

A Parity Check Analog Decoder for Molecular Communication Based on Biological Circuits

Alessio Marcone^{*‡}, Massimiliano Pierobon[‡], and Maurizio Magarini^{*}

^{*}Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, I-20133 Milano, Italy

[‡]Department of Computer Science and Engineering

University of Nebraska-Lincoln, Lincoln, Nebraska 68588 USA

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

Abstract—Molecular Communication (MC) is an enabling paradigm for the interconnection of future devices and networks in the biological environment, with applications ranging from bio-medicine to environmental monitoring and control. The engineering of biological circuits, which allows to manipulate the molecular information processing abilities of biological cells, is a candidate technology for the realization of MC-enabled devices. In this paper, inspired by recent studies favoring the efficiency of analog computation over digital in biological cells, an analog decoder design is proposed based on biological circuit components. In particular, this decoder computes the a-posteriori log-likelihood ratio of parity-check-encoded bits from a binary-modulated concentration of molecules. The proposed design implements the required L -value and the box-plus operations entirely in the biochemical domain by using activation and repression of gene expression, and reactions of molecular species. Each component of the circuit is designed and tuned in this paper by comparing the resulting functionality with that of the corresponding analytical expression. Despite evident differences with classical electronics, biochemical simulation data of the resulting biological circuit demonstrate very close performance in terms of Mean Squared Error (MSE) and Bit Error Rate (BER), and validate the proposed approach for the future realization of MC components.

I. INTRODUCTION

The design and characterization of communication systems based on the exchange of molecules, directly inspired by biological processes, is a fast growing field within the communications and computer network engineering communities, encouraged by the need to interconnect devices with increasing miniaturization, ubiquity, and biocompatibility [1], [2]. These novel Molecular Communication (MC) systems further expand the Internet of Things concept to cover domains, the biological and nanoscale, where classical communication solutions show limitations [3]. MC-enabled communications are envisioned at the basis of the access, control, and collaborative processing of devices with various capabilities, such as bio-sensing, stimulation or actuation of biochemical processes, or even augmentation of the functionalities of the human body [4]. Despite the current advancements in the communication theoretical studies, which also stimulated an ongoing standardization effort [5], unified and coherent technologies to engineer and optimize systems, devices, and components capable of molecular communications are currently missing.

Recent advances in synthetic biology are providing new tools for the design, realization, and control of biological processes through the programming of cells' genetic code [6]. In particular, the theory of biological circuits, based on networks of DNA genes linked together by activation and repression mechanisms that regulate their expression into proteins, provides basic components and processes to design functionalities and behaviors in cells, mostly bacteria, by following an engineering-based approach [7], [8]. Despite a trend towards the development of biological circuits with digital-like logics [9], recent research suggests that functions based on analog computation are significantly more efficient [10].

The engineering of cell-cell communications is one of the latest frontiers in synthetic biology [11], where cells can be abstracted as the transmitters and receivers in molecular communication systems [12]. Recent literature in MC is exploring the possibility of utilizing biological circuits to realize molecular communication functionalities. In [13] the minimal subset of biological circuit elements necessary to emit and receive an analog-modulated MC signal, which propagates between cells through diffusion, is modeled through chemical reaction rate equations. General guidelines and modeling strategies to design an MC transceiver with biological circuits able to receive, process, and retransmit binary information by utilizing bacteria are included in [14], based on digital-like biological circuit functionalities and M -ary molecule concentration modulation coupled with a hard threshold detection.

Inspired by the aforementioned research on analog computation in engineered cells, in this paper we propose a design based on biological circuits that implements the analog decoding of binary information from a modulated molecule concentration. In particular, based on the seminal work in [15] on analog decoding of block and convolutional codes with non-linear electrical networks, we propose to exploit the analog computing functionalities of biological circuits to obtain as output the a-posteriori log-likelihood ratio (reliability) of a transmitted encoded bit sequence, given a received binary-modulated molecule concentration as input. To provide a first proof-of-concept demonstration of such a design, we make use of the simplest block code, namely, a parity-check code with a block length of 3 bits, and we implement the required L -value and box-plus operations through biological circuit

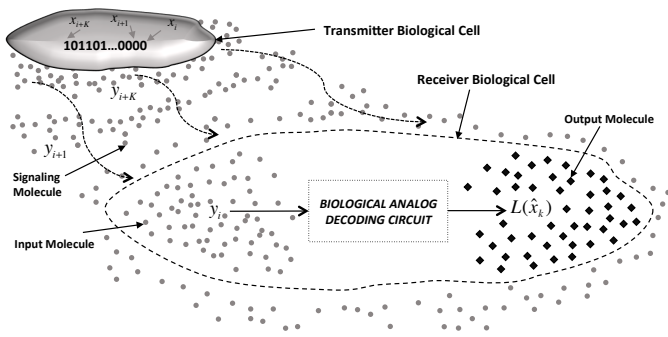


Figure 1. Sketch of the proposed MC decoder engineered in a biological cell. components, designed and tuned according to the desired output. Biochemical simulation data of the resulting biological circuit demonstrate very close performance to an electrical network implementation [15] in terms of Mean Squared Error (MSE) and Bit Error Rate (BER).

The rest of the paper is organized as follows. In Sec. II we introduce the biological analog decoding circuit based on the a-posteriori log-likelihood computation of the first bit of the code block. In Sec. III we give a brief introduction to biological circuits and their main components. In Sec. IV, Sec. V, and Sec. VI we detail the design of the required L -value, delay line, and box-plus operations, respectively. In Sec. VII we finalize the a-posteriori log-likelihood computation from the contributions of the previous elements, while in Sec. VIII we present the simulation-based validation. Finally, in Sec. IX we conclude the paper.

II. ANALOG DECODING FOR MOLECULAR COMMUNICATION

In this paper, we aim to design a biological circuit able to compute the a-posteriori log-likelihood ratio $L(\hat{x}_k)$ from a received noisy input signal $y_i, y_{i+1}, \dots, y_{i+K}$ modulated according to block-encoded bits $x_i, x_{i+1}, \dots, x_{i+K}$, where K is the block size. The sign of this log-likelihood ratio corresponds to the optimal decision on the transmitted bit, while its magnitude measures the reliability of this decision. This is defined as follows [15]:

$$L(\hat{x}_k) = \log \frac{P(x_k = 1 | y_i, y_{i+1}, \dots, y_{i+K})}{P(x_k = 0 | y_i, y_{i+1}, \dots, y_{i+K})}, \quad (1)$$

In Fig. 1 we show the scenario referenced in this paper, where a receiver cell containing the biological decoding circuit detailed in the following provides a concentration of output molecules equal to the a-posteriori log-likelihood ratio $L(\hat{x}_k)$ for the k -th bit of the code block from the modulated concentration $y_i, y_{i+1}, \dots, y_{i+K}$ of input molecules around the cell at specific time instants. The concentration of signaling molecules is modulated by a transmitter cell, *i.e.*, one or more cells containing a biological circuit for transmission, such as in [13], [14], according to block-encoded bits $x_i, x_{i+1}, \dots, x_{i+K}$ [15], and propagates through the space via diffusion until reaching the cell containing the biological decoding circuit. The proof-of-concept design of the biological

decoding circuit presented in this paper is based on the following assumptions:

- The concentrations of the molecular species, *i.e.*, input, output molecules, and proteins involved in the biological circuit, are considered homogeneous at any time instant inside and around the cell. This approximates the behavior of the system when we sample these molecule concentrations at steady state. Although this does not hold true for the concentration of the diffusing molecules from the transmitter to the receiver, we assume a distance between the transmitter cell and the receiver cell much larger than the size of the receiver cell itself. Given the Green's function of the diffusion equation [16], the longer the distance, and therefore the propagation time, the more homogeneous the impulse response is with respect to the space.
- 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 receiver cell membrane, therefore resulting in a concentration inside the cell that is the same as around the cell, or they can be molecules resulting from a chemical reception process at the receiver cell [13]. The latter process would be activated by ligand-receptor binding reception either at the membrane or inside the cell [13], and result in a concentration of input molecules of the biological decoder that is proportional to the concentration of signaling molecules around the cell.
- The modulated concentration $y_i, y_{i+1}, \dots, y_{i+K}$ at specific time instants is affected by Additive White Gaussian Noise (AWGN). This assumption is justified by the fact that the noise contribution from the biological circuits, including those that might be present at the transmitter, can be modeled according to the steady state approximation of the Langevin equation (see Appendix A in [17]), resulting in white Gaussian contributions to each circuit output [18]. Moreover, for a sufficient number of emitted molecules and for a sufficiently long time interval between the samples, the diffusion process is independent and has a Gaussian noise contribution to the input molecule concentration [19].
- Since the molecule concentration cannot assume negative values, it is modulated at the transmitter by following a binary unipolar pulse amplitude scheme. We assume that each input molecule concentration sample y_i can assume the (positive) values a_0 or a_1 (plus AWGN) depending whether the corresponding bit $x_i = 0$ or 1, respectively.
- In the rest of the paper, for the purpose of presenting a proof-of-concept biological analog decoding circuit, we reference to the simplest block code scenario that supports analog decoding [15], namely, the Single Parity Check (SPC) code with block length $K = 3$. Moreover, given the complexity of the resulting biological circuit, the preliminary design in this paper realizes only the computation of the a-posteriori log-likelihood ratio of the first bit of the block. The contribution in this paper can be extended to more complex block codes by stemming from our methodology. In fact, parity check is the fundamental decoding operation at the basis of more complex coding schemes, such as low-density parity-check.

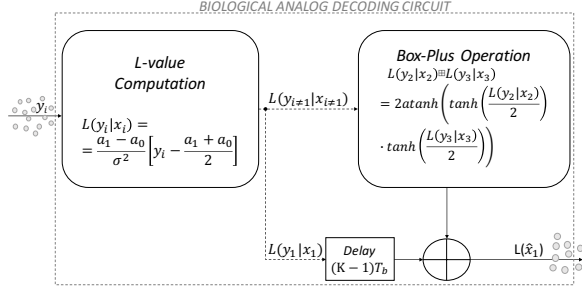


Figure 2. Biological decoding circuit

- 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 [20]. 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/m³].

As a consequence of the aforementioned assumptions, the formula in (1) to compute the a-posteriori log-likelihood ratio (*L*-value) of the first transmitted bit $L(\hat{x}_1)$ becomes [15]:

$$L(\hat{x}_1) = L(y_1|x_1) + (L(y_2|x_2) \boxplus L(y_3|x_3)), \quad (2)$$

where x_2 and x_3 represent the remaining channel bits of that same codeword. $L(y_i|x_i)$, $i \in \{1, 2, 3\}$, gives the conditioned *L*-value, which is the *L*-value of the received concentration y_i conditioned to the transmitted bit x_i , and \boxplus indicates the box-plus operation, which is defined as follows:

$$L(y_2|x_2) \boxplus L(y_3|x_3) = 2 \operatorname{atanh} \left(\tanh \left(\frac{L(y_2|x_2)}{2} \right) \cdot \tanh \left(\frac{L(y_3|x_3)}{2} \right) \right). \quad (3)$$

The expression of $L(y_i|x_i)$ is derived by considering the aforementioned concentration values a_0 and a_1 corresponding to bit 0 and 1, respectively. In contrast to [15], where a binary antipodal pulse amplitude scheme is employed, the resulting detection threshold is in our case equal to $\frac{a_1 + a_0}{2}$. As a consequence, conditional *L*-value expression is

$$L(y_i|x_i) = \log \frac{\exp \left(-\frac{(y_i - a_1)^2}{2\sigma^2} \right)}{\exp \left(-\frac{(y_i - a_0)^2}{2\sigma^2} \right)} = \frac{(a_1 - a_0)}{\sigma^2} \left[y_i - \frac{(a_1 + a_0)}{2} \right], \quad (4)$$

where σ^2 is the variance of the aforementioned AWGN present in the received molecule concentration y_i .

Hagenauer *et al.* implemented (2) with analog Very-Large-Scale Integration (VLSI) circuits, by exploiting the nonlinearities of a modified Gilbert cell [15]. In the following, we present an implementation of the same expression by using biological circuit elements. As shown in Fig. 2, our biological analog decoding circuit is composed of three main elements, namely, the *L*-value computation, which implements (4), the box-plus operation, which realizes the expression in (3), and a delay line (active only when $t = lKT_b$, where T_b is the

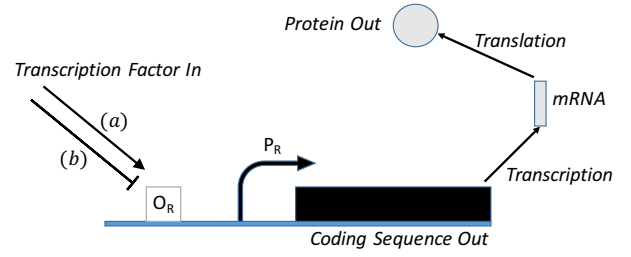


Figure 3. Gene expression: (a) Activation (\downarrow), (b) Repression (\uparrow)

bit interval, K is the block size, equal to 3, $l = 0, 1, 2, \dots$, which is needed to isolate the result of the first term of the sum in (2) from the second term, computed at a later time. In the following, after a brief overview of the main biological circuit components, we detail our biological circuit design.

III. COMPONENTS OF A BIOLOGICAL CIRCUIT

A *biological circuit* is a network of *genes* and *chemical reactions* that work together to implement a specific biological function [7].

A. Gene

A gene is the fundamental unit of information, composed of an *operator region* (O_R), a *promoter region* (P_R), and a *coding sequence*. Most genes are a stretch of DNA that codes for a *protein* molecule, a sequence of amino acids, through the fundamental processes of *transcription* and *translation* as shown in Fig. 3, where the gene is *activated/repressed* by a *transcription factor protein In*.

- Transcription is triggered by the enzyme, a specific type of protein, *RNA polymerase (RNAP)* that binds to the promoter region of the considered gene, starting the production of the *messenger RNA (mRNA)* molecule. This latter molecule is used to carry the genetic information encoded in the coding sequence of the gene to the *ribosome*, the protein assembly machinery. The ability of RNAP to bind to a promoter site can be either enhanced or lowered by other proteins called *transcription factors (activators/repressors)*.
- Activators bind to the operator region near the promoter site up-regulating the transcription of the subsequent coding sequence by increasing the RNAP molecule binding rate.
- Repressors reduce RNAP binding rate either by changing the conformation of the RNAP molecule or obstructing the binding sites of the promoter region, therefore down-regulating the transcription of the subsequent coding sequence.
- Translation is the final step in protein production. Ribosomes translate genes starting from mRNA. They are able to recognize and bind to the mRNA molecule by means of *Ribosome Binding Sites (RBSs)*, which are special sequences of nucleotides in the mRNA strand. Once the ribosome binds to the RBS of an mRNA molecule, it starts synthesizing the corresponding *protein* from component amino acids.
- As in [21], an activated gene expression can be modeled through a Hill function. As a consequence, the rate of output protein *Out* is expressed as

$$\frac{d[Out]}{dt} = k' + \operatorname{MAX} \left(\frac{[In]^n}{K_p^n + [In]^n} \right) - k_{deg}[Out], \quad (5)$$

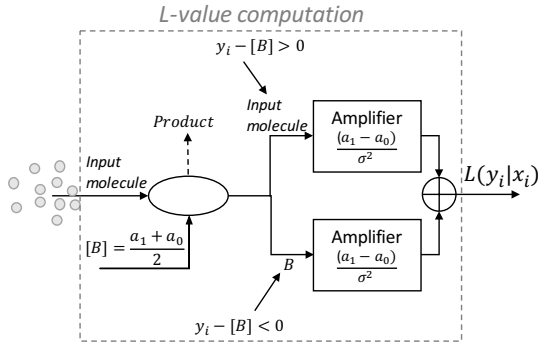


Figure 4. L -value computation

where k' is the basal rate of production, *i.e.* gene expression in the absence of input transcription factors, MAX is a constant defining the maximum value at the output, K_p is the input concentration for which the output expression rate is half of the maximum value, n is the Hill coefficient and the bracketed term is the Hill function. k_{deg} is the degradation rate of the output proteins. In the following, we assume to have non-constitutive promoters, *i.e.* $k' = 0$, which means that there is gene expression only when external activating signals are present [21]. In (5) and hereafter, the square brackets notation stands for protein concentration. From [21], repression can be obtained by exchanging the roles of K_p and In in (5).

B. Mass Action Chemical Reaction

A mass action chemical reaction is a process that converts a set of one or more input molecules (*reactants*) into another set of one or more output molecules (*products*). Reactions may proceed in the forward or reverse direction until they go to completion or reach equilibrium. We assume unbalanced reactions where the forward reaction rate is much greater than the backward. The forward rate is proportional to the reactants concentrations. In this paper we will consider the following:

- **Subtraction Operation:** We consider two reactants and one product. The product molecule concentration is the minimum among the initial reactant concentrations. Therefore, for the mass conservation law, the concentration of the remaining reactant ("*survivor molecule*") is just the difference between the two initial concentrations.
- **Storage Operation:** We consider two reactant species, one is the ligand, the molecule whose concentration value needs to be stored, and the other is the receptor protein, synthesized by a specific gene. The two species react by binding and producing a concentration of complexes proportional to the concentration of ligands if the initial receptor concentration is set sufficiently high.

IV. BIOLOGICAL L -VALUE COMPUTATION

Figure 4 reports the sequence of biological operations needed to compute the L -value $L(y_i | x_i)$ given the received noisy modulated concentration y_i .

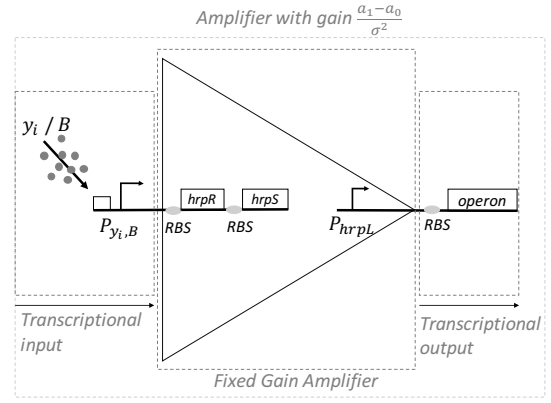


Figure 5. Fixed Gain biological amplifier

A. Mass-Action Reaction

The first element in Fig. 4 represents a mass-action reaction [7] for Subtraction Operation between the input molecules having concentration y_i and the molecules of species B . The molecule species B has 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 (4). Once the reaction has occurred, the remaining concentration of any of the two molecule species is equal to the value $|y_i - \frac{(a_1 + a_0)}{2}|$, with the following two possibilities:

- If the input molecules survive, it means that $y_i - \frac{(a_1 + a_0)}{2} > 0$ and, therefore, the received noisy concentration y_i is above the threshold. In this case, the "*survivor molecule*" concentration value should be interpreted as *positive*.
- If the molecules of species B survive, it means that $y_i - \frac{(a_1 + a_0)}{2} < 0$, and the received noisy concentration y_i is below the threshold. In this case, the "*survivor molecule*" concentration value should be interpreted as *negative*.

In order to distinguish between the two cases, and therefore propagate the sign through the biological circuit, two branches have been introduced in the block diagram, as shown in Fig. 4. The upper branch is for the *positive* concentration and the lower branch is for the *negative* concentration.

B. Amplifiers with Gain $\frac{(a_1 - a_0)}{\sigma^2}$

From (4), to complete the calculation of $L(y_i | x_i)$, we need to multiply the difference $y_i - [B]$ by the quantity $\frac{(a_1 - a_0)}{\sigma^2}$. For this, we use fixed gain amplifiers with gain $\frac{(a_1 - a_0)}{\sigma^2}$. The amplifier can be realized as proposed in [22], where it is shown that the transcriptional-output/transcriptional-input relationship, *i.e.*, the amplifier transfer function, is a linear amplification whose magnitude depends on the translational strength of the RBSs sequence in front of a coding sequence of a protein called *hrpS*, as in Fig. 5.

At the output of each amplifier we have an operon in place of a single gene, which is needed for the subsequent processing. An operon is a group of genes controlled by the same promoter and expressing proteins at the same concentration rate, as shown in Fig. 6. In our design, operons in both amplifiers shown in Fig. 4 encode for the proteins $\{t_2, z_3, u_2, u_3\}$, while they differ for the last two, namely,

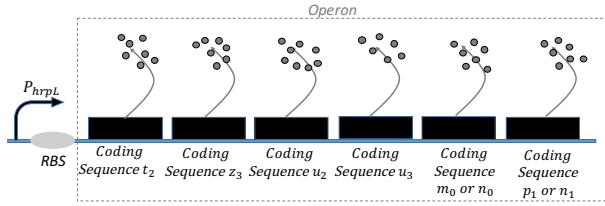


Figure 6. Operon schematic

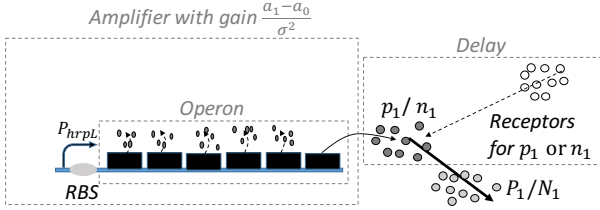


Figure 7. Storage Operation for molecules P_1 and N_1

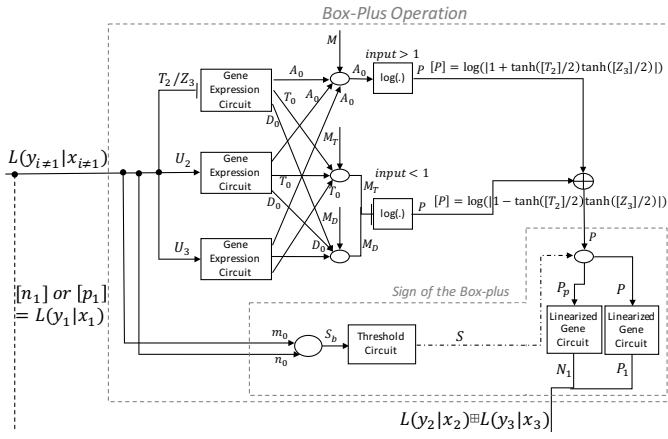


Figure 8. Box-plus operation

m_0, p_1 in the upper branch amplifier, and n_0, n_1 in the lower branch amplifier. All protein species expressed within the i -th bit time interval will reach the same concentration, equal to $L(y_i|x_i)$. The specific role of each protein will be detailed in the following.

V. BIOLOGICAL DELAY LINE

By analogy with electrical circuits, a delay line, *i.e.*, the lower branch with the Delay block in Fig. 2, aids in the computation of (2), where the conditional L -value of the first bit of the block $L(y_1|x_1)$ needs to be isolated and stored for the subsequent sum operation once the box-plus operation on the second and third bits is complete.

By using biological circuits, delay lines can be realized by means of receptors as illustrated in Fig. 7, through the Storage Operation, described in Sec. III.B, at the first bit interval, where n_1 and p_1 generated by the amplifiers are the ligands, and create very stable complexes, *i.e.*, N_1 and P_1 , with receptors, keeping their concentration values for $2T_b$, until the box-plus result is ready. For this, we assume that P_1 and N_1 degrade with very small degradation rate k_{deg} . This implements a delay line where the storage function is realized by the cell itself and the delay is provided by the stability

of the created complexes. On the contrary, the second and the third bits will be processed by the Box Plus Operation element.

VI. BIOLOGICAL BOX-PLUS OPERATION

A description of the biological circuit blocks and their operations inside the Box Plus Operation element, whose biological circuit is reported in Fig. 8, is given in the following.

A. Gene Expression Circuit

To implement the calculation of the expression in (3), it is necessary to compute the hyperbolic tangent, which is a basic function for the box-plus operation. For this, by stemming from the Hill function model in (5), we design a gene having a promoter regulated by activation. In particular, we optimize the parameters n and K_p in (5) to provide the relation between the rate of output protein Out and the input transcription factor In as close as possible to a hyperbolic tangent. If we optimize in the mean-squared error (MSE) sense by setting $MAX=1$ and varying both n and K_p values by 0.1 steps, we find that the minimum MSE is achieved for $n = 2$ and $K_p = 1$ in (5), expressed as follows:

$$\frac{d[Out]}{dt} = \frac{[In]^2}{1 + [In]^2} \cong \tanh\left(\frac{[In]}{2}\right) \quad (6)$$

Since the expression in (5) contains also the term $-k_{deg}[Out]$, based on the aforementioned assumption of steady state in the sampling of the molecule concentration values in the designed biological circuit, we obtain the following:

$$\begin{aligned} \frac{d[Out]}{dt} &= \frac{[In]^2}{1 + [In]^2} - k_{deg}[Out] = 0 \Rightarrow \\ \Rightarrow [Out] &= \frac{1}{k_{deg}} \frac{[In]^2}{1 + [In]^2} \cong \frac{1}{k_{deg}} \tanh\left(\frac{[In]}{2}\right), \end{aligned}$$

which results in $[Out] = \frac{d[Out]}{dt}$ in the case when $k_{deg} = 1$.

B. The Logarithmic Function

The expression in (3) includes also the inverse hyperbolic tangent function, which we realize by stemming from the following trigonometric identity:

$$2 \operatorname{atanh}(d) = \log\left(\frac{1+d}{1-d}\right) = \log(|1+d|) - \log(|1-d|). \quad (7)$$

By applying (7), we rewrite the inverse hyperbolic tangent in terms of logarithms. This operation is made by the two $\log(\cdot)$ blocks in Fig. 8, where the parameter d in (7) becomes $d = \tanh\left(\frac{L(y_2|x_2)}{2}\right) \tanh\left(\frac{L(y_3|x_3)}{2}\right)$, thus resulting in the same expression as in (3).

While the implementation of the arguments $1+d$ and $1-d$ is detailed in Sec. VI-C, here we focus on the realization of the logarithmic function. As shown in Fig. 8, since molecule concentrations cannot assume negative values, two different biological circuit blocks are used to approximate the logarithmic function for either input > 1 or input < 1 , detailed next.

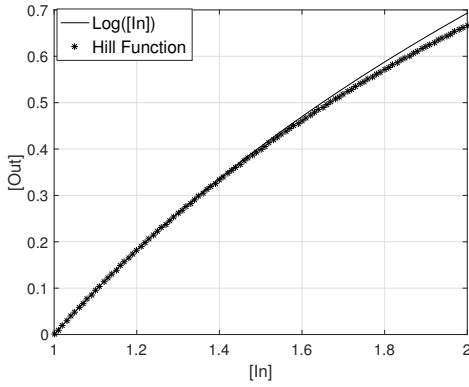


Figure 9. Comparison between the logarithmic and Hill functions with $n = 1$, $K_p = 2$ and $MAX = 2$

1) *Logarithm approximation for input greater than 1:*

From (7), since $d > 0$, which is exactly our case since concentrations are positive, the argument $(1+d)$ of the first log is always greater than 1. In Fig. 8, the block that realizes this operation is the upper $\log(\cdot)$ block that takes A_0 molecules as input and gives P molecules as output, where, as detailed in the following, the concentration $[A_0] = 1 + d$. For this, we use an activated gene as modeled by the Hill function in (5), where $[In] = [A_0]$ and $[Out] = [P]$. In order to approximate the logarithmic function with the Hill function, first, we realize a horizontal shift of our Hill function by means of a mass action reaction for Subtraction Operation between the input protein A_0 and the protein M , which is set to a concentration equal to 1. This gives the difference $[A_0]^{out} = [A_0]^{in} - 1$, as explained in Sec. III. The concentration $[A_0]^{out}$ is then fed into the $\log(\cdot)$ block. Second, after an MSE optimization to match the designed biological circuit block with the positive logarithmic function, we identify the following parameter values: $n = 1$, $K_p = 2$ and $MAX = 2$. In this minimization, we took into account that $1 < [In] = [A_0] = |1 + d| < 2$, without considering the shift, since $d = \tanh\left(\frac{L(y_2|x_2)}{2}\right) \tanh\left(\frac{L(y_3|x_3)}{2}\right)$. A comparison between the two functions is shown in Fig. 9, where we observe that a very good approximation of the logarithmic function is achieved for input values $[In] \in [1, 2]$.

2) *Logarithm approximation for input smaller than 1:* The argument $(1-d)$ of the second log in (7) is always smaller than 1. Hence, the second $\log(\cdot)$ block in Fig. 8 has to approximate the absolute value of the negative part of the logarithm. We realize this through the complementary Hill function of a repressed gene. An MSE optimization to match the gene expression with the negative logarithmic function is realized by setting $n = 1$, $K_p = 0.05$ and $MAX = |\log(10^{-3})|$, where 10^{-3} represents the supposed minimum value that the argument $[In]$ can assume, as shown in Fig. 10. However, for values close to 1, there is a small but non negligible error. In order to alleviate this problem we design a control in the expression of the output protein P (see Fig. 8) with two, in place of just one, different repressive transcription factors, M_T and M_D . This approach results in a bivariate Hill function [23] that describes the rate of output protein

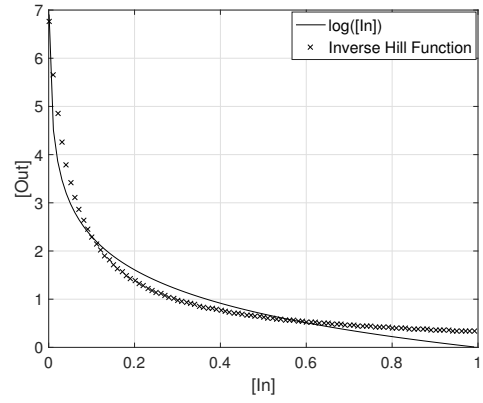


Figure 10. Comparison between a logarithmic function and a complementary Hill Function with $n = 1$, $K_p = 0.05$ and $MAX = |\log(10^{-3})|$

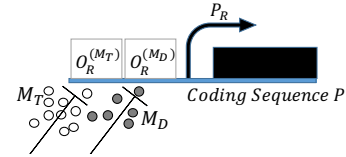


Figure 11. Logarithmic function

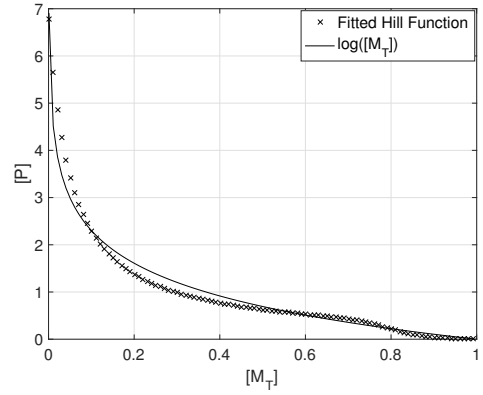


Figure 12. Comparison between a logarithmic function and a complementary bi-variate Hill Function with $n_1 = n = 1$, $\beta_1 + \beta_2 = MAX = |\log(10^{-3})|$, $k_1 = K_p = 0.05$, $n_2 = 20$ and $k_2 = 0.7$

production as a function of $[M_T]$ and $[M_D]$ as follows:

$$\frac{d[P]}{dt} = H([M_T], [M_D]) = \frac{\beta_1 + \beta_2}{1 + \left(\frac{[M_T]}{k_1}\right)^{n_1} + \left(\frac{[M_D]}{k_2}\right)^{n_2}}, \quad (8)$$

where β_i , $i = 1, 2$, are the maximum expressions in absence of the repressors, while n_i and k_i are the Hill coefficient and Hill constant related to the corresponding operator region, respectively, as sketched in Fig. 11. Although the bivariate Hill function in (8) is valid for any combination of the inputs $[M_T]$ and $[M_D]$, as explained in Sec. VI-C, in our biological circuit design these will always have the same value, resulting in the final behavior of the gene expression $[P]$ versus $[M_T] = [M_D]$ shown in Fig. 12, where it is compared to the desired curve of the logarithmic function. The MSE optimal parameters to achieve this behavior are $n_1 = n = 1$, $\beta_1 + \beta_2 = MAX = |\log(10^{-3})|$, $k_1 = K_p = 0.05$, $n_2 = 20$ and $k_2 = 0.7$.

C. Details on the Box-plus Operation

The design in Fig. 8 realized with the components described above, and leading to the box-plus operation described in (3), is detailed next. The upper Gene Expression Circuit takes as input two repressive transcription factors, namely, the molecules T_2 , whose concentration value is $L(y_2|x_2)$, stored at the second bit interval T_b according to the Storage Operation process described in Sec. III, and the molecules Z_3 , whose concentration value is $L(y_3|x_3)$, similarly stored at the third bit interval T_b . T_2 and Z_3 are the complexes resulting from the proteins t_2 and z_3 , respectively. Since this gene is an operon containing the coding sequences of the proteins A_0 , T_0 , and D_0 , the output concentrations will result from (8) by setting $\beta_1 + \beta_2 = 1$, $n_1 = n_2 = n_g$ and $k_1 = k_2 = k_g$, meaning that the operator regions for the two transcription factors have the same characteristics, expressed as

$$\frac{d[A_0]}{dt} = H([T_2], [Z_3]) = \frac{1}{1 + \left(\frac{[T_2]}{k_g}\right)^{n_g} + \left(\frac{[Z_3]}{k_g}\right)^{n_g}}. \quad (9)$$

By analyzing the expression in (9), we observe that it resembles the following product:

$$\begin{aligned} \tilde{H}([T_2], [Z_3]) &= \frac{1}{1 + \left(\frac{[T_2]}{k_g}\right)^{n_g}} \cdot \frac{1}{1 + \left(\frac{[Z_3]}{k_g}\right)^{n_g}} \\ &= \frac{1}{1 + \left(\frac{[T_2]}{k_g}\right)^{n_g} + \left(\frac{[Z_3]}{k_g}\right)^{n_g} + \left(\frac{[T_2][Z_3]}{k_g^2}\right)^{n_g}}, \quad (10) \end{aligned}$$

except for the cross term at the denominator $\left(\frac{[T_2][Z_3]}{k_g^2}\right)^{n_g}$. At the same time, the considerations made for the Hill function in Section VI-A can be extended to the complementary Hill function. Again, using the same notation as in (6), for $K_p = 1$ and $n = 2$, we obtain

$$\frac{d[Out]}{dt} = \frac{1}{1 + [In]^2} \cong 1 - \tanh\left(\frac{[In]}{2}\right). \quad (11)$$

This means that $\frac{1}{1+[T_2]^2} \cong 1 - \tanh\left(\frac{[T_2]}{2}\right)$ and $\frac{1}{1+[Z_3]^2} \cong 1 - \tanh\left(\frac{[Z_3]}{2}\right)$, therefore $\tilde{H}([T_2], [Z_3]) \cong (1 - \tanh\left(\frac{[T_2]}{2}\right))(1 - \tanh\left(\frac{[Z_3]}{2}\right))$ when $k_g = 1$ and $n_g = 2$. If then we consider the error between H and \tilde{H} negligible, our gene controlled by two different repressors gives as output something very similar to $(1 - \tanh\left(\frac{[T_2]}{2}\right))(1 - \tanh\left(\frac{[Z_3]}{2}\right))$.

By expanding the presented product $\frac{d[A_0]}{dt} = H([T_2], [Z_3]) = 1 - \tanh\left(\frac{[T_2]}{2}\right) - \tanh\left(\frac{[Z_3]}{2}\right) + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)$, we realize that, in steady state, $[A_0]$ contains the desired quantity. In conclusion, for retrieving $\tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)$, we just need to subtract the quantity $1 - \tanh\left(\frac{[T_2]}{2}\right) - \tanh\left(\frac{[Z_3]}{2}\right)$ to $[A_0]$. To do that, we exploit the other two Gene Expression Circuits, as shown in Fig. 8, which have both a promoter regulated by an activator. These circuits take as input the molecules U_2 and U_3 , output

complexes of a Storage Operation on u_2 and u_3 at the second and third bit interval, respectively. U_2 and U_3 have concentration values $L(y_2|x_2)$ and $L(y_3|x_3)$, respectively. Both these Gene Expression Circuits give as output (operons) the molecules A_0, T_0, D_0 whose concentrations are related to the input through the Hill function in (6).

From Section IV-B, $[u_2] = [t_2]$ and $[u_3] = [z_3]$, and consequently $[U_2] = [T_2]$ and $[U_3] = [Z_3]$. The output concentrations of the two genes will be $\frac{[U_2]^n}{K_p^n + [U_2]^n} = \frac{[T_2]^n}{K_p^n + [T_2]^n} \cong \tanh\left(\frac{[T_2]}{2}\right)$ and $\frac{[U_3]^n}{K_p^n + [U_3]^n} = \frac{[Z_3]^n}{K_p^n + [Z_3]^n} \cong \tanh\left(\frac{[Z_3]}{2}\right)$ for $K_p = 1$ and $n = 2$. Since the output species are the same for all genes, their concentrations sum up leading to

$$\begin{aligned} [A_0] &= [T_0] = [D_0] = 1 - \tanh\left(\frac{[T_2]}{2}\right) - \tanh\left(\frac{[Z_3]}{2}\right) \\ &+ \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right) + \tanh\left(\frac{[T_2]}{2}\right) + \tanh\left(\frac{[Z_3]}{2}\right) \\ &= 1 + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right), \quad (12) \end{aligned}$$

which is exactly the argument of the first \log in (7). Note that the expression on the left hand side should be a temporal derivative that has the same value of the concentration itself, according to our steady state assumption.

To obtain the argument of the second \log in (7), we use the molecules T_0 and D_0 , which end up producing two transcription factors with identical concentrations as input to the $\log(\cdot)$. This results in the better fit of the Hill function to the logarithmic function detailed in (8) and shown in Fig. 12. From now on, we focus on protein T_0 since D_0 goes through the same processing.

At the output of the Gene Expression Circuits, T_0 reacts with the molecule M_T having concentration 2. In this way, at the end of the reaction we obtain the desired value $[M_T] = 2 - (1 + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)) = 1 - \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)$. In the end, we are going to use A_0 as the input to the upper $\log(\cdot)$ block, and M_T, M_D as the inputs to the lower $\log(\cdot)$ block.

The two $\log(\cdot)$ blocks should have as output the same molecules P , since from (7) the second \log is preceded by a minus sign that reverses its actual sign (negative because $1 - d < 1$). Ideally, *i.e.*, by neglecting the approximations, at the output of the $\log(\cdot)$ blocks we have

$$\begin{aligned} [P] &= \log\left(1 + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &- \log\left(1 - \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &= 2 \operatorname{atanh}\left(\tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &= 2 \operatorname{atanh}\left(\left|\tanh\left(\frac{L(y_2|x_2)}{2}\right)\tanh\left(\frac{L(y_3|x_3)}{2}\right)\right|\right), \quad (13) \end{aligned}$$

which is exactly the magnitude, or absolute value $|\cdot|$, of the desired box-plus operation expressed in (3).

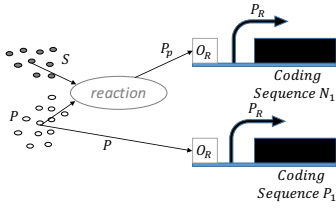
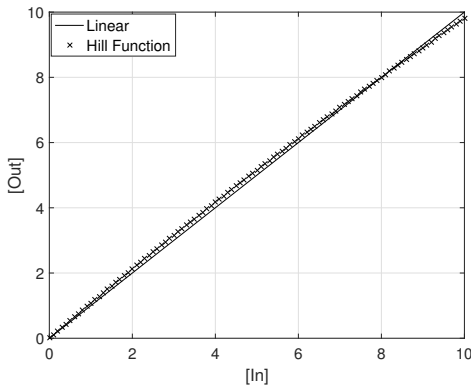


Figure 13. Sign of the box-plus expression.

D. Sign of the Box-plus

The Sign of the Box-plus block in Fig. 8 has the specific purpose of computing the sign of the box-plus operation expressed in (3). The sign of the box-plus is negative only when $L(y_2|x_2)$ and $L(y_3|x_3)$ have opposite signs, which means that both n_0 and m_0 proteins are expressed. These proteins bind into the complex S_b , leading to production of the sign protein S through the Threshold Circuit. The latter is a simple gene circuit with a very steep response that gives a high output whenever there is input. If the protein S is present, it reacts with the $\log(\cdot)$ block output protein P through a ligand-receptor binding, creating a complex that binds to a Linearized Gene Circuit, detailed next, with n_1 as output (*negative* protein). If the protein S is not present, the protein P alone binds to another Linearized Gene Circuit that expresses the protein p_1 (*positive* protein). In Fig. 8, at the output of the Linearized Gene Circuits, the complexes N_1 and P_1 are shown. They result from a Storage Operation on n_1 and p_1 , respectively, at the third bit interval T_b .

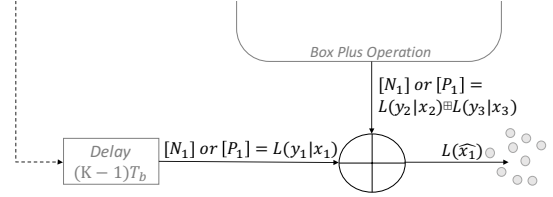
The aforementioned Linearized Gene Circuits, whose biological schematic is shown in Fig. 13, are obtained through promoters with a linear response in a given interval $[0, \delta]$. For this, we need to set $MAX \gg \delta$, $K_p \gg \delta$ and $n = 1$, where MAX , K_p and n are defined in Sec. III. For example, if we set $\delta = 10$, $MAX = 100$ and $K_p = 92$, we obtain a good approximation, see Fig. 14.


 Figure 14. Comparison between a linear function and a Hill function with $MAX = 100$, $K_p = 92$ and $n = 1$ for $[In] \in [0, 10]$

VII. SUM OPERATION TO COMPUTE $L(\hat{x}_1)$

To obtain the complete expression of a-posteriori log-likelihood ratio (L -value) $L(\hat{x}_1)$ in (3), as shown in Fig. 15 we sum the output of the biological L -value computation from the first received modulated concentration, *i.e.*, $L(y_1|x_1)$, which

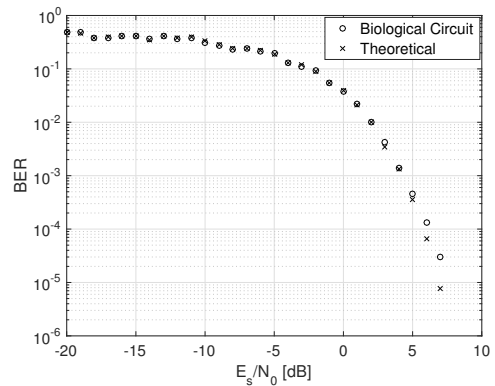
is stored in the concentration of the Biological Delay Line complexes containing N_1 or P_1 as described in Sec. V, with the result of the box-plus operation after the Sign Inversion Block, as described in Sec. VI-D, in terms of N_1 or P_1 , whose concentration is equal to the expression in (3). For this, we assume that the output molecules N_1 or P_1 from the Sign Inversion Block react with the same receptors utilized in the Biological Delay Line, thus forming N_1 or P_1 complexes.


 Figure 15. Arithmetic sum to compute the a-posteriori L -value.

In the following, N_1 is the *negative* complex and P_1 is the *positive* complex. We assume that N_1 complexes and P_1 complexes react with each other. If $L(y_1|x_1)$ and $L(y_2|x_2) \boxplus L(y_3|x_3)$ are of different species, at the end of the reaction, we are left with the *positive* P_1 or *negative* N_1 complex, whose concentration is the a-posteriori L -value $L(\hat{x}_1)$ (Subtraction Operation). On the contrary, in the case $L(y_1|x_1)$ and $L(y_2|x_2) \boxplus L(y_3|x_3)$ are of the same species, we assume that they will not react, and their concentrations will sum, leading to the desired result.

VIII. NUMERICAL RESULTS

To test the performance of the proposed biological circuit design, we implemented the block diagram in Fig. 2 in the Matlab SimBiology environment, generalized to compute the L -value for each of the three bits. The value of the bit interval has been tuned such that all the biological circuits achieve the steady state condition, approximated by the condition that all the output concentrations should be above 99% of their ideal steady state value. The AWGN is generated by *randn* and summed to the transmitted channel bits according to the assumptions in Sec. II to get the received modulated concentrations y_i . The sequence of y_i is then passed as input to our code, which computes an estimate of $L(\hat{x}_k)$, $k = 1, 2, 3$, after the third bit time interval of each codeword.


 Figure 16. Bit Error Rate computation for $\frac{E_s}{N_0} \in [-20, 7]$ dB

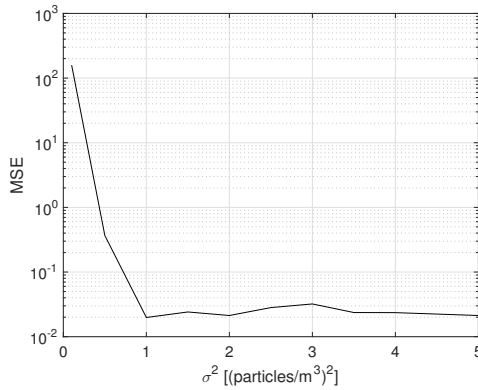


Figure 17. MSE for $\sigma^2 = \{0.1, 0.5, 1, 1.5, 2, \dots, 5\}$ [(particles/m³)²]

To have a comparison between the theoretical performance achieved by the ideal L -value computation and that given by our circuit, we run Monte Carlo simulations to measure the BER versus signal-to-noise ratio $\frac{E_s}{N_0} = \frac{1}{2\sigma^2}$. These results, shown in Fig. 16, are obtained by deciding on the received bit according to the L -value sign, *i.e.*, Maximum A-Posteriori detection. The number of transmitted codewords N used to test the performance is variable and depends on the E_s/N_0 value. It has been chosen as a trade-off between computational time and reliability of the estimate. The number of codewords for BER calculation varies from $N = 100$ codewords at $E_s/N_0 = -20$ dB to $N = 3 \cdot 10^6$ at $E_s/N_0 = 7$ dB.

It is also of interest to analyze the error of our biological circuit in terms of L -values. We calculated $L(\hat{x}_k)$ for 90 transmitted channel bits, belonging to $N = 30$ different codewords, impaired by Gaussian noise with variance σ^2 , and we measured the MSE with respect to the L -values provided by our biological circuit for $\sigma^2 \in [10^{-1}, 5]$, as reported in Fig. 17. A higher error for smaller noise values can be observed, since high values of log-likelihood ratios are obtained leading to $\tanh(L(y_i|x_i)) \cong 1$. From (13), this results into very large output values of the logarithmic function which are not well approximated by our biological circuit elements. In any case, even when the MSE for the L -values is large, the performance of our circuit in terms of BER are very close to those resulting from an electrical circuit implementation [15].

IX. CONCLUSION

In this paper, inspired by recent studies favoring the efficiency of analog computation over digital in biological cells, we proposed an analog decoder design for Molecular Communication (MC) based on the genetic engineering of biological circuits, realized entirely in the biochemical domain by using activation and repression of gene expression, and reactions of molecular species. Biochemical simulation data of the resulting biological circuit demonstrate very close performance to an electrical network implementation in terms of BER and low MSE with respect to the L -values computed with electrical circuits for channels in conditions of relatively high noise. This decoder presented in this paper is intended as a proof-of-concept design methodology for utilizing biological

circuit components to design functionalities in the MC domain, with potential use in the engineering of future devices for the Internet of Things in biological environments. For this reason, the main focus has been devoted to the analysis of the analog decoder by considering only a single transmitter-receiver system, leaving a more realistic study with multiple transmitter and receiver cells to future work.

ACKNOWLEDGMENT

This work was supported by the US NSF through grant MCB-1449014, and the NSF EPSCoR EPS-1004094.

REFERENCES

- [1] E. Sazonov and M. R. Neuman, Eds., *Wearable Sensors: Fundamentals, Implementation and Applications*. Elsevier, 2014.
- [2] S. Andreescu and O. A. Sadik, "Trends and challenges in biochemical sensors for clinical and environmental monitoring," *Pure Appl. Chem.*, vol. 76, no. 4, pp. 861–878, 2004.
- [3] 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.
- [4] B. Atakan, O. B. Akan, and S. Balasubramaniam, "Body area nanonetworks with molecular communications in nanomedicine," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 28–34, January 2012.
- [5] IEEE P1906.1/D2.0, *IEEE Draft Recommended Practice for Nanoscale and Molecular Communication Framework*. IEEE Standards, 2015.
- [6] 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.
- [7] C. J. Myers, *Engineering genetic Circuits*. Chapman & Hall, 2009.
- [8] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, and I. Netravali, "Genetic circuit building blocks for cellular computation, communications, and signal processing," *Natural Computing*, vol. 2, no. 1, pp. 47–84, March 2003.
- [9] 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.
- [10] R. Sarpeshkar, "Analog synthetic biology," *Philos Trans A Math Phys Eng Sci.*, vol. 372, no. 2012, p. 20130110, 2014.
- [11] S. Payne and L. You, "Engineered cell-cell communication and its applications," *Adv Biochem Eng Biotechnol*, vol. 146, pp. 97–121, 2014.
- [12] D. B. Menendez, V. R. Senthivel, and M. Isalan, "Sender-receiver systems and applying information theory for quantitative synthetic biology," *Curr Opin Biotechnol*, vol. 31C, pp. 101–107, March 2015.
- [13] 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.
- [14] B. Unluturk, A. Bicen, and I. Akyildiz, "Genetically engineered bacteria-based biotransceivers for molecular communication," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1271–1281, 2015.
- [15] J. Hagenauer, E. Offer, C. Measson, and M. Morz, "Decoding and equalization with analog non-linear networks," *European Transactions on Telecommunications (ETT)*, 1999.
- [16] A. Mandelis, *Diffusion-wave fields: mathematical methods and Green functions*. Springer-Verlag, 2001.
- [17] 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.
- [18] G. Tkačik, A. M. Walczak, and W. Bialek, "Optimizing information flow in small genetic networks," *Phys Rev E Stat Nonlin Soft Matter Phys.*, vol. 80, no. 3 Pt 1, p. 031920, September 2009.
- [19] D. Kilinc and O. B. Akan, "Receiver design for molecular communication," *IEEE Journal On Selected Areas In Communications/Supplement - Part 2*, vol. 31, no. 12, pp. 705–714, December 2013.
- [20] B. Alberts, A. Johnson, and *et al.* J. Lewis, *Molecular Biology of the Cell. 4th edition*. New York: Garland Science, 2002.
- [21] J. Ang, E. Harris, B. J. Hussey, R. Kil, and D. R. McMillen, "Tuning Response Curves for Synthetic Biology," *ACS Synthetic Biology*, 2013.
- [22] 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.
- [23] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.