An Iterative Approach for Estimating Information Exchange in Cell-to-cell 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, charper@cse.unl.edu, maurizio.magarini@polimi.it, pierobon@cse.unl.edu

***Abstract*—The characterization of biological systems in terms of communication performance is currently limited. In this paper, an iterative method is presented for evaluating mutual information in molecular communication systems composed of genetically engineered cells, based on an indirect non-parametric estimation of probability density functions. Results are presented based on the application of this method to proof-of-concept *in-silico* data generated via stochastic simulation.**

I. Introduction

Molecular Communication (MC) is an emerging technology directly inspired by natural communications between cells in biology [1], [2]. In MC, information is encoded into and decoded from molecules, rather than electromagnetic waves, thus opening the road to exploiting biological means to enable communication among small-scale devices, or nanomachines.

Understanding living cells from an information and communication theoretical perspective is one of the challenges to gain insights into the fundamentals of MC system engineering. The objective of this work is to minimize the upper bound on the Mutual Information (MI), which is usually adopted to measure the exchange of information in communication theory [3], in a communication system composed of two genetically engineered cells [4]. The proposed approach is based on an iterative method that minimizes an upper bound of the MI as a difference of entropies, in agreement with [5].

II. System Model

We consider the same biological system model presented in [4]. In particular, two *Escherichia Coli (E. Coli)* bacteria cells are modeled as the transmitter and receiver, respectively. Different concentration values of the protein called Isopropyl β-D-1-Thiogalactopyranoside (IPTG) are injected in the environment, each one identifiable with the message $X_{source}(t) = x_{source}\mathcal{H}(t)$, where $\mathcal{H}(t)$ is the Heaviside step function.

A change of IPTG concentration causes a series of chemical reactions [4] and consequent variations in the concentration of biological molecules inside the transmitter cell, which emits in response Acyl-Homoserine Lactone (AHL) in the environment. AHL is a small molecule employed for signaling in biological systems, also known as quorum sensing autoinducer. The emitted concentration of AHL molecules then propagates by means of diffusion (diffusive channel) to the receiver cell. The receiving cell captures the propagated AHL molecules by means of a number of receptors. After a series of chemical reactions inside the receiver cell, a concentration of Green Fluorescent Protein (GFP), a protein visible at the microscope, is produced as the received message $Y_{dest}(t)$ [4]. While it is in general a function of time t, in this work we consider solely the information content of the maximum difference between the maximum and the minimum value of $Y_{dest}(t)$. This corresponds to measuring the maximum distinctness of the signal that allows for a better estimation of the MI. When propagating from $X_{source}(t)$ to $Y_{dest}(t)$, the information signal attenuates and accumulates noise, generally modeled by a Poisson process [4]. With the goal of realizing a discrete uniform input distribution, the range of inputs X_{source} has been varied from 1.5×10^6 nM to 1.5×10^8 nM with a step size 1.5×10^6 nM.

III. Information Theory Background

Two fundamental concepts of this study are the entropy and the MI. The former defines the probabilistic behavior of a source of information and gives a limit to the complexity below which a signal can not be compressed. The MI of two random variables is a measure of the mutual dependence between the two variables. More specifically, it quantifies the amount of information expressed in bits obtained about one random variable, through the other random variable.

In this work, the MI is calculated as [3]

$$I(X_{source}; Y_{dest}) = H(X_{source}) - H(X_{source}|Y_{dest}) \quad (1)$$

where $H(X_{source})$ is the entropy of the input variable X_{source} and $H(X_{source}|Y_{dest})$ is the conditioned entropy of the input X_{source} given the output Y_{dest}. The main motivation for using this formulation is that of performing an optimization on the input distribution, as illustrated in the next section.

IV. Proposed Iterative Method for the Minimization of the Upper Bound on the MI

An iterative method is here illustrated to minimize the upper bound on the MI with a constraint on the input concentration of the engineered cell-to-cell communication system described in Sec. II. The Nelder-Mead iterative method [6] is a numerical

Algorithm 1: Iterative algorithm to minimize the upper bound on the MI

Procedure: Minimization of $I(X,Y)$;
Require: $I(X,Y) = H(X) - H(X|Y)$ **and** $H(X) \geq H(X|Y)$;
while $I_{new} < I_{old}$ **do**
  Calculate $p(X)$ from $mean$, $variance$, $skewness$, $kurtosis$.
  Pruning technique on X, Y.
  Doane's formula to know the number of bins for the $p(X)_{new}$ after pruning, for $p(Y)$ and for $p(X|y)$.
  Calculate $H(X)$ for uniform quantization from $p(X)_{new}$.
  Calculate $p(X|y)$ and from that $H(X|y)$ is found.
  Weight $H(X|y)$ with $p(Y)$ to obtain $H(X|Y)$.
  $I_{old} = I(X,Y)$
  $I(X,Y) = H(X) - H(X|Y)$
  $I_{new} = I(X,Y)$
  $mean = mean_{new}$, $variance = variance_{new}$, $skewness = skewness_{new}$, $kurtosis = kurtosis_{new}$

approach used to find the minimum or maximum of an objective function in a multidimensional space. According to our approach, we use a class of probability density functions (pdfs) parameterized in a multidimensional space that is defined by the first four moments, *i.e.* the mean, the variance, the skewness, and the kurtosis. The main advantage of using this approach is that it is already implemented in the built-in MATLAB function *fminsearch* [7]. The pseudo-code of the iterative procedure is described in Algorithm 1, where the subscripts of the variables are removed for simplicity.

The numerical results presented here are based on Gillespie's stochastic simulation [8], implemented using MATLAB Simbiology [4]. The MI is estimated through the histogram technique, a simplified version of the Kernel Density Estimation (KDE) belonging to indirect non-parametric estimation methods. Non-parametric statistics is based on either being distribution-free, or having a specified distribution but with unspecified distribution parameters. The histograms used to estimate the pdfs have been generated according to the Doane's formula as in [9], which defines the minimum number of bins that are required for a wanted precision. Also, we have used the pruning technique to fit the input data to the pdf that results at each step of the iterative algorithm. This pruning is initially applied to the uniform pdf associated to the input concentrations considered in our Simbiology simulations. The four moments from which the algorithm calculates the initial pdf $p(X)$ are those associated with the uniform distribution. So, they are derived by assuming all the input concentrations equiprobables. The corresponding $p(Y)$ is estimated by following the steps of the algorithm, unless for the pruning technique, that is not necessary in case of uniform distribution.

In this case, the minimization of the upper bound on the MI leads to an estimation of the best upper bound to the capacity of the channel. Channel capacity is defined by $C = \max_{X \in \mathcal{P}(\mathcal{X})} I(X_{source}, Y_{dest})$ and it can be obtained only if all the possible pdfs for the concentration of the IPTG proteins are explored at the input, which is clearly not feasible in practice.

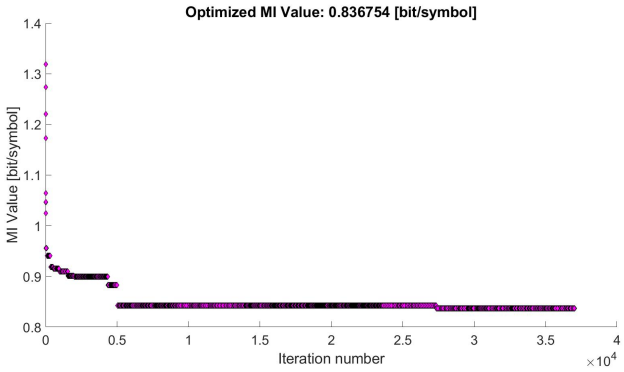

Fig. 1. Result of the minimization with the iterative algorithm.

The histogram method, used to estimate the MI, has some statistical limitations, as explained in [5], since it provides an upper bound to the mutual information. In fact, the Nelder-Mead iterative method is a heuristic technique that can lead to convergence on non-stationary points. Another important aspect is the class of pdfs used at the input $X_{source}(t)$, which should be further investigated for the goal of minimizing the MI for this particular communication system.

V. Results and Conclusion

In-silico simulations were done by repeating 100 times each value of IPTG concentration in order to observe the corresponding values of Y_{dest}. Figure 1 shows a first result achieved with the application of the proposed iterative algorithm. The distance between the transmitter cell and the receiver has been chosen equal to 30 μm, as in [4]. As it can be observed, after a fast initial decrease, the upper bound on the MI converges to a stable value after 5×10^3 iterations. This is a first step in the research of bounds to the MI of engineered cell-to-cell MC, where a lower bound should also be investigated.

References

[1] I. F. Akyildiz, J. M. Jornet, and M. Pierobon, "Nanonetworks: A new frontier in communications," *Communications of the ACMs*, vol. 54, no. 11, pp. 84–89, November 2011.

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

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

[4] C. Harper, M. Pierobon, and M. Magarini, "Estimating information exchange performance of engineered cell-to-cell molecular communications: A computational approach," in *Proc. IEEE INFOCOM*, 2018, pp. 1–9.

[5] D. McAllester and K. Statos, "Formal limitations on the measurement of mutual information," *arXiv preprint arXiv:1811.04251*, 2018.

[6] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 01 1965.

[7] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[8] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, pp. 2340–2361, 1977.

[9] Z. Sakkaff, A. Immaneni, and M. Pierobon, "Applying molecular communication theory to estimate information loss in cell signal transduction: an approach based on cancer transcriptomics," in *Proc. 5th ACM Intern. Conf. Nanoscale Computing Commun.*, 2018, pp. 1–16.