

# The Gaussian Approximation in Soft Detection for Molecular Communication Via Biological Circuits

Alessio Marcone<sup>\*†</sup>, Massimiliano Pierobon<sup>‡</sup>, and Maurizio Magarini<sup>\*</sup>

<sup>\*</sup>Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, 20133 Milano, Italy

<sup>‡</sup>Department of Computer Science and Engineering

University of Nebraska-Lincoln, Lincoln, Nebraska 68588 USA

Email: alessio.marcone@mail.polimi.it, pierobon@cse.unl.edu, maurizio.magarini@polimi.it

**Abstract**—The programming of biological cells by genetic circuit engineering is enabling the development of man-made devices and systems in the biochemical environment, with applications in the areas of biomedicine, security, and environmental sensing and control, amongst others. The exchange of information through biochemical reactions and molecule transport, *i.e.*, Molecular Communication (MC), stands as one of the foundational paradigms for the design and characterization of these systems. In a previous work, the same authors developed an analog soft decoder design for MC based on biological circuits inspired by the analog information processing in biochemical reactions. While such a design was optimized for an MC channel affected by Gaussian noise, realistic noise models in molecule transport processes and biochemical reactions tend to deviate from this assumption. In this paper, these models are discussed, together with the validity of their Gaussian approximations, in light of the performance of the log-likelihood ratio calculation of the aforementioned design, numerically evaluated through biochemical simulation. These models, which are directly derived from the theory of molecular diffusion and stochastic chemical reaction analysis, are formulated with a general validity with respect to any future MC system design based on biological circuits.

**Index Terms**—Molecular communication; diffusion channel; soft detection; synthetic biology; biological circuit; Poisson noise; Langevin equation; biochemical simulation

## I. INTRODUCTION

Synthetic biology is a novel cutting edge discipline that promises to provide tools for the engineering community to implement devices and programs in the biological environment [1]. Biological circuits are the first attempt to a forward engineering approach based on the property of DNA genes to enable the flow of information signals through their mutual activation and repression [2]. Biological circuits are directly compared to electrical circuits in terms of functionality and energy consumption [3], and are currently studied as the basis of novel applications in the areas of biomedicine, security, and environmental sensing and control, amongst others.

Within the synthetic biology community, the engineering of cell-to-cell communication has recently gained particular interest [4], where the exchange of information, mostly between microbes, is at the basis of many functionalities ranging from studies in evolution and ecology, to distributed bio-computation. Molecular Communication (MC) [5], [6], a novel

paradigm in communication engineering that investigates information exchange by molecular means, has been recently proposed as a unifying discipline to guide the characterization, modeling and design of these cell communication systems [7], [8], towards a future Internet of Things applied to the biological and nanoscale domains [9].

In the direction of the design of novel MC components based on biological circuits, we previously proposed an analog soft detection decoder entirely in the biochemical domain [10]. In particular, inspired by a recent study on the efficiency of analog computation in biological circuits [3], and based on the seminal work in [11] on analog decoding of block and convolutional codes with non-linear electrical networks, we presented a proof-of-concept demonstration of the design of a biological circuit to decode parity-check-binary-encoded molecule concentrations into log-likelihood bit values (L-values). While biochemical simulations data of the resulting biological circuits demonstrate very close behavior to the ideal scheme, the underlying assumption of a Gaussian white noise seems to have a limited validity, where realistic noise models in biochemical reactions and molecule transport processes tend to deviate from this assumption.

In this paper, we consider a system where the aforementioned analog soft detection decoder receives modulated bits from a diffusion-based channel. From this, we discuss the most widely recognized noise models applicable to this system, together with the validity of their Gaussian approximations. In addition, biochemical simulation results are provided to compare the performance of the system in different noise scenarios. These models, which are directly derived from the theory of molecular diffusion and stochastic chemical reaction analysis, are formulated with a general validity with respect to any future MC system design based on biological circuits.

The rest of the paper is organized as follows. In Sec. II-A we introduce the main processes at the basis of our MC system based on biological circuits and molecule diffusion. In Sec. III we detail the Poisson-based noise model of molecular diffusion, and derive a Gaussian noise approximation. In Sec. IV, we detail a stochastic model of the noise in biochemical reactions based on Gaussian statistics, while in Sec. V numerical results of biochemical simulation are presented to compare the system performance in different noise scenarios

in terms of soft bit detection. Finally, in Sec. VI we conclude the paper.

## II. A SOFT DETECTION MOLECULAR COMMUNICATION SYSTEM

The reference MC system considered in this paper is shown in Fig. 1. A sequence of molecule concentrations, modulated according to transmitted bits  $x_i$  and propagating via free diffusion, or **Diffusion Channel**, reaches an engineered cell with concentration values  $y_i$ , where a **Biological Circuit** operates the soft detection of the transmitted bits  $x_i$ . The modulated concentrations are thought to be transmitted by another engineered cell where a biological circuit operates the modulation and emission of molecules, as described in [7], [8]. This system is derived from the proof-of-concept design we introduced in [10], where we presented a more complex biological circuit able to decode parity-check-encoded molecule concentrations with analog biological circuit components, which were recently promoted as more efficient [3] that their digital counterparts [12]. In this previous contribution, the biological circuit was designed by optimizing its performance with respect to a generic Gaussian noise affecting the received molecule concentration. While this approach simplifies the design of the circuit components, a more realistic noise model would necessarily consider all the sources contributing to the **Diffusion Noise** and **Biological Circuit Noise**, and, as we detail in the following, deviate from the Gaussian assumption. *The objective of this paper is to reason on the performance of a simplified version of the biological circuit in [10], where we only consider the soft (analog) detection of uncoded bits, under the condition of widely recognized realistic noise sources in diffusion and biological circuits Vs their approximation with a Gaussian model.* The approach presented in this paper can be extended in the future to more complex systems where molecule diffusion and biological circuits are considered. Next, we detail the main elements of the reference MC system in Fig. 1.

### A. Diffusion Channel

As shown in Fig. 1, information is transmitted through modulated molecule concentrations that reach the destination through a molecular diffusion process. The motion of each molecule is described by Brownian random walk where, assuming independent movement of each molecule, molecular diffusion can be modeled by the Fick's laws characterized by a homogeneous diffusion coefficient  $D$  both in space and time [13]. We make the following assumptions:

- The transmitter, located at a distance  $r_{rx}$  from the engineered cell in Fig. 1, modulates and emits molecules according to the bit sequence  $x_i$ . Each molecule emission happens instantaneously at the beginning of a bit time interval  $T$  and emits a concentration of  $Q_1$  or  $Q_0$  [molecules/unit volume] if a bit 1 or 0 is to be transmitted, respectively. Since in nature cells do not usually utilize a zero release rate of signaling molecules [4], which is also in agreement with the evidence of basal expression rate that characterizes many

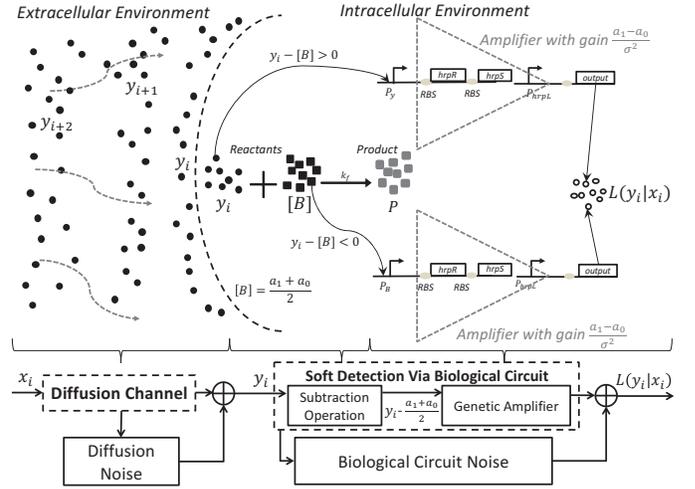


Fig. 1. Schematic of a system for MC soft detection via biological circuits.

DNA genes in biological circuits [14], we consider positive values for  $Q_1$  and  $Q_0$ .

- The received modulated molecule concentrations  $y_i$  are the samples of the molecule concentration  $y(t)$  at the location of the engineered cell in Fig. 1, considered homogeneous around and inside the cell. The sampling time  $\bar{t}_i$  is equal to the expected time of maximum of each modulated concentration after propagation.

As a consequence of these assumptions, taking into account the diffusion channel model [15], the received molecule concentration  $y_i$  has an average that is expressed as follows:

$$E[y_i] = \frac{Q_{1/0} e^{-r^2/(4D(\bar{t}_i - iT))}}{(4\pi D(\bar{t}_i - iT))^{3/2}} = a_{1/0}, \quad (1)$$

which assumes the two values, namely,  $a_1$  or  $a_0$ . In this paper, we assume that these values account for not only the average contribution of the  $i$ -th transmitted bit to the received concentration, but also the average contribution of all the possible combinations of previous bits, *i.e.*, the intersymbol interference (ISI).  $D$  is the diffusion coefficient, and the time instant  $\bar{t}_i$  has to result into sampling the propagated concentration at the so-called Pulse Delay  $t_d$  [13], the time when the Green's function of the diffusion equation [16] is at its maximum, expressed as follows:

$$\bar{t}_i = t_d + iT = \frac{r_{rx}^2}{6D} + iT. \quad (2)$$

### B. Soft Detection Via Biological Circuits

In this paper we utilize analog biological circuit components to compute soft values, or the a-posteriori probability (APP) that the received concentration  $y_i$  corresponds to a transmitted bit 0 over a transmitter bit 1, in the form of log-likelihood ratio (LLR) of  $x_i$  being 0 to the APP of it being 1. The LLR  $L(x_i|y_i)$  of a transmitted bit  $x_i$  given the set of all the received concentrations  $y_i$  can be calculated according to

$$L(x_i|y_i) = \ln \frac{P(x_i = 0|y_i)}{P(x_i = 1|y_i)}. \quad (3)$$

In case the concentrations  $y_k, k \neq i$  are not correlated with the bit  $x_i$ , the computation of the LLR of  $x_i$  depends only on  $y_i$  and, therefore, the expression becomes

$$\begin{aligned} L(x_i|y_i) &= \ln \frac{P(x_i=0|y_i)}{P(x_i=1|y_i)} = \ln \frac{f(y_i|x_i=0) \cdot P(x_i=0)}{f(y_i|x_i=1) \cdot P(x_i=1)} \\ &= \ln \frac{f(y_i|x_i=0)}{f(y_i|x_i=1)} + \ln \frac{P(x_i=0)}{P(x_i=1)} = L(y_i|x_i) + L(x_i). \end{aligned} \quad (4)$$

In our case  $L(x_i)=0$  since 0 and 1 have the same a-priori probability. For the considered two-level transmission with concentrations  $a_0$  and  $a_1$  over an additive white Gaussian noise (AWGN) channel we get [11]

$$L(y_i|x_i) = \ln \frac{\exp\left(-\frac{(y_i-a_0)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(y_i-a_1)^2}{2\sigma^2}\right)} = \frac{(a_0-a_1)}{\sigma^2} \left[ y_i - \frac{(a_1+a_0)}{2} \right], \quad (5)$$

where the same variance  $\sigma^2$  is assumed for both possible transmitted bits. The sign of  $L(x_i|y_i)$  gives the decision about the transmitted symbol and its magnitude gives the reliability of this decision. As it can be observed from the above equation, when the received signal concentration  $y_i$  is lower than the threshold  $y_{th} = (a_1 + a_0)/2$ , defined by the arithmetic mean of the concentrations  $a_0$  and  $a_1$ , the LLR is positive and, therefore, the decision about the transmitted bit is 0. Viceversa, when  $y_i$  is higher than  $y_{th}$ , the LLR is negative and, therefore, the decision about the transmitted bit is 1.

As shown in Fig. 1, the computation of (5) is implemented through a biological circuit of our design. This implementation is based on the following assumptions, which we have justified in [10]:

- The concentrations of the molecular species, *i.e.*, received molecules, proteins, and other molecules involved in the biological circuit, are considered homogeneous at any time instant inside and around the cell.
- The input molecules of the biological decoding circuit can be either the same incoming signaling molecules emitted by the transmitter cell, and able to cross the cell membrane, therefore resulting in the same concentration inside the cell as around the cell, or they can be molecules resulting from a chemical reception process at the receiver cell [7].
- All the concentration values expressed in the following are intended as normalized with respect to the average number of intracellular signaling molecules, typically equal to 1000 molecules per cell [17]. In the case of an *E coli* bacterium, a usual chassis in synthetic biology, this corresponds to a concentration of  $1 \mu\text{M} = 6 \cdot 10^{20}$  [molecules/m<sup>3</sup>].

As a consequence of these assumptions, the proposed biological circuit, which implements (5), is composed of two main blocks, namely, the subtraction operation, and the genetic amplifier, detailed next.

The Subtraction Operation block is realized through a mass action chemical reaction that converts the propagated molecules of the received concentration  $y_i$  and the molecules of species  $B$  into product molecule species  $P$ . The molecule

species  $B$  is kept by the biological circuit to a fixed concentration equal to  $\frac{a_1+a_0}{2}$ , which is the quantity to be subtracted from  $y_i$  to get  $L(y_i|x_i)$  according to (5). This reaction occurs with a forward rate  $k_f$  until the remaining concentration of any of the two molecule species is equal to the value  $\left| y_i - \frac{(a_1+a_0)}{2} \right|$ . In particular, on the one hand, if  $y_i - \frac{(a_1+a_0)}{2} > 0$ , then the molecule of species  $B$  are completely consumed, and some propagated molecules of the received concentration  $y_i$  are left, with a concentration value equal to the aforementioned subtraction. On the other hand, if  $y_i - \frac{(a_1+a_0)}{2} < 0$ , the propagated molecules of the received concentration  $y_i$  will be consumed, while some molecule of species  $B$  are left, again with a concentration value equal to the subtraction. These two cases feed alternatively two genetic amplifiers, as shown in Fig. 1. The upper branch is for the *positive* subtracted value and the lower branch is for the *negative* subtracted value. The subtraction operation block is modeled by the following chemical reaction rate equation (RRE) [2]:

$$\frac{d[P]}{dt} = k_f y_i [B] - k_r [P], \quad (6)$$

where  $k_r$  is the reverse rate. Here we assume reactions where the forward rate is much greater than the reverse rate, as in [18].

The Genetic Amplifier block is inspired by the biological circuit design presented in [19], and it implements a multiplication of the subtraction operation block output by  $\frac{a_0-a_1}{\sigma^2}$ , as in (5). In particular, we adapted the fixed gain amplifier design from [19] to include in its linear range all the possible values of the subtraction operation block output  $\left| y_i - \frac{(a_1+a_0)}{2} \right|$ , and its gain to the aforementioned  $\frac{a_0-a_1}{\sigma^2}$ , by tuning biological circuit parameters.

A biological circuit is generally defined as a genetic regulatory network [2] embedded in a biological cell, where DNA genes are linked together by chemical reactions that result in activation and repression mechanisms that regulate their expression into proteins, which are biological macromolecules. By means of synthetic biology techniques, genes and chemical reactions can be arranged to implement a specific biological function. A gene in a biological circuit is composed of a *promoter region* (e.g.,  $P_y$ ,  $P_B$ , and  $P_{harpL}$  in Fig. 1) which recruits the RNA Polymerase holoenzyme (RNAP) to start the transcription of the rest of the gene into a messenger RNA molecule (mRNA), one or more *coding sequences* (e.g.,  $P_y$ ,  $P_B$ , and  $P_{harpL}$  in Fig. 1), which contain the information to build a protein from component amino acids that is transcribed into the mRNA. In addition, the Ribosome Binding Sites, (*RBSs* in Fig. 1) are DNA sequences placed before the coding sequences, transcribed as well into the mRNA, that recruit the ribosome, which then translates the coding sequence from the mRNA into the final protein. Amongst others, promoters can be inducible, which happens if special proteins, or activators, bind to the operator region near the promoter site, up-regulating the transcription of the subsequent coding sequence by increasing the RNAP recruiting rate. Activators in the circuit in Fig. 1 are the molecules  $Y$  of the received

concentration  $y_i$ , the molecules  $B$ , and the molecules  $RS$ , the latter being a synergistic activator complex formed by the proteins expressed from the coding sequences  $hrpR$  and  $hrpS$ , shown in the figure.

While the fixed gain amplifier in [19] is modeled in the steady state regime, resulting in a simplified expression in terms of Hill functions [14], in this paper, in order to properly express and simulate the biological circuit noise, as detailed in the following, we express our version of the biological circuit in terms of mass action kinetics, which translates into the following RREs:

$$\begin{aligned} \frac{d[P_{xx}]}{dt} &= [S_{xx}] - K_{a_{xx}}[P_{xx}][xx][RNAP] \\ \frac{d[xx]}{dt} &= [S_{xx}] - K_{a_{xx}}[P_{xx}][xx][RNAP] \\ \frac{d[S_{xx}]}{dt} &= -K_{a_{xx}}[P_{xx}][xx][RNAP] + [S_{xx}], \end{aligned} \quad (7)$$

where these equations are replicated for  $xx$  equal to  $Y$  and  $B$ , and  $[P_{xx}]$ ,  $[S_{xx}]$ , and  $[xx]$  are time-varying concentration of free promoters, promoters bound to RNAP, and molecules  $Y$  or  $B$ , where  $[Y] = y_i$ . Without loss of generality, the concentration of  $[RNAP]$ , which is relatively abundant in a cell, is approximated as a constant [2]. The constant parameter  $K_{a_{xx}}$  is the equilibrium constant of the reaction where the promoter recruits the RNAP and the activator  $xx$ ,  $Y$  or  $B$ , respectively. Similarly, the following RREs complete the chemical kinetic model of the two genetic amplifiers in Fig. 1:

$$\begin{aligned} \frac{d[RS]}{dt} &= n_p k_{a_y} [S_y] + n_p k_{a_B} [S_B] + \\ &\quad - n_{RS} (K_{a_{hrpL}} [P_{hrpL}] [RS] - [S_{hrpL}]), \\ \frac{d[P_{hrpL}]}{dt} &= [S_{hrpL}] - K_{a_{hrpL}} [P_{hrpL}] [RS]^{n_{RS}} [RNAP], \\ \frac{d[S_{hrpL}]}{dt} &= K_{a_{hrpL}} [P_{hrpL}] [RS]^{n_{RS}} [RNAP] - [S_{hrpL}], \\ L(x_i|y_i) &= n_{pl} k_{a_{hrpL}} [S_{hrpL}], \end{aligned} \quad (8)$$

where  $[RS]$  is the concentration of the aforementioned activator complexes, and  $n_p$ ,  $k_{a_y}$ , and  $k_{a_B}$  are the number and the rate of the expression of the activator complexes  $RS$  from the gene with promoter  $P_Y$  and  $P_B$ , respectively. The time varying concentrations  $[P_{xx}]$ ,  $[S_{xx}]$ , and  $[xx]$  and the constant parameter  $K_{a_{hrpL}}$  have similar definition as above, while  $n_{RS}$  expresses the number of activator complexes necessary to activate the promoter  $P_{hrpL}$ , and  $n_{pl}$  and  $k_{a_{hrpL}}$  are the number and rate of expression of the molecules whose concentration rate is equal to the LLR  $L(x_i|y_i)$ . Most of the aforementioned parameters are experimentally evaluated in [19] with the exception of  $n_p$ ,  $k_{a_y}$ , and  $k_{a_B}$  that depend on the RBS strength and have been tuned to obtain the desired gain, as suggested in [19].

### III. DIFFUSION NOISE

#### A. The Poisson Model

In a realistic scenario, the expression in (1) corresponds to the expected value of the concentration at a given time  $t$  at

distance  $r_{rx}$ . Molecules, in fact, move with Brownian motion and each molecule is either inside or outside the engineered cell at a given instant. For this reason the number of observed (received) molecules follows a binomial distribution that can be approximated as a Poisson counting process [18], [20]. If the probability  $P_{ob}(t)$  that a given molecule is observed within the receiver volume is small and the number of transmitted molecules  $N_{tx}$  is large, the number of received molecules at time  $t$ ,  $N_{rx}(t)$ , can be approximated as a Poisson variable with distribution

$$P(N_{rx}(t) = N_y) = \frac{(E[N_{rx}(t)])^{N_y} \exp(-E[N_{rx}(t)])}{N_y!}, \quad (9)$$

where  $E[N_{rx}(t)]$  is the expected number of received molecules [18], [20]. In Sec. II-A, we have assumed that the received molecule concentration is sampled in the time instants defined by (2). As a consequence,

$$E[N_{rx}(\bar{t}_i)] = a_{1/0} \cdot V, \quad (10)$$

where  $V$  is the volume of the engineered cell in Fig. 1. We obtain two different expected molecule numbers, namely,  $N_0$  and  $N_1$  depending whether the transmitted bit  $a_i$  is equal to  $a_0$  or  $a_1$ , respectively. The probability mass function (pmf) of the number of received molecules at time  $\bar{t}_i$ , conditioned to the transmitted bit, is expressed as

$$P(N_{rx}(\bar{t}_i) = N_y | x_i = 0) = \frac{(N_0)^{N_y} \exp(-N_0)}{N_y!}, \quad (11)$$

$$P(N_{rx}(\bar{t}_i) = N_y | x_i = 1) = \frac{(N_1)^{N_y} \exp(-N_1)}{N_y!}. \quad (12)$$

The variables  $y_i$  and  $N_{rx}$  are linearly dependent, hence it is straightforward to compute the pmf for  $y_i$  conditioned to the transmitted bit as

$$P(y_i | x_i = 0/1) = \frac{(N_{0/1})^{y_i V} \exp(-N_{0/1})}{(y_i V)!}. \quad (13)$$

In this case, since the statistics of the noise is non-Gaussian, the LLR calculation in (5) at the basis of our biological circuit implementation is no longer optimal. By comparison, the optimal calculation of the LLR  $L(y_i|x_i)$  can be obtained by replacing (11) and (12) in (4), resulting into

$$L(y_i|x_i) = \ln \frac{(N_0)^{y_i V} \exp(-N_0)}{(N_1)^{y_i V} \exp(-N_1)} = y_i V \ln \left( \frac{N_0}{N_1} \right) - (N_0 - N_1), \quad (14)$$

where, in this case, the threshold is  $y_{th} = (N_0 - N_1) / (V(\ln N_0 - \ln N_1))$ . Since our biological circuit described in Sec. II-B implements the calculation of (5), in the presence of a realistic diffusion noise source, the LLR calculation will deviate from the ideal in (14) by an error, which will be explored in Sec. V.

#### B. The Gaussian Approximation

When  $P_{ob}(t)$  is not close to one or zero and  $N_{tx}$  is sufficiently large, the binomial distribution for the received number of molecules can be approximated with a Gaussian distribution with mean  $E[N_{rx}(t)]$  and variance  $\sigma^2(t) = E[N_{rx}(t)](1 - P_{ob}(t))$  [18]. We can therefore write the following expression:

$$P(N_{rx}(t) = N_y) = \frac{\exp\left(-\frac{(N_y - E[N_{rx}(t)])^2}{2E[N_{rx}(t)](1 - P_{ob}(t))}\right)}{\sqrt{2\pi E[N_{rx}(t)](1 - P_{ob}(t))}}, \quad (15)$$

from which

$$P(y_i) = \frac{\exp\left(-\frac{(y_i V - E[N_{rx}(t)])^2}{2E[N_{rx}(t)](1 - P_{ob}(t))}\right)}{\sqrt{2\pi E[N_{rx}(t)](1 - P_{ob}(t))}}. \quad (16)$$

Again,  $E[N_{rx}(t = t_d)]$  assumes values  $N_0$  and  $N_1$  when bit 0 or 1 is transmitted, respectively, leading to different variance values  $\sigma_0^2$  and  $\sigma_1^2$  for transmission of different bits.

As a consequence of this Gaussian approximation for the diffusion noise, the calculation of the LLR  $L(y_i|x_i)$  becomes as follows:

$$\begin{aligned} L(y_i|x_i) &= \frac{\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(y_i V - N_0)^2}{2\sigma_0^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(y_i V - N_1)^2}{2\sigma_1^2}\right)} = \\ &= \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \frac{1}{2\sigma_1^2\sigma_0^2} \{V^2(\sigma_0^2 - \sigma_1^2)y_i^2 + \\ &\quad -2V(N_1\sigma_0^2 - N_0\sigma_1^2)y_i + (N_1^2\sigma_0^2 - N_0^2\sigma_1^2)\}. \quad (17) \end{aligned}$$

It is worth observing that in comparison to (5), where the equal variance assumption due to the AWGN channel model is done, here two different variances are considered in order to have a more realistic approximation of the Poisson noise. Since the computation of this quantity in the biological domain is far from being trivial, we could linearize this function around the value of the decision threshold (for a hard decision decoding) that differs from the arithmetic mean since the variances of the noise in the two cases are different. The new threshold can be computed either as the point where the two Gaussian functions meet or as the point where the LLR  $L(y_i|x_i)$  is zero. After resolution of a quadratic equation, we get two opposite solutions but, since concentrations are only positive valued, we neglect the negative term to have

$$\begin{aligned} y_{th} &= \frac{1}{V} \frac{1}{\sigma_0^2 - \sigma_1^2} \left\{ N_1\sigma_0^2 - N_0\sigma_1^2 \right. \\ &\quad \left. + \sigma_0\sigma_1 \sqrt{(N_0 - N_1)^2 - 2(\sigma_0^2 - \sigma_1^2) \ln(\sigma_1/\sigma_0)} \right\}. \quad (18) \end{aligned}$$

By truncating the Taylor expansion of  $L(y_i|x_i)$  to the first term, we get the linearized LLR  $L_l(y_i|x_i)$  around  $y_{th}$  as

$$L_l(y_i|x_i) = \left[ \frac{V^2(\sigma_0^2 - \sigma_1^2)}{\sigma_0^2\sigma_1^2} y_{th} - \frac{V(N_1\sigma_0^2 - N_0\sigma_1^2)}{\sigma_0^2\sigma_1^2} \right] (y_i - y_{th}). \quad (19)$$

As in (5) and (14), the expression  $(y_i - y_{th})$  is preserved. In Fig. 2 it is shown a comparison between the quadratic function  $L(y_i|x_i)$  and its linearization  $L_l(y_i|x_i)$ . The circle represents the range of values of  $y_i$  for which we are mostly interested in a very good approximation, where the resulting reliability values are low, because  $y_i$  is close to the threshold. The approximation achieved in that range is nearly perfect. We can easily adapt the concentration  $[B]$  in the subtraction operation and the gain of the genetic amplifier in the biological decoding circuit to compute the new LLR in (19), while maintaining the same design.

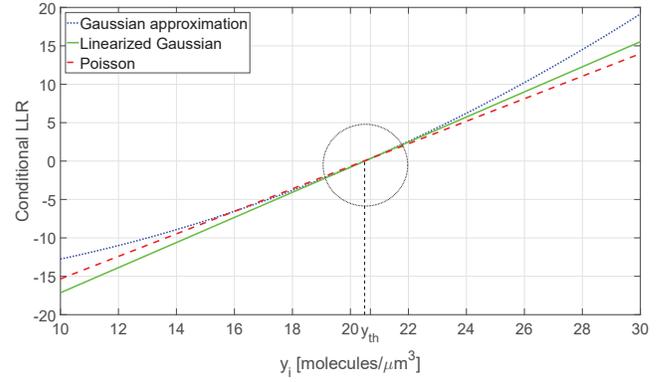


Fig. 2. Comparison between the quadratic function  $L(y_i|x_i)$  (red) and the linearized  $L_l(y_i|x_i)$  (blue stars) for  $a_0 = 22 \text{ molecules}/\mu\text{m}^3$ ,  $a_1 = 20 \text{ molecules}/\mu\text{m}^3$ ,  $V = 10 \mu\text{m}^3$ ,  $P_{ob}(t_d) = 0.1$ .

#### IV. BIOLOGICAL CIRCUIT NOISE

The randomness inherently present in the chemical reactions underlying a biological circuit generates noise in their input-output response. For this reason, the deterministic model of our proposed biological circuit expressed in terms of RREs in (7) and (8) is only accounting for the average behavior of the biological circuit, and a proper stochastic model should be utilized to represent the chemical reaction noise. In this paper, in place of more general formulations based on the Chemical Master Equation [21], or the  $\tau$ -leaping approximate stochastic method [22], which, similarly to the diffusion noise, is based on a Poisson counting process, we make use of the Chemical Langevin Equation (CLE) [23] formulation. For this, we can rewrite the biological circuit model expressed in Section II-B through the RREs by adding the noise contribution as a Gaussian Process [23], as detailed in the following.

The CLE can be expressed for each chemical reaction in (7) and (8) as follows [23]:

$$\frac{dX_i(t)}{dt} = \sum_{m=1}^M V_{im} a_m(\mathbf{X}(t)) + \sum_{m=1}^M V_{im} \sqrt{a_m(\mathbf{X}(t))} \Gamma_m(t), \quad (20)$$

where  $\mathbf{X}(t)$  is the number of molecules for each species at time  $t$ ,  $a_m(\mathbf{X}(t))$  is the propensity function for a chemical reaction  $m$ , which can be computed as in [2], and  $V_{im}$  is the stoichiometric coefficient, which corresponds to the number of molecules for each species that the reaction  $m$  operates when it occurs. The terms  $dX_i(t)/dt$  and  $V_{im} a_m(\mathbf{X}(t))$  in (20) correspond to the left hand side and each term of the sum at the right hand side of each equation in (7) and (8), respectively, where in place of species concentrations we have the number of molecules of each species inside the volume  $V$  (concentrations can be then computed by dividing each molecule number by the volume  $V$ ).  $\Gamma_m(t)$  is a white noise process for the reaction  $m$  with variance equal to 1, statistically independent from the white noise processes of other reactions.

#### V. NUMERICAL RESULTS

In this section, we present the bit error rate (BER) performance achieved using the LLR computation of transmitted bits based on the Poisson model for the noise and that achieved using its Gaussian approximation.

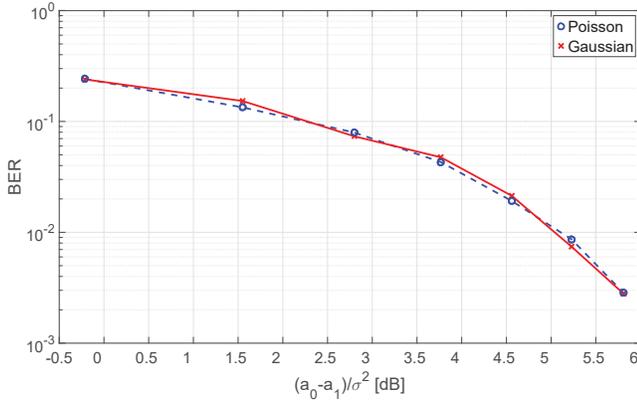


Fig. 3. BER versus  $a_0 - a_1$  for  $a_0 = 22 \text{ molecules}/\mu\text{m}^3$ ,  $a_1 = 20 \text{ molecules}/\mu\text{m}^3$ ,  $V = 10 \mu\text{m}^3$ ,  $P_{ob}(t_d) = 0.1$ .

We first analyze the LLR functions associated with the two noise distributions. In Fig. 2 it is shown the LLR given in (14) and in (17) for the Poisson and for its Gaussian quadratic approximation, respectively, as a function of the input received concentration of molecules  $y_i$ . The plots have been obtained using the parameters reported in the figure caption. Figure 2 also includes the linearized LLR given in (19) for the Gaussian quadratic approximation around  $y_{th} \approx 20.49$ , which results from (18) using the chosen parameters. As it can be observed, the linearized expression provides a nearly perfect approximation for the range of values of interest.

The BER performance achieved in case of binary transmission using the Poisson model for the noise is here compared to that obtained using the Gaussian approximation as a function of the ratio  $(a_0 - a_1)/\sigma^2$ , where  $\sigma^2 = (a_1 + a_0)/2$  denotes the average power of the noise that affects the two concentration levels  $a_0$  and  $a_1$ . The LLR is here computed by simulating the Soft Detection Via Biological Circuit block that is enclosed in the dashed rectangular box of Fig. 1 within the Matlab SimBiology environment, through a stochastic simulation according to the Langevin Equation model detailed in Sec. IV. As it can be observed from the results in Fig. 3, the Gaussian approximation gives virtually the same BER performance as that obtained using Poisson distribution, thus justifying its use for evaluating performance of molecular communication systems.

## VI. CONCLUSION

In this paper, we focused on the impact of approximated noise models on the bit error rate performance of analog soft detection for MC based on biological circuits. Soft detection in the context of MC has been recently proposed by the same authors in previous work using a Gaussian model for the noise sources, while in reality diffusion processes and chemical reactions tend to deviate from this assumption. Starting from the analysis of diffusion processes and biochemical reactions, we discussed Poisson models and Gaussian-approximated models for noise distributions. We then derived corresponding expressions to realize soft detection, and evaluate the resulting performance using computer simulations. Numerical results show that the soft detection computations based on Poisson and Gaussian models provide similar BER performance, thus

validating the Gaussian approximation in the design of MC systems based on biological circuits.

## ACKNOWLEDGMENT

This work was supported by the US National Science Foundation (NSF) through grant MCB-1449014, and the NSF EPSCoR First Award EPS-1004094.

## REFERENCES

- [1] L. J. Kahl and D. Endy, "A survey of enabling technologies in synthetic biology," *Journal of Biol. Eng.*, vol. 7, no. 1, p. 13, May 2013.
- [2] C. J. Myers, *Engineering genetic circuits*. Chapman & Hall/CRC, Mathematical and Computational Biology Series, 2009.
- [3] R. Sarpeshkar, "Analog synthetic biology," *Philos Trans A Math Phys Eng Sci.*, vol. 372, no. 2012, p. 20130110, 2014.
- [4] S. Payne and L. You, "Engineered cell-cell communication and its applications," *Adv Biochem Eng Biotechnol*, vol. 146, pp. 97–121, 2014.
- [5] 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, Nov. 2011.
- [6] N. Farsad, H. B. 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, Third Quarter 2016.
- [7] M. Pierobon, "A systems-theoretic model of a biological circuit for molecular communication in nanonetworks," *Nano Communication Networks (Elsevier)*, vol. 5, no. 1-2, pp. 25–34, March-June 2014.
- [8] B. D. Unluturk, A. O. Bicen, and I. F. Akyildiz, "Genetically engineered bacteria-based biotransceivers for molecular communication," *IEEE Trans. on Communications*, vol. 63, no. 4, pp. 1271–1281, 2015.
- [9] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, March 2015.
- [10] A. Marcone, M. Pierobon, and M. Magarini, "A parity check analog decoder for molecular communication based on biological circuits," in *In Proc. of the IEEE International Conference on Computer Communications (INFOCOM)*, May 2017, pp. 1–9.
- [11] J. Hagenauer, E. Offer, C. Measson, and M. Morz, "Decoding and equalization with analog non-linear networks," *European Transactions on Telecommunications (ETT)*, 1999.
- [12] B. Wang, R. I. Kitney, N. Joly, and M. Buck, "Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology," *Nature Communications*, vol. 2, no. 508, pp. 1–9, 2011.
- [13] I. Llatser, E. Alarcón, and M. Pierobon, "Diffusion-based channel characterization in molecular nanonetworks," in *Conference on Computer Communications Workshops*. IEEE, 2011, pp. 467–472.
- [14] U. Alon, *An Introduction to Systems Biology - Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.
- [15] M. Pierobon and I. F. Akyildiz, "A statistical-physical model of interference in diffusion-based molecular nanonetworks," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 2085–2095, June 2014.
- [16] A. Mandelis, *Diffusion-wave fields: mathematical methods and Green functions*. Springer-Verlag, 2001.
- [17] B. Alberts, A. Johnson, and *et al.* J. Lewis, *Molecular Biology of the Cell. 4th edition*. New York: Garland Science, 2002.
- [18] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Transactions on NanoBioscience*, 2014.
- [19] B. Wang, M. Barahona, and M. Buck, "Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks," *Nucleic Acids Research*, 2014.
- [20] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2532–2547, June 2011.
- [21] P. Smadbeck and Y. N. Kaznessis, "Chemical master equation closure for computer-aided synthetic biology," *Methods Mol Biol.*, vol. 1244, pp. 179–191, 2015.
- [22] S. S. Andrews, T. Dinh, and A. P. Arkin, *Stochastic models of biological processes*. Meyers, Robert (Ed.), Springer, NY, 2009, vol. 9.
- [23] D. T. Gillespie, "The chemical langevin equation," *Journal of Chemical Physics*, vol. 113, no. 1, pp. 297–306, Jul. 2000.