



**HAL**  
open science

## Estimation and Optimization for Molecular Communications with a Coexistence Constraint

Malcolm Egan, Valeria Loscrì, Ido Nevat, Trung Q Duong, Marco Di Renzo

► **To cite this version:**

Malcolm Egan, Valeria Loscrì, Ido Nevat, Trung Q Duong, Marco Di Renzo. Estimation and Optimization for Molecular Communications with a Coexistence Constraint. NanoCom 2019 - Sixth Annual ACM International Conference on Nanoscale Computing and Communication, Sep 2019, Dublin, Ireland. pp.1-6, 10.1145/3345312.3345472 . hal-02178421

**HAL Id: hal-02178421**

**<https://hal.science/hal-02178421>**

Submitted on 11 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation and Optimization for Molecular Communications with a Coexistence Constraint

Malcolm Egan, Valeria Loscri, Ido Nevat, Trung Q. Duong and Marco Di Renzo

**Abstract**—At present, molecular communications has largely been studied in isolation; that is, without the presence of any external biological systems. However, for applications in nanomedicine and advanced manufacturing, this assumption of isolation is unlikely to always hold. A key question is therefore how to design signaling strategies such that the desired information can be communicated, while not disrupting the function of the external biological system. In this paper, we study this problem by establishing a coexistence constraint based on the Kullback-Leibler divergence between the distribution of molecules on the surface of the biological system, with and without the presence of the molecular communication system. We then address the problem of how the transmitting device can estimate these distributions from local observations of its environment. Based on this, we optimize the modulation scheme used by the transmitter subject to the coexistence constraint. Tradeoffs between the coexistence constraint, the probability of error, and the number of samples used to estimate distributions needed to compute the coexistence constraint are then studied numerically.

## I. INTRODUCTION

The use of molecular communication—where information is encoded in the type, quantity, or timing of molecules—has been proposed to form the basis of new technologies for advanced manufacturing and nanomedicine, amongst others [1], [2]. A common feature of these applications is that the molecular communication system is embedded in a complex biochemical system. For example, in bacteria bioreactors [3] the introduction of new chemical species communicating control information may promote the production of desired synthetic molecules. However, these new species may also increase the metabolic load, which reduces the growth of the colony (e.g., [4]).

In order to minimize the impact of molecular communication networks on any external biochemical system—or support *coexistence*—there is a need to carefully design the communication strategy. One way of quantifying the perturbation due to the presence of a molecular communication system is through changes to steady state concentrations [5]. As such, an important coexistence constraint is that changes to expected steady state concentrations are bounded.

M. Egan is with the CITI Lab, 69621 Villeurbanne, France which is a joint laboratory in Université de Lyon, INSA-Lyon, and INRIA (e-mail: malcom.egan@inria.fr). V. Loscri is with INRIA Lille-Nord Europe, France. I. Nevat is with TUMCREATE, Singapore. T.Q. Duong is with the Institute of Electronics, Communications and Information Technology in Queen’s University Belfast, Belfast BT7 1NN, U.K. M. Di Renzo is with the Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec, Univ Paris Sud, Université Paris-Saclay, 91192 Gif-sur-Yvette, France.

At present, there is limited work addressing this coexistence problem in molecular communications (an extensive survey of recent work is available in [6]). In our previous work [5], the problem was formalized and, for simple chemical systems in the environment, information molecule selection strategies were proposed. In particular, these strategies ensured that the steady state concentrations in the biological system did not change.

However, it may be difficult in practice to find information molecules that do not change steady state concentrations in complex external biochemical systems. In this case, it is necessary to design the communication strategy to relax the conditions imposed in [5] and consider coexistence with bounded changes to expected steady state concentrations.

In this paper, we study molecular communication systems that do not perturb the expected steady state concentrations of an external biochemical system by more than a fixed value  $\delta$ . We first observe that this difference in expected steady state concentrations can be bounded by the Kullback-Leibler divergence between the distribution of molecules on the surface of the biochemical system, with and without a molecular communication system present. This provides a means of selecting a communication strategy that satisfies the coexistence constraint.

A key challenge in ensuring satisfaction of the coexistence constraint is how to obtain an estimate of the distribution for the number of molecules on the surface of the biochemical system. This is complicated due to the fact that receptors on the transmitter to detect the presence of molecules may be non-specific. This means that each receptor can bind with more than one type of species. As the biological system may be sensitive to only a subset of chemical species in the environment, it is necessary for the transmitter to correctly estimate the distribution for a particular type given the receptor observations.

We address the problem of obtaining estimates of concentration distributions given the observations from non-specific receptors. To do this, we use a recent technique from the stochastic chemical reaction network (SCRN) literature [7]. In particular, based on the observations at the receptors, a SCRN is constructed which produces samples from the posterior distribution of concentrations of molecules in the environment. This approach has the benefit that it is based on a bio-circuit and requires very few resources in order to perform the sampling procedure. We remark that our approach complements recent work of developing bio-circuits to perform coding and decoding [8] by providing a means of obtaining side

information about the biological environment of the molecular communication system.

With the samples from the concentration posterior distributions in hand, it is then possible to construct estimators for these distributions. As such, the coexistence constraint can be computed for a given communication strategy. In particular, we use the estimates to optimize molecular communication systems where information is encoded in the rate of production of molecules. This modulation scheme has been also studied in [9], albeit when a coexistence constraint is not present. For the binary case, we then numerically optimize the rates of production for each symbol to minimize the probability of error subject to the coexistence constraint. In this setting, we demonstrate the tradeoffs between the coexistence constraint and the probability of error, along with the impact of the number of samples used to estimate distributions of concentrations.

## II. SYSTEM MODEL

### A. Molecular Communication Network Model

Consider a molecular communication system consisting of a single transmitter and a single receiver. We assume that the devices are synchronized and that the transmitter uses a binary modulation scheme, where at the beginning of each time slot a random quantity,  $X$ , of molecules are emitted. The random variable  $X$  follows a Poisson distribution with parameter  $\lambda_m$ ,  $m \in \{0, 1\}$ , where  $m$  corresponds to the bit that is to be modulated. We assume that  $\lambda_1 > \lambda_0 \geq 0$ .

Each emitted molecule from the transmitter reaches the receiver within a time period  $T$  with probability  $p$ . We assume that transmissions are spaced sufficiently far apart in time that there is no inter-symbol interference. As such, the only molecules that arrive at the receiver within the time slot are those that are also transmitted within the same time slot. Hence, the number of molecules that reach the receiver in a given time slot is also Poisson with parameter  $p\lambda_m$ ,  $m \in \{0, 1\}$ .

We remark that the probability  $p$  that each molecule reaches the receiver in a given time slot is determined by the fluid medium. For example, if the molecules diffuse according to one-dimensional Brownian motion and there are no other biochemical processes in the environment, then the probability  $p$  is determined by the cumulative distribution function of an inverse Gaussian random variable [10]. Moreover,  $p$  may also be affected by the ability of the receiver to absorb information-carrying molecules.

With probability  $1-p$ , a molecule emitted by the transmitter does not reach the receiver within time  $T$ . In this paper, we make the conservative assumption that all molecules emitted by the molecular communication system that do not reach the receiver bind to the biological system, which is detailed next.

### B. Biological System Model

The biological system model that we consider in this paper is motivated the biochemical processes present inside the bacteria *Escherichia coli*. This kind of system exploits a sensing mechanism to adapt to the concentration of certain chemical

species in the environment. In the case of *Escherichia coli*, this sensing mechanism allows for the bacteria to move towards food sources or away from repellants [11].

In this system, when molecules from the environment bind with receptors on the surface of the biological system, an adaptation process begins. This process ultimately adjusts the steady state concentration<sup>1</sup> of a given chemical species, denoted by  $F$ , within the biological system. We assume that the biological system is sensitive to both information molecules from the communication system and also other chemical species that are naturally present in the environment. We denote by  $Z$  the total concentration of all molecules for which the biological system is sensitive. It is assumed that the concentration of  $Z$  is stationary in space when the molecular communication is not present. This means that estimates for this distribution obtained in one region are also valid in other regions.

In order to model the kinetics of the biological system, we adopt a stochastic chemical reaction network model. In particular, there are assumed to be  $s_0$  chemical species (including  $F$ ) and  $r_0$  chemical reactions. In general, the network of chemical reactions can be written as [12],

$$\sum_{i=1}^{s_0} \nu_{ik} S_i \rightarrow \sum_{i=1}^{s_0} \nu'_{ik} S_i, \quad k = 1, \dots, r_0, \quad (1)$$

where  $\nu_{ik}, \nu'_{ik} \in \mathbb{Z}_{\geq 0}$ . We denote  $\nu_k = [\nu_{1k}, \dots, \nu_{s_0k}]^T$  and  $\nu'_k = [\nu'_{1k}, \dots, \nu'_{s_0k}]^T$ .

At time  $t$ , the number of molecules for each species is described by a component of the random vector  $\mathbf{X}(t)$ . To account for the fact that reactions require individual molecules of each chemical species to be in close proximity, the rate that the  $k$ -th reaction occurs when the quantity of molecules for each species is given by  $\mathbf{x}$  is denoted by  $\lambda_k(\mathbf{x})$ , called the *propensity*. If the  $k$ -th reaction occurs at time  $t$ , the new state is then

$$\mathbf{X}(t) = \mathbf{X}(t-) + \nu'_k - \nu_k. \quad (2)$$

Following [12], the number of times the  $k$ -th reaction occurs by time  $t$  is given by the counting process

$$R_k(t) = Y_k \left( \int_0^t \lambda_k(\mathbf{X}(s)) ds \right), \quad (3)$$

where the  $Y_k$ ,  $k = 1, \dots, r_0$  are independent unit rate Poisson processes. We assume stochastic mass action kinetics, which means that for  $x \in \mathbb{Z}_{\geq 0}^{s_0}$  the propensities are given by

$$\lambda_k(x) = \kappa_k \left( \prod_{l=1}^{s_0} \nu_{lk}! \right) \binom{x}{\nu_k}, \quad k = 1, \dots, r_0, \quad (4)$$

where  $\kappa_k > 0$  is the rate constant for the  $k$ -th reaction.

As such, the state of the system at time  $t$  is given by

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{k=1}^{r_0} R_k(t) (\nu'_k - \nu_k) \quad (5)$$

<sup>1</sup>The steady state concentration occurs when the concentration  $\mathbf{x}$  satisfies  $\frac{d\mathbf{x}}{dt} = 0$ .

and the concentration dynamics forms a Markov chain.

We assume the initial state of the biological system  $\mathbf{X}(0)$  is fixed except for the species corresponding to  $Z$  and that a stationary distribution exists. Note that conditions for the existence of a stationary distribution can be found in [12]. As such, the stationary distribution in a given time slot depends on  $Z$ . We therefore denote the steady-state concentration of  $F$  by  $\Phi_F(Z)$ .

In the sequel, we will also assume that the distribution for the concentration of  $Z$  in the absence of the molecular communication system is Poisson distributed with parameter  $q_Z$ . This can be justified as the Poisson distribution is known to be the stationary distribution of many stochastic chemical reaction networks [12].

### C. Coexistence Constraint

Due to the importance of steady state concentrations on biological function, we formalize the coexistence constraint as follows. Recall that  $\Phi_F(Z)$  is the random variable corresponding to the steady state concentration of the chemical species  $F$  affecting the function of the biological system. The concentration  $\Phi_F(Z)$  has distribution  $P_{\Phi_F(Z)}$ , where  $Z$  is the quantity of molecules bound to the biological system in a given time slot.

Let  $P_Z$  be the distribution of molecules at the biological system when the communication system is present and  $Q_Z$  be the distribution when the communication system is not present. One means of ensuring coexistence is therefore to ensure that for some  $\delta > 0$ ,

$$D(Q_Z||P_Z) < \delta, \quad (6)$$

where  $D(\cdot||\cdot)$  denotes the Kullback-Leibler divergence. To justify this constraint let  $x_0$  denote the empty symbol where the transmitter does not send anything. When  $\Phi_F(Z)$  has bounded support with maximum value  $\Phi_{\max}$ , it then follows that

$$\begin{aligned} & |\mathbb{E}_{\Phi_F(Z)}[\Phi_F(Z)] - \mathbb{E}_{\Phi_F(Z)|X=x_0}[\Phi_F(Z)]| \\ & \leq \Phi_{\max} \|P_{\Phi_F(Z)} - P_{\Phi_F(Z)|X=x_0}\|_1 \\ & \leq \Phi_{\max} \sqrt{2D(P_{\Phi_F(Z)|X=x_0}||P_{\Phi_F(Z)})} \\ & \leq \Phi_{\max} \sqrt{2D(Q_Z||P_Z)}. \end{aligned} \quad (7)$$

The second inequality follows from Pinsker's inequality and the third inequality follows from the data processing inequality for the Kullback-Leibler divergence [13]. The difference in expectations in (7) directly constrains the variations in the steady state concentration of  $F$ . As such, the coexistence constraint in (6) forms a reasonable measure of changes to the function of the biological system.

## III. ESTIMATION FOR DISTRIBUTIONS OF BACKGROUND CONCENTRATIONS

In order to design communication strategies such that the coexistence constraint in (6) holds, we require the distribution of concentrations for chemical species in the environment. In order to estimate these distributions, we require samples.

However, in general receptors that bind to molecules in the environment are non-specific. This means that more than one chemical species in the environment can bind with each receptor.

In this section, we develop a means of computing the coexistence constraint for the molecular communication system detailed in Section II with non-specific receptors on the surface of the transmitter. In this setting, the number of each type of molecule produced by the receptors depends linearly on the number of molecules that bind with the receptors. In particular,

$$\mathbf{r} = \mathbf{O}\mathbf{l}, \quad (8)$$

where  $\mathbf{r} \in \mathbb{Z}_{\geq 0}^{n_R}$  is the number of each type of receptor molecule produced, which can be directly observed within the transmitting device. The vector  $\mathbf{l} \in \mathbb{Z}_{\geq 0}^{n_L}$  is the number of each type of species in the environment that bind to receptors on the transmitter. The matrix  $\mathbf{O}$  is an  $n_R \times n_L$  observation matrix determining the production of receptor molecules given the number of molecules from the environment that bind with each receptor.

### A. Posterior Via Chemical Reaction Networks

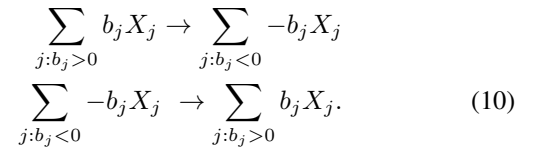
We now detail the procedure to obtain samples of the posterior  $\Pr(\mathbf{l} = \mathbf{u}|\mathbf{r})$  developed in [7]. More precisely, this posterior is obtained under the Bayesian methodology, assuming the product-Poisson prior for the number of molecules in the environment that bind with the receptors, given by

$$\Pr(\mathbf{l} = \mathbf{u}) = \prod_{i=1}^{n_L} e^{-\tilde{q}_i} \frac{\tilde{q}_i^{u_i}}{u_i!}, \quad (9)$$

where  $\mathbf{l} \in \mathbb{Z}_{\geq 0}^{n_L}$ ,  $\mathbf{r} \in \mathbb{Z}_{\geq 0}^{n_R}$ ,  $\tilde{\mathbf{q}} = [\tilde{q}_1, \dots, \tilde{q}_{n_L}]^T$  are the positive rate parameters.

To begin, we require the notion of a tidy observation matrix  $\mathbf{O}$ . The matrix  $\mathbf{O}$  is tidy if there exists a  $n_L \times n_R$  matrix  $\mathbf{M}$  such that  $\mathbf{O}\mathbf{M} = \mathbf{I}_{n_R \times n_R}$  and  $\mathbf{M} \in \mathbb{Z}_{\geq 0}^{n_L \times n_R}$ . The chemical reaction network which will compute the samples from the distribution  $\Pr(\mathbf{l}|\mathbf{r})$  is defined as follows. Note that  $\ker(\mathbf{O})$  denotes the nullspace of the matrix  $\mathbf{O}$ .

**Definition 1.** Consider the observation matrix  $\mathbf{O} = (o_{ij})_{n_R \times n_L}$  with rational entries  $o_{ij} \in \mathbb{Q}$  and a basis  $\mathbf{B}$  for  $\mathbb{Z}^n \cap \ker(\mathbf{O})$ . The chemical reaction network  $\text{Proj}(\mathbf{O}, \mathbf{B})$  is defined by the species  $X_1, \dots, X_n$  and for each  $\mathbf{b} \in \mathbf{B}$  (viewed as columns of  $\mathbf{B}$ ), the reactions



For each reaction, we denote the molecular weight for species  $i$  on the reactant side of reaction  $k$  by  $\nu_{ik}(\mathbf{B})$  the molecular weight for species  $i$  on the product side of reaction  $k$  by  $\nu'_{ik}(\mathbf{B})$ , analogously to the notation used in the model of the biological system. The vectors of molecular weights for reaction  $k$  are denoted by  $\nu_k(\mathbf{B})$  and  $\nu'_k(\mathbf{B})$ , respectively.

Note that the basis can be constructed as follows. Let  $\mathbf{B}'$  be a basis for the nullspace of  $\mathbf{O}$  with rational coefficients. Then,  $\mathbf{B}$  can be obtained by scaling the columns of  $\mathbf{B}'$  such that each element is an integer.

To illustrate the construction of the chemical reactions in Definition 1, consider the following example given in [7].

**Example 1.** Consider an observation matrix

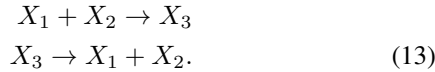
$$\mathbf{O} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad (11)$$

Note that  $\mathbf{O}$  is tidy since there exists

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (12)$$

such that  $\mathbf{O}\mathbf{M} = \mathbf{I}_{n_R \times n_R}$ .

Observe that a basis for the kernel is given by  $(1, 1, -1)^T$ , which has integer coefficients. The chemical reactions in Definition 1 are then given by



A realization of the trajectories for this SCRN with rate coefficients  $\kappa_{q,1} = q_3$  for the first reaction in (13) and  $\kappa_{q,2} = q_1 q_2$  for the second reaction in (13), where  $\mathbf{q} = [2, 2, 2]^T$  is shown in Fig. 1. For the purposes of illustration, initial concentrations  $\mathbf{X}(0) = [30, 50, 0]^T$ . These trajectories were obtained via the stochastic simulation method known as the Gillespie algorithm [14]. Observe the convergence of the trajectories to a stationary distribution.

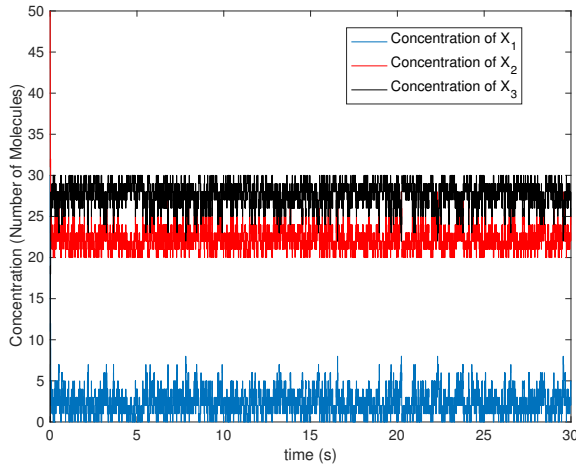


Fig. 1. Plot of a realization of the concentrations of  $X_1, X_2, X_3$  from Examples 1

A main result in [7] is that the distribution of the posterior  $\Pr(\mathbf{l}|\mathbf{r})$  with the product-Poisson prior can be sampled from the concentrations of species  $X_1, \dots, X_n$  in Definition 1 at a sufficiently large time  $t$ .

**Theorem 1.** Consider the chemical reaction system  $(\text{Proj}(\mathbf{O}, \mathbf{B}), \kappa)$  with stochastic mass action kinetics, associated message matrix  $\mathbf{M}$ , and the initial state  $\mathbf{n}(0) = \sum_{i=1}^{n_R} r_i \mathbf{m}_i$ , where<sup>2</sup>  $\kappa_k = \tilde{\mathbf{q}}^{\nu'_k(\mathbf{B})}$ . Then, the Bayesian posterior  $\Pr(\mathbf{l}|\mathbf{r})$  with the product-Poisson prior is the stationary distribution of the stochastic chemical reaction network.

Theorem 1 provides a means of obtaining samples from the posterior distribution  $P_{\mathbf{l}|\mathbf{r}}$ . This will be used in the sequel to obtain an estimate of  $P_1$  and to compute the coexistence constraint in (6). Ultimately, this will form a basis to optimize the modulation scheme in Section IV.

#### B. Distribution of $Z$ and the Coexistence Constraint

The chemical reaction network described in Section III-A provides a means of obtaining samples of the distribution of molecules  $\mathbf{l}$  outside the device conditioned on the observations  $\mathbf{r}$ . As detailed in Section II, we assume that one element of  $\mathbf{l}$  will correspond to the species  $Z$  to which the biological system is sensitive. Without loss of generality, we will take  $l_1 = Z$ . As such, the estimate of the distribution of  $P_{\mathbf{l}|\mathbf{r}}$  can then be used to obtain an estimate for the distribution of  $Z$  when the molecular communication system is not present, denoted by  $Z_{back}$ . This will be used to select a modulation scheme satisfying the coexistence constraint in (6).

We adopt the following approach to estimating the distribution of  $P_{Z_{back}}$ . For each observation  $\mathbf{r}$ , the stochastic chemical reaction network in Section III-A produces a sample from the posterior distribution  $P_{Z_{back}|\mathbf{r}}$ . To proceed, we recall the assumption from Section II that  $P_{Z_{back}}$  is a Poisson distribution, albeit with an unknown parameter  $q_Z$ .

To obtain an estimate of  $q_Z$ , for each observation  $\mathbf{r}$  we first estimate the posterior distribution  $P_{Z_{back}|\mathbf{r}}$  from  $K$  samples via the empirical distribution

$$\hat{P}_{Z_{back}|\mathbf{r}}(z) = \frac{1}{K} \sum_{i=1}^K \mathbf{1}_{\{Z_{r,i} \leq z\}}, \quad (14)$$

where each  $Z_{r,i}$  is a random variable drawn independently from  $P_{Z_{back}|\mathbf{r}}$ . The true value of  $Z_{back}$  corresponding to the observation  $\mathbf{r}$  is then estimated via the MAP estimator

$$\hat{Z}_{back} = \arg \max_z \hat{P}_{Z_{back}|\mathbf{r}}(z). \quad (15)$$

This procedure is then repeated for every realization of  $\mathbf{r}$ , from which we obtain a sequence of  $M$  estimates  $\hat{Z}_{back,1}, \dots, \hat{Z}_{back,M}$ . These estimates are used as a proxy for the true value of the realization of  $Z_{back}$  resulting in the observation  $\mathbf{r}$ . Under the assumption that  $Z_{back}$  is Poisson distributed, the parameter  $q_Z$  can then be obtained via the maximum likelihood estimate

$$\hat{q}_Z = \frac{1}{M} \sum_{k=1}^M \hat{Z}_{back,k}. \quad (16)$$

<sup>2</sup>To avoid any confusion, note that the values of  $\nu'_k(\mathbf{B})$  here correspond to the reaction network in Definition 1 and not the biological system.

The estimated distribution of  $Z_{back}$  is then

$$\hat{P}_{Z_{back}}(z) = e^{-\hat{q}_Z} \frac{(\hat{q}_Z)^z}{z!}. \quad (17)$$

Table I demonstrates the effect of increasing the number of samples  $M$  in order to estimate  $q_Z$ . In the table,  $q$  denotes the true value of the parameter and  $q_{prior}$  denotes the parameter used in the prior. The estimate of the posterior distribution for each observation  $\mathbf{r}$  were obtained using 10 independent observations of the SCRN at  $t = 30$  s. Observe that obtaining the samples from the posterior improves the estimate of  $q$  over the prior. As expected, the standard deviation (Std.) also decreases as the number of samples increases. The gap between the estimate and the true parameter is due to the fact that the posterior is obtained from only one observation of  $\mathbf{r}$ . Typically this can be improved by using Bayesian updating where after each observation  $\mathbf{r}$  the prior is replaced by the new posterior distribution. However, as the posterior distribution is not a product-Poisson prior it cannot be directly applied. It is an avenue of future work to address this issue.

TABLE I  
ESTIMATION OF  $\hat{q}_Z$

$q$	$q_{prior}$	Samples $M$	Average $\hat{q}_Z$	Std. $\hat{q}_Z$
4	2	10	2.98	0.57
4	2	20	2.95	0.33
4	2	50	2.94	0.27
4	2	100	2.94	0.17
4	3	10	3.4	0.52
4	3	20	3.43	0.39
4	3	50	3.43	0.26
4	3	50	3.41	0.19

With the estimated distribution  $\hat{P}_{Z_{back}}$  in hand, we can now compute the coexistence constraint for a given distribution of information molecules at the biological system. Let  $Z_{info}$  be the distribution of molecules that are transmitted by the molecular communication system that arrive at the biological system. Note that  $Z_{info}|_{X=0} \sim \text{Pois}((1-p)\lambda_0)$  and  $Z_{info}|_{X=1} \sim \text{Pois}((1-p)\lambda_1)$ , under the assumption in Section II that all emitted molecules by the communication system that do not reach the receiver are bound to the biological system. The Kullback-Leibler divergence required to compute the coexistence constraint is then given by

$$D(\hat{P}_{Z_{back}} || \hat{P}_{Z_{back}+Z_{info}}) = \sum_{j \in \mathbb{Z}_{\geq 0}} \hat{P}_{Z_{back}}(j) \log \left( \frac{\hat{P}_{Z_{back}}(j)}{\sum_{k \in \mathbb{Z}_{\geq 0}} P_{Z_{info}}(j-k) \hat{P}_{Z_{back}}(k)} \right). \quad (18)$$

#### IV. OPTIMIZING RELIABILITY FOR BINARY MODULATION

In this section, we address the problem of optimizing the modulation scheme for the molecular communication system detailed in Section II subject to the coexistence constraint in (6). Denote the probability of error given transmission rates

$\lambda_0, \lambda_1$  by  $P_e(\lambda_0, \lambda_1)$ . Then, we seek to solve the optimization problem

$$\min_{\lambda_{\min} \leq \lambda_0 < \lambda_1 \leq \lambda_{\max}} P_e(\lambda_0, \lambda_1) \quad (19)$$

subject to the constraint

$$D(\hat{P}_{Z_{back}} || \hat{P}_{Z_{back}+Z_{info}}) < \delta. \quad (20)$$

Note that since the transmitter emits a Poisson distributed number of molecules, it follows that the optimal decision rule is obtained from the binary hypothesis test

$$\begin{aligned} \mathcal{H}_0 : Y &\sim \text{Pois}(p\lambda_0) \\ \mathcal{H}_1 : Y &\sim \text{Pois}(p\lambda_1). \end{aligned} \quad (21)$$

It is then straightforward to obtain the decision rule via the likelihood ratio test [15]. In particular, given an observation of  $y$  molecules at the receiver, the decision rule is

$$y \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \frac{p(\lambda_1 - \lambda_0)}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} =: \tau. \quad (22)$$

Under this decision rule, the probability of error is given by

$$P_e(\lambda_0, \lambda_1) = \frac{1}{2} \sum_{i=0}^{\lfloor \tau \rfloor} e^{-\lambda_1 p} \frac{(\lambda_1 p)^i}{i!} + \frac{1}{2} \left( 1 - \sum_{i=0}^{\lfloor \tau \rfloor} e^{-\lambda_0 p} \frac{(\lambda_0 p)^i}{i!} \right). \quad (23)$$

The probability of error subject to the coexistence constraint can then be solved numerically by optimizing  $\lambda_1, \lambda_0 \in \mathbb{R}_+$ . Next, we examine the tradeoffs introduced by the coexistence constraint and the estimation procedure detailed in Section III.

#### A. Numerical Results

We first investigate the tradeoff between the probability of error and the coexistence constraint under the assumption the distributions  $P_{Z_{back}}$  and  $P_{Z_{back}+Z_{info}}$  are perfectly known. This provides insight into the impact of the coexistence constraint on communication.

Fig. 2 plots the probability of error in (23) for varying values of  $\delta$  in (6) and probability  $p$  that a given molecule reaches the receiver. In this figure, it is assumed that the transmitter knows the true distribution of  $Z_{back}$ , which is Poisson with parameter  $q = 2$ . The values of  $\lambda_0, \lambda_1$  are constrained to lie in  $[0, 20]$ . The key observation is that as the value of  $\delta$  increases the probability of error reduces. However, an error floor is reached due to the presence of the upper bound on the value of  $\lambda_1$ . This upper bound incorporates constraints on the ability for the transmitter to generate the required number of molecules that are to be emitted. Moreover, observe that as the success probability  $p$  is increased, the error floor significantly decreases. As such, improving the quality of the channel can lead to very large reductions in the probability of error.

The estimation procedure in Section III will inevitably lead to estimation errors in the parameter  $q_Z$ . As such, it is important to understand the effect of the estimation error on

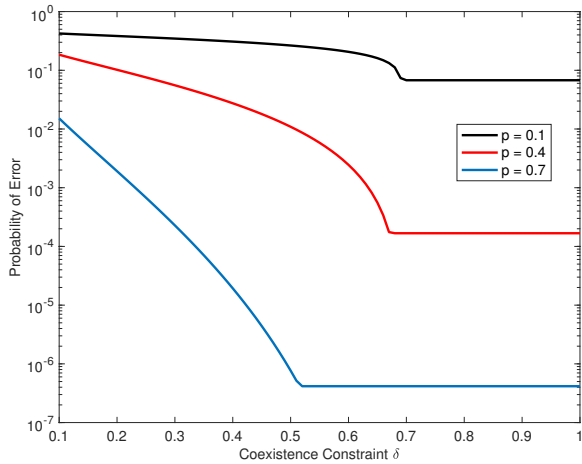


Fig. 2. Plot of the probability of error for varying values of the coexistence constraint  $\delta$ .

the probability of error and the coexistence constraint. To this end, Fig. 3 plots the difference between the Kullback-Leibler divergence with the optimized modulation scheme (based on the estimated  $q_Z$ ) and  $\delta$  using the true distribution of  $Z_{back}$ , called the tolerance. In the figure,  $\delta = 0.2$  and  $q = 2$ . Observe that for  $\hat{q}_Z$  (estimated  $q_Z$ ) larger than 2, the coexistence constraint is not satisfied. As such, in scenarios where the estimated  $\hat{q}_Z$  can be significantly larger than the true value  $q$  it is highly desirable to optimize the modulation scheme using a smaller value of  $\delta$ .

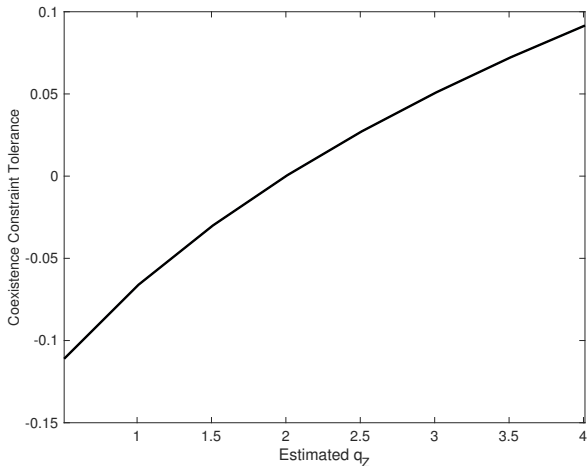


Fig. 3. Plot of the tolerance in the coexistence constraint for varying values of the estimated parameter  $q_Z$ , with the true value  $q = 2$ .

## V. CONCLUSION

While a range of different techniques have been proposed to improve the performance of molecular communication systems, almost all assume that there is no external biological

systems present. When such biological systems are in fact present—likely in applications such as nanomedicine and advanced manufacturing—the design of the molecular communication system must be adapted to ensure that biological function is not disrupted. In this paper, we have studied molecular communication systems subject to an additional coexistence constraint. This coexistence constraint is formalized in terms of a Kullback-Leibler divergence between the distribution on the number of molecules that can bind to the biological system, with and without the presence of the communication system.

We have shown that this constraint can significantly reduce the performance of the molecular communication system by introducing error floors in the probability of error. We have also made initial steps towards resolving the problem of how to estimate the distributions on the number of molecules. In particular, by adapting techniques from stochastic chemical reaction networks, part of the estimation problem can be performed by a bio-circuit. This is highly desirable from the perspective of reducing the complexity of the transmitting device.

## REFERENCES

- [1] T. Nakano, A. Eckford, and T. Haraguchi, *Molecular Communication*. Cambridge University Press, 2013.
- [2] Y. Chahibi, “Molecular communication for drug delivery systems: a survey,” *Nano Communication Networks*, vol. 11, pp. 90–102, 2017.
- [3] D. Das and T. Veziroglu, “Hydrogen production by biological processes: a survey of literature,” *International Journal of Hydrogen Energy*, vol. 26, no. 1, pp. 13–28, 2001.
- [4] A. San and R. Maclean, “Fitness costs of plasmids: a limit to plasmid transmission,” *Microbiology spectrum*, vol. 5, no. 5, 2017.
- [5] M. Egan, T. Mai, T. Duong, and M. Di Renzo, “Coexistence in molecular communications,” *Nano Communication Networks*, vol. 16, pp. 37–44, 2018.
- [6] F. Farsad, H. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, “A comprehensive survey of recent advancements in molecular communication,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1887–1919, 2016.
- [7] M. Virinchi, A. Behera, and M. Gopalkrishnan, “A stochastic molecular scheme for an artificial cell to infer its environment from partial observations,” in *Proc. International Conference on DNA-Based Computers*, 2017.
- [8] A. Marcone, M. Pierobon, and M. Magarini, “A parity check analog decoder for molecular communication based on biological circuits,” in *Proceedings of Conference on Computer Communications (INFOCOM)*, 2017.
- [9] R. Mosayebi, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, “Receivers for diffusion-based molecular communication: Exploiting memory and sampling rate,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 12, pp. 2368–2380, 2014.
- [10] K. Srinivas, A. Eckford, and R. Adve, “Molecular communication in fluid media: the additive inverse Gaussian noise channel,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4678–4692, 2012.
- [11] U. Alon, M. Surette, N. Barkai, and S. Leibler, “Robustness in bacterial chemotaxis,” *Nature*, vol. 397, no. 6715, pp. 168–171, 1999.
- [12] D. Anderson, G. Craciun, and T. Kurtz, “Product-form stationary distributions for deficiency zero chemical reaction networks,” *Bulletin of Mathematical Biology*, vol. 72, no. 8, pp. 1947–1970, 2010.
- [13] T. Cover and J. Thomas, *Elements of Information Theory, Second Edition*. John Wiley and Sons, Inc., 2006.
- [14] D. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [15] H. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 1994.