Optimizing Information Transfer Through Chemical Channels in Molecular Communication

Francesca Ratti*†, Colton Harper*§, Maurizio Magarini†, and Massimiliano Pierobon§
†Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, I-20133 Milano, Italy
§Department of Computer Science and Engineering
University of Nebraska-Lincoln, Lincoln, Nebraska 68588 USA
Email: francesca.ratti@polimi.it, colton.harper@huskers.unl.edu, maurizio.magarini@polimi.it, pierobon@cse.unl.edu

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685247

***Abstract*—The optimization of information transfer through molecule diffusion and chemical reactions is one of the leading research directions in Molecular Communication (MC) theory. The highly nonlinear nature of the processes underlying these channels poses challenges in adopting analytical approaches for their information-theoretic modeling and analysis. In this paper, a novel iterative methodology is proposed to numerically estimate achievable information rates. Based on the Nelder-Mead optimization, this methodology does not necessitate analytical formulations of MC components and their stochastic behavior, and, when applied to well-known scenarios, it demonstrates consistent results with theoretical bounds and superior performance to prior literature. A numerical example that abstracts communications between genetically engineered cells via simulation is presented and discussed in light of possible future applications to support the design and engineering of realistic MC systems.**

***Index Terms*—Molecular Communication, Chemical Reaction Channel, Diffusion Channel, Iterative Algorithm, Mutual Information, Achievable Information Rate**

I. Introduction

Molecular Communication (MC) is a field of research that focuses on the study of information propagation through molecules and chemical reactions, including natural communications in biology, from the perspectives of information theory and communication engineering [1], [2]. Due to their system-evolution-dependent stochastic nature, most of the chemical and biological MC channels exhibit highly nonlinear input-output behavior [3]. Analytical formulas to estimate MC channel capacity can be found only in very specific cases or under strong assumptions, and because of the complex characteristics of such channels, most realistic chemical and biological MC channels do not yet have even a full statistical characterization in current literature [1]. For this reason, unlike quintessential communication channels, defining the capacity of MC channels is non-trivial.

A number of papers in the literature try to address the issue of finding ideal conditions to transfer reliable information in MC systems. The authors of [4] build an MC system where the channel is approximated as slotted binary and evaluate the achievable capacity. In [5], an information-theoretical approach for estimating the capacity of a molecular channel between two nanomachines is developed. The impact of relay/cooperative nanomachines on capacity for a diffusive mobile MC system is studied in [6]. The authors of [7] derive a closed-form expression for the capacity of a noisy MC channel, while in [8] enzymatic reaction cycles are exploited to improve the upper bound of Mutual Information (MI) for a diffusion-based communication system. In [9], an optimization of capacity bounds is performed for stationary and ergodic discrete-time channels with memory. In [10], the authors study the performance of the MC system in terms of reliable information exchange for the Poisson channel with finite-state memory. Another paper [11] presents the evaluation of the channel capacity for an MC system model that considers both the diffusion-based channel and the ligand-based receiver.

In MC the practical transmission of information is realized by means of different modulation schemes according to the physical/chemical property of the molecules that is being changed, *i.e.*, concentration, type, timing, and space [12]. The most common form is concentration-based modulation, where the information is encoded in the number of particles released at the transmitter. For this type of modulation, a constraint on the maximum number of released particles arises in a natural way. Under this constraint, from information theory we know that the capacity-achieving input distribution is discrete, and can be efficiently obtained by implementing a numerical approach based on the Blahut-Arimoto (BA) algorithm [13]. In the context of a biochemical scenario, a variant of the BA has been proposed in [14] with application to cellular signaling. The latter is based on an a-priori estimation of the equivocation probability from the data, where analytical expression of the probability mass function is obtained using logistic regression.

As in [14], in the present work we develop a methodology to address the problem of estimating the maximum reliable information (*i.e.*, the MI) that can be transmitted over a chemical MC channel *with unknown statistical model*. This consists of a numerical approach where a relatively small amount of data (simulated or experimental) is exploited to build the estimate of the probability distributions of the transmitted and received messages, which in turn are utilized to estimate the MI. Our methodology is validated in this paper for the case of a discrete-time amplitude-constrained Additive White Gaussian Noise (AWGN) channel, and compared to known results from the literature [15]. We then employ the proposed methodology to maximize the MI in a simulation-based scenario abstracting an MC channel between two genetically engineered cells [3].

Our method shows that the same channel could achieve better performance in terms of information transfer by choosing a particular input distribution, and quantifies this achievable performance in terms of exchanged bits of information.

This paper is organized as follows. In Sec. II we detail the model of an MC channel with chemical reactions and diffusion, which abstracts information propagation in biology, and we motivate and detail our proposed methodology based on the Nelder-Mead iterative algorithm, then validated for the case of AWGN channel and compared to existing literature. In Sec. III we present and discuss the results of our algorithm applied to a particular simulation-based case study of chemical MC channel. Finally, in Sec. IV we conclude the paper.

II. System Model and Methodology

In this section, we detail the proposed iterative algorithm to optimize reliable information transfer through chemical channels in an MC system, with reference to Fig. 1.

A. A Chemical Molecular Communication Channel

In the scope of this paper, we consider a molecular channel based on chemical reactions and free diffusion in a fluid environment, which abstracts the basics of molecular information propagation through/among biological cells [16]. In particular, the molecules of reacting chemical species are confined into multiple volumes (cells), *e.g.*, Ω_1 and Ω_2 in Fig. 1, and some of these species can cross the boundaries (cell membrane) of these volumes and diffuse between them. Inside each of these volumes, we assume that the molecules are well-stirred, resulting in no bulk diffusion of any of the chemical species. Without loss of generality, we assume that the **Transmitter** is emitting into the first volume of the channel a number $N_{Tx}(t)$ of information-bearing molecules as a function of the time t that encodes a Transmitted information Message X. Once emitted, the molecules are in general involved into chains of chemical **Reactions** (cell pathways), each propagating the information message X from a number of reactant species to a number of product species, eventually crossing the volume boundaries and **Diffusing** into other volumes, or being detected at the **Receiver**, which decodes the Received information Message Y from the number $N_{Rx}(t)$ of the molecules as function of the time t.

In agreement with [17], for a time interval τ sufficiently small as to consider that chemical reactions are statistically independent from each other, a chemical **Reaction** is mathematically expressed in this paper as follows:

$$N_i(t+\tau) = N_i(t) + \sum_{j=1}^{M} \nu_{ji} \mathcal{P}_j \left(a_j(\mathbf{N}(t)), \tau\right) , \qquad (1)$$

where $\mathcal{P}_j\left(a_j(\mathbf{N}(t)), \tau\right)$ is a Poisson random variable with average value $a_j(\mathbf{N}(t))\tau$, and ν_{ji} is the stoichiometric coefficient equal to the changes in the number of molecules of type i that the chemical reaction j operates when it occurs. The parameter $a_j(\mathbf{N}(t))$ is called propensity function of the chemical reaction j, and it corresponds to the probability that the reaction j

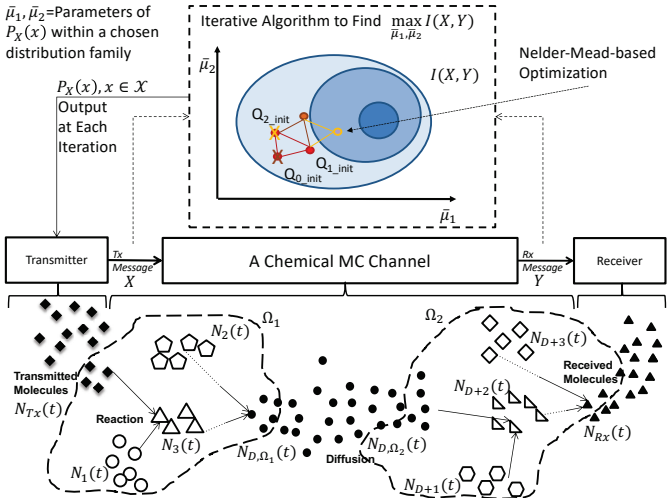


Fig. 1: Block scheme of the system model and methodology proposed in this paper.

occurs in an infinitesimal time interval after time t, given the values in $\mathbf{N}(t)$, *i.e.*, the set of all the numbers of molecules for each type. The propensity function $a_j(\mathbf{N}(t))$ for a chemical reaction j of the type considered in this paper is computed as follows [17]:

$$a_j(\mathbf{N}(t)) = \Omega k_j \prod_{N_i \in \mathcal{R}_j} N_i(t) \,, \qquad (2)$$

where $\mathcal{R}_j$ is the set of reactant species for the chemical reaction j, k_j is the reaction rate, and Ω is the volume that contains these reactants species, Ω_1 or Ω_2 in Fig. 1.

The propagation of the molecules between the volumes through diffusion can be modeled at different precision and scale [1]. For simplicity, and in agreement with [3], in this paper the **Diffusion** is modeled through the probability of capture by a spherical absorber under the assumptions that the diffusion distance between the two volumes is relatively short with respect to the average radius of the destination volume (according to the direction of propagation of the transmitted message) [18], and most of the molecules that diffuse in the destination volume are there involved in chemical reactions. Under this assumption, the probability that each molecule emitted by a volume is captured by an adjacent volume is equal to $P_{cap} = r_c/R$, where r_c is the average radius of the receiver cell, and R is the average distance between the center of the destination volume Ω_2 and the boundary of the volume Ω_1, and we consider the diffusion process as instantaneous with respect to the evolution of the diffusing molecule species D. If we consider a number of molecules $N_{D,\Omega_1}(t)$ that can reach the volume Ω_2 via diffusion, each with probability P_{cap}, the number $N_{D,\Omega_2}(t)$ is distributed according to a Binomial [13] as

$$P\left(N_{D,\Omega_2}(t)\right) = Bin(P_{cap}, N_{D,\Omega_1}(t)) \,. \qquad (3)$$

B. Iterative and Derivative-free Algorithm for Optimizing Information Transfer

According to information theory [19], our goal of optimizing the information transfer through this channel translates into

the following:

$$I_{opt}\left(X,Y\right)=\max_{P_X(x),x\in\mathcal{X}} I\left(X,Y\right)\,, \tag{4}$$

where $I\left(X,Y\right)$ is the MI between the transmitted message X and the corresponding received message Y from the system in Fig. 1, whose maximum is found with respect to a particular $P_X\left(x\right), x\in\mathcal{X}$, which is the probability distribution of the transmitted message X within the set $\mathcal{X}$ of all admissible messages. The MI is computed as follows [19]:

$$I\left(X,Y\right)=\mathbb{E}\left[\frac{P_{X|Y}\left(X\mid Y\right)}{P_X(X)}\right]=\mathbb{E}\left[\frac{P_{Y|X}\left(Y\mid X\right)}{P_Y(Y)}\right]\,. \tag{5}$$

Main Idea. While analytical solutions to the maximization in (4) are possible only for very special cases [19], the non-linear and non-smooth nature of the types of channel models described in Sec.II-A necessitates a numerical solution, *i.e.*, with an iterative algorithm. The BA is a class of these types of algorithms often utilized for this task, which necessitates the knowledge of the analytical expressions of the probabilities of the equivocation $P_{X|Y}\left(x\mid Y\right), x\in\mathcal{X}$, or equivalently, the expression of the channel law $P_{Y|X}\left(y\mid X\right), y\in\mathcal{Y}$. In particular, BA has already been successfully applied to basic MC channels [13], with analytical expressions of the aforementioned probabilities. As a consequence of the mathematical models detailed in Sec. II-A, which in general are composed of nested Poisson and Binomial processes according to the chemical reactions and molecule diffusion through which the transmitted message propagates, there is no straightforward way to analytically express these probabilities. A solution called SLEMI is proposed in [14] where in a context of information propagation within a single cell, expressions for these probabilities are found by fitting experimental data through logistic curve regression. A BA algorithm is then applied to numerically find a solution to (4). *In this paper, we propose instead a derivative-free iterative methodology based on the* ***Nelder-Mead*** *[20] optimization, which does not necessitate analytical expressions of the aforementioned probabilities, but just numerical values estimated from the data (simulated, in the case of this paper).* Given the similarity of the overall context and goal, we will compare the performance of our methodology with SLEMI based on the BA from [14].

The Nelder-Mead algorithm is a method to search for the minimum value of a nonlinear function of *n* variables based on the concept of a simplex [20]. The algorithm evaluates the output of the function at a set of $n+1$ test points and then performs the simplex in the factor-space, and continually forms new simplices by reflecting one point in the hyperplane of the remaining points [20]. It has been shown that this algorithm does not always converge to a minimum if the objective function is not strictly convex [21].

In Fig. 1 (upper) we show the first steps of the Nelder-Mead algorithm. The blue round-shaped curves represent the contour lines of the function to be maximized (objective function). It assumes higher values as the blue becomes more intense. If we assume that the objective function depends on two variables $\bar{\mu}_1$ and $\bar{\mu}_2$, by following the methodology of the algorithm, we need the simplex to be composed of $n+1$ points Q_0, Q_1, Q_2. The red dots represent the initial simplex $Q_{0_{init}}$, $Q_{1_{init}}$, $Q_{2_{init}}$. The first sample $Q_{0_{init}}$ is defined by the initial value of the variables $\bar{\mu}_{1_{init}}$ and $\bar{\mu}_{2_{init}}$, while $Q_{1_{init}}, Q_{2_{init}}$ are created by adding 5% of each component $Q_{i_{init}}$ to the initial guess $Q_{0_{init}}$. Then, the first step of the algorithm investigates which point of the simplex is farthest from the function's maximum value. It is reflected and then discarded. The green dot represents the new point of the simplex during the first iteration, that is the substitute of $Q_{0_{init}}$. The same procedure is also proposed in the next iterations of the method. Subsequently, $Q_{2_{init}}$ is discarded in favor of the yellow point in Fig. 1, being now the farthest from the maximum.

Our Methodology based on the Nelder-Mead algorithm has the following characteristics:

- The objective function to be maximized is the MI of the chemical channel, *i.e.*, $F_{\text{obj}} = I\left(X,Y\right)$, defined in (5), which is a function of the probability $P_X\left(x\right), x\in\mathcal{X}$ of the transmitted message X.
- The probability $P_X\left(x\right)$ is on its turn defined by the values of a set of parameters μ_k, $k=1,\ldots,K$, which define the algorithm search space, as in Fig. 1 (upper).
- The numerical values of $P_{X|Y}\left(x\mid y\right), x\in\mathcal{X}, y\in\mathcal{Y}$ are recomputed at each iteration of the Nelder-Mead algorithm from the data by pruning the dataset and estimating the probability via histograms, as explained in the following.

Search Parameters for $P_X\left(x\right)$. The number of parameters K depends on the considered family of distributions for $P_X\left(x\right)$. The family of distributions we have chosen for this work is continuous, the so-called *Pearson* [22]. The reason behind this choice is that the Pearson distribution family can easily be modeled via its first four moments, namely the mean μ, variance σ^2, skewness $\int$, and kurtosis κ. This represents a good trade-off between degrees of freedom in the shape of the distribution and number of variables to be optimized in the iterative procedure. Thus, the pruning at the *i*th iteration is performed based on these *i*th four moments as determined by the Nelder-Mead algorithm.

Histogram Estimation by Data Pruning. To obtain numerical values of $P_{X|Y}\left(x\mid y\right), x\in\mathcal{X}, y\in\mathcal{Y}$ at each iteration according to the current probability $P_X\left(x\right)$ (defined by the current values of μ, σ^2, $\int$, and κ), instead of obtaining additional data, thanks to the low sensitivity of our iterative methodology to the dataset size, as can be appreciated from the numerical results, we are able to reuse the same dataset via *pruning*. Pruning algorithms are often used to reduce the size of databases by removing unused information, or to reduce the size of neural networks [23], [24]. For our case, we utilize the following formula:

$$k_i=\left\lceil M-\frac{M}{p_{\max}}p_i\right\rceil\,, \tag{6}$$

where k_i is the number of datapoints (repetition of an experiment/simulation) corresponding to the *i*th input message to be removed from the dataset, p_i is the probability of that

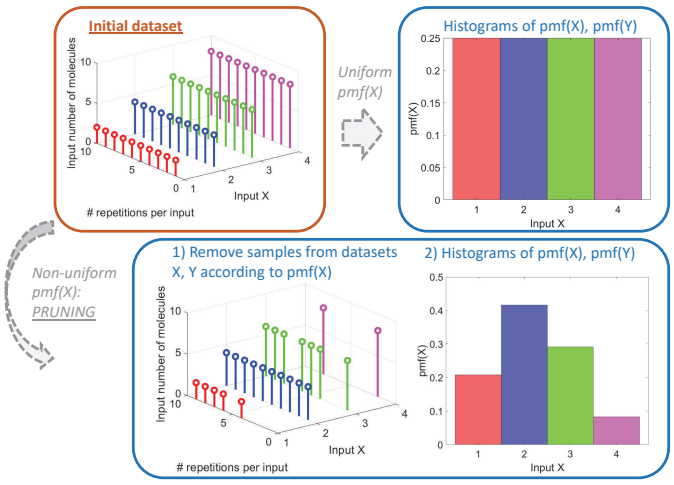


Fig. 2: Example of pruning and consequent histogram of the probability. Only the $\mathcal{X}$ dataset is shown for simplicity.

same input message, $p_{\max}$ is the highest probability within the input distribution $P_X(x)$, and M is the initial total number of datapoints of each input. When removing a datapoint, it is implicit that we remove both the input and its corresponding outcome from the dataset. Subsequently, from the pruned dataset, we estimate via histograms [25] the probabilities $P_X(X)$ and $P_Y(Y)$. Then, for each $P_{X|Y}(x \mid y)$, a histogram is built by setting the height of each bin as the number of all the x values that correspond to the those y values grouped into the same bin of the histogram representing $P_Y(Y)$.

Figure 2 is a visual example of pruning. On the x-axis we have the different input symbols (*i.e.*, 4), the y-axis represents the number of times the experiment (*i.e.*, the transmission of the input symbol) is repeated (*i.e.*, 10). We assume that the dataset contains the same number of datapoints per input symbol. When the ith $P_X(x), x \in \mathcal{X}$ is non-uniform, then for each message x a number of datapoints is removed randomly according to (6). Then, the histograms of $P_X(X)$ and $P_Y(Y)$ are estimated from the modified dataset.

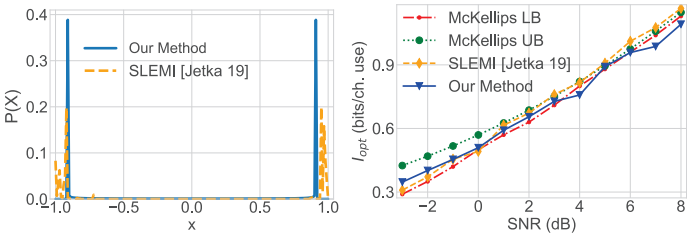
While we consider the width of the bins of each histogram as constant over the considered range of data, *i.e.*, $\mathcal{X}$ or $\mathcal{Y}$, the number of bins of the histograms representing the probabilities is modified at each iteration by following Doane's rule [26], which defines a lower bound on the required number of bins to reach a desired precision. Furthermore, we impose all the histograms at the same iteration to have the same bin width. The pseudocode in Algorithm 1 summarizes the main steps of our proposed methodology.

Validation and Comparison with SLEMI [14]. For the purpose of validation, we apply our method to an AWGN channel $Y = X + Z$, where Z is a normally-distributed noise, with a constraint on the amplitude A, in line with [15]. Then, the obtained $P_Y(Y)$ is a mixture of a uniform distribution in the interval $|y| \leq A$ and a "split and scaled" Gaussian density in the interval $|y| > A$, as in [15].

We perform the analysis for different Signal-to-Noise Ratios (SNRs), by following the same assumptions as in [15], so that we can compare the results. In particular, we fix the noise

Algorithm 1: Iterative procedure to optimize $I(X,Y)$

Procedure: Maximization of $F_{\text{obj}} = I(X,Y)$;
while $F_{obj_{NEW}} < F_{obj_{OLD}}$ **do**
  Calculation of $P(X)$ from its parameters.
  Pruning technique on X, Y datasets to fit the current $P(X)$.
  Doane's rule to define the optimal number of bins for the $P(X)$, $P(Y)$, and $P(X \mid y)$.
  Histograms for estimating $P(X)$, $P(Y)$, and $P(X \mid y)$.
  $F_{\text{obj}_{\text{OLD}}} = F_{\text{obj}}$.
  $F_{\text{obj}} = \mathbb{E}\left[\frac{P_{X|Y}(X|Y)}{P_X(X)}\right]$
  $F_{\text{obj}_{\text{NEW}}} = F_{\text{obj}}$.
  Update of the parameters characterizing $P(X)$.



(a) (b)

Fig. 3: AWGN with amplitude constraint. (a) $P_X(X)$ corresponding to $I_{opt}(X)$ with our method and with SLEMI [14]. (b) Corresponding $I_{opt}(X)$ obtained as function of the algorithm iterations and validation with McKellips' upper and lower bound for different SNRs.

power to be equal to 1. Thus, the SNR of the channel coincides with the power of the signal $P = A^2$. We consider values for A such that the SNR ranges from -3 to 8 dB, as in [15].

We generate a dataset via simulation with repetitions such that we obtain a uniform distribution of the input (*i.e.*, the transmitted symbols) with uniformly and monotonically increasing values. The input-output dataset is composed of 50000 input-output pairs, with an input range spanning from -2.5 to 2.5, 125 unique input values, and 400 repetitions for each unique input value.

As in [15], we compare our optimized $I_{opt}(X,Y)$ value for each SNR to the channel capacity bounds obtained using the McKellips' formulas [27]. Figure 3b shows that the $I_{opt}(X,Y)$ approaches the McKellips' lower bound on the channel capacity, supporting the validity of our methodology. Furthermore, the yellow dashed line represents the channel capacity value for different SNRs obtained when applying SLEMI [14]. This value and the value obtained with our methodology are comparable for low values of SNR (≤ 5 dB), while for higher values the SLEMI estimation is above the McKellips' upper bound. Since the objective function in [14] consists of a double maximization, for non-optimal parameters it can be considered a lower bound on the channel capacity value. For this reason, we can reasonably say that when the SLEMI optimal value is above the McKellips' upper bound, it overestimates the channel capacity.

The solid blue line in Fig. 3a shows the $P_X(X)$ that

corresponds to the $I_{opt}(X,Y)$. This figure corresponds to an amplitude $A = 1$. The shape of the $P_X(X)$ is nearly the same for all the considered values of SNR, *i.e.*, two mass points with non-zero probability near A and $-A$. While we do not have a proof that this distribution containing only two mass points is effectively optimal [28], this result is consistent with the literature [29], [30], thus reinforcing the validity of our method. Notice that for cases where the SNR is sufficiently high, we hypothesize that the capacity-achieving distribution may include more non-zero and equispaced mass points across the input distribution, as demonstrated in [29]. The yellow dashed line corresponds to the optimal $P_X(X)$ estimated through the SLEMI [14] for the same A, which has a similar shape to the one obtained from our methodology.

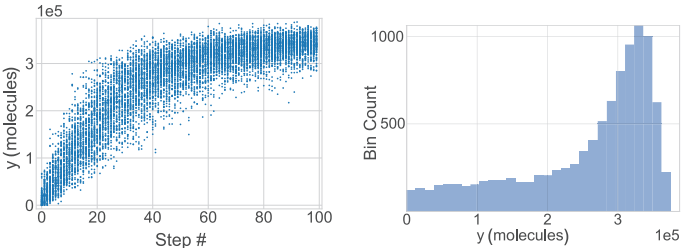
III. Case Study: An Engineered Cell-to-Cell Molecular Communication System

A. LuxR-LuxI-Based Channel

As a first case study of the types of channels detailed in Sec. II, in this paper we apply our developed methodology to estimate the optimized information transfer through the engineered cell-to-cell communication system presented in [3]. In particular, this simulated system is based on the LuxR-LuxI cell communication modules, where synthetic biology components are derived from the artificial sender and receiver design included in the experimental work in [16]. The system is composed of two biological cells (two volumes) located in an infinite extracellular environment. One of the two cells acts as a transmitter, where the source message x is a modulated number $N_{Tx}(t)$ of molecules of β-D-1-Thiogalactopyranoside (IPTG) [3]. Through a chain of well-defined chemical reactions of the type modeled in (1) and (2), involving proteins and DNA genes, the IPTG is transformed into the transmitted signaling molecule Acyl-Homoserine Lactone (AHL), which can cross the cell membranes (volume boundaries) and diffuse in the extracellular space. A receiver cell captures diffusing AHL molecules according to the model in (3) and, after a chain of other chemical reactions, returns a number $N_{Rx}(t)$ of Green Fluorescent Protein (GFP) molecules, from which the received information message y is decoded (as the maximum of $N_{Rx}(t)$ [3]). Further details on the modeling and simulation of this system according to the formulation in Sec. II are here omitted for space constraints, and can be found in [3].

B. Numerical Results

The aforementioned LuxR-LuxI-Based Channel is modeled and simulated *in-silico* via MATLAB SimBiology, as detailed in [3]. The input $N_{Tx}(t)$ varies from 3.6×10^5 IPTG molecules, to 3.6×10^7 molecules with a step size 3.6×10^5 molecules (100 different input messages). This range is chosen by calculating the average μ_Y and the variance σ_Y^2 of the output dataset, and by observing their behavior for different input values, as suggested in [3]. Our results show that for the output corresponding to input values ranging from 0 to about 1.4×10^7 molecules there is a linear increase of the average and variance, while for higher values there is a saturation of



(a) (b)

Fig. 4: (a) Output obtained from a uniform input with 100 monotonically increasing IPTG molecule counts and 100 simulation repetitions per input. (b) Corresponding histogram of the output.

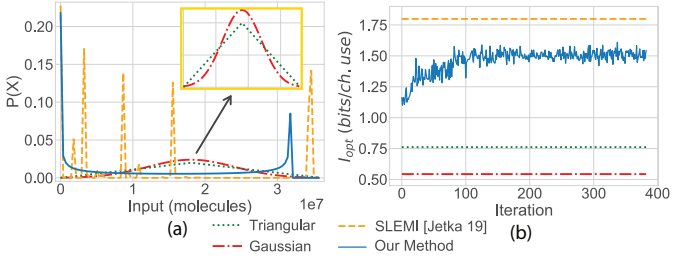


Fig. 5: LuxR-LuxI-based channel. (a) $P_X(X)$ corresponding to $Iopt(X)$ with our method and with SLEMI [14], compared to a Gaussian and Triangular distributions. (b) Corresponding values of obtained $I_{opt}(X)$ as function of algorithm iterations.

the average value and the associated variance, which remains almost constant up to 3.6×10^7 molecules. We repeat the simulation for each input value 100 times. Thereby, we obtain a 100×100 input-output dataset, which unpruned corresponds to a uniform input distribution. From a communication perspective, this is equivalent to a discrete memoryless channel with Concentration Shift Keying (CSK) [12].

Figure 4a illustrates the unpruned output dataset generated from the uniform input dataset. The input is propagated through the reaction and diffusion processes composing the MC channel. Figure 4b depicts the histogram of the channel output considered in Fig. 4a.

The blue continuous line in Fig. 5b shows the progressive optimization of $I_{opt}(X,Y)$ operated by our methodology, as detailed in Sec. II-B. The $I_{opt}(X,Y)$ value generally increases with the number of iterations of the algorithm, until it converges. The $I_{opt}(X,Y)$ value we obtain is 1.60 (bit/ch. use). The corresponding input distribution $P_X(X)$ from the Pearson family is represented in Fig. 5a with the blue continuous line. As in the AWGN case, the optimal $P_X(X)$ has two peaks at the extreme values of the input dataset. Thus, the method privileges the two symbols farthest from each other in this case as well. The four moments corresponding to this specific distribution of the Pearson family are $\mu = 1.60 \times 10^7$, $\sigma^2 = 1.15 \times 10^7$, $\int = -4.03 \times 10^{-17}$, and $\kappa = 1.46$.

The MI value for this channel is evaluated in [3] by considering a Gaussian and a Triangular input distribution. For the sake of comparison, we plot our results against the MI value for these distributions. Fig. 5a represents the optimal

distribution found by our method, as well as the Gaussian (red dot dashed line) and Triangular (green dotted line) input distributions. The MI values obtained from the considered distributions are plotted in Fig. 5b. The $I_{opt}(X,Y)$ obtained by the proposed method, 1.60 (bits/ch. use), is greater than the MI values obtained by Gaussian and Triangular input distributions (*i.e.,* 0.55 and 0.76 (bit/ch. use), respectively).

The yellow dashed lines in Fig. 5a and b represent the optimal input distribution and corresponding capacity value, 1.80 (bit/ch. use), obtained when applying SLEMI [14] to the molecular dataset. The channel capacity value obtained by SLEMI is slightly higher than that obtained by the proposed method. Although the input distribution found by SLEMI has more than two peaks, most of them are concentrated near the extremes of the input ranges, similar to the $P_X(X)$ observed when using the proposed methodology. Additional investigation into why SLEMI provides a higher bound in some cases is left to future work.

IV. Conclusion

This paper has focused on the problem of optimizing the information transfer (*i.e.*, the MI) in an MC system where the combined effects of chemical reactions and molecule diffusion hinders the application of any information-theoretic analytical methodology. In addition, the practical and realistic constraint on a maximum value of transmitted molecules makes the problem of estimating the MI even more challenging. To address these issues, a novel iterative methodology based on the Nelder-Mead algorithm, which does not require any analytical formulations of MC components and their stochastic behavior, is proposed to estimate the optimal MI and is validated against known results from the literature. Numerical results obtained by applying this methodology to a simulation-based scenario abstracting an MC channel between genetically engineered cells reveal the viability of information transfer optimization via proper design of the distribution of the input messages. Future application scenarios could make use of our methodology to not only optimize chemical and biological experimental design, but also support synthetic biology in the engineering of communication systems and networks that work and interact in fully biochemical environments.

Acknowledgement

This material is based upon work supported by the U.S. National Science Foundation under Grant No. CCF-1816969.

References

[1] I. F. Akyildiz, M. Pierobon, and S. Balasubramaniam, "An information theoretic framework to analyze molecular communication systems based on statistical mechanics," *Proc. IEEE*, vol. 107, no. 7, pp. 1230–1255, July 2019.

[2] O. B. Akan, H. Ramezani, T. Khan, N. A. Abbasi, and M. Kuscu, "Fundamentals of molecular information and communication science," *Proc. IEEE*, vol. 105, no. 2, pp. 306–318, 2017.

[3] C. Harper, M. Pierobon, and M. Magarini, "Estimating information exchange performance of engineered cell-to-cell molecular communications: a computational approach," in *IEEE INFOCOM*, 2018, pp. 729–737.

[4] T. Nakano, Y. Okaie, and J.-Q. Liu, "Channel model and capacity analysis of molecular communication with brownian motion," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 797–800, 2012.

[5] B. Atakan and O. B. Akan, "On channel capacity and error compensation in molecular communication," in *Transactions Comput. Syst. Biol. X*. Springer, 2008, pp. 59–80.

[6] N. Varshney, A. Patel, W. Haselmayr, A. K. Jagannatham, P. K. Varshney, and W. Guo, "Impact of cooperation in flow-induced diffusive mobile molecular communication," in *IEEE ASILOMAR*, 2018, pp. 494–498.

[7] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 942–954, 2013.

[8] H. Awan and C. T. Chou, "Improving the capacity of molecular communication using enzymatic reaction cycles," *IEEE Trans. Nanobiosci.*, vol. 16, no. 8, pp. 744–754, 2017.

[9] P. Sadeghi, P. O. Vontobel, and R. Shams, "Optimization of information rate upper and lower bounds for channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 663–688, 2009.

[10] F. Ratti, F. Vakilipoor, H. Awan, and M. Magarini, "Bounds on the constrained capacity for the diffusive poisson molecular channel with memory," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, 2021.

[11] Q. Liu and K. Yang, "Channel capacity analysis of a diffusion-based molecular communication system with ligand receptors," *Int. J. Commun. Syst.*, vol. 28, no. 8, pp. 1508–1520, 2015.

[12] M. Ş. Kuran, H. B. Yilmaz, I. Demirkol, N. Farsad, and A. Goldsmith, "A survey on modulation techniques in molecular communication via diffusion," *IEEE Commun. Surveys Tuts.*, 2020.

[13] N. Farsad, W. Chuang, A. Goldsmith, C. Komninakis, M. Médard, C. Rose, L. Vandenberghe, E. E. Wesel, and R. D. Wesel, "Capacities and optimal input distributions for particle-intensity channels," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 6, no. 3, pp. 220–232, 2020.

[14] T. Jetka, K. Nienałtowski, T. Winarski, S. Błoński, and M. Komorowski, "Information-theoretic analysis of multivariate single-cell signaling responses," *PLoS Comput. Biol.*, vol. 15, no. 7, p. e1007132, 2019.

[15] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, 2017.

[16] T. Ramalho, A. Meyer, A. Mückl, K. Kapsner, U. Gerland, and F. C. Simmel, "Single Cell Analysis of a Bacterial Sender-Receiver System," *PLOS ONE*, vol. 11, no. 1, p. e0145829, 2016. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0145829

[17] D. T. Gillespie, "The chemical langevin equation," *The Journal of Chemical Physics*, vol. 113, pp. 297–306, 2000.

[18] H. Berg, *Random walks in biology*. Princeton University Press, 1993.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[20] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[21] K. I. McKinnon, "Convergence of the nelder–mead simplex method to a nonstationary point," *SIAM J. Optim.*, vol. 9, no. 1, pp. 148–158, 1998.

[22] P. Kerl, "Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material," *The Royal Society of London*, vol. 186, pp. 343–414, 1895.

[23] R. Reed, "Pruning algorithms-a survey," *IEEE Trans. Neural Netw.*, vol. 4, no. 5, pp. 740–747, 1993.

[24] M. Augasta and T. Kathirvalavakumar, "Pruning algorithms of neural networks—a comparative study," *Open Computer Science*, vol. 3, no. 3, pp. 105–115, 2013.

[25] A. Stuart, K. Ord, and S. Arnold, "Kendall's advanced theory of statistics, volume 2a: classical inference," 2009.

[26] D. P. Doane, "Aesthetic frequency classifications," *The American Statistician*, vol. 30, no. 4, pp. 181–183, 1976.

[27] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *ISIT*. IEEE, 2004, pp. 348–348.

[28] D. Xiao, L. Wang, D. Song, and R. D. Wesel, "Finite-support capacity-approaching distributions for awgn channels," in *2020 IEEE ITW*. IEEE, 2021, pp. 1–5.

[29] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[30] A. Favano, M. Ferrari, M. Magarini, and L. Barletta, "The capacity of the amplitude-constrained vector Gaussian channel," *arXiv preprint arXiv:2101.08643*, 2021.