Parity-Check Coding Based on Genetic Circuits for Engineered Molecular Communication Between Biological Cells

Alessio Marcone, Massimiliano Pierobon, *Member, IEEE*, and Maurizio Magarini, *Member, IEEE*

***Abstract*— Synthetic biology, through genetic circuit engineering in biological cells, is paving the way toward the realization of programmable man-made living devices, able to naturally operate within normally less accessible domains, *i.e.*, the biological and the nanoscale. The control of the information processing and exchange between these engineered-cell devices, based on molecules and biochemical reactions, *i.e.*, molecular communication (MC), will be enabling technologies for the emerging paradigm of the Internet of Bio-Nano Things, with applications ranging from tissue engineering to bioremediation. In this paper, the design of genetic circuits to enable MC links between engineered cells is proposed by stemming from techniques for information coding and inspired by recent studies favoring the efficiency of analog computation over digital in biological cells. In particular, the design of a joint encoder-modulator for the transmission of binary-modulated molecule concentration is coupled with a decoder that computes the *a-posteriori* log-likelihood ratio of the information bits from the propagated concentration. These functionalities are implemented entirely in the biochemical domain through activation and repression of genes, and biochemical reactions, rather than classical electrical circuits. Biochemical simulations are used to evaluate the proposed design against a theoretical encoder/decoder implementation taking into account impairments introduced by diffusion noise.**

***Index Terms*— Molecular communication, synthetic biology, genetic circuit, parity-check encoding, analog decoding, Internet of Bio-Nano Things, Hill's function, mass action kinetics, soft bit, biochemical simulation.**

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], [5], [43]. In particular, the modeling of natural Molecular Communication (MC) processes in cells with communication engineering frameworks and tools is gaining growing interest in recent years. Significant examples range from the modeling of bacteria conjugation and electron transfer [32], to the characterization of signal transduction pathways [22], [47] and metabolic regulation [41]. The possibility of mimicking or even directly exploiting and reengineering these biochemical communication functionalities is opening the road for the development of systems that will further expand the Internet of Things concept to cover domains, the biological and nanoscale, where classical communication solutions show limitations [2], [14]. More precisely, 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 [10], [21], or even augmentation of the functionalities of the human body [7].

Despite the current advancements in the communication theoretic studies, which also stimulated an ongoing standardization effort [17], unified and coherent technologies to engineer and optimize systems, devices, and components capable of molecular communications are currently missing. Recent literature in MC systems is exploring the possibility of utilizing engineered cells to realize MC functionalities, where cells can be abstracted as transmitters and receivers [31]. 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 [19]. In particular, the theory of genetic 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 a forward engineering approach [33], [53]. The engineering of MC components and systems in cells through synthetic biology has gained particular interest in the last couple of years [37], mostly through the manipulation of natural MC processes such as bacterial quorum sensing. While complete experimental characterization and standardization of these components are still open challenges, it is today possible

Manuscript received January 22, 2018; revised June 1, 2018; accepted July 12, 2018. Date of publication July 24, 2018; date of current version December 14, 2018. This material is based upon work supported by the National Science Foundation under Grant No. CCF-1816969. The associate editor coordinating the review of this paper and approving it for publication was Y. Deng. *(Corresponding author: Massimiliano Pierobon.)*

A. Marcone was with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy. He is now with the Standardization Department, Qualcomm Inc., 90411 Nuremberg, Germany (e-mail: amarcone@qti.qualcomm.com).

M. Pierobon is with the Department of Computer Science and Engineering, University of Nebraska–Lincoln, Lincoln, NE 68588 USA (e-mail: pierobon@cse.unl.edu).

M. Magarini is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy, and also with the Consorzio Nazionale Interuniveritario per le Telecomunicazioni, 20133 Milan, Italy (e-mail: maurizio.magarini@polimi.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2859308

to theoretically study and predict in-silico the behavior of engineered genetic circuits of great complexity [33]. In this direction, in [38] the minimal subset of genetic 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 genetic circuits able to receive, process, and retransmit binary information by utilizing bacteria are included in [49], based on digital-like genetic circuit functionalities and M-ary molecule concentration modulation coupled with hard threshold detection.

In this paper, we propose the design, simulation, and characterization of genetic circuits for MC to realize the encoding of molecular information in a cell, its modulation through the emission of information-bearing molecules, and subsequent decoding of the transmitted information, motivated as a natural extension of the aforementioned studies of genetic circuits to realize MC components to improve the communication performance of engineered cells. In [24], 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 a simple block code scheme, by utilizing the components and rules of genetic circuit theory. From a practical application standpoint, a circuit to encode information, such as that proposed in this paper, would enhance the reliability of MC channels in scenarios where these are harnessed to control the propagation of information messages in biological environments, for applications ranging from controlled cell patterning for tissue engineering, to amorphous computing [9], [36], [37], [46], [52], [54]. Moreover, our design is general enough to find application into engineered cell-to-cell communications that have been demonstrated with cells from various kingdoms [45].

Despite a trend towards the development of genetic circuits with digital-like logics [51], recent research suggests that functions based on analog computation in cells are significantly more efficient [42]. Inspired by this recent research, and based on the seminal work in [15] on soft decoding of block and convolutional codes with non-linear electrical networks, we propose to exploit the analog computing functionalities of genetic circuits to obtain as output the a-posteriori log-likelihood ratio values (L-values), which provide the reliability of each decoded bit. Soft decoding, implementing the computation of the logarithm of the a-posteriori probabilities of the information bits, is realized by analog processing of a modulated molecule concentration *i.e.*, molecules utilized in nature for intercellular information exchange, such as in bacterial quorum sensing. It is worth observing that L-value computation based on analog filtering in MC literature has been previously considered in [8] and [13]. However, our work is the first to consider a detailed design of transmitter and receiver based on genetic circuits components, and is based on modulating information on molecule concentrations rather than in a deterministic molecule number, which is more realistic in a biological scenario. In addition, our computation extends the use of L-values to the calculus of the reliability of information bits for a block code, while only uncoded transmission is considered in [8] and [13]. In our work, to provide a first proof-of-concept demonstration, we make use of the simplest block code, namely, a Single Parity-Check (SPC) code and provide a specific design for block length $K = 3$ bits, where the computation of the L-value is implemented through genetic circuit components, designed and tuned according to the desired output. Biochemical simulation data of the resulting genetic circuit demonstrate very close performance to an electrical network implementation [15] in terms of Mean Squared Error (MSE) and Bit Error Rate (BER). While partial results of this work were presented in our previous conference publications [28], [29], this paper contains a comprehensive end-to-end analysis of this system, which includes both encoding and decoding genetic circuits [26], [27]. In this journal paper, we provide the complete design of the cell-to-cell communication system and obtain ideal constraints on the genetic circuit parameters to ensure the validity of a Gaussian approximation of the noise in the diffusion-based channel, as well as the steady-state approximation of gene expression. It is worth observing that, in comparison to the existing models for MC based on the use of biological circuits such as, for example, [12], [49], the proposed end-to-end system shares the same white Gaussian noise assumption, *i.e.*, independent noise affecting the received signal samples, while its main novelty is in the introduction of memory among bits transmitted in different intervals and the implementation of soft decision decoding at the receiver.

The rest of the paper is organized as follows. In Sec. II we introduce the model of the considered MC system, which is composed of the modulated parity-check encoder, the diffusive channel that transports the information and adds the noise, and the biological analog decoder 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 genetic circuits, their main components, and mathematical models. In Sec. IV we detail the encoder design by describing the necessary genetic circuit components and the tuning of their parameters. We detail the design of the required L-value, delay line, and box-plus operations in Sec. V, Sec. VI, and Sec. VII, respectively. In Sec. VIII we finalize the a-posteriori log-likelihood computation from the contributions of the previous elements. The implementation of the genetic circuits in a simulation environment along with the corresponding numerical results are presented in Sec. IX. Finally, in Sec. X we conclude the paper.

II. A Molecular Communication System Between Biological Cells With Parity-Check Coding

In this paper, we define a biological cell, *i.e.,* cell, as a finite environment that contains n chemical species $\{S_1, \ldots, S_N\}$ linked together by chemical reactions [33]. From this definition, it is possible to characterize a cell according to its state $\mathbf{s}(t) = \{s_n(t)\}_{n=1}^{N}$ as a function of the time t, where the single element $s_i(t)$ 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-

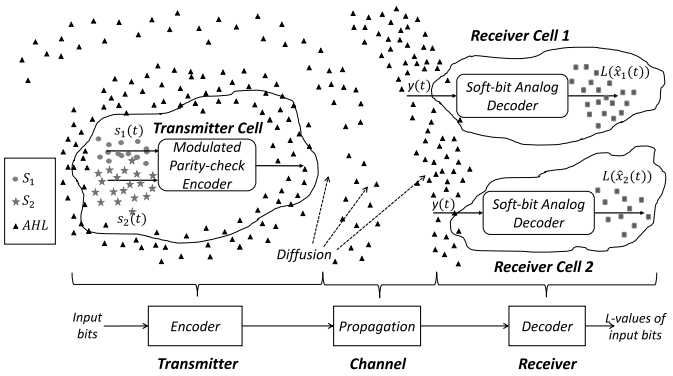


Fig. 1. Pictorial representation of the considered MC scenario.

made device present in the external environment, is a valid abstraction of an MC system.

With reference to Fig. 1, where different molecular species are indicated with different shapes, we consider for simplicity a cell, *i.e.*, **Transmitter**, that contains only two species, namely, S_1 and S_2, whose concentrations define the cell state at time t, $\mathbf{s}(t) = (s_1(t), s_2(t))$. The concentration $s_n(t)$ can be approximated as a continuous-time binary variable with value 1 if the nth 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 MC system between cells would be based on modulating the release of signaling molecules to the external environment, and would rely on their diffusion, *i.e.*, **Channel**, to propagate this information to one or more recipients at remote locations within the same fluid [52]. 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 [34]. A single signaling molecule species is considered for simplicity in this paper, but this can be easily generalizable to multiple non-interacting species. Finally, the recipients are other cells, *i.e.*, **Receiver**s, able to modulate their own state according to the information about the transmitted cell state carried by the modulated concentration $y(t)$ of signaling molecules at different time instants, present at their locations. The rest of the paper is based on the following assumption:

- The concentrations of the molecular species 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 received molecule concentration after diffusion, the longer the distance, and therefore the propagation time, the more homogeneous the impulse response is with respect to the space [23].

In this paper, we propose the design of a system to be deployed on board of cells where an *Encoder* module at the transmitter, and subsequently a *Decoder* module at the receiver/s, convert the aforementioned transmitter cell states, abstracted by information bits, into a-posteriori L-values, or soft bits at the receiver/s, after their *Propagation* through the diffusion channel. In the following, we provide the functional descriptions of these modules along with their main assumptions.

A. Modulated Parity-Check Encoder

The 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 [24]. More precisely, our modulation scheme is based on the following:

- The transmitter encodes, modulates and emits molecules according to the aforementioned concentrations $\mathbf{s}(t)$. Each molecule emission happens instantaneously at the beginning of a bit time interval T_b 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 [37], which is also in agreement with the evidence of basal expression rate that characterizes many DNA genes in genetic circuits [4], including those responsible for the release of the aforementioned AHL molecules [34], we consider positive values for Q_1 and Q_0.

The Brownian motion underlying the diffusion-based propagation of these molecules could lead to noise at reception that is theoretically modeled by a Poisson distribution [35], which is prone to consequent errors in understanding the transmitter cell state. To alleviate this issue, we propose to design a *Parity-Check Encoder* that takes as input the concentrations $\mathbf{s}(t)$ and gives as output the encoded bits x_i, which are then *modulated* into molecule emission symbols $x_i(t)$ according to the aforementioned two-level scheme. 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.*, Low-Density Parity-Check (LDPC) or Hamming codes [24], that allow the recipient not only to detect but also correct errors.

B. Diffusion

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 [23]. We make the following assumption:

- The extracellular space is an unbounded fluid medium where molecules freely diffuse. The distance between the transmitter and the receiver is much longer than the diameter of the cells, therefore resulting in an approximation of the cells as points at their respective locations. As a consequence, the presence of the cells does not interfere with the diffusion of AHL molecules.

- The received modulated molecule concentrations y_i are the samples of the molecule concentration $y(t)$ at the location of the engineered cell, 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 [40], the received molecule concentration y_i has an average that is expressed as follows:

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

which can be equal to one of 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 ith transmitted bit to the received concentration, but also the average contribution of all the possible combinations of previous bits, *i.e.*, the Inter-Symbol 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 [23], the time when the Green's function of the diffusion equation [25] is at its maximum, expressed as follows:

$$\bar{t}_i = t_d + iT_b = \frac{r_{rx}^2}{6D} + iT_b. \tag{2}$$

The ith sample y_i of the concentration at the receiving cell can be mathematically written as

$$y_i = a_{0/1} + n_i, \tag{3}$$

where n_i is an additive noise with variance σ^2. In what follows, we assume that the stochastic process n_i is described as an Additive White Gaussian Noise (AWGN). This assumption is justified by the fact that the noise contribution from the genetic 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 [40]), resulting in white Gaussian contributions to each circuit output [48]. 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 [20]. From a quantitative analysis point of view, a check of the AWGN assumption can be found in [27] where, starting from the analysis of diffusion processes and biochemical reactions, the Poisson model and the Gaussian approximation are compared in the evaluation of BER when optimal soft-decision decoding is implemented. By means of computer simulations, it is shown in [27] that similar BER performance are obtained for both the two cases of noise by the proposed biologic decoding circuits, thus validating the commonly adopted white Gaussian assumption for the noise affecting the received signal in diffusion-based MC systems. Moreover, it is worth observing that the AWGN assumption also holds in other nano-scale MC systems, as demonstrated in [11] for a flow-induced MC system where molecular transmitter and receiver are placed in chambers and communicate over a microfluidic channel containing fluid flow.

C. Soft-Bit Analog Decoder

One of the main contributions of this paper is the design of a genetic circuit that implements a soft-bit analog decoder, able to compute the a-posteriori log-likelihood ratio $L(\hat{x}_k)$ from a received noisy input signal $y_i, y_{i+1}, \ldots, y_{i+K}$ modulated according to block-encoded bits $x_i, x_{i+1}, \ldots, x_{i+K}$, 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]:

$$L(\hat{x}_k) = \log \frac{P(x_k = 1 | y_i, y_{i+1}, \ldots, y_{i+K})}{P(x_k = 0 | y_i, y_{i+1}, \ldots, y_{i+K})}, \tag{4}$$

In Fig. 1 it is shown that the kth receiver cell, which contains the soft-bit analog decoder detailed in the following, provides a concentration of output molecules equal to the a-posteriori log-likelihood ratio $L(\hat{x}_k)$ for the kth bit of the transmitted codeword that gives rise to the modulated concentrations $y_i, y_{i+1}, \ldots, y_{i+K}$ of input molecules around the receiving cell at specific time instants. The proof-of-concept design of the soft-bit analog decoder presented in this paper is based on the following assumptions:

- 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 [38]. The latter process would be activated by ligand-receptor binding reception either at the membrane or inside the cell [38], and result in a concentration of input molecules of the biological decoder that is proportional to the concentration of signaling molecules around the cell.
- 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 SPC code with block length $K = 3$. Moreover, given the complexity of the resulting genetic 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 LDPC.

As a consequence of the aforementioned assumptions, the formula in (4) to compute the L-value of the first transmitted bit $L(\hat{x}_1)$ becomes [15]:

$$L(\hat{x}_1) = L(y_1|x_1) + \left(L(y_2|x_2) \boxplus L(y_3|x_3)\right), \tag{5}$$

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

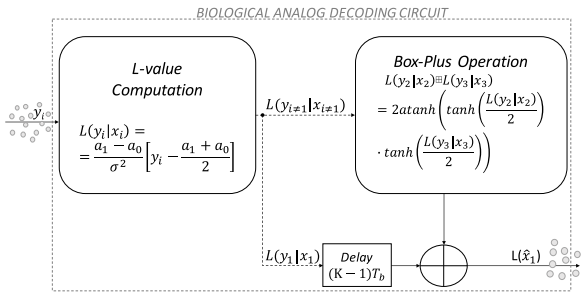


Fig. 2. Functional diagram of the biological decoding circuit.

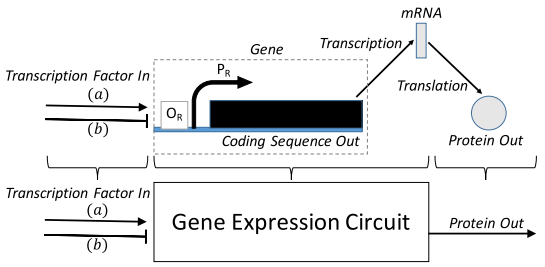


Fig. 3. Gene expression scheme: (a) Activation (↓), (b) Repression (⊥).

box-plus operation, which is defined as follows:

$$L(y_2|x_2) \boxplus L(y_3|x_3) = 2\text{atanh}\left(\tanh\left(\frac{L(y_2|x_2)}{2}\right)\cdot\tanh\left(\frac{L(y_3|x_3)}{2}\right)\right). \quad (6)$$

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

$$L(y_i|x_i) = \log\frac{\exp\left(-\frac{(y_i-a_1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(y_i-a_0)^2}{2\sigma^2}\right)} = \frac{(a_1-a_0)}{\sigma^2}\left[y_i - \frac{(a_1+a_0)}{2}\right], \quad (7)$$

where σ^2 is the variance of the aforementioned AWGN given in (3).

Hagenauer *et al.* [15] implemented (5) with analog Very-Large-Scale Integration (VLSI) circuits, by exploiting the non-linearities of a modified Gilbert cell. In the following, we present an implementation of the same expression by using genetic 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 (7), the box-plus operation, which realizes the expression in (6), and a delay line (active only when $t = lKT_b$, where T_b is the bit interval, K is the block size, equal to 3, $l = 0, 1, 2, \ldots$), which is needed to isolate the result of the first term of the sum in (5) from the second term, computed at a later time. In the following, after a brief overview of the main genetic circuit components, we detail our genetic circuit design.

III. Components of a Genetic Circuit

A *genetic circuit* is a network of *chemical reactions* involving *genes* and other molecular species that work together to implement a specific biological function [33].

A. Gene Expression

As shown in Fig. 3, 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 [33]:

- **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).
- **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.
- **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 [34].
- **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], [6]. This widely accepted approximation in biological circuit modeling is based on the assumption that these processes are in a steady state (equilibrium), and that this steady state is reached in a shorter time interval with respect to the time scale of the other processes underlying the biological circuit. For an *E. coli* bacterium, this time interval is in the order of seconds [4], which is compatible with our bit

time interval of $T_b = 5\,s$ (the reference time for sampling all the concentrations in the system) in our numerical examples. In addition, our numerical results found a good agreement between the performance of our biological circuit design based on the Hill function, and simulated according to chemical kinetics on Matlab Simbiology, and an ideal computation of the L-value using analytical formulas, as detailed in Sec. IX. According to the Hill function, the rate $\frac{d[Out]}{dt}$ of output protein, in case of activation, is expressed as

$$\frac{d[Out]}{dt} = k' + MAX\left(\frac{([In]/K)^n}{1+([In]/K)^n}\right) - k_{deg}[Out], \quad (8)$$

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 [6]. In (8) and hereafter, the square brackets notation stands for molecule concentration. According to [6], in case of repression, the rate $\frac{d[Out]}{dt}$ of output protein is

$$\frac{d[Out]}{dt} = MAX\left(\frac{1}{1+([In]/K)^n}\right) - k_{deg}[Out]. \quad (9)$$

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** [4] as follows

$$\frac{d[Out]}{dt} = \frac{\sum_i MAX_i([In_i]/K_i)^{n_i}}{1+\sum_i([In_i]/K_i)^{m_i}} - k_{deg}[Out], \quad (10)$$

where In_i refers to the ith transcription factor, and $n_i = m_i$ if the ith 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 (8), (9), and (10) 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 (10) (general case) to zero, the steady state output concentration $[Out]^{SS}$ is found with the following expression:

$$[Out]^{SS} = \frac{1}{k_{deg}} \frac{\sum_i MAX_i([In_i]/K_i)^{n_i}}{1+\sum_i([In_i]/K_i)^{m_i}}, \quad (11)$$

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.

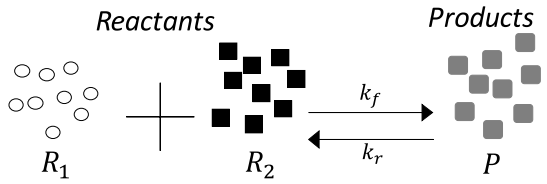


Fig. 4. Mass action chemical reaction between two generic molecular species $R_1 + R_2 \underset{k_r}{\overset{k_f}{\rightleftharpoons}} P$.

Within the scope of this paper, we assume unbalanced reactions where the forward reaction rate is much greater than the reverse rate, as in [35]. An example of a reaction with two reactant species and one product species is shown in Fig. 4. 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.
- **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. 4 is expressed through a reaction rate equation [33] as

$$\frac{d[P]}{dt} = k_f[R_1][R_2] - k_r[P]. \quad (12)$$

- **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. Modulated Encoder Design Based on Genetic Circuits

The proposed modulated encoder design encodes two bits representing the state of the cell at time t, *i.e.* $\mathbf{s}(t) = (s_1(t), s_2(t))$, with an SPC code characterized by codeword length $K = 3$ bits, and modulates the transmission of the codeword bits x_i with two levels, namely, a_0 and a_1, for the bit 0 and 1.

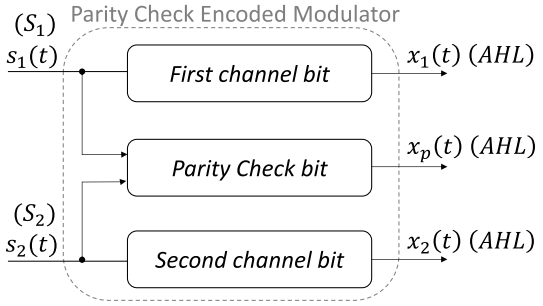


Fig. 5. Functional diagram of the modulated parity check encoder.

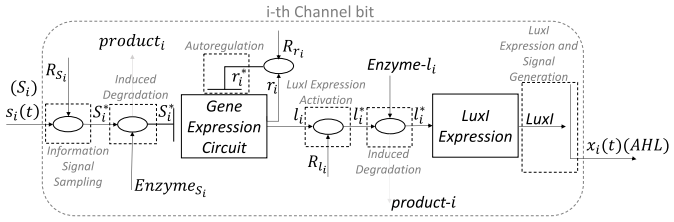


Fig. 6. Functional diagram of the first and second channel bit production with $i = 1$ and $i = 2$, respectively.

The proposed design is realized through three main branching genetic circuits as shown in Fig. 5. 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 [44]. In order to gain insight into these processes, in the following we consider the ith branch, $i = 1, 2$, that produces $x_i(t)$ as reported in Fig. 6. There are two main types of components in our circuit:

- Gene Expression (**rectangular blocks** in Fig. 6), 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. 6), 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 $s_1(t)$ and $s_2(t)$ 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 $t = \overline{t}$, we must consider $\mathbf{s}(t = \overline{t})$ which means we have to freeze the state of the cell at that given instant. As in digital communication systems, we achieve that by sampling. At $t = \overline{t}$ we just inject a sufficiently high number of corepressors R_{x_i} that, through the Transcription Factor Activation Reaction, activate the molecular concentration of input (repressive) transcription factors $s_i(t = \overline{t})$. As a result we drop the time dependency notation in Fig. 6 after the sampling operation, with s_i^* being the sample of $s_i(t)$ at time $\overline{t}$. The value s_i^* is the concentration of the activated species S_i^* obtained from the Transcription Factor Activation Reaction

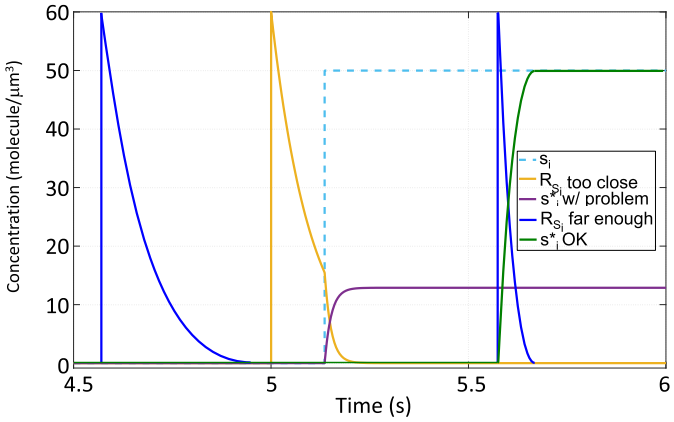


Fig. 7. Sampling time $\overline{t}$ close to a raising edge of the variable $s_i(t)$. In the simulation, the variable $s_i(t)$ has been assumed to be a square wave with high level 50 molecule/μm^3 and the injection of the corepressors R_{S_i} has been considered as an instantaneous event. The case where $t = \overline{t} = 5\,s$ shows the problem of being too close to the raising edge of $s_i(t)$.

$S_i + R_{S_i} \rightarrow S_i^*$, where we have the particular case in which $R_1 \equiv S_i$, $R_2 \equiv R_{S_i}$, 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).

Unlike digital communications with electrical circuits, in the context of genetic 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" [16]. For this reason, and also for the stochasticity of the state of the cell, an issue can arise if the sampling instant $\overline{t}$ coincides (or is very close to) a raising or falling edge of the variable $s_i(t)$, as illustrated in Fig. 7. In the first interval (for $t \geq 5\,s$ but before s_i goes up), R_{S_i} molecules degrade just because of natural degradation since there are no S_i molecules around ($x_i(t = \overline{t}) = 0$). During the sampling process, s_i changes its state causing the residual corepressors R_{S_i} to bind (notice the increased consumption speed since the corepressors are both sequestered by the reaction with s_i and naturally degraded) and produce some complexes S_i^* with concentration s_i^*. In the end, s_i^* will be different from what it was supposed to be, as shown by the purple curve in Fig. 7. When sampling, in fact, we aim to have $s_i^* = s_i(t = \overline{t})$, achieved when $\overline{t}$ 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 s_i^* at the input of the Gene Expression Circuit as a consequence of the random parameters ϕ_1 and ϕ_2, as explained in Sec. IX.

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. 3b. This gene contains two different

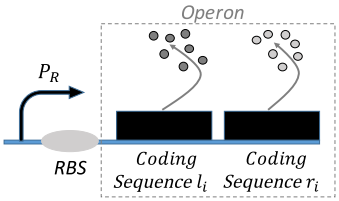


Fig. 8. Scheme of the operon at the Gene Expression Circuit block.

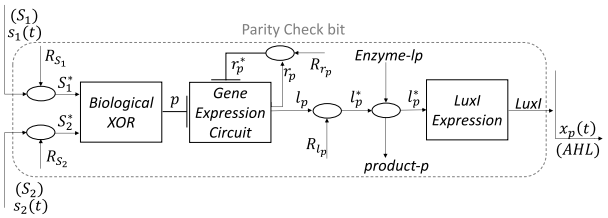


Fig. 9. Functional diagram of the parity check bit production.

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. 8 for the Gene Expression Circuit block.

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 s_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 s_i^* is high, the Gene Expression Circuit activity is not completely repressed but is downscaled with respect to the maximum value, obtained when s_i^* is low. Therefore $[l_i]^{SS}$ reports the information on the value of s_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 s_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 (s_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 (s_i^* low), and a_1 (transmitted signal for the bit 1) when $[l_i^*]^{SS}$ is at its minimum (s_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 [18], 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:

$$x_i(t) = \frac{dA}{dt} = k_0 I(t) = k_0 I^{SS} = k_0 Q_{0/1}, \tag{13}$$

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.6) 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, ISI (at least in transmission) is mitigated. Additionally, if l_i^* complexes are not degraded before transmission of the ith 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.

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

The realization of the parity check modulated emission $x_p(t)$ has some differences worth being analyzed separately. Details on the Parity Check bit are reported in Fig. 9. Here, the samples s_1^* and s_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 x_p. The Biological XOR block realizes the XOR summation between the information bits s_1^* and s_2^* and it is here designed and modeled as suggested by Myers in [33]. 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 to be tuned in order to transmit the parity check symbol only after the modulated symbols $x_1(t)$ and $x_2(t)$.

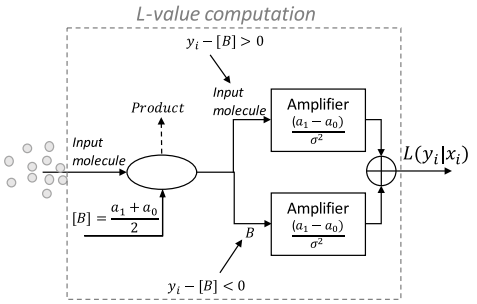


Fig. 10. Functional diagram of the L-value computation.

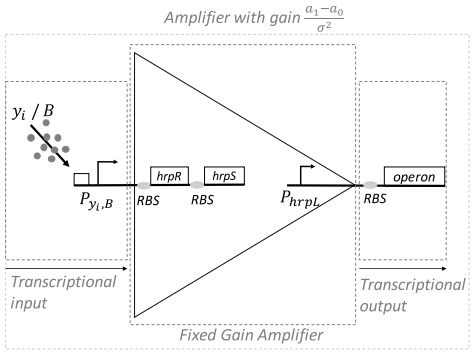


Fig. 11. Scheme of the fixed gain biological amplifier.

V. Genetic Circuit for L-Value Computation

Fig. 10 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.

A. Mass-Action Reaction

The first element in Fig. 10 represents a mass-action reaction [33] 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 (7). Once the reaction has occurred, the remaining concentration of any of the two molecule species is equal to the value $\left|y_i - \frac{(a_1+a_0)}{2}\right|$, 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 genetic circuit, two branches have been introduced in the block diagram, as shown in Fig. 10. The upper branch is for the *positive* concentration and the lower branch is for the *negative* concentration.

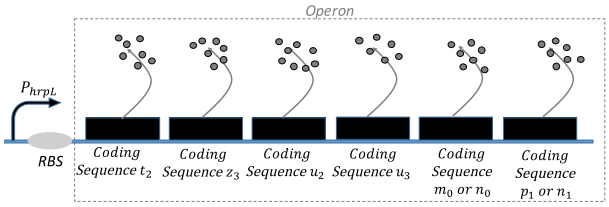


Fig. 12. Scheme of the operon at the output of each amplifier shown in Fig. 11.

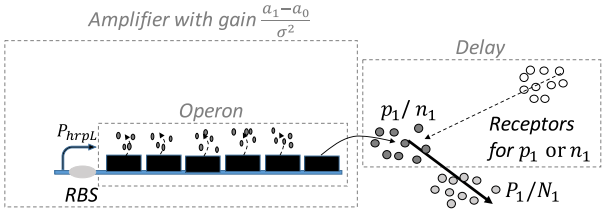


Fig. 13. Storage Operation for molecules P_1 and N_1.

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

From (7), 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 [50], 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. 11. As shown in the figure, the genetic amplifier is designed by using the orthogonal genetic components (hrpR, hrpS and PhrpL), from the hrp (hypersensitive response and pathogenicity) gene regulatory module from plant pathogen.

At the output of each amplifier we have an operon in place of a single gene, which is needed for the subsequent processing. These operons, shown in Fig. 10, encode for the proteins $\{t_2, z_3, u_2, u_3\}$, while they differ for the last two, namely, 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 ith 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.

VI. Genetic Circuit for the 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 (5), 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 genetic circuits, delay lines can be realized by means of receptors as illustrated in Fig. 13, 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 $2T_b$,

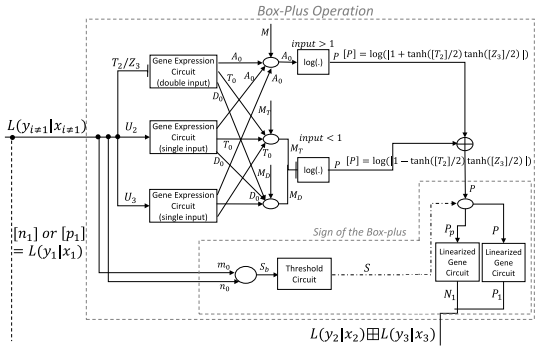


Fig. 14. Functional diagram of the Box-plus operation.

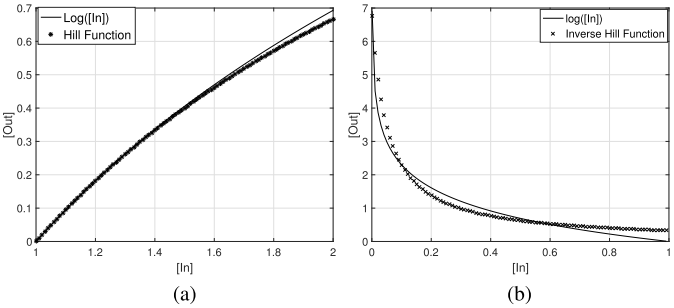


Fig. 15. Comparisons showing good agreements between (a) the logarithmic and Hill functions with $n = 1$, $K = 2$ and $MAX = 2$, and (b) a logarithmic function and a complementary Hill Function with $n = 1$, $K_p = 0.05$ and $MAX = |\log(10^{-3})|$, respectively.

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.

VII. Genetic Circuit for the Box-Plus Operation

A description of the genetic circuit blocks and their operations inside the Box Plus Operation element, whose genetic circuit is reported in Fig. 14, is given in the following.

A. Gene Expression Circuit

To implement the calculation of the expression in (6), it is necessary to compute the hyperbolic tangent and its inverse, which are basic functions for the box-plus operation. For this, by stemming from the multi-dimensional Hill function model defined in (10), having (8) as one-dimensional special case, our design consists of an upper gene regulated by repression and two lower genes regulated by activation, as shown in Fig. 14. In particular, for the one-dimensional case we optimize the parameters n and K in (8) 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 $K = 1$ in (8), expressed as follows:

$$\frac{d[Out]}{dt} = \frac{[In]^2}{1+[In]^2} \cong \tanh\left(\frac{[In]}{2}\right). \tag{14}$$

Since the expression in (8) 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 genetic circuit, we obtain the following:

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

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

B. The Logarithmic Function

The expression in (6) includes also the inverse hyperbolic tangent function, which we realize by stemming from the following trigonometric identity:

$$2\,\mathrm{atanh}(d) = \log\left(\frac{1+d}{1-d}\right) = \log(|1+d|) - \log(|1-d|). \tag{15}$$

By applying (15), we rewrite the inverse hyperbolic tangent in terms of logarithms. This operation is made by the two $\log(\cdot)$ blocks in Fig. 14, where the parameter d in (15) 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 (6).

While the implementation of the arguments $1+d$ and $1-d$ is detailed in Sec. VII-C, here we focus on the realization of the logarithmic function. As shown in Fig. 14, since molecule concentrations cannot assume negative values, two different genetic circuit blocks are used to approximate the logarithmic function for either input > 1 or input < 1, detailed next.

1) Logarithm Approximation for Input Greater Than 1: From (15), 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. 14, 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 (8), 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 genetic circuit block with the positive logarithmic function, we identify the following parameter values: $n = 1$, $K = 2$ 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. 15(a), where we observe that a very good approximation of the logarithmic function is achieved for input values $[In] \in [1, 2]$.

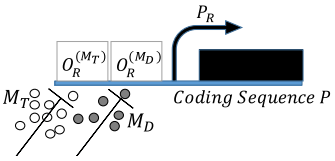


Fig. 16. Scheme of the gene that realizes the logarithmic function.

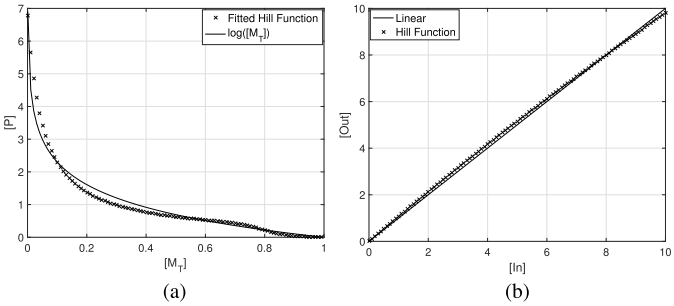
2) Logarithm Approximation for Input Smaller Than 1: The argument $(1-d)$ of the second log in (15) is always smaller than 1. Hence, the second $\log(\cdot)$ block in Fig. 14 has to approximate the absolute value of the negative part of the logarithm. We realize this through the complementary Hill function 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. 15(b). 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. 14) with two, in place of just one, different repressive transcription factors, M_T and M_D. This approach results in a bivariate Hill function [4] that describes the rate of output protein production as a function of $[M_T]$ and $[M_D]$ as follows:

$$\frac{d[P]}{dt} = H([M_T],[M_D]) = \frac{\beta_1+\beta_2}{1+\left(\frac{[M_T]}{k_1}\right)^{n_1}+\left(\frac{[M_D]}{k_2}\right)^{n_2}}, \quad (16)$$

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. 16. Although the bivariate Hill function in (16) is valid for any combination of the inputs $[M_T]$ and $[M_D]$, as explained in Sec. VII-C, in our genetic 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. 17(a), where it is compared to the desired curve of the logarithmic function. The MSE optimal parameters to achieve this behavior are $n_1 = n = 1$, $\beta_1 + \beta_2 = MAX = |\log(10^{-3})|$, $k_1 = K = 0.05$, $n_2 = 20$ and $k_2 = 0.7$.

C. Details on the Box-Plus Operation

The design in Fig. 14 realized with the components described above, and leading to the box-plus operation described in (6), 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 (16) 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



(a) (b)

Fig. 17. Comparison showing good agreement between (a) a logarithmic function and a complementary bi-variate Hill Function with $n_1 = n = 1$, $\beta_1 + \beta_2 = MAX = |\log(10^{-3})|$, $k_1 = K = 0.05$, $n_2 = 20$ and $k_2 = 0.7$, and (b) a linear function and a Hill function with $MAX = 100$, $K = 92$ and $n = 1$ for $[In] \in [0, 10]$, respectively.

$$\frac{d[A_0]}{dt} = H([T_2],[Z_3]) = \frac{1}{1+\left(\frac{[T_2]}{k_g}\right)^{n_g}+\left(\frac{[Z_3]}{k_g}\right)^{n_g}}. \quad (17)$$

By analyzing the expression in (17), we observe that it resembles the following product:

$$\begin{aligned}\tilde{H}([T_2],[Z_3]) &= \frac{1}{1+\left(\frac{[T_2]}{k_g}\right)^{n_g}} \cdot \frac{1}{1+\left(\frac{[Z_3]}{k_g}\right)^{n_g}} \\ &= \frac{1}{1+\left(\frac{[T_2]}{k_g}\right)^{n_g}+\left(\frac{[Z_3]}{k_g}\right)^{n_g}+\left(\frac{[T_2][Z_3]}{k_g^2}\right)^{n_g}},\end{aligned} \quad (18)$$

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 VII-A can be extended to the complementary Hill function. Again, using the same notation as in (14), for $K = 1$ and $n = 2$, we obtain

$$\frac{d[Out]}{dt} = \frac{1}{1+[In]^2} \cong 1 - \tanh\left(\frac{[In]}{2}\right). \quad (19)$$

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. 14,

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 (14).

From Section V-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 $\frac{[U_2]^n}{K^n+[U_2]^n} = \frac{[T_2]^n}{K^n+[T_2]^n} \cong \tanh\left(\frac{[T_2]}{2}\right)$ and $\frac{[U_3]^n}{K^n+[U_3]^n} = \frac{[Z_3]^n}{K^n+[Z_3]^n} \cong \tanh\left(\frac{[Z_3]}{2}\right)$ for $K_p = 1$ and $n = 2$. Since the output species are the same for all genes, their concentrations sum up leading to

$$\begin{aligned}[A_0] = [T_0] = [D_0] &= 1 - \tanh\left(\frac{[T_2]}{2}\right) - \tanh\left(\frac{[Z_3]}{2}\right) \\ &+ \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right) \\ &+ \tanh\left(\frac{[T_2]}{2}\right) + \tanh\left(\frac{[Z_3]}{2}\right) \\ &= 1 + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right), \end{aligned} \tag{20}$$

which is exactly the argument of the first log in (15). 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 (15), 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 (16) and shown in Fig. 17(a). 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 (15) 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

$$\begin{aligned}[P] &= \log\left(1 + \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &\quad - \log\left(1 - \tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &= 2\,\mathrm{atanh}\left(\tanh\left(\frac{[T_2]}{2}\right)\tanh\left(\frac{[Z_3]}{2}\right)\right) \\ &= 2\,\mathrm{atanh}\left(\left|\tanh\left(\frac{L(y_2|x_2)}{2}\right)\tanh\left(\frac{L(y_3|x_3)}{2}\right)\right|\right), \end{aligned} \tag{21}$$

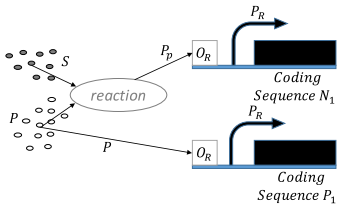


Fig. 18. Scheme of the chemical reactions and gene that compute the sign of the box-plus expression.

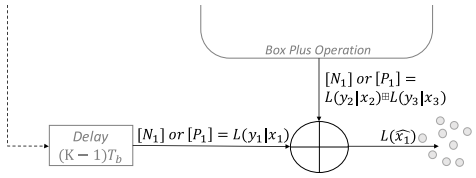


Fig. 19. Arithmetic sum to compute the a-posteriori L-value.

which is exactly the magnitude, or absolute value $|\cdot|$, of the desired box-plus operation expressed in (6).

D. Sign of the Box-Plus

The Sign of the Box-plus block in Fig. 14 has the specific purpose of computing the sign of the box-plus operation expressed in (6). 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. 14, 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. 18, are obtained through promoters with a linear response in a given interval $[0, \delta]$. For this, we need to set $MAX \gg \delta$, $K \gg \delta$ and $n = 1$, where MAX, K and n are defined in Sec. III. For example, if we set $\delta = 10$, $MAX = 100$ and $K = 92$, we obtain a good approximation, see Fig. 17(b).

VIII. Genetic Circuit for the 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 (6), as shown in Fig. 19 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. VI, with

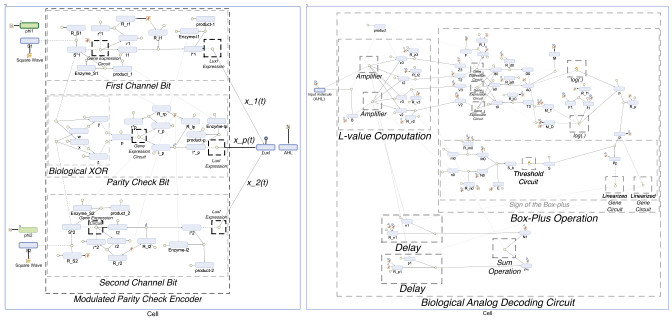


Fig. 20. Simbiology model of the molecular communication system between biological cells with parity-check coding presented in this paper.

the result of the box-plus operation after the Sign Inversion Block, as described in Sec. VII-D, in terms of N_1 or P_1, whose concentration is equal to the expression in (6). 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.

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.

IX. SimBiology Implementation and Numerical Results

The scheme in Fig. 5 has been replicated in Simbiology, a Matlab package for simulation of biological networks, with the result reported in Fig. 20. 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 [3]. In the case of an *E. coli* bacterium, a usual chassis in synthetic biology, this corresponds to a concentration of 1 molecule/μm^3 in a volume $V \cong 10\mu$m^3 [34]. From Fig. 20 it is possible to distinguish three main blocks [30]:

- 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. 20 have concentrations x_1 and x_2 defined by continuous time Square Waves with high level $x_i = 50$, $i = 1, 2$, in order to simulate the binary nature of the concentrations $s_1(t)$ and $s_2(t)$ that continuously define the state of the cell. The Square Wave is characterized by a random initial phase (ϕ_1 and ϕ_2 for s_1 and s_2, respectively) so as to model the random state of the cell when the coding process begins.

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 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. 6 and 9 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 (10). 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 (8).

The parameters of the Gene Expression Circuit have been chosen considering $s_1(t) = 50$ when the species S_1 is present. As in Fig. 22(a), even if $s_1(t) = 50$ at the sampling time ($s_1^* = 50$), 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 $s_1^* = 0$) and $k_0 I^{SS} = Q_1 \cong 20$ when $[l_1^*]^{SS} \cong 2$ (obtained when $s_1^* = 50$), as in Fig. 22(b). While a thorough discussion is left to future work, these values allow for an optimal Gaussian approximation of the molecule counting noise in a diffusion-based channel [39].

Finally, as explained in Sec. IV-B.2, 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 (s_1^* high). The result for $k_{1_G} = 1$ is observable in Fig. 22(a). In order to have a quite strong repression, whatever the value of s_1^*, we need $[r_1^*]^{SS} \geq 2$ and this is guaranteed since, even

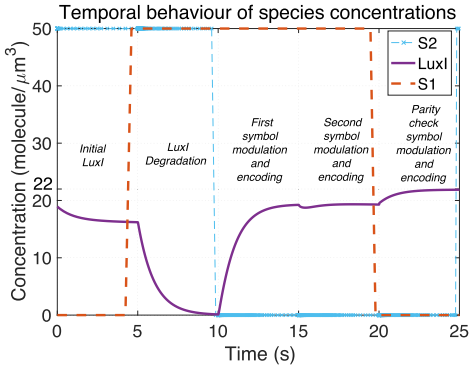


Fig. 21. Simulation of the value of *LuxI* concentration at the transmitter given the input concentrations of the molecular species S_1 S_2 over time.

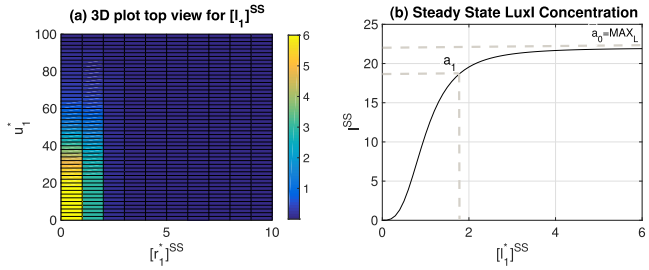


Fig. 22. Steady state concentration (a) $[l_1]^{SS}$ for the Gene Expression Circuit and (b) I^{SS} for the LuxI Expression.

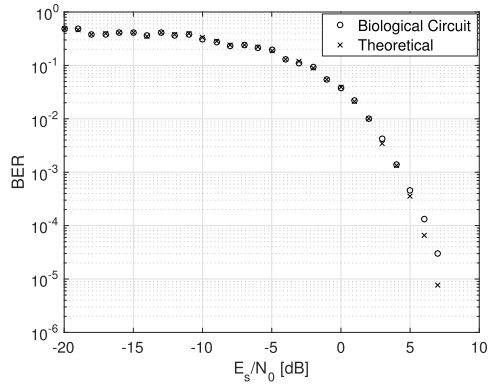


Fig. 23. BER results for a signal-to-noise ratio $\frac{E_s}{N_0} \epsilon [-20, 7]\, dB$, which show a good agreement between the performance of the simulated biological circuit and an exact L-value computation using analytical formulas.

in the case of maximum gene expression input $s_1^* = 50$, the output of the operon (Fig. 8) would be $[r_1^*]^{SS} \equiv [l_1^*]^{SS} \cong 2$.

Simulation curves as function of time are reported in Fig. 21. For a better visual result, only the main species S_1, S_2 and $LuxI$ with concentrations $s_1(t)$, $s_2(t)$ and $I(t)$, respectively, are shown. Notice that, from Eq. (13), $I(t)$ and $x_i(t)$ are directly related, hence $I(t)$ behavior over time gives us information about the transmitted symbols.

The sampling of $s_1(t)$ and $s_2(t)$ occurs at $t = \overline{t} = 5\, s$. In that instant, the state of the cell $(s_1(t = \overline{t}), s_2(t = \overline{t})) = (50, 50)$ so the parity check bit should be 0. Looking at the variable $LuxI$ in Fig. 21, we realize the information has been encoded and modulated correctly. The encoding process begins

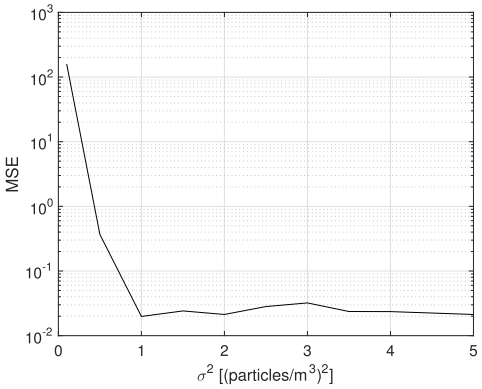


Fig. 24. MSE of the performance of the simulated biological circuit with respect to an exact L-value computation using analytical formulas for Gaussian noise with variance $\sigma^2 = \{0.1, 0.5, 1, 1.5, 2, \ldots, 5\}$ $[(particles/m^3)^2]$.

at $t = 10\, s$, using the $5\, s$ time gap to degrade any possible residual LuxI molecules from the previous codeword. The bit time interval is set to $T_b = 5\, s$ in order to get the steady state expression of the LuxI Expression block, *i.e.* $Q_0 = 22$ when $s_i^* = 0$ and $Q_1 \cong 20$ when $s_i^* = 50$ $(i = 1, 2)$. Observing the LuxI concentration in $t = 15$ s, $t = 20$ s and $t = 25$ s, we get the modulated channel symbols (Q_1, Q_1, Q_0) corresponding to the encoded bits (110), consistently with what we desired to transmit.

To test the performance of the proposed genetic 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 genetic 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-C 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.

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. 23, 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 genetic 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 genetic circuit for $\sigma^2 \epsilon [10^{-1}, 5]$, as reported in Fig. 24.

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 (21), this results into very large output values of the logarithmic function which are not well approximated by our genetic 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].

Finally, it is important to notice that the ideal time bit interval $T_b = 5\ s$, chosen to ensure the steady state expression of the LuxI Expression block according to the parameters in Table I, results in an overall bit-rate of the communication system of 0.2 bps (or 2/3*0.2 = 0.13 bps, if we exclude parity bits). While outside of the scope of this paper, the low-level design of our biological circuits would proceed by choosing the biological circuit components that are in agreement with the constraints given in our theoretical analysis. We anticipate that this will mainly depend on the time for the Hill function of each gene to reach the needed steady state of protein concentration starting from the state at the previous time interval (bit time interval or sampling interval depending on its location in the circuit), as expressed in (11). This can range from seconds to minutes or even hours, depending on the particular gene, and the slowest gene to reach the needed steady state could constitute the rate limiting step of the overall communication system, and have most of the influence in the choice of the optimal T_b.

X. Conclusion

In this paper, inspired by recent studies favoring the efficiency of analog computation over digital one in biological cells, we proposed the design and simulation of a biological modulated parity-check encoder and of the associated analog decoder for molecular communication in a biological cell. The design is based on the genetic engineering of genetic circuits, realized entirely in the biochemical domain by using activation and repression of gene expression, and reactions of molecular species. Genetic circuits have 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. Encoder and decoder presented in this paper are intended as a proof-of-concept design methodology for utilizing genetic 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. Biochemical simulation data of the resulting genetic 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.

References

[1] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Comput. Netw.*, vol. 52, no. 12, pp. 2260–2279, Aug. 2008.

[2] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The Internet of bio-nano things," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 32–40, Mar. 2015.

[3] B. Alberts, A. Johnson, and L. J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. New York, NY, USA: Garland, 2002.

[4] U. Alon, *An Introduction To Systems Biology: Design Principles of Biological Circuits*. London, U.K.: Chapman & Hall, 2006.

[5] 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.

[6] J. Ang, E. Harris, B. J. Hussey, R. Kil, and D. R. McMillen, "Tuning response curves for synthetic biology," *ACS Synth. Biol.*, vol. 2, no. 10, pp. 547–567, Aug. 2013.

[7] B. Atakan, O. B. Akan, and S. Balasubramaniam, "Body area nanonetworks with molecular communications in nanomedicine," *IEEE Commun. Mag.*, vol. 50, no. 1, pp. 28–34, Jan. 2012.

[8] H. Awan and C. T. Chou, "Generalized solution for the demodulation of reaction shift keying signals in molecular communication networks," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 715–727, Feb. 2017.

[9] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, Apr. 2005.

[10] 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.

[11] A. O. Bicen and I. F. Akyildiz, "End-to-end propagation noise and memory analysis for molecular communication over microfluidic channels," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2432–2443, Jul. 2014.

[12] A. O. Bicen, C. M. Austin, I. F. Akyildiz, and C. R. Forest, "Efficient sampling of bacterial signal transduction for detection of pulse-amplitude modulated molecular signals," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 4, pp. 505–517, Aug. 2015.

[13] C. T. Chou, "Maximum *a-posteriori* decoding for diffusion-based molecular communication using analog filters," *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 1054–1067, Nov. 2015.

[14] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 1887–1919, 3rd Quart., 2016.

[15] J. Hagenauer, E. Offer, C. Mèasson, and M. Mörz, "Decoding and equalization with analog non-linear networks," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 659–680, 1999.

[16] S. Haykin and M. Moher, *Introduction to Analog & Digital Communications*. Hoboken, NJ, USA: Wiley, 2007.

[17] *IEEE Draft Recommended Practice for Nanoscale and Molecular Communication Framework*, IEEE Standard IEEE P1906.1/D2.0, 2015.

[18] B. Ingalls, *Mathematical Modelling in Systems Biology: An Introduction*. Cambridge, MA, USA: MIT Press, 2012.

[19] L. J. Kahl and D. Endy, "A survey of enabling technologies in synthetic biology," *J. Biol. Eng.*, vol. 7, no. 1, p. 13, May 2013.

[20] D. Kilinc and O. B. Akan, "Receiver design for molecular communication," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 705–714, Dec. 2013.

[21] J. R. Kirby, "Synthetic biology: Designer bacteria degrades toxin," *Nature Chem. Biol.*, vol. 6, pp. 398–399, Jun. 2010.

[22] M. Laddomada and M. Pierobon, "A crosstalk-based linear filter in biochemical signal transduction pathways for the Internet of bio-things," in *Proc. IEEE ICASSP*, Apr. 2015, pp. 5520–5524.

[23] I. Llatser, A. Cabellos-Aparicio, M. Pierobon, and E. Alarcón, "Detection techniques for diffusion-based molecular communication," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 726–734, Dec. 2013.

[24] Y. Lu, M. D. Higgins, and M. S. Leeson, "Comparison of channel coding schemes for molecular communications systems," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3991–4001, Nov. 2015.

[25] A. Mandelis, *Diffusion-Wave Fields: Mathematical Methods and Green Functions*. New York, NY, USA: Springer-Verlag, 2001.

[26] A. Marcone, "Encoding and soft decoding in molecular communication based on biological circuits," M.S. thesis, Dept. Elettron., Inf. Bioing., Politecnico Milano, Milan, Italy, 2016.

[27] A. Marcone, M. Pierobon, and M. Magarini, "The Gaussian approximation in soft detection for molecular communication via biological circuits," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–6.
[28] A. Marcone, M. Pierobon, and M. Magarini, "A biological circuit design for modulated parity-check encoding in molecular communication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[29] A. Marcone, M. Pierobon, and M. Magarini, "A parity check analog decoder for molecular communication based on biological circuits," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.
[30] Mathworks. (2016). *SimBiology Release 2016b Documentation*. [Online]. Available: https://www.mathworks.com/help/simbio/index.html
[31] D. B. Menendez, V. R. Senthivel, and M. Isalan, "Sender–receiver systems and applying information theory for quantitative synthetic biology," *Current Opinion Biotechnol.*, vol. 31, pp. 101–107, Feb. 2015.
[32] N. Michelusi, S. Pirbadian, M. Y. El-Naggar, and U. Mitra, "A stochastic model for electron transfer in bacterial cables," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 12, pp. 2402–2416, Dec. 2014.
[33] C. J. Myers, *Engineering Genetic Circuits*. London, U.K.: Chapman & Hall, 2009.
[34] D. L. Nelson, A. L. Lehninger, and M. M. Cox, *Lehninger Principles of Biochemistry*. New York, NY, USA: Macmillan, 2008.
[35] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Trans. Nanobiosci.*, vol. 13, no. 1, pp. 31–43, Mar. 2014.
[36] M. E. Ortiz and D. Endy, "Engineered cell-cell communication via DNA messaging," *J. Biol. Eng.*, vol. 6, p. 16, Dec. 2012.
[37] S. Payne and L. You, "Engineered cell–cell communication and its applications," in *Productive Biofilms* (Advances in Biochemical Engineering/Biotechnology), vol. 146. Cham, Switzerland: Springer, 2014, pp. 97–121.
[38] M. Pierobon, "A systems-theoretic model of a biological circuit for molecular communication in nanonetworks," *Nano Commun. Netw.*, vol. 5, nos. 1–2, pp. 25–34, Mar./Jun. 2014.
[39] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2532–2547, Jun. 2011.
[40] M. Pierobon and I. F. Akyildiz, "A statistical–physical model of interference in diffusion-based molecular nanonetworks," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2085–2095, Jun. 2014.
[41] M. Pierobon, Z. Sakkaff, J. L. Catlett, and N. R. Buan, "Mutual information upper bound of molecular communication based on cell metabolism," in *Proc. IEEE SPAWC*, Jul. 2016, pp. 1–6.
[42] R. Sarpeshkar, "Analog synthetic biology," *Philos. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 372, no. 2012, p. 20130110, 2014.
[43] E. Sazonov and M. R. Neuman, Eds., *Wearable Sensors: Fundamentals, Implementation and Applications*. Amsterdam, The Netherlands: Elsevier, 2014.
[44] J. T. Sexton and J. J. Tabor, "Multiplexed bacterial cell-cell communication via a genetically encoded CRISPRI-based multiplexer-demultiplexer circuit," in *Proc. 3rd ACM NanoCom*, 2016, Art. no. 12.
[45] S. Hennig, G. Rödel, and K. Ostermann, "Artificial cell-cell communication as an emerging tool in synthetic biology applications," *J. Biol. Eng.*, vol. 9, p. 13, Aug. 2015.
[46] A. Tasmir, J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires,'" *Nature*, vol. 469, pp. 212–215, Jan. 2011.
[47] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7358–7382, Dec. 2016.
[48] G. Tkaćik, A. M. Walczak, and W. Bialek, "Optimizing information flow in small genetic networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 3, p. 031920, Sep. 2009.
[49] B. D. Unluturk, A. O. Bicen, and I. F. Akyildiz, "Genetically engineered bacteria-based biotransceivers for molecular communication," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1271–1281, Apr. 2015.
[50] B. Wang, M. Barahona, and M. Buck, "Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks," *Nucl. Acids Res.*, vol. 42, no. 14, pp. 9484–9492, 2014.
[51] B. Wang, R. I. Kitney, N. Joly, and M. Buck, "Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology," *Nature Commun.*, vol. 2, Oct. 2011, Art. no. 508.
[52] R. Weiss and T. F. Knight, "Engineered communications for microbial robotics," in *DNA Computing* (Lecture Notes in Computer Science), vol. 2054, A. Condon and G. Rozenberg, Eds. Berlin, Germany: Springer, 2000.
[53] R. Weiss *et al.*, "Genetic circuit building blocks for cellular computation, communications, and signal processing," *Natural Comput.*, vol. 2, no. 1, pp. 47–84, Mar. 2003.
[54] L. You, R. X. Cox, III, R. Weiss, and F. H. Arnold, "Programmed population control by cell–cell communication and regulated killing," *Nature*, vol. 428, pp. 868–871, Apr. 2004.

Alessio Marcone received the M.S. degree in telecommunication engineering from the Politecnico di Milano, Milan, Italy, in 2016. He was a Researcher in molecular communications with the University of Nebraska-Lincoln, USA. He has been with the Standardization Department, Qualcomm Inc., since 2017. He has authored or co-authored several papers in molecular communication and synthetic biology. His interests include millimeter wave and sub-6GHz 5G systems, multiantenna communication techniques, and 5G systems modeling and standardization. He was a recipient of the 2017 IEEE INFOCOM Best Paper Runner-up Award.

Massimiliano Pierobon (S'09–M'13) received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2013. Since 2013, he has been an Assistant Professor with the Department of Computer Science and Engineering, University of Nebraska–Lincoln (UNL), NE, USA, where he also holds a courtesy appointment at the Department of Biochemistry. His research interests are in molecular communication theory, nanonetworks, intra-body networks, communication engineering applied to synthetic biology, and the Internet of Bio-Nano Things. He received the 2011 Georgia Tech BWN Lab Researcher of the Year Award, the 2013 IEEE COMMUNICATIONS LETTERS Exemplary Reviewer Award, the UNL CSE Upper and Graduate Level Teaching Award in 2016 and 2017, the 2017 IEEE INFOCOM Best Paper Runner-up Award, and the Best In-Session Presentation Award. He is currently the PI of the NSF project WetComm: Foundations of Wet Communication Theory and the Co-PI of the NSF project Redox-enabled Bio-Electronics for Molecular Communication and Memory (RE-BIONICS) and has been the Lead PI of the NSF project TelePathy: Telecommunication Systems Modeling and Engineering of Cell Communication Pathways. He has been an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS since 2013 and a Co-Editor-in-Chief of *Nano Communication Networks* (Elsevier) since 2017.

Maurizio Magarini (M'16) received the M.S. and Ph.D. degrees in electronic engineering from the Politecnico di Milano, Milan, Italy, in 1994 and 1999, respectively. He was a Research Associate with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, from 1999 to 2001. From 2001 to 2018, he was an Assistant Professor with the Politecnico di Milano, where he has been an Associate Professor since 2018. From 2008 to 2009, he was on a sabbatical leave at Bell Labs, Alcatel-Lucent, Holmdel, NJ, USA. He has authored or co-authored over 100 journal and conference papers. He has been involved in several European and national research projects. His research interests are in the broad area of communication and information theory, including synchronization, channel estimation, and equalization and coding applied to wireless and optical communication systems. Most recent research activities have focused on molecular communications, massive MIMO, study of waveforms for 5G cellular systems, wireless sensor networks for mission critical applications, and wireless networks using unmanned aerial vehicles and high-altitude platforms. In 1994, he was granted the TELECOM Italia Scholarship Award for his M.S. thesis.