}$ $\left(P^{\theta}_{Y_i|Y_{i-1}}\right)$, it implies that the following equations have a unique solution.

$$\begin{cases}\boldsymbol{\mu}\mathbf{P}_{Y_i|Y_{i-1}}=\boldsymbol{\mu}\\ \boldsymbol{\mu}\mathbf{1}^{\mathsf{T}}=1,\end{cases} \tag{54}$$

where $\boldsymbol{\mu}\in\{\boldsymbol{\pi},\boldsymbol{\pi}^{\theta}\}$, $\mathbf{1}=(1,1,\ldots,1)$ is a $(N+1)$ dimensional row vector.

The equations in (54) can be reorganized as

$$\begin{cases}\boldsymbol{\pi}^{\theta}\mathbf{P}^{\theta}_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}=\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}\\ \left(\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}\right)\mathbf{1}^{\mathsf{T}}=0\end{cases} \tag{55}$$

Taking (49) into the first equation in (55), we have

$$(1-\theta)\boldsymbol{\pi}^{\theta}\mathbf{P}_{Y_i|Y_{i-1}}+\theta\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}=\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}. \tag{56}$$

Let $\mathbf{I}$ denote the $(N+1)$ dimensional identical matrix. The equations in (56) can be expressed as

$$\begin{cases}\left(\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}\right)\boldsymbol{\Psi}_{\theta}=\theta\left(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\right)\\ \left(\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}\right)\mathbf{1}^{\mathsf{T}}=0.\end{cases} \tag{57}$$

where $\boldsymbol{\Psi}_{\theta}\triangleq\left(\mathbf{I}-(1-\theta)\mathbf{P}_{Y_i|Y_{i-1}}\right)$ is singular as $\theta\downarrow 0$.

Based on the fact that (54) has a unique solution, the expanded matrix $[\boldsymbol{\Psi}_{\theta},\mathbf{1}^{\mathsf{T}}]$ has rank $(N+1)$.

By adding $\mathbf{1}^{\mathsf{T}}$ to the first column of $\boldsymbol{\Psi}_{\theta}$, the second equation in (57) can be embedded in the first one and lead to the following equation.

$$\left(\boldsymbol{\pi}^{\theta}-\boldsymbol{\pi}\right)\widetilde{\boldsymbol{\Psi}}=\theta\left(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\right), \tag{58}$$

where $\widetilde{\boldsymbol{\Psi}}_{\theta}\triangleq\mathbf{W}-(1-\theta)\mathbf{P}_{Y_i|Y_{i-1}}$ with $\mathbf{W}$ defined as

$$\mathbf{W}=\begin{bmatrix}2&0&0&\ldots&0\\1&1&0&\ldots&0\\1&0&1&\ldots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&0&0&\ldots&1\end{bmatrix}. \tag{59}$$

Hence we have

$$\boldsymbol{\pi}^{\theta}=\boldsymbol{\pi}+\theta\left(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\right)\widetilde{\boldsymbol{\Psi}}_{\theta}^{-1}. \tag{60}$$

Remark 4: It can be verified that the eigenvalues of $\mathbf{P}_{Y_i|Y_{i-1}}$ is a convex combination of the eigenvalues of $\mathbf{P}_{Y_i|Y_{i-1},X_i=x_i}$, in same order, with the coefficient $P_X(x_i)$ for $x_i\in\mathcal{S}$. According to (5) and (6), eigenvalues of $\mathbf{P}_{Y_i|Y_{i-1}}$ are no greater than 1, and only one of them equals 1. Therefore, the eigenvalues of the matrix $\left(\mathbf{W}-(1-\theta)\mathbf{P}_{Y_i|Y_{i-1}}\right)$ are all positive, i.e., it is a nonsingular matrix and hence invertible.

Taking (50) and (60) into the definition of weak derivative in (26), one can obtain the expression shown in (61), as shown at the top of the next page.

Based on Lemma 2, it can be deduced that

$$\begin{aligned}I'_{\infty}(P'_X;P_X)&=\pi\left(\mathbf{P}'_{Y_i|Y_{i-1}}-\mathbf{P}_{Y_i|Y_{i-1}}\right)\widetilde{\boldsymbol{\Psi}}_0^{-1}\mathbf{R}\left(P_{Y_i|Y_{i-1}},\mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}\\&\quad+\pi\mathbf{R}\left(P_{Y_i|Y_{i-1}},\mathcal{S}'\right)\left(\mathbf{p}'_X\right)^{\mathsf{T}}-\pi\mathbf{R}\left(P_{Y_i|Y_{i-1}}\right)\mathbf{p}_X^{\mathsf{T}}\\&=\pi\left(\mathbf{P}'_{Y_i|Y_{i-1}}-\mathbf{W}\right)\widetilde{\boldsymbol{\Psi}}_0^{-1}\mathbf{R}\left(P_{Y_i|Y_{i-1}},\mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}\\&\quad+\pi\mathbf{R}\left(P_{Y_i|Y_{i-1}},\mathcal{S}'\right)\left(\mathbf{p}'_X\right)^{\mathsf{T}}.\end{aligned} \tag{62}$$

$$
\begin{aligned}
I'_{\infty}(P'_X;P_X)=&\lim_{\theta\downarrow 0}\frac{1}{\theta}\Bigg\{(1-\theta)\Big[\boldsymbol{\pi}+\theta\Big(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\Big)\widetilde{\boldsymbol{\Psi}}_{\theta}^{-1}\Big]\Big[\mathbf{R}\big(P_{Y_i|Y_{i-1}},\mathcal{S}\big)\mathbf{p}_X^{\mathsf{T}}+\mathbf{D}\Big(P_{Y_i|Y_{i-1}}\Big\|P^{\theta}_{Y_i|Y_{i-1}}\Big)\Big]\\
&+\theta\Big[\boldsymbol{\pi}+\theta\Big(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\Big)\widetilde{\boldsymbol{\Psi}}_{\theta}^{-1}\Big]\mathbf{R}\Big(P^{\theta}_{Y_i|Y_{i-1}},\mathcal{S}'\Big)\big(\mathbf{p}'_X\big)^{\mathsf{T}}-\boldsymbol{\pi}\mathbf{R}\big(P_{Y_i|Y_{i-1}}\big)\mathbf{p}_X^{\mathsf{T}}\Bigg\}\\
=&\lim_{\theta\downarrow 0}\frac{(1-\theta)\boldsymbol{\pi}\mathbf{D}\Big(P_{Y_i|Y_{i-1}}\Big\|P^{\theta}_{Y_i|Y_{i-1}}\Big)}{\theta}+\lim_{\theta\downarrow 0}\Bigg\{(1-\theta)\Big(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\Big)\widetilde{\boldsymbol{\Psi}}_{\theta}^{-1}\Big[\mathbf{R}\big(P_{Y_i|Y_{i-1}},\mathcal{S}\big)\mathbf{p}_X^{\mathsf{T}}\\
&+\mathbf{D}\Big(P_{Y_i|Y_{i-1}}\Big\|P^{\theta}_{Y_i|Y_{i-1}}\Big)\Big]+\Big[\boldsymbol{\pi}+\theta\Big(\boldsymbol{\pi}^{\theta}\mathbf{P}'_{Y_i|Y_{i-1}}-\boldsymbol{\pi}\mathbf{P}_{Y_i|Y_{i-1}}\Big)\widetilde{\boldsymbol{\Psi}}_{\theta}^{-1}\Big]\mathbf{R}\Big(P^{\theta}_{Y_i|Y_{i-1}},\mathcal{S}'\Big)\big(\mathbf{p}'_X\big)^{\mathsf{T}}\Bigg\}\\
&-\boldsymbol{\pi}\mathbf{R}\big(P_{Y_i|Y_{i-1}}\big)\mathbf{p}_X^{\mathsf{T}} \qquad (61)
\end{aligned}
$$

APPENDIX B
PROOF OF THE SUFFICIENT AND NECESSARY CONDITIONS

Define the Lagrangian

$$L(P_X)=I_{\infty}(P_X)-\vartheta\left(\mathbb{E}_{P_X}(X)-\bar{c}\right), \qquad (63)$$

where $\vartheta\geq 0$ is the Lagrangian multiplier.

By the method of Lagrangian multipliers [37], P_X^* is optimal if and only if P_X^* and ϑ are such that

1) $\vartheta\left(\mathbb{E}_{P_X^*}(X^*)-\bar{c}\right)=0$;
2) $L(P_X^*)\geq L(P_X)$ for all $P_X\in\mathcal{P}(c_{\mathrm{L}},c_{\mathrm{H}},\bar{c})$.

Since I_{∞} is a concave function on the convex set $\mathcal{P}(c_{\mathrm{L}},c_{\mathrm{H}},\bar{c})$ [31], L is also a concave function and the maximizer of (19) must exist.

For any legal input distribution P_X, the weak derivative of L at P_X^* is calculated as

$$
\begin{aligned}
L'(P_X;P_X^*)=&\lim_{\theta\downarrow 0}\frac{L\big((1-\theta)P_X^*+\theta P_X\big)-L(P_X^*)}{\theta}\\
=&I'_{\infty}(P_X;P_X^*)-\vartheta\left(\mathbb{E}_{P_X}(X)-\mathbb{E}_{P_X^*}(X^*)\right). \qquad (64)
\end{aligned}
$$

Taking (27) into (64), then

$$
\begin{aligned}
L'(P_X;P_X^*)=&\boldsymbol{\pi}^*\big(\mathbf{P}_{Y_i|Y_{i-1}}-\mathbf{W}\big)\left(\widetilde{\boldsymbol{\Psi}}_0^*\right)^{-1}\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}\\
&+\boldsymbol{\pi}^*\mathbf{R}\left(P^*_{Y_i|Y_{i-1}},\mathcal{S}\right)\mathbf{p}_X^{\mathsf{T}}\\
&-\vartheta\left(\mathbb{E}_{P_X}(X)-\mathbb{E}_{P_X^*}(X^*)\right), \qquad (65)
\end{aligned}
$$

where $\widetilde{\boldsymbol{\Psi}}_0^*=\mathbf{W}-\mathbf{P}^*_{Y_i|Y_{i-1}}$, $\mathbf{R}^*$ is the shorthand of $\mathbf{R}\left(P^*_{Y_i|Y_{i-1}},\mathcal{S}^*\right)$.

Generally, it is assumed that $\bar{c}<c_{\mathrm{H}}$ since $\bar{c}=c_{\mathrm{H}}$ is a special case that can be proved in the same way by setting $\vartheta=0$.

A. NECESSITY

Suppose P_X^* is the capacity-achieving distribution. According to the condition 2) listed at the beginning of this section, the weak derivative $L'(P_X;P_X^*)$ should be non-positive for any $P_X\in\mathcal{P}$, namely,

$$L'(P_X;P_X^*)\leq 0. \qquad (66)$$

Assume that δ_c is a Dirac measure defined as

$$\delta_c(x)=\begin{cases}1 & x=c\\ 0 & x\neq c.\end{cases} \qquad (67)$$

For $c\in[c_{\mathrm{L}},\bar{c}]$, the distribution δ_c is certainly an element of the feasible set $\mathcal{P}$. Let $\mathbf{Q}_c\triangleq\mathbf{P}(T;c)=\mathbf{P}_{Y_i|Y_{i-1},X_i=c}$. Based on the fact that $\mathbb{E}_{\delta_c}(X)=c$, we have

$$
\begin{aligned}
L'(\delta_c;P_X^*)=&\boldsymbol{\pi}^*\left(\mathbf{Q}_c-\mathbf{W}\right)\left(\widetilde{\boldsymbol{\Psi}}_0^*\right)^{-1}\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}+\boldsymbol{\pi}^*\mathbf{R}\left(P^*_{Y_i|Y_{i-1}},c\right)\\
&-\vartheta\left(\mathbb{E}_{\delta_c}(X)-\mathbb{E}_{P_X^*}(X^*)\right),\\
=&\boldsymbol{\pi}^*\left(\mathbf{Q}_c-\mathbf{W}\right)\left(\widetilde{\boldsymbol{\Psi}}_0^*\right)^{-1}\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}+\boldsymbol{\pi}^*\mathbf{R}\left(P^*_{Y_i|Y_{i-1}},c\right)\\
&-\vartheta\left(c-\bar{c}\right)\leq 0, \qquad (68)
\end{aligned}
$$

where the last equality is derived from condition 1).

For $c_{\mathrm{L}}\leq s<\bar{c}\leq c\leq c_{\mathrm{H}}$, define a Bernoulli distribution B_{c_1,c_2} as

$$B_{s,c}=(1-\theta)\delta_s+\theta\delta_c,\ 0<\theta\leq\frac{\bar{c}-s}{c-s}. \qquad (69)$$

Note that $L'(P_X;P_X^*)$ is a linear function of P_X as P_X^* is given. Then the weak derivative $L'(B_{s,c};P_X^*)$ can be regarded as a convex combination which is

$$L'(B_{s,c};P_X^*)=(1-\theta)L'(\delta_s;P_X^*)+\theta L'(\delta_c;P_X^*)\leq 0. \qquad (70)$$

Without loss of generality, assume s is the solution to $L'(\delta_s;P_X^*)=0$ for $s\in[c_{\mathrm{L}},\bar{c})$. Then

$$L'(B_{s,c})=\theta L'(\delta_c;P_X^*)\leq 0,\quad c\in[\bar{c},c_{\mathrm{H}}]. \qquad (71)$$

Since $\theta>0$, for $c\in[c_{\mathrm{L}},c_{\mathrm{H}}]$, we have

$$
\begin{aligned}
L'(\delta_c;P_X^*)=&\boldsymbol{\pi}^*\left(\mathbf{Q}_c-\mathbf{W}\right)\left(\widetilde{\boldsymbol{\Psi}}_0^*\right)^{-1}\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}+\boldsymbol{\pi}^*\mathbf{R}\left(P^*_{Y_i|Y_{i-1}},c\right)\\
&-\vartheta\left(c-\bar{c}\right)\leq 0. \qquad (72)
\end{aligned}
$$

The weighted sum of $I(\delta_c;P_X^*)$ with the coefficient $P_X^*(c)$ for $c\in\mathcal{S}^*$ is

$$
\begin{aligned}
&\sum_{c\in\mathcal{S}^*}P_X^*(c)L'(\delta_c;P_X^*)\\
&=\boldsymbol{\pi}^*\left(\mathbf{P}^*_{Y_i|Y_{i-1}}-\mathbf{W}\right)\left(\widetilde{\boldsymbol{\Psi}}_0^*\right)^{-1}\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}+\boldsymbol{\pi}^*\mathbf{R}^*\left(\mathbf{p}_X^*\right)^{\mathsf{T}}
\end{aligned}
$$

$$
\begin{aligned}
&- \vartheta \left(\sum_{c \in \mathcal{S}^*} P_X^*(c)c - \bar{c} \right) \\
&\overset{(a)}{=} -\boldsymbol{\pi}^* \mathbf{R}^* \left(\mathbf{p}_X^* \right)^{\mathsf{T}} + \boldsymbol{\pi}^* \mathbf{R}^* \left(\mathbf{p}_X^* \right)^{\mathsf{T}} \\
&= 0, \qquad (73)
\end{aligned}
$$

where (a) is established since $\mathbf{P}^*_{Y_i|Y_{i-1}} - \mathbf{W} = -\widetilde{\boldsymbol{\Psi}}_0^*$ and $\sum_{c \in \mathcal{S}^*} P_X^*(c)c = \mathbb{E}_{P_X^*}(X^*) = \bar{c}$.

Based on the non-positivity of $L'(\delta_c; P_X^*)$, it is proved that

$$L'(\delta_c; P_X^*) = 0, \ c \in \mathcal{S}^*. \qquad (74)$$

B. SUFFICIENCY

Suppose the inequality in (28) is satisfied, i.e.,

$$I(\delta_c; P_X^*) \leq 0, \ c \in [c_{\mathrm{L}}, c_{\mathrm{H}}]. \qquad (75)$$

Suppose $\mathcal{S}^* = \left\{ s_1^*, \ldots, s_M^* \right\}$. With a slight abuse of notation, let C^* be a random variable following the distribution P_X^*. One can calculate the expectation of $L'(\delta_{C^*}; P_X^*)$ as

$$
\begin{aligned}
\mathbb{E}_{P_X^*} \left(L'(\delta_{C^*}; P_X^*) \right) &= \sum_{i=1}^{M} P_X^*(s_i^*) L'(\delta_{s_i^*}; P_X^*) \\
&= \boldsymbol{\pi}^* \left(\mathbf{P}^*_{Y_i|Y_{i-1}} - \mathbf{W} \right) \left(\widetilde{\boldsymbol{\Psi}}_0^* \right)^{-1} \mathbf{R}^* \left(\mathbf{p}_X^* \right)^{\mathsf{T}} \\
&\quad + \boldsymbol{\pi}^* \mathbf{R}^* \left(\mathbf{p}_X^* \right)^{\mathsf{T}} - \vartheta \left(\mathbb{E}_{P_X^*}(C^*) - \bar{c} \right) \\
&= -\vartheta \left(\mathbb{E}_{P_X^*}(C^*) - \bar{c} \right) \\
&\geq 0, \qquad (76)
\end{aligned}
$$

where the last inequality holds because $P_X^* \in \mathcal{P}$.

Combining (75) and (76), it is proved that

$$\vartheta \left(\mathbb{E}_{P_X^*}(X^*) - \bar{c} \right) = \vartheta \left(\mathbb{E}_{P_X^*}(C^*) - \bar{c} \right) = 0, \qquad (77)$$

which leads to the condition 1) listed at the beginning of Appendix B.

Suppose P_X is an arbitrary distribution in $\mathcal{P}$, and the corresponding support set is denoted by $\mathcal{S} = \{s_1, \ldots, s_m\}$. Let $\mathbf{p}_X = (p_1, \ldots, p_2)$ be the probability vector of P_X, where $p_j = P_X(s_j)$, $j = 1, \ldots, m$. Thus, we have

$$P_X = \sum_{j=1}^{m} p_j \delta_{s_j}. \qquad (78)$$

By linearity of $L'(P_X; P_X^*)$, it can be deduced that

$$L'(P_X; P_X^*) = \sum_{j=1}^{m} p_j L'(\delta_{s_j}; P_X^*) \leq 0, \qquad (79)$$

which leads to the condition 2).

In conclusion, as (28) is satisfied, condition 1) and 2) are both established. Hence P_X^* achieves the capacity.

REFERENCES

[1] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication—A tutorial review," 2018, *arXiv:1812.05492*. [Online]. Available: https://arxiv.org/abs/1812.05492

[2] M. Kuscu and O. B. Akan, "On the physical design of molecular communication receiver based on nanoscale biosensors," *IEEE Sensors J.*, vol. 16, no. 8, pp. 2228–2243, Sep. 2016.

[3] V. Singh and I. Nemenman, "Simple biochemical networks allow accurate sensing of multiple ligands with a single receptor," *PLoS Comput. Biol.*, vol. 13, no. 4, 2017, Art. no. e1005490.

[4] A. Ahmadzadeh, H. Arjmandi, A. Burkovski, and R. Schober, "Comprehensive reactive receiver modeling for diffusive molecular communication systems: Reversible binding, molecule degradation, and finite number of receptors," *IEEE Trans. Nanobiosci.*, vol. 15, no. 7, pp. 713–727, Oct. 2016.

[5] M. M. Al-Zu'bi and A. S. Mohan, "Modeling of ligand-receptor protein interaction in biodegradable spherical bounded biological microenvironments," *IEEE Access*, vol. 6, pp. 25007–25018, 2018.

[6] M. Pierobon and I. F. Akyildiz, "Noise analysis in ligand-binding reception for molecular communication in nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4168–4182, Sep. 2011.

[7] M. Kuscu and O. B. Akan, "Maximum likelihood detection with ligand receptors for diffusion-based molecular communications in Internet of bio-nano things," *IEEE Trans. Nanobiosci.*, vol. 17, no. 1, pp. 44–54, Jan. 2018.

[8] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, "Capacity of discrete molecular diffusion channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul./Aug. 2011, pp. 723–727.

[9] M. Tahmasbi and F. Fekri, "On the capacity achieving probability measures for molecular receivers," in *Proc. IEEE Inf. Theory Workshop-Fall (ITW)*, Oct. 2015, pp. 109–113.

[10] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7358–7382, Dec. 2016.

[11] P. J. Thomas and A. W. Eckford, "Shannon capacity of signal transduction for multiple independent receptors," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1804–1808.

[12] A. W. Eckford and P. J. Thomas, "The channel capacity of channelrhodopsin and other intensity-driven signal transduction receptors," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 1, pp. 27–38, Mar. 2018.

[13] H. G. Bafghi, A. Gohari, M. Mirmohseni, G. Aminian, and M. Nasiri-Kenari, "Diffusion-based molecular communication with limited molecule production rate," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 2, pp. 61–72, Jun. 2018.

[14] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.

[15] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[16] P. O. Vontobel, A. Kavcic, D. M. Arnold, and H.-A. Loeliger, "A generalization of the Blahut–Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[17] I. Naiss and H. H. Permuter, "Extension of the Blahut–Arimoto algorithm for maximizing directed information," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 204–222, Jan. 2013.

[18] R. D. Wesel, E. E. Wesel, L. Vandenberghe, C. Komninakis, and M. Medard, "Efficient binomial channel capacity computation with an application to molecular communication," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–5.

[19] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience*, vol. 2. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2007.

[20] K.-W. Yau and R. C. Hardie, "Phototransduction motifs and variations," *Cell*, vol. 139, no. 2, pp. 246–264, 2009.

[21] J. Banchereau, F. Briere, C. Caux, J. Davoust, S. Lebecque, Y.-J. Liu, B. Pulendran, and K. Palucka, "Immunobiology of dendritic cells," *Annu. Rev. Immunol.*, vol. 18, no. 1, pp. 767–811, 2000.

[22] E. T. Roussos, J. S. Condeelis, and A. Patsialou, "Chemotaxis in cancer," *Nature Rev. Cancer*, vol. 11, no. 8, pp. 573–587, 2011.

[23] U. A. K. Chude-Okonkwo, R. Malekian, and B. T. Maharaj, "Biologically inspired bio-cyber interface architecture and model for Internet of bio-nanothings applications," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3444–3455, Aug. 2016.

[24] U. A. K. Chude-Okonkwo, R. Malekian, B. T. Maharaj, and A. V. Vasilakos, "Molecular communication and nanonetwork for targeted drug delivery: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 3046–3096, 4th Quart., 2017.
[25] S. M. J. L. van den Wildenberg, B. Prevo, and E. J. G. Peterman, "A brief introduction to single-molecule fluorescence methods," in *Single Molecule Analysis: Methods and Protocols*, E. J. G. Peterman and G. J. L. Wuite, Eds. Totowa, NJ, USA: Humana Press, 2011, pp. 81–99.
[26] C. Eleftheriou, F. Cesca, L. Maragliano, F. Benfenati, and J. F. Maya-Vetencourt, "Optogenetic modulation of intracellular signalling and transcription: Focus on neuronal plasticity," *J. Exp. Neurosci.*, vol. 11, pp. 1–16, Apr. 2017.
[27] L. Khaloopour, M. Mirmohseni, and M. Nasiri-Kenari, "An adaptive pulse-width modulation for limited molecule production and storage," in *Proc. Iran Workshop Commun. Inf. Theory (IWCIT)*, Apr. 2018, pp. 1–6.
[28] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.
[29] A. J. Gibb, "Glutamate unbinding reveals new insights into NMDA receptor activation," *J. Physiol.*, vol. 574, p. 329, Jul. 2006.
[30] N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models," *Nature Commun.*, vol. 6, Jul. 2015, Art. no. 7653.
[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[32] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 265–271, May 1980.
[33] J. E. Martínez-Legaz, "On Weierstrass extreme value theorem," *Optim. Lett.*, vol. 8, no. 1, pp. 391–393, 2014.
[34] H. Li, S. M. Moser, and D. Guo, "Capacity of the memoryless additive inverse Gaussian noise channel," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 12, pp. 2315–2329, Dec. 2014.
[35] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336–2351, Jul. 2005.
[36] J. D. Clements, A. Feltz, Y. Sahara, and G. L. Westbrook, "Activation kinetics of AMPA receptor channels reveal the number of functional agonist binding sites," *J. Neurosci.*, vol. 18, no. 1, pp. 119–127, 1998.
[37] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ, USA: Wiley, 1997.
[38] P. A. Stavrou, C. D. Charalambous, and C. K. Kourtellaris, "Sequential necessary and sufficient conditions for capacity achieving distributions of channels with memory and feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7095–7115, Nov. 2017.

JIANFENG SUN received the B.Eng. degree from Northeastern University at Qinhuangdao, Qinhuangdao, China, in 2011. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His research interests include channel modeling and channel capacity analysis in molecular communication.

HUI LI received the B.Eng. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 1999 and 2004, respectively. He joined the Faculty of the University of Science and Technology of China, Hefei, in 2004, where he is currently an Associate Professor with the Department of Electronic Engineering and Information Science. He has held visiting positions with Northwestern University, Evanston, IL, USA, in 2011 and 2012, respectively. His research interests include wireless communication and information theory.