A Biological Circuit Design for Modulated Parity-check Encoding in Molecular Communication

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*—Regarded as one of the future enabling technologies of the Internet of Things at the biological and nanoscale domains, Molecular Communication (MC) promises to enable applications in healthcare, environmental protection, and bioremediation, amongst others. Since MC is directly inspired by communication processes in biological cells, the engineering of biological circuits through cells' genetic code manipulation, which enables access to the cells' information processing abilities, is a candidate technology for the future realization of MC components. In this paper, inspired by previous research on channel coding schemes for MC and biological circuits for cell communications, a joint encoder and modulator design is proposed for the transmission of cellular information through signaling molecules. In particular, the information encoding and modulation are based on a binary parity check scheme, and they are implemented by interconnecting biological circuit components based on gene expression and mass action reactions. Each component is mathematically modeled and tuned according to the desired output. The implementation of the biological circuit in a simulation environment is then presented along with the corresponding numerical results, which validate the proposed design by showing agreement with an ideal encoding and modulator scheme.**

***Index Terms*—Molecular communication; biological circuit; parity-check encoding; Hill function; chemical reaction modeling; biochemical simulation**

I. Introduction

Recent novel advances in the study of communication systems based on molecule exchange have contributed to the research field of Molecular Communication (MC) [1], [6], regarded as one of the future enabling technologies of the Internet of Things at the biological and nanoscale domains [2]. Future applications of MC-capable systems include the networking of wearable and implantable devices for healthcare [22], environmental protection [5], and bioremediation [10], amongst others.

The concept of MC is directly inspired by communications among cells in the biological environment [1], where molecular information exchange underlies cells' interactions and coordination of uni- and multi-cellular organisms, populations, and multi-species consortia, and participates in most of the major cellular functionalities such as cell growth, proliferation and apoptosis. As a consequence, the modeling of natural MC processes in cells with communication engineering frameworks and tools has gained a growing interest in recent years. Examples range from the modeling of bacteria conjugation and electron transfer [14], to the characterization of signal transduction pathways [11], [24] and metabolic regulation [21].

The discipline of synthetic biology is providing the engineering community with novel tools and techniques to tap into cells and their functionalities for the design, realization, and control of biological processes [9]. In particular, the theory of biological circuits, based on the manipulation of DNA genes linked together by activation and repression mechanisms, enables engineers to design, and program, cellular functionalities and behaviors based on a predefined set of basic components and rules [15]. While complete experimental characterization and standardization of these components are still open challenges, it is today possible to theoretically study and predict in-silico the behavior of engineered biological circuits of great complexity [15], even if their experimental implementation is still limited to simpler cases.

The engineering of MC components and systems in cells through synthetic biology has gained particular interest in the last couple of years [18], mostly through the manipulation of natural MC processes such as bacterial quorum sensing. Recent literature in MC is exploring the possibility of utilizing biological circuits to realize MC functionalities. In [19] 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 and analytically characterized. 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 [25], based on digital-like biological circuit functionalities and M-ary molecule concentration modulation coupled with hard threshold detection.

In this paper, we propose the design, simulation, and characterization of a biological circuit for MC to realize the binary parity-check encoding of molecular information in a cell, and its modulation through the emission of signaling molecules, *i.e.*, molecules utilized in nature for intercellular information exchange, such as in bacterial quorum sensing [18]. The synthetic biology design of a binary encoding and modulation system of cellular information is motivated as a natural extension of the aforementioned studies of biological circuits for MC. In [12], classical channel coding schemes have been considered within the MC realm, and characterized

on the basis of their feasibility and performance. Inspired by this work, we detail here an implementation of the simplest coding scheme, namely, the binary parity check coding, and subsequent binary modulation of output signaling molecules, by utilizing the component and rules of biological circuit theory. In particular, our design implements the simplest binary parity-check code, *i.e.*, with a block length of 3 bits, tuned in its parameters according to the desired output. Biochemical simulation data of the resulting biological circuit demonstrate very close behavior to the ideal scheme.

The rest of the paper is organized as follows. In Sec. II we introduce the modulated parity-check encoder for MC and motivate its implementation to realize cell-to-cell communications. In Sec. III we detail the main components of a biological circuit and their mathematical models. In Sec. IV the encoder design is detailed by describing the necessary biological circuit components and the tuning of their parameters, while in Sec. V the implementation of the biological circuit in a simulation environment is presented along with the corresponding numerical results. Finally, in Sec. VI we conclude the paper.

II. A Modulated Parity-check Encoder for Molecular Communication

A cell can be defined as a finite environment that contains n chemical species $\{S_1, ..., S_n\}$ linked together by chemical reactions [15]. From this definition, it is possible to characterize a cell according to its state $\mathbf{u}(t) = \{u_i(t)\}_{i=1}^{n}$ as a function of the time t, where the single element $u_i(t)$ represents the molecular concentration of the species S_i at time t. The chemical communication of information about this state, either complete or partial, to a recipient outside of the cell, *e.g.*, another cell or man-made device present in the external environment, is a valid abstraction of an MC system.

With reference to Fig. 1, we consider for simplicity a cell, *i.e.*, **Transmitter Cell**, that contains only two species, namely, S_1 and S_2, whose concentrations define the cell state at time t, $\mathbf{u}(t) = (u_1(t), u_2(t))$. The concentration $u_i(t)$, $i = 1, 2$, can be approximated as a continuous-time binary variable with value 1 if the i-th species is present, and 0 if absent. These variables contain a portion of information, in bits, on the aforementioned cell state at time t. If we reasonably assume that the cell is immersed in a fluidic environment, the simplest cell-based MC system would be based on the release of signaling molecules to the external environment, and would rely on their Brownian motion to propagate this information to a recipient at a remote location within the same fluid. To reduce the complexity, a general trend in nature is to rely on a limited subset of signaling molecule species compared to the state-defining species, such as the N-acyl homoserine lactones (AHL) in Gram-negative bacteria [16].

This aforementioned signaling species is generally released by natural cells according to a continuous-time molecule release rate signal, usually abstracted in MC through the common On-Off-Keying (OOK) modulation scheme [12]. The Brownian motion propagation of these molecules could lead to noise at reception that is theoretically modeled by a Poisson

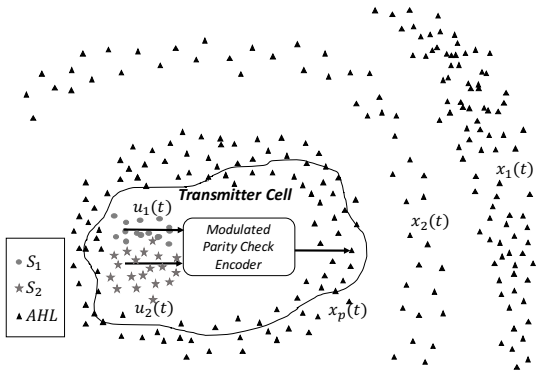

Fig. 1. Pictorial representation of the considered MC scenario.

distribution [17], which is prone to consequent errors in understanding the Transmitter Cell state. To alleviate this issue, we propose to design a Modulated Parity-Check Encoder that takes as input the information bits in $\mathbf{u}(t)$ and gives as output two-level modulated and encoded symbols $x_i(t)$. The proposed parity check code allows just for error detection. Nevertheless, it could be used as basic building block to construct more complex codes, *i.e.* LDPC or Hamming codes [12], that allow the recipient not only to detect but also correct errors. In addition, our encoder design is realized with biochemical processes at the basis of the biological circuit theory, namely, gene expression and mass action chemical reaction. In nature, cells do not usually utilize a zero release rate of signaling molecules [18], necessary for OOK modulation. This is also in agreement with the evidence of basal expression rate that characterizes many DNA genes [3], including those responsible for the release of the aforementioned AHL molecules [16]. In our work, we consider that the transmitted encoded symbols $x_i(t)$ will assume in general the positive values a_0, a_1 when bit 0, 1 are transmitted, respectively. Hence, a direct mapping between bits and symbols is realized, and OOK is here replaced with a two-level signaling modulation of the transmitted encoded symbols, which is more realistic. In addition, using a tunable level a_0 in place of no transmission for bit 0, as in OOK, allows for higher flexibility in the design of coded modulation schemes.

III. A Biological Circuit

A *biological circuit* is a network of *genes* and *chemical reactions* that implement a specific biological function [15].

A. Gene Expression

As shown in Fig. 2, a gene is 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, expressed from the gene through the fundamental processes of *transcription* and *translation*. Protein expression can be up or down-regulated by a *transcription factor* protein *In*, *activator* (a) or *repressor* (b), respectively. When the gene expresses proteins independently from transcription factors, it is said to have a *constitutive* promoter. Protein expression is based on [15]:

- **Transcription** is initiated by the enzyme (a specific type of protein) *RNA polymerase (RNAP)* that binds to the promoter

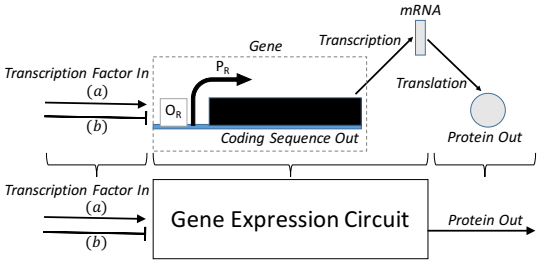

Fig. 2. Gene expression scheme: (a) Activation (↓), (b) Repression (⊥).

region of the considered gene, starting the production of the *messenger RNA (mRNA)* molecules. These latter molecules are used to carry the genetic information encoded in the coding sequence of the gene to the *ribosome*, the protein assembly machinery.

- **Repression** is present when repressors obstruct the binding sites of the promoter region and down-regulate the transcription of the subsequent coding sequence by reducing RNAP binding rate. *Corepressors* in prokaryotes, like the *E. coli*, are small molecules that bind and activate a repressor transcription factor [16].
- **Activation** happens if activators bind to the operator region near the promoter site up-regulating the transcription of the subsequent coding sequence by increasing the RNAP binding rate. *Inducers* are small molecules that bind and activate activators. There are, in fact, activators and repressors that, without the respective inducers and corepressors, bind poorly to the operator region causing no actual change in the transcription rate.
- **Translation** is performed through the ribosomes, which are able to recognize and bind to the mRNA molecules by means of *Ribosome Binding Sites (RBSs)*, special sequences of nucleotides in the mRNA strand. Once a ribosome binds to the RBS of an mRNA molecule, it completes the synthesis of the corresponding protein by assembling together the component amino acids.

The aforementioned processes of transcription, activation, repression, and translation for protein synthesis are generally modeled as a single event by using the so-called **Hill function** ([4],[3]). As a consequence, the rate $\frac{d[Out]}{dt}$ of output protein, in case of activation, is expressed as

$$\frac{d[Out]}{dt} = k' + MAX\left(\frac{([In]/K)^n}{1+([In]/K)^n}\right) - k_{deg}[Out], \quad (1)$$

where $[Out]$ is the concentration of the output protein Out, 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 rate value at the output, K 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, which we define also as the output production rate. k_{deg} is the degradation rate of the output proteins, defined in the following. In this paper, we assume to have non-leaky promoters, *i.e.* $k' = 0$, which means that there is gene expression only when external activating signals are present [4]. In (1) and hereafter, the square brackets notation

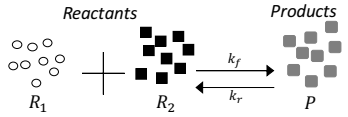

Fig. 3. Mass Action Chemical Reaction scheme between two generic molecular species $R_1 + R_2 \underset{k_r}{\overset{k_f}{\rightleftarrows}} P$.

stands for molecule concentration. According to [4], in case of repression, the rate $\frac{d[Out]}{dt}$ of output protein is

$$\frac{d[Out]}{dt} = MAX\left(\frac{1}{1+([In]/K)^n}\right) - k_{deg}[Out]. \quad (2)$$

In general, genes can be regulated by more than one transcription factor. In such a case, gene expression can be described by a **multi-dimensional Hill function** [3] as follows

$$\frac{d[Out]}{dt} = \frac{\sum_i MAX_i([In_i]/K_i)^{n_i}}{1+\sum_i([In_i]/K_i)^{m_i}} - k_{deg}[Out], \quad (3)$$

where In_i refers to the i-th transcription factor, and $n_i = m_i$ if the i-th transcription factor is an activator, while $n_i = 0$ and $m_i > 0$ if it is a repressor.

It is possible to convert the differential equations in (1), (2), and (3) to more practical non-differential expressions by simply considering the steady state condition. In the steady state, the degradation rate equals the production rate, yielding no more temporal variation of the output concentration. By equating (3) (general case) to zero, the steady state output concentration $[Out]^{SS}$ is found with the following expression:

$$[Out]^{SS} = \frac{1}{k_{deg}} \frac{\sum_i MAX_i([In_i]/K_i)^{n_i}}{1+\sum_i([In_i]/K_i)^{m_i}}, \quad (4)$$

which corresponds exactly to the production rate if $k_{deg} = 1\,s^{-1}$.

B. Mass Action Chemical Reaction

A Mass Action Chemical Reaction is a process that converts one or more input molecules (*reactants*) into one or more output molecules (*products*). Reactions may proceed in forward or reverse directions, which are characterized by forward (k_f) and reverse (k_r) reaction rates, respectively. Within the scope of this paper, we assume unbalanced reactions where the forward reaction rate is much greater than the reverse rate, as in [17]. An example of a reaction with two reactant species and one product species is shown in Fig. 3. In this work, we will consider the following Mass Action Chemical Reactions:

- **Transcription Factor Activation Reaction:** We consider two reactant species, a repressor (activator) transcription factor and its corresponding corepressor (inducer). The corepressors (inducers) bind to specific sites on the particular transcription factor proteins and produce a steady state concentration of *complexes* (activated transcription factors) equal to the concentration of transcription factors if the initial corepressors (inducers) concentration is sufficiently high. It is generally assumed that a single transcription factor molecule can bind only one corepressor or inducer.

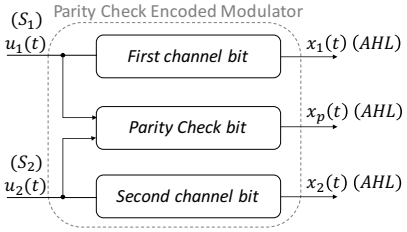

Fig. 4. Modulated Parity Check Encoder.

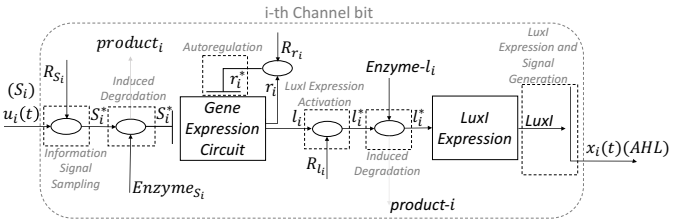

Fig. 5. First and Second Channel bit production with $i = 1$ and $i = 2$, respectively.

- **Degradation Reaction:** We consider complexes and enzymes as reactants, and at least two product species. The reactant complexes are bound by the enzymes and split into simpler molecular species (the products). The mathematical model for the example of Fig. 3 is expressed through a reaction rate equation [15] as

$$\frac{d[P]}{dt} = k_f[R_1][R_2] - k_r[P]. \tag{5}$$

IV. Modulated Encoder Design based on Biological Circuits

The modulated encoder design proposed in this paper is sketched in Fig. 1. It encodes two bits representing the state of the cell at time t, *i.e.*, $\mathbf{u}(t) = (u_1(t), u_2(t))$, with a Single Parity Check (SPC) code characterized by codeword length $K = 3$ bits, and modulates the transmission of the codeword bits with two levels, namely, a_0 and a_1, for the bit 0 and 1.

The proposed design is realized through three main branching biological circuits as shown in Fig. 4. Each branch encodes and modulates a codeword symbol, among which $x_p(t)$ is the parity check symbol. All the symbols are transmitted using the same molecular species AHL to reduce the complexity of communications and the energy expenditures [23]. In order to gain insight into these processes, in the following we consider the i-th branch, $i = 1, 2$, that produces $x_i(t)$ as reported in Fig. 5. There are two main types of components in our circuit:

- Gene Expression (**rectangular blocks** in Fig. 5), where, as detailed in Sec. III-A, a gene repressed/activated by some transcription factor species (input) synthesizes proteins (output) which may subsequently be transcription factors for other rectangular blocks.
- Mass Action Chemical Reaction (**oval blocks** in Fig. 5), which corresponds to either a Degradation Reaction or a Transcription Factor Activation Reaction, as in Sec. III-B.

A. The Sampling of the Information Signal

The concentrations $u_1(t)$ and $u_2(t)$ are continuous time binary variables that carry the information on the state of the cell over time. In order to send this information to neighboring cells at time $t = \bar{t}$, we must consider $\mathbf{u}(t = \bar{t})$ which means we have to freeze the state of the cell at that given instant. As in telecommunication, we achieve that by sampling. At $t = \bar{t}$ we just inject a sufficiently high number of corepressors R_{u_i} that, through the Transcription Factor Activation Reaction, activate the molecular concentration of input (repressive) transcription factors $u_i(t = \bar{t})$. As a result we drop the time dependency notation in Fig. 5 after the sampling operation, with u_i^* being the sample of $u_i(t)$ at time $\bar{t}$. u_i^* is the concentration of the activated species S_i^* obtained from the Transcription Factor Activation Reaction $S_i + R_{u_i} \rightarrow S_i^*$, where we have the particular case in which $R_1 \equiv S_i$, $R_2 \equiv R_{u_i}$, and $P \equiv S_i^*$. The subscript i is hereafter considered to assume either the value 1 (First Channel bit) or 2 (Second Channel bit).

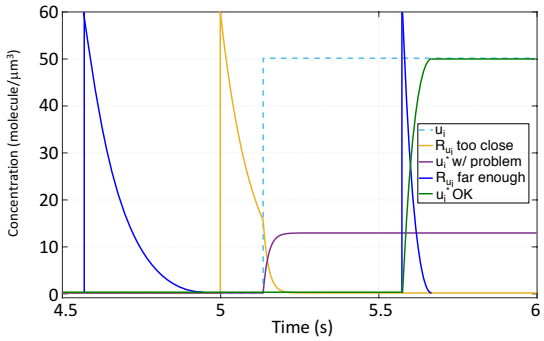

Fig. 6. Sampling time $\bar{t}$ close to a raising edge of the variable $u_i(t)$. In the simulation, the variable $u_i(t)$ has been assumed to be a square wave with high level 50 molecule/μm^3 and the injection of the corepressors R_{u_i} has been considered as an instantaneous event. The case where $t = \bar{t} = 5\ s$ shows the problem of being too close to the raising edge of $u_i(t)$.

Unlike digital communications with electrical circuits, in the context of biological circuits we cannot approximate sampling operations as instantaneous processes. Even if we consider the injection of molecules as instantaneous, the degradation time generally happens at a time scale comparable to the rest of the biological processes, such as gene expression. This phenomenon, in the context of digital communication, is known as the "aperture effect" [7]. For this reason and also for the stochasticity of the state of the cell, an issue can arise if the sampling instant $\bar{t}$ coincides (or is very close to) a raising or falling edge of the variable $u_i(t)$, as illustrated in Fig. 6. In the first interval (for $t \geq 5\ s$ but before u_i goes up), R_{u_i} molecules degrade just because of natural degradation since there are no S_i molecules around ($u_i(t = \bar{t}) = 0$). During the sampling process, u_i changes its state causing the residual corepressors R_{u_i} to bind (notice the increased consumation speed since the corepressors are both sequestered by the reaction with u_i and naturally degraded) and produce some complexes S_i^* with concentration u_i^*. In the end, u_i^* will be different from what it was supposed to be, as shown by the purple curve in Fig. 6. When sampling, in fact, we aim to have $u_i^* = u_i(t = \bar{t})$, achieved when $\bar{t}$ is not close to any edge. This issue is taken into account in the simulation of the modulated encoder, and it results in a random contribution to the concentration u_i^* at the input of the Gene Expression Circuit as a consequence of the random parameters phi_1 and phi_2, as explained in Sec. V.

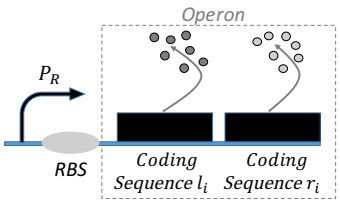

Fig. 7. Operon scheme.

B. Production of the Signaling Molecules AHL for the First and Second Channel bits

1) Gene Expression Circuit: Once produced, the complexes S_i^* repress the promoter of the gene in the Gene Expression Circuit block in Fig. 2b. This gene contains two different coding sequences (operon), therefore encoding for two different output protein species, l_i and r_i. An operon is a group of coding sequences controlled by the same promoter and expressing the same protein concentration if their RBSs have similar binding affinity with the ribosome. A simple representation is shown in Fig. 7.

The species l_i is a transcription factor used to activate the subsequent LuxI Expression block, and its steady state concentration $[l_i]^{SS}$ value ranges between a minimum and a maximum value (depending on u_i^*) without ever being zero. A zero $[l_i]^{SS}$ means, in fact, no activation of the LuxI Expression, hence no production of signaling molecules AHL yielding an unwanted OOK modulation. For this reason, when u_i^* is high, the Gene Expression Circuit activity is not completely repressed but is downscaled with respect to the maximum value, obtained when u_i^* is low. Therefore $[l_i]^{SS}$ reports the information on the value of u_i^*. The species r_i works as an autoregulator of the gene expression, as detailed next.

2) Gene Expression Circuit Autoregulation and Activation of the LuxI Expression: The species l_i is not directly used as input of the LuxI Expression block because otherwise we would not be able to control the transmission time, since AHL molecules would be continuously expressed. Proteins l_i start activating LuxI Expression only when they are activated by some inducers R_{l_i} through the Transcription Factor Activation Reaction. These molecules will be injected only when the first symbol has to be transmitted on the channel. However, since $[l_i]^{SS}$ is a crucial quantity carrying information about u_i^*, we have to maintain this value throughout the activation process even if it is not known a priori since it depends on the state of the cell at time $\bar{t}$. It follows that we are forced to inject a quite high amount of molecules R_{l_i} in order to be sure that we will not lose the concentration information even when the Gene Expression Circuit output is maximum (u_i^* low). High R_{l_i} concentration, however, leads to errors if during the binding process the l_i proteins keep being produced because the residual R_{l_i} molecules will bind to the newly produced l_i proteins. As a consequence, the transcription factors r_i^* are in place to completely repress the production of new l_i.

3) LuxI Expression and Signal Generation: The last block, LuxI Expression, is a simple gene activated by l_i^* and with LuxI as coding sequence. The parameters of the related promoter are engineered to give a steady state output concentration ($[LuxI]^{SS} = I^{SS}$) equal to the value a_0 (transmitted signal for the bit 0) when $[l_i^*]^{SS}$ assumes its maximum value (u_i^* low), and a_1 (transmitted signal for the bit 1) when $[l_i^*]^{SS}$ is at its minimum (u_i^* high).

Finally, the LuxI enzymes activate the production of the signaling molecules AHL whose temporal variation is our transmitted signal $x_i(t)$. As in [8], if A is the concentration of AHL, $I(t)$ the concentration of LuxI as a function of the time t and I^{SS} its steady-state concentration, we get the following differential equation:

$$x_i(t) = \frac{dA}{dt} = k_0 I(t) = k_0 I^{SS} = k_0 a_{0/1}, \tag{6}$$

where $k_0 = 1\ s^{-1}$.

4) Induced Degradation of Molecular Species: After transmission of the signal $x_i(t)$, enzymes with a degradation rate (Enzyme-l_i in Fig.5) are injected to react with and degrade l_i^*. This operation is needed for two reasons. Firstly, LuxI production by that particular LuxI Expression block has to be stopped when the following symbol has to be sent on the channel, since AHL molecules are used for transmission of all the channel symbols. That way, InterSymbol Interference (ISI) (at least in transmission) is mitigated. Additionally, if l_i^* complexes are not degraded before transmission of the i-th channel bit of the successive codeword, InterBlock Interference (IBI) (in transmission) might occur. IBI occurs also if the enzymes do not degrade. In both cases I_{SS} would be impaired and a wrong signal $x_i(t)$ would be produced. For the same reason, S_i^* complexes have to be degraded. This is achieved with the $Enzyme_{S_i}$ molecules that are injected only after the parity bit has been encoded and modulated.

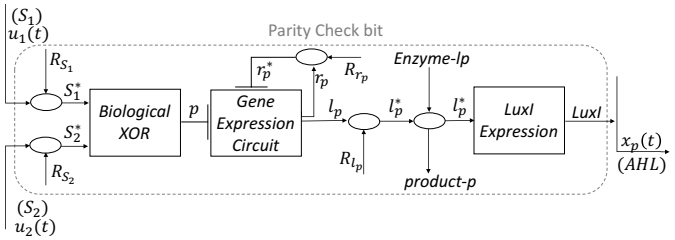

Fig. 8. Parity Check bit.

C. Production of the Signaling Molecules AHL for the Parity Check Bit

The realization of the parity check bit $x_p(t)$ has some differences worth being analyzed separately. Details on the Parity Check bit are reported in Fig. 8.

Here, the samples u_1^* and u_2^* do not act directly on the Gene Expression Circuit block but, instead, are processed by the Biological XOR block to produce the parity check bit p. The Biological XOR block realizes the XOR summation between the information bits u_1^* and u_2^* and it is here designed and modeled as suggested by Myers in [15]. Once p is produced, it goes through the same processing as for the first and second bits, this time to obtain $x_p(t)$. Here, the injection times have

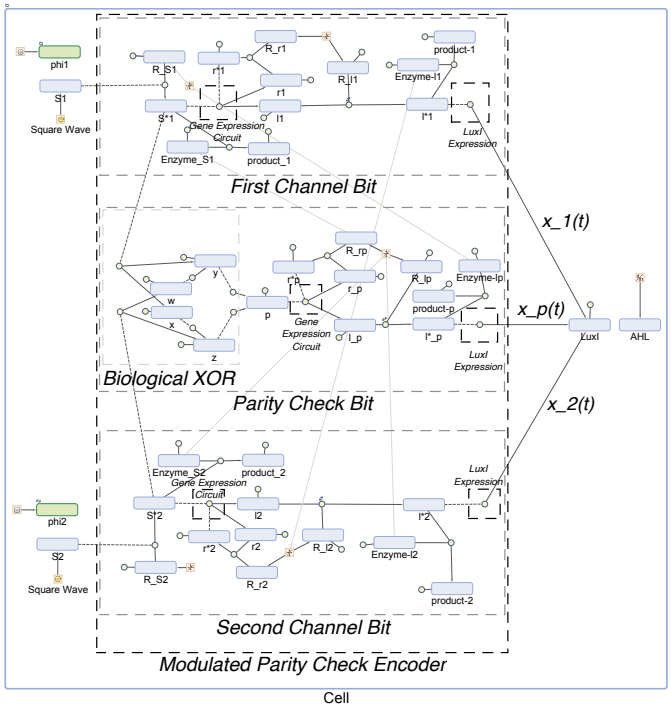

Fig. 9. Simbiology model of the Biological Coding Circuit.

to be tuned in order to transmit the parity check symbol only after $x_1(t)$ and $x_2(t)$.

V. SimBiology Implementation and Simulation Results

The scheme in Fig. 4 has been replicated in Simbiology, a Matlab package for simulation of biological networks, with the result reported in Fig. 9. The concentrations have been considered normalized to the value 1 molecule/μm^3 in a volume $V \cong 10\mu$m^3 [16]. From Fig. 9 it is possible to distinguish three main blocks [13]:

- The **Rectangular Blocks** represent the species involved in the reaction network and their concentrations. The only exception are the green blocks phi_1 and phi_2 that are constant parameters with random value, as explained in the following.
- The **Circular Blocks** represent generic chemical reactions stemming from the simple Degradation and Transcription Factor Activation Reactions to the more complex Repression and Activation of Gene Expression.
- The **Square Blocks** represent events, value assignments or mathematical rules.

The rectangular blue blocks S_1 and S_2 on the left side of Fig. 9 have concentrations u_1 and u_2 defined by continuous time Square Waves with high level $u_i = 50$, $i = 1, 2$, in order to simulate the binary nature of the concentrations $u_1(t)$ and $u_2(t)$ that continuously define the state of the cell. The Square Wave is characterized by a random initial phase (phi_1 and phi_2 for u_1 and u_2, respectively) so as to model the random state of the cell when the coding process begins.

The main Circular Blocks are highlighted inside dashed boxes. The Circular Block inside the dashed box Gene Expression Circuit simulates a repressed gene expression, where the transcription and translation processes are seen as one step process mathematically described by the Hill function, as detailed in Sec. III-A. The main parameters that characterize this Circular Block in First Channel Bit are reported in Table I.A. The Circular Block inside the dashed box LuxI Expression models the activated LuxI Expression in Figs. 5 and 8 and the related parameters are shown in Table I.B. Notation is the same as in Sec. III-A except for the subscripts G and L used to distinguish the Gene Expression Circuit and LuxI Expression parameters, respectively. Since the promoter of the Gene Expression Circuit is repressed by two different transcription factors, S_1^* and r_1^*, it is modeled by a multi-dimensional Hill function, as in (3). The promoter of the LuxI Expression is instead activated by just one transcription factor species l_1^*, hence the Reaction Rate is defined by the one-dimensional Hill function as in (1).

TABLE I
(A) Parameters of the Gene Expression Circuit block, (B) Parameters of the LuxI Expression block

	(A) Gene Expression Circuit	(B) LuxI Expression
Reaction Rate	$\frac{MAX_G}{1+\left(\frac{u_1^*}{k_G}\right)^{n_G}+\left(\frac{r_1^*}{k_{1_G}}\right)^{n_G}}$	$\frac{MAX_L(l_1^*/k_L)^{n_L}}{1+(l_1^*/k_L)^{n_L}}$
$MAX_{G/L}$	6	22
$k_{G/L}$	40	1
$n_{G/L}$	5	3
k_{1_G}	1	–

The parameters of the Gene Expression Circuit have been chosen considering $u_1(t) = 50$ when the species S_1 is present. As in Fig. 10(a), even if $u_1(t) = 50$ at the sampling time ($u_1^* = 50$), assuming $[r_1^*]^{SS} = 0$, the gene is not completely repressed hence, in steady state, $[l_1]^{SS}$ (and likewise $[l_1^*]^{SS}$ for the Transcription Factor Activation Reaction) will be different from zero and able to activate the subsequent LuxI Expression.

The parameters of the LuxI Expression have been engineered to give a steady state output $k_0 I^{SS} = a_0 = MAX_L = 22$ when $[l_1^*]^{SS} = 6 = MAX_G$ (which, in turn, means $u_1^* = 0$) and $k_0 I^{SS} = a_1 \cong 20$ when $[l_1^*]^{SS} \cong 2$ (obtained when $u_1^* = 50$), as in Fig. 10(b). While a thorough discussion is left to future work, these values allow an optimal Gaussian approximation of the molecule counting noise in a diffusion-based channel [20].

Finally, as explained in Sec. IV-B2, r_1^* has to strongly repress the promoter of the Gene Expression Circuit. Since $[r_1^*]^{SS} \equiv [l_1^*]^{SS}$, a small value for k_{1_G} has to be chosen in order to get an efficient repression even when $[r_1^*]^{SS}$ is small (u_1^* high). The result for $k_{1_G} = 1$ is observable in Fig. 10(a). In order to have a quite strong repression, whatever the value of u_1^*, we need $[r_1^*]^{SS} \geq 2$ and this is guaranteed since, even in the case of maximum gene expression input $u_1^* = 50$, the output of the operon (Fig. 7) would be $[r_1^*]^{SS} \equiv [l_1^*]^{SS} \cong 2$.

Simulation curves as function of time are reported in Fig. 11. For a better visual result, only the main species S_1, S_2 and $LuxI$ with concentrations $u_1(t)$, $u_2(t)$ and $I(t)$, respectively, are shown. Notice that, from Eq. (6), $I(t)$ and $x_i(t)$ are directly related, hence $I(t)$ behavior over time gives us information

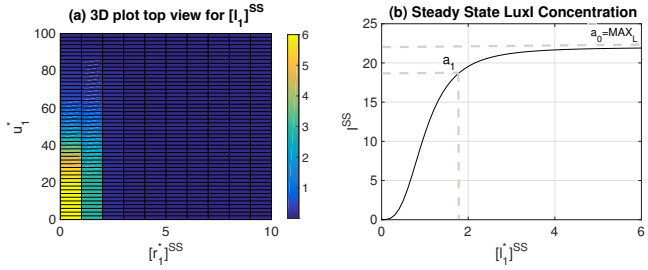

Fig. 10. Steady state concentration (a) $[l_1]^{SS}$ for the Gene Expression Circuit and (b) I^{SS} for the LuxI Expression.

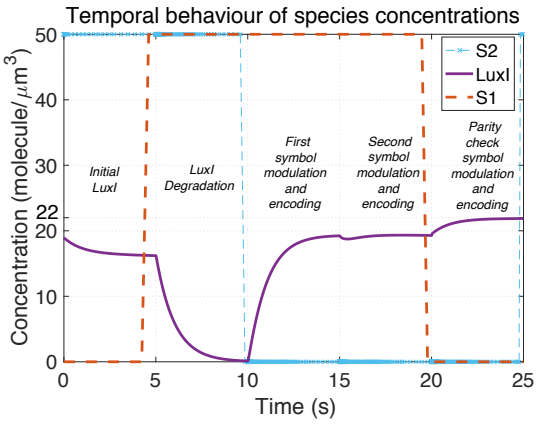

Fig. 11. S_1 (red), S_2 (blue) and *LuxI* (purple) concentrations over time.

about the transmitted symbols. The sampling of $u_1(t)$ and $u_2(t)$ occurs at $t = \bar{t} = 5\,s$. In that instant, the state of the cell $(u_1(t = \bar{t}), u_2(t = \bar{t})) = (50, 50)$ so the parity check bit should be 0. Looking at the variable $LuxI$ in Fig. 11, we realize the information has been encoded and modulated correctly. The encoding process begins at $t = 10\,s$, using the $5\,s$ time gap to degrade any possible residual LuxI molecules from the previous codeword. The symbol time is set to $5\,s$ in order to get the steady state expression of the LuxI Expression block, *i.e.* $a_0 = 22$ when $u_i^* = 0$ and $a_1 \cong 20$ when $u_i^* = 50$ $(i = 1, 2)$. Observing the LuxI concentration in $t = 15\,s$, $t = 20\,s$ and $t = 25\,s$, we get the modulated channel symbols (a_1, a_1, a_0) corresponding to the encoded bits (110), consistently with what we desired to transmit.

VI. Conclusion

In this paper, we propose the design and simulation of a biological modulated parity-check encoder of molecular information in a cell. The biological circuit has been first designed by using concepts provided by synthetic biology and then simulated using SimBiology, which provides programmatic tools to model, simulate, and analyze dynamic systems. The modulated parity-check encoder is able to read and encode the molecular information through serialization of a naturally parallel information. The encoder presented in this paper is intended as a proof-of-concept design methodology that could be replicated to build more complex encoding schemes, with potential use in the engineering of future devices for the Internet of Things in biological environments.

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] I. F. Akyildiz, F. Brunetti, and C. Blazquez, "Nanonetworks: A new communication paradigm at molecular level," *Computer Networks (Elsevier) Journal*, vol. 52, no. 12, pp. 2260–2279, August 2008.
[2] 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.
[3] U. Alon, *An Introduction To Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.
[4] J. Ang, E. Harris, B. J. Hussey, R. Kil, and D. R. McMillen, "Tuning Response Curves for Synthetic Biology," *ACS Synthetic Biology*, vol. 2, no. 10, pp. 547–567, August 2013.
[5] L. T. Bereza-Malcolm, G. Mann, and A. E. Franks, "Environmental sensing of heavy metals through whole cell microbial biosensors: A synthetic biology approach," *ACS Synth Biol.*, vol. 4, no. 5, pp. 535–546, May 2015.
[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] S. Haykin and M. Moher, "Introduction to analog & digital communications, hoboken," 2007.
[8] B. Ingalls, *Mathematical Modelling in Systems Biology: An Introduction*, 2012.
[9] 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.
[10] J. R. Kirby, "Synthetic biology: Designer bacteria degrades toxin," *Nature Chemical Biology*, vol. 6, pp. 398–399, 2010.
[11] M. Laddomada and M. Pierobon, "A crosstalk-based linear filter in biochemical signal transduction pathways for the internet of bio-things," in *Proc. of the IEEE ICASSP*, April 2015.
[12] Y. Lu, M. D. Higgins, and M. S. Leeson, "Comparison of channel coding schemes for molecular communications systems," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 3991–4001, November 2015.
[13] Mathworks, "SimBiology Release 2016b Documentation," https://www.mathworks.com/help/simbio/index.html, 2016, [Online].
[14] N. Michelusi, S. Pirbadian, M. Y. El-Naggar, and U. Mitra, "A stochastic model for electron transfer in bacterial cables," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 12, pp. 2402–2416, Dec. 2014.
[15] C. J. Myers, *Engineering Genetic Circuits*. Chapman & Hall, 2009.
[16] D. L. Nelson, A. L. Lehninger, and M. M. Cox, *Lehninger Principles of Biochemistry*. Macmillan, 2008.
[17] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Transactions on NanoBioscience*, 2014.
[18] S. Payne and L. You, "Engineered cell-cell communication and its applications," *Adv Biochem Eng Biotechnol*, vol. 146, pp. 97–121, 2014.
[19] 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.
[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] M. Pierobon, Z. Sakkaff, J. L. Catlett, and N. R. Buan, "Mutual information upper bound of molecular communication based on cell metabolism," in *Proc. of the IEEE SPAWC*, July 2016.
[22] E. Sazonov and M. R. Neuman, Eds., *Wearable Sensors: Fundamentals, Implementation and Applications*. Elsevier, 2014.
[23] J. T. Sexton and J. J. Tabor, "Multiplexed bacterial cell-cell communication via a genetically encoded crispri-based multiplexer-demultiplexer circuit," in *Proceedings of the 3rd ACM NanoCom*. ACM, 2016, p. 12.
[24] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7358–7382, December 2016.
[25] 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.