A Molecular Communication Scheme to Estimate the State of Biochemical Processes on a Lab-on-a-Chip

Bayram Cevdet Akdeniz, Malcolm Egan

A Molecular Communication Scheme to Estimate the State of Biochemical Processes on a Lab-on-a-Chip

Bayram Cevdet Akdeniz
University of Oslo, Centre for Bioinformatics
Oslo, Norway
b.c.akdeniz@mn.uio.no

Malcolm Egan
Univ. Lyon, INSA Lyon, Inria, CITI
Lyon, France
malcom.egan@inria.fr

ABSTRACT

A key application of advanced spectroscopy methods is to estimate equilibrium states of biochemical processes *in situ* and *in vivo*. Nevertheless, an often present difficulty is the requirement that the biochemical process and its environment (such as the substrate) satisfy special conditions. One means of resolving this issue is to communicate information about the equilibrium states of the biochemical process to another location, supported via microfluidic channels within a lab-on-a-chip. In this paper, we develop a signaling strategy and estimation algorithms for equilibrium states of a biochemical process. For a toggle-switch circuit model important in cellular differentiation studies, we study via simulation the tradeoff between the rate of obtaining spectroscopy measurements and the estimation error, providing insights into requirements of spectroscopy devices for high-throughput biological assays.

CCS CONCEPTS

• **Applied computing** → *Health care information systems*; • **Computing methodologies** → Model verification and validation; **Mixture modeling**.

KEYWORDS

molecular communication, microfludics, lab-on-a-chip

1 INTRODUCTION

An important question in molecular biology is the dependence of dynamics for biochemical processes on environmental conditions, such as temperature of chemical composition. In practice, to address this question, it is necessary to perform an assay; i.e., measurements of chemical concentrations under varying conditions [14].

Ideally, biochemical assays are performed *in situ* or *in vivo*, allowing observation of the true dynamics. Such measurements require advanced spectroscopy methods, including Raman or atomic force microscopy (see, e.g., [24, 32]). However, these methods often require special environmental conditions, such as well-defined, flat and non-reactive substrates [13].

When it is not feasible for in situ spectroscopy, a natural solution is to communicate information about the equilibrium state of the process to another location and perform measurements there. One approach to implementing such distributed observation is to house the biochemical process and spectroscopy on a single chip, known as a lab-on-a-chip (LoC) or micro total analysis system [10]. In this setting, communication of information about the equilibrium state of the biochemical process can be supported via microfluidic channels.

A number of previous works have investigated microfluidic systems from a molecular communication perspective. In [20], the pulse width, pulse delay and pulse amplitude were studied under a convection-diffusion-reaction channel. In [3, 8], microfluidic circuits were developed in order to produce a desired pulse shape at the transmitter and perform detection at the receiver. In [7, 17], droplet-based microfluidic channels were characterized and the information capacity evaluated. Further work investigating the information capacity of microfluidic channels has been carried out in [4, 5, 11, 15, 19], often under the assumption of laminar flow modeled by convection-diffusion equations.

In this paper, we consider a biochemical process under observation contained in a chamber, connected to spectroscopy chamber—where the quantity of a given chemical species can be directly observed—via a microfluidic channel, illustrated in Fig. 1. Based on the observations from the spectroscopy chamber, the aim is to reliably estimate the equilibrium state of the biochemical process under investigation.
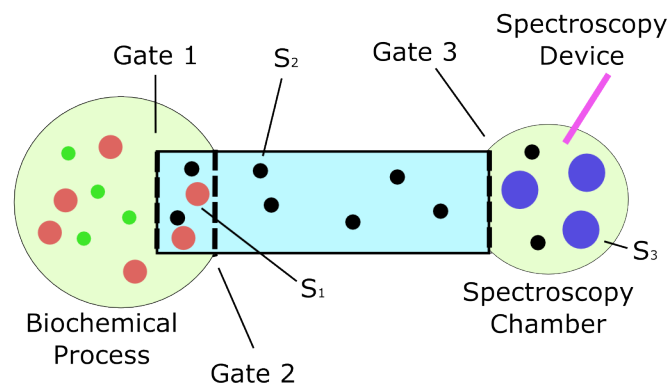
Figure 1: System Model.

While microfluidic channels often support laminar flow, it is also true that the small scale of the channel can facilitate rapid diffusion [29]. For example, diffusion-based microfluidics has a long history in validating stochastic diffusion models [27]. As such, we focus microfluidic channels with dynamics purely driven by reactions and diffusion, without convection.

The absence of convection often means that, under certain reversibility conditions on the chemical reactions in the system, that the statistics for the quantity of molecules in the spectroscopy

chamber converges to an steady state (not to be confused with the equilibrium state of the biochemical process under investigation). Moreover, this steady state can be readily characterized, depending only on the total quantity of molecules in the system, the reaction rates, and the volume [2]; all of which can be readily estimated.

As such, it is feasible to utilize observations from the spectroscopy chamber in order to estimate the equilibrium state of the biochemical process when the set of equilibrium states is unknown. In order to estimate these unknown equilibrium states, we develop an estimation procedure by exploiting a Gaussian mixture model. This is feasible due to the fact, shown in the sequel, that the steady state statistics for the quantity of molecules in the spectroscopy chamber can be well approximated via a Gaussian distribution.

We illustrate our approach via an example based on a self-actuating toggle switch circuit, which plays an important role in cellular differentiation [21]. A key feature of the toggle switch circuit is that it can admit tristability; that is, the circuit can produce three distinct equilibrium states. We numerically investigate the performance of our approach in estimation of the set of unknown states. In particular, we obtain tradeoffs between the number of required samples on the probability an equilibrium state is misidentified (called a clustering error), and on the mean-square error of the estimated equilibrium states. These results provide guidelines for the required rate of spectroscopy measurements for high-throughput biological assays.

2 SYSTEM MODEL

2.1 Biochemical Assay Setup

Consider a microfluidic chip consisting of two chambers and a microchannel, as illustrated in Fig. 1. The first chamber houses the biochemical process under investigation. This chamber is separated from the microchannel by two gates. The first gate is opened once the biochemical process reaches an equilibrium state, allowing molecules S_1 from the first chamber to diffuse inside. The second gate acts as a filter, allowing only a further information-carrying molecule S_2, distinct from any of the molecules produced by the biochemical process, to pass through. After a short period of time, the first gate is closed, allowing no further molecules from the biochemical process to enter the beginning of the microchannel. We expect that such gates can be implemented using microfiltration techniques [31].

When molecules of species S_1 from the biochemical process enter the beginning of the microchannel, they are able to react to form information-carrying molecules of species S_2 (detailed in Sec. 2.2). The molecules of S_2 can then pass through the second gate and diffuse through the remainder of the microchannel towards the spectroscopy chamber.

Within the spectroscopy chamber, the molecules of S_2 can then react to form molecules of species S_3 (detailed in Sec. 2.2). Between the microchannel and the spectroscopy chamber is a third gate, which prevents molecules of species S_3 to pass back into the microchannel, while allowing molecules of species S_2.

The spectroscopy chamber is equipped with a spectroscopy device able to passively count how many molecules of species S_3 are present. The precise form of the spectroscopy device depends on the species S_3. The spectroscopy device then passes the count to an external processor, which is used to estimate the quantity of molecules of S_1 which were present initially at the beginning of the microchannel between the first and second gates.

A concrete example of this setup can be developed for detection of autoinducer molecules arising in bacteria colonies for the purpose of quorum sensing. In this case, each bacteria forms a biochemical switch; able to produce either a small or large quantity of autoinducer molecules depending on the density of the bacteria colony.

In order to observe the quantity of autoinducer, Raman spectroscopy has been exploited in [24] for bacteria colonies *in situ*. This previous work does not consider distributed spectroscopy; i.e., the spectroscopy is performed in the same chamber as the biochemical process. Nevertheless, it is feasible that, equipped with a LoC, the spectroscopy could be performed in a separate chamber. In fact, Raman spectroscopy has been proposed as a means of cooperative in-vivo sensing in [16]. We also highlight that a distributed implementation is likely to be necessary for other systems, where the spectroscopy method may perturb the biochemical process or the detection chamber requires special preparation (such as in force spectroscopy [13]).

2.2 Dynamics of the Microfluidic Channel

The portion of the first chamber used to produce signaling molecules, the microchannel, and the detection chamber are assumed to form a domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ with smooth boundary $\partial\Omega$. Consider the discretization of Ω into N volume elements (voxels) each of volume V_{vox}, with the subdomain forming voxel i denoted by $\mathcal{V}_i$, $i = 1, \ldots, N$. Here, volume is interpreted as length in $\mathbb{R}^1$, area in $\mathbb{R}^2$, and volume in $\mathbb{R}^3$.

The portion of the first chamber housing the biochemical process used to produce the signaling molecules has volume V_{Tx}. The production of signaling molecules is achieved via the unimolecular reactions

$$S_1 \rightarrow S_2$$
$$S_2 \rightarrow S_1. \tag{1}$$

In particular, the transmitter produces information-carrying molecules of species S_2 by the first reaction in (1). In general, a more complex reaction pathway may be present; however, we assume in this case that the intermediate reactions occur rapidly as is common, for example, in enzyme-based reactions. We also note that unimolecular reactions are capable of modeling the dynamics of a range of biochemical systems [28].

We assume that molecules of species S_1 produced in the transmitter are not capable of diffusing into the channel. On the other hand, this is possible for species S_2.

In the detection chamber, with volume V_{Rx} (not necessarily the same as V_{Tx}), molecules of species S_2 are able to generate molecules of species S_3 via

$$S_2 \rightarrow S_3$$
$$S_3 \rightarrow S_2. \tag{2}$$

The molecules of S_3 are then used for the purpose of spectroscopy (e.g., Raman or force spectroscopy) in the detection chamber. We assume that the spectroscopy is passive, which means that no molecules of S_3 are removed by the detection process. As such, the detection process can be viewed as a passive receiver, which

has been considered in previous work on molecular communications in different contexts (see, e.g., [23]).

In order to capture the effect of small quantities of each chemical species in the system (i.e., S_1, S_2, S_3), we consider a stochastic model for the dynamics. To formally describe the setup, we introduce the following notation. Let $M_i^l(t)$, $l = 1, 2, 3$, $i = 1, \ldots, N$ denote the random variable for the number of molecules of species S_1, S_2 or S_3 in voxel i at time t. Denote $M_i(t) = [M_i^1(t), M_i^2(t), , M_i^3(t)]$ as the state vector in voxel i and the matrix consisting of all state vectors by $M(t) = [M_1(t), \ldots, M_N(t)]$. The probability that $M(t)$ has value m at time t is then denoted by

$$P(m, t) = \Pr(M(t) = m | M(0) = m_0), \qquad (3)$$

where $M(0)$ is the initial quantity of molecules of each species in each voxel. In the present context, the $M(0)$ is dependent on the equilibrium state of the biochemical process under observation.

Since each reaction is unimolecular, it follows that in each reaction the number of molecules of the three species involved can only increase or decrease by one. Let 1_i^l be the state where the number of molecules in all voxels is zero, except for species l in voxel i. That is, $M(t) + 1_i^l$ means that the number of molecules of species l in voxel i is increased by one.

A popular model for stochastic dynamics of molecules is the reaction-diffusion master equation (RDME) [18], also utilized in the context of molecular communications in [6]. In this model, the diffusive jump rate is denoted by κ_{ij}^l for each individual molecules of the l-th species moving from voxel j into voxel i, with $\kappa_{ii} = 0$, $i = 1, \ldots, N$. In particular, the probability per unit time that a molecule of S_l diffuses from voxel j to voxel i at time t is given by $\kappa_{ij}^l M_j^l(t)$. We expect that in many microfluidic systems, κ_{ij}^l is constant for a given species S_l over all voxels i, j. Nevertheless, it is also possible to consider spatially inhomogeneous diffusion [2].

In the case of mass-action kinetics and first-order reactions, the probability per unit time that a molecule of S_l in voxel i reacts at time t is given by $a_i^l M_i^l(t)$ with rate constants a_i^l. In general, the reaction rate is dependent on the voxel index. The net change of each chemical species due to the reaction with substrate S_l is expressed via the column vector $\nu_l \in \mathbb{N}^2$. The term $\nu_l 1_i$ indicates that $M(t)$ changes by ν_l in the i-th voxel. In order to model production of S_1, S_2, S_3 in the transmitter and receiver, we assume that for the voxels i comprising the transmitter and the receiver $a_i^l = a^l$, while $a_i^l = 0$ for voxels comprising the channel.

In the RMDE model, the probability distribution $P(m, t)$ evolves according to the system of differential equations given by

$$\frac{dP(m, t)}{dt}$$
$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{2} \left(\kappa_{ij}^l (m_j^l + 1) P(m + 1_j^l - 1_i^l, t) \right.$$
$$\left. - \kappa_{ji}^l m_i^l P(m, t) \right) + \sum_{i=1}^{N} \sum_{l=1}^{2} \left(a_i^l (m_i^l + 1) P(m - \nu_l 1_i, t) \right.$$
$$\left. - a_i^l m_i^l P(m, t) \right). \qquad (4)$$

The sums in (4) correspond to net changes in probability per unit time due to diffusion and reactions, respectively. We refer the reader to [30] for further details.

The system of ordinary differential equations in (4) corresponds to the Kolmogorov forward equation for a continuous-time Markov chain; that is, the evolution of the system state is Markovian. In our setting, the Markov chain corresponding to the RDME is irreducible and positive recurrent. Therefore, a stationary distribution exists and is given by [9]

$$\pi(m) = \lim_{t \to \infty} \Pr(M(t) = m | M(0) = m_0). \qquad (5)$$

3 MOLECULAR COMMUNICATION SCHEME

3.1 Signaling

We are concerned with the scenario that the system is reset for each experiment; i.e., when the biochemical process is exposed to new environmental conditions, such as concentrations of input molecules or temperatures. In particular, no molecules of species S_1, S_2 or S_3 are present in the channel nor the spectroscopy chamber at time $t = 0$.

After an external stimulus is introduced, we assume that the biochemical process converges rapidly to an equilibrium state indexed by an element of $\{1, \ldots, K\}$. For example, in the case of a unique equilibrium state (e.g., if the process is governed by a deficiency zero chemical reaction network [12]), $K = 1$. On the other hand, if the biochemical process is bistable (e.g., in quorum sensing [22]), $K = 2$. Larger numbers of equilibrium states can occur if the underlying chemical reaction network has more complex structure [21], as will be considered in Sec. 5.

An equilibrium state corresponds to a concentration of S_1 denoted by C_i, $i = 1, \ldots, K$. As such, the quantity of S_1 in the entrance to the microchannel is given by

$$\Delta^i = C_i V_{Tx}. \qquad (6)$$

The index of the equilibrium state is denoted by $i \in \{1, \ldots, K\}$.

At a time δ shortly after the biochemical process reaches equilibrium, a quantity of S_1 is present in the beginning of the microchannel, depending on equilibrium state of the biochemical process. In particular, assuming that the equilibrium state of the biochemical process is indexed by i,

$$N_{Tx,1}(\delta) = N_{Tx,1} + \Delta^i, \qquad (7)$$

where Δ^i corresponds to the quantity of molecules of S_1 in the beginning of the microchannel, as defined in (6). We note that $\delta > 0$ is chosen to be a sufficiently small period of time such that no reactions occur in the beginning of the microchannel.

The key idea behind our approach is that for sufficiently large T_s, the total number of molecules of species S_2 and S_3 in the detection chamber at the time of sampling will be approximately drawn from the stationary distribution of the RDME. As such, if the stationary distribution is known, then reliable detection of the equilibrium state for the biochemical process can be obtained.

3.2 Statistics for the Quantity of S_3

We now seek to obtain a good approximation for the statistics for the quantity of S_3; i.e., $N_3(T_s)$. Suppose that the measurement process results in the production of Δ^i at the transmitter. Then, by the

analytical and empirical evaluation in [2], the following assertion provides an accurate characterization for sufficiently large T_s.

ASSERTION 1. *Let $N_3(T_s|i)$ denote the number of molecules of* S_3 *in the system at time T_s given an equilibrium state of the biochemical system i corresponding to a measurement of Δ^i defined in (6). Then,*

$$N_3(T_s|i) \sim \mathcal{N}(\mu^i, \mu^i), \tag{8}$$

where $\mu^i > 0$ is a known constant, only dependent on the volume of the enclosing container and not the specific geometry, and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian law with mean μ and variance σ^2. In particular,

$$\mu^i = \frac{a^1}{a^2} \frac{\Delta^i \frac{V_{\mathrm{Rx}}}{NV_{\mathrm{vox}}}}{1 + \frac{a^1}{a^2} \frac{V_{\mathrm{Tx}} + V_{\mathrm{Rx}}}{NV_{\mathrm{vox}}}}. \tag{9}$$

Under the assumption that $a^1 = a^2$ and $V_{\mathrm{Tx}} = V_{\mathrm{Rx}}$,

$$\mu^i = \frac{\Delta^i V_{\mathrm{Rx}}}{NV_{\mathrm{vox}} + 2V_{\mathrm{Rx}}}. \tag{10}$$

So far, we have assumed that T_s is sufficiently large such that the steady state of the RDME is approximately reached. However, it is also desirable to obtain multiple observations per symbol in order to improve the detection performance. In this case, the time interval between samples is smaller and independence cannot be guaranteed from mixing properties of the RDME Markov chain. Nevertheless the convergence is rapid, as can be verified numerically.

Let N_s be the maximum number of independent samples that the receiver can observe per measurement; e.g., due to time constraints. Under the assumption of independent observations and the Gaussian law in Assertion 1, it is possible to reduce the variance by averaging these observations via

$$N_{\mathrm{Rx},3} = \frac{1}{N_s} \sum_{j=1}^{N_s} N_{\mathrm{Rx},3}(t_j|i) \tag{11}$$

where $t_j \in \{t_1, t_2, \ldots, t_{N_s}\}$, where $t_j > T_s$. Then the corresponding distribution of the averaged observations $N_{\mathrm{Rx},3}(T_s|i)$ can be written as

$$N_{\mathrm{Rx},3}(T_s|i) \sim \mathcal{N}\left(\mu^i, \frac{\mu^i}{N_s}\right). \tag{12}$$

4 BIOLOGICAL ASSAYS VIA EQUILIBRIUM SIGNALING

The aim of this work is to establish a communication mechanism to detect the equilibrium state of the biochemical process from a distributed detection chamber. While the problem bears similarities with standard formulations of molecular communication systems—namely, that an element discrete set of states (or messages) is to be detected—there are two key differences.

One difference is that each equilibrium state is not in general equally likely to arise. This is due to the fact that the equilibrium state depends on the environmental input of the biochemical process and different equilibrium states are associated to a larger set of the inputs. The consequence of this is that the average probability

of error is not an appropriate performance metric. Instead, the error for each equilibrium state is a more useful metric.

Perhaps the most significant difference is that the set of equilibrium states is not necessarily known before the experiment begins. This may be due to limited pre-trials, which were only able to identify a subset of the equilibrium states of the biochemical process. In this section, we develop a detection strategy, when both the number and value of the states are not known.

4.1 Equilibrium State Estimation

Even if none of the equilibrium states are known in a system, using Assertion 1, the concentration of the biochemical system can be estimated. We seek to both identify the set of equilibrium states and also estimate the equilibrium states arising from each measurement. To do so, we model the observations from a number of experiments (corresponding to changes in the environment) as a Gaussian mixture [25]. This is motivated by the fact that the observation in the detection chamber is Gaussian conditioned on the equilibrium state, with statistics given in (12).

Recall that a Gaussian mixture model is defined by a probability density function of the form

$$f_{GM}(x; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma})) = \sum_{i=1}^{N} \pi_i g(x; \mu_i, \sigma_i^2), \tag{13}$$

where

$$g(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{14}$$

In the Gaussian mixture model, the parameters $\boldsymbol{\mu}$ correspond to the estimated values of the equilibrium states, $\boldsymbol{\sigma}$ corresponds to the standard deviation associated to each cluster, and $\boldsymbol{\pi}$ corresponds to the weight associated to each cluster, which is related to how likely it is that a point belongs to each cluster.

To estimate the Gaussian mixture model parameters, the standard method is based on the expectation-maximization algorithm. Suppose D measurements—corresponding to different perturbations of the biochemical process, distinct from the number of samples N_s—are taken. Then under the assumption there are K clusters, the inner loop in Algorithm 1 details the expectation-maximization procedure (i.e., the E-step and the M-step).

The parameters $\hat{\gamma}_j^i$ estimate the likelihood that measurement i is associated to cluster j, which provides the basis for clustering the observations. In particular, measurement i is associated to the cluster $j = \arg\max_k \hat{\gamma}_k^i$.

However, when the biochemical process under observation is not well understood, it may not be clear how many equilibrium points and hence which cluster size K should be chosen. In order to select the cluster size K, we exploit the Akaike information criterion (AIC) [1], detailed in line 9 of Algorithm 1. In particular, AIC is given by

$$AIC = 2K - 2L \tag{15}$$

where L represents the maximized log likelihood function given in line 8 of Algorithm 1. As detailed in Algorithm 1, the GMM is applied for different values of K (corresponding to a different number of equilibrium states). The value of K then corresponds to the minimum AIC.

The AIC can be replaced by the Bayesian Information Criterion (BIC) [26], given by

$$BIC = K \log(D) - 2L. \tag{16}$$

This can be implemented in Algorithm 1 by replacing the AIC in line 9 with the BIC.

Algorithm 1 GMM Clustering Algorithm

1: For $K = 1 : K_{max}$
2: **Initialize:** $\hat{\mu}_{1:K}, \hat{\sigma}^2_{1:K}, \hat{\pi}_{1:K}$
3: While not converged
4: $\phi_j \sim \mathcal{N}(\hat{\mu}_j, \hat{\sigma}^2_j) \qquad j = 1 : K$
5: E-step : $\hat{\gamma}^i_j = \frac{\hat{\pi}_j \phi_j(x_i)}{\sum_{k=1}^K \hat{\pi}_k \phi_k(x_i)} \qquad j = 1 : K, \quad i = 1 : D$
6: M-step : $\hat{\mu}_j = \frac{\sum_{i=1}^D \hat{\gamma}^i_j x_i}{\sum_{i=1}^D \hat{\gamma}^i_j}$

$\qquad \hat{\sigma}^2_j = \frac{\sum_{i=1}^D \hat{\gamma}^i_j (x_i - \hat{\mu}_j)^2}{\sum_{i=1}^D \hat{\gamma}^i_j} \qquad j = 1 : K$

$\qquad \pi_j = \frac{\sum_{i=1}^D \hat{\gamma}^i_j}{D}$
7: End while
8: $L = \sum_{i=1}^D \log \sum_j^K \pi_j \phi_j(x_i)$
9: $AIC_K = 2K - 2L$
10: End for
11: Return $K_{opt} = \min_K AIC_K, \quad \hat{\mu}_{1:K_{opt}}$

---

5 EVALUATION IN REALISTIC BIOCHEMICAL PROCESSES

5.1 Self Activating Toggle Switch Circuits

Gene regulatory networks play a key role in cellular differentiation. In particular, the decision networks often heavily rely on a regulatory motif in which the proteins of two genes act as mutually inhibiting transcription factors [21]. It is often the case that one or both genes also act as self-activating transcription factors, leading to a self-activating toggle switch. A key example is the (CDX2-OCT3/4) pair [21], which is associated with the differentiation of embryonic stem cells.

An important feature of self-activating toggle switches is that they can admit tristability. That is, three distinct stable equilibrium states are present. Moreover, the occurrence of distinct equilibrium states varies from one equilibrium state to another. As a consequence, some equilibrium states can be much rarer than others.

We consider an experiment where the goal is to identify the equilibrium states corresponding to different initial conditions of a self-activating toggle switch, when none, some, or all equilibrium states are known. In particular, the system consists of two transcription factors and two noncompetitive binding sites, which facilitate transcription of proteins A and B.

The self-activating toggle switch is governed by the system of ordinary differential equations [21]

$$\frac{dc_A}{dt} = G_A(c_A, c_B) - k_A c_A$$
$$\frac{dc_B}{dt} = g_B(c_B, c_A) - k_B c_B, \tag{17}$$

where the mean transcription rate of X is given by

$$G_X(c_A, c_B)$$
$$= (g_{X,AB} + g_{X,A} H^-(c_B)) H^+(c_A) + g_{X,B} + g_{X,0} H^-(c_B), \tag{18}$$

with transcription rates for each promoter state given by $g_{X,A} \gg g_{X,0}, g_{X,B} \ll g_{X,0}, g_{X,AB} \approx g_{X,0}$, and the dynamics are governed by the Hill functions

$$H^-(c_A) = \frac{1}{1 + c_A^{n_A}}$$
$$H^+(c_A) = 1 - H^-(c_A). \tag{19}$$

Fig. 2a plots the equilibrium states arising from the model in (17) with parameters detailed in the figure. Observe that the system is tristable, with three equilibrium states that occur with different frequencies. Moreover, each equilibrium state does not occur with the same frequency.

We now evaluate the performance of each algorithm in Section 4 for the biochemical process defined by (17) with parameters the same as given in Fig. 2. Fig. 2b shows the probability that an observation is incorrectly clustered. Since, no equilibrium states are initially known to the receiver, In order to estimate the equilibrium states and identify which equilibrium state each observation should be assigned to, we utilize Algorithm 1. In this case, the clustering error is the proportion of points that are not associated with the cluster with mean closest to the true equilibrium state for the point. In Fig. 2c the estimation error of the equilibrium states are given. Observe that the estimation error for all states as the number of samples increases. As expected, the estimation error measured by the normalized mean-square error (NMSE) decreases as the number of samples increases.
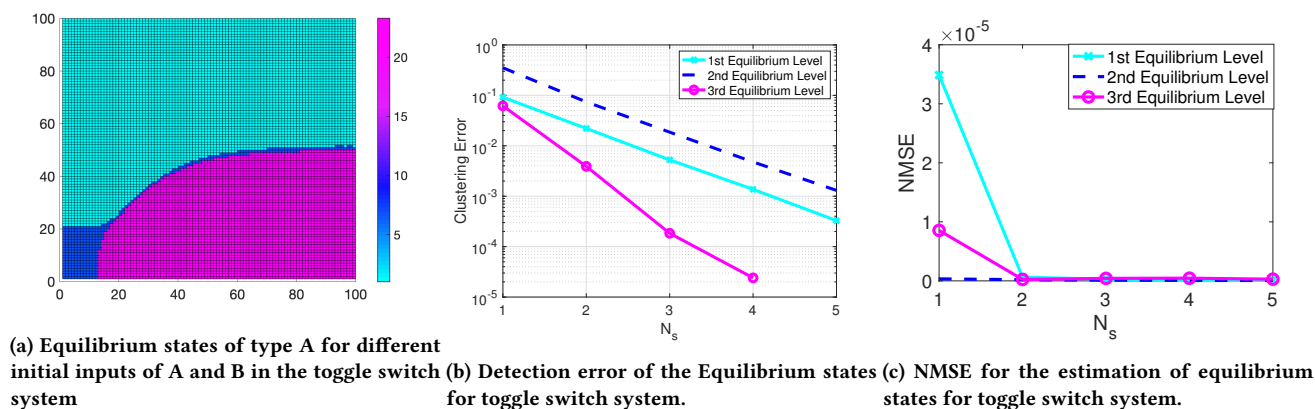
6 CONCLUSION

Advanced spectroscopy methods provide a means of passively estimating the number of molecules of a given chemical species. One application is to investigate the equilibrium behavior of a range of biochemical processes. However, spectroscopy devices may require special conditions to operate, which are not consistent with normal behavior of the biochemical process. In this case, it is desirable to exploit LoC technology to measure chemical signals in a location away from where the biochemical process.

In order to do so, it is necessary to introduce a communication channel and a means of characterizing the statistics for the number of molecules in the external spectroscopy chamber. In this paper, we have proposed an approach exploiting the steady state behavior of reaction-diffusion systems. We have shown that this approach yields reliable estimates of equilibrium states, even if the full set of possible equilibrium states is not known.

REFERENCES

[1] H. Akaike. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (1974), 716–723.

[2] B. Akdeniz, M. Egan, and B.Q. Tang. 2020. Equilibrium signaling: molecular communication robust to geometry uncertainties. Accepted for publication in IEEE Transactions on Communications, preprint available https://hal.archives-ouvertes.fr/hal-02536318.

[3] D. Bi, Y. Deng, M. Pierobon, and A. Nallanathan. 2019. Chemical reactions-based microfluidic transmitter and receiver for molecular communication. *arXiv:1908.03441v1* (2019).

(a) Equilibrium states of type A for different initial inputs of A and B in the toggle switch system

(b) Detection error of the Equilibrium states for toggle switch system.

(c) NMSE for the estimation of equilibrium states for toggle switch system.

Figure 2: Detection error of the equilibrium states for Self Activating Toggle Switch Circuits with $n_A = 4$, $k_A = 4s^{-1}$, $k_B = 4s^{-1}$, $g_{A,A} = 5s^{-1}$, $g_{A,0} = 1s^{-1}$, $g_{A,B} = 0.1s^{-1}$, $g_{A,AB} = g_{A,0}$, $g_{B,A} = 0.1s^{-1}$, $g_{B,0} = 1s^{-1}$, $g_{B,B} = 10s^{-1}$ $g_{B,BA} = g_{B,0}$ **resulting in three different equilibrium points** $\mu^1 = 1.01$, $\mu^2 = 6.64$ $\mu^3 = 23.29$.

[4] A.O. Bicen and I.F. Akyildiz. 2013. System-theoretic analysis and least-squares design of microfluidic channels for flow-induced molecular communication. *IEEE Transactions on Signal Processing* 61, 20 (October 2013), 5000–5013.

[5] A Ozan Bicen, Janne J Lehtomäki, and Ian F Akyildiz. 2018. Shannon meets Fick on the microfluidic channel: diffusion limit to sum broadcast capacity for molecular communication. *IEEE Transactions on NanoBioscience* 17, 1 (January 2018), 88–94.

[6] C.T. Chou. 2013. Extended master equation models for molecular communication networks. *IEEE Transactions on NanoBioscience* 12, 2 (2013), 79–92.

[7] E. De Leo, L. Donvito, L. Gallucio, A. Lombardo, G. Morabito, and L.M. Zanoli. 2013. Communications and switching in microfluidic systems: pure hydrodynamic control for networking labs-on-a-chip. *IEEE Transactions on Communications* 61, 11 (November 2013), 4663–4677.

[8] Y. Deng, M. Pierobon, and A. Nallanathan. 2017. A microfluidic feed forward loop pulse generator for molecular communication. In *IEEE Global Communications Conference (GLOBECOM)*.

[9] S.N. Ethier and T.G. Kurtz. 2009. *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

[10] R.B. Fair. 2007. Digital microfluidics: is a true lab-on-a-chip possible? *Microfluidics and Nanofluidics* 3, 3 (2007), 245–281.

[11] N. Farsad, A.W. Eckford, S. Hiyama, and Y. Moritani. 2012. On-chip molecular communication: analysis and design. *IEEE Transactions on NanoBioscience* 11, 3 (September 2012), 304–314.

[12] M. Feinberg. 1987. Chemical reaction network structure and the stability of complex isothermal reactors–I. The deficiency zero and deficiency one theorems. *Chemical Engineering Science* 42, 10 (1987), 2229–2268.

[13] L. Ferrari, J. Kaufmann, F. Winnefeld, and J. Plank. 2010. Interaction of cement model systems with superplasticizers investigated by atomic force microscopy, zeta potential, and adsorption measures. *Journal of Colloid and Interface Science* 347 (2010), 15–24.

[14] F. Freyer, J.A. Roberts, P. Ritter, and M. Breakspear. 2012. A canonical model of multistability and scale-invariance in biological systems. *PLoS Computational Biology* 8, 8 (2012).

[15] L. Galluccio et al. 2018. Capacity of a binary droplet-based microfluidic channel with memory and anticipation for flow-induced molecular communications. *IEEE Transactions on Communications* 66, 1 (January 2018), 194–208.

[16] Hongzhi Guo, Josep Miquel Jornet, Qiaoqiang Gan, and Zhi Sun. 2017. Cooperative Raman spectroscopy for real-time in vivo nano-biosensing. *IEEE Transactions on NanoBioscience* 16, 7 (2017), 571–584.

[17] M. Hamidović et al. 2019. Information encoding in droplet-based microfluidic systems: first practical study. In *Proc. ACM International Conference on Nanoscale Computing and Communication*.

[18] S.A. Isaacson. 2009. The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. *SIAM J. Appl. Math.* 70, 1 (2009), 77–111.

[19] B. Krishnaswamy et al. 2013. Time-elapse communication: bacterial communication on a microfluidic chip. *IEEE Transactions on Communications* 61, 12 (December 2013), 5139–5151.

[20] M. Kuscu and O.B. Akan. 2018. Modeling convection-diffusion-reaction systems for microfluidic molecular communications with surface-based receivers in internet of bio-nano things. *PLoS One* 13, 2 (2018).

[21] M. Lu et al. 2013. Tristability in cancer-associated microRNA-TF chimera toggle switch. *J. Phys. Chem. B* 117 (2013), 13164–13174.

[22] M.B. Miller and B.L. Bassler. 2001. Quorum sensing in bacteria. *Annual Reviews in Microbiology* 55, 1 (2001), 165–199.

[23] Adam Noel, Karen C Cheung, and Robert Schober. 2014. Optimal receiver design for diffusive molecular communication with flow and additive noise. *IEEE transactions on nanobioscience* 13, 3 (2014), 350–362.

[24] W.F. Pearman, M. Lawrence-Snyder, S.M. Angel, and A.W. Decho. 2007. Surface-enhanced Raman spectroscopy for *in situ* measurements of signaling molecules (autoinducers) relevant to bacteria quorum sensing. *Applied Spectroscopy* 61, 12 (2007), 1295–1300.

[25] D.A. Reynolds. 2009. Gaussian mixture models. *Encyclopedia of Biometrics* 741 (2009).

[26] G. Schwarz. 1978. Estimating the dimension of a model. *Annals of Statistics* 6, 2 (1978), 461–464.

[27] E. Seitaridou et al. 2007. Measuring flux distributions for diffusion in the small-numbers limit. *The Journal of Physical Chemistry B* 111, 9 (2007), 2288–2292.

[28] D. Soloveichik, G. Seelig, and E. Winfree. 2010. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences* 107, 12 (2010), 5393–5398.

[29] R.M. Van Dam. 2006. Solvent-resistant elastomeric microfluidic devices and applications. *Diss. California Institute of Technology* (2006).

[30] N.G. van Kampen. 2007. *Stochastic Processes in Physics and Chemistry*. Elsevier Science & Technology Books.

[31] L.J. Zeman and A. Zydney. 2017. *Microfiltration and Ultrafiltration: Principles and Applications*. CRC Press.

[32] J. Zlatanova et al. 2000. Single molecule force spectroscopy in biology using the atomic force microscope. *Progress in Biophysics and Molecular Biology* 74, 1-2 (2000), 37–61.