TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES
Trans. Emerging Tel. Tech. 0000; **00**:1–14
DOI: 10.1002/ett

RESEARCH ARTICLE

Throughput and Energy Efficiency Based Packet Size Optimization of ARQ Protocols in Bacterial Quorum Communications

Chenyao Bai[1]*, Mark. S. Leeson[1], Matthew D. Higgins[2] and Yi Lu[1]

[1]School of Engineering, University of Warwick, Coventry, CV4 7AL, UK,
[2]WMG, International Manufacturing Centre, University of Warwick, Coventry, CV4 7AL, UK

ABSTRACT

The discovery that bacteria use signalling molecules, which are released into the environment, to communicate with each other changed our general perception of organisms inhabiting the world. Nowadays, the term quorum sensing (QS) is used to describe the phenomenon whereby a coordinated population response is controlled by exchanging specific diffusible chemical signals called autoinducers, enabling a cluster of bacteria to regulate their gene expression and behaviour collectively and synchronously. Bacteria assess their own population and coordinate their actions through the synthesis, accumulation and subsequent sensing of autoinducers. In this work, a bacterial quorum communication system is introduced, which contains two clusters of bacteria, specifically *Vibrio fischeri*, a gram-negative marine bacterium, as the transmitter node and receiver node, and the diffusive channel. The transmitted information is encoded into the concentration of autoinducers, with binary representation, and then it is divided into frames for transmission. Automatic Repeat reQuest (ARQ) protocols are applied to achieve better reliability. In addition, this paper addresses the question of optimal frame size for data communication in this channel capacity and energy constrained bacterial quorum communication system. The optimal fixed frame length is determined for a set of channel parameters by maximizing the throughput and energy efficiency matrix. Copyright © 0000 John Wiley & Sons, Ltd.

*Correspondence
School of Engineering, University of Warwick, Coventry, CV4 7AL, UK. E-mail: Chenyao.Bai@warwick.ac.uk

1. INTRODUCTION

Bacterial communications occur through a process called quorum sensing (QS), using signalling molecules, called autoinducers, which are released into their immediate environment [1]. QS describes the phenomenon whereby the accumulation of signalling molecules in the surrounding environment enables a single cell to sense the number of bacteria, or cell density, to produce a coordinated response from the whole population [1]. If the molecular signal concentration in the medium exceeds a certain threshold, an individual bacterium in a population releases more molecules into the environment [2], increasing the density of signalling molecules so that the concentration of external autoinducers is correlated with the bacterial cell population density. QS was first discovered via the marine bacterium *V. fischeri*, which produces luminescence when the local bacterial population is high [1][2]. Bacteria can alter the target gene expressions by changing the signalling molecular concentration, enabling coordinated behaviours, such as swarming motility and biofilm maturation based on the local bacterial population density [1]. Different bacterial species use different classes of signalling molecules to communicate and a single bacterial species may have more than one QS system, more than one signalling molecule and the ability to respond to different classes of autoinducers [3]. In this paper, a diffusive bacterial communication network between two populations of *V. fischeri* is considered. Transmitted information is represented by the signalling molecules concentration encoded into data frames; the release of molecules represents a binary "1" and no release represents a binary "0".

Noise resulting from gene expressions at the intracellular level and the diffusion of autoinducers presents a major challenge to the performance of natural and engineered QS networks [4]. For the bacterial communication system proposed in this paper, the effect of intersymbol interference (ISI) caused by molecular diffusion is a major impairment source. This may result in data packet corruption and out-of-sequence delivery, making it necessary to apply error detection rules and Automatic Repeat reQuest

Prepared using ettauth.cls [Version: 2012/06/19 v2.10]

(ARQ) mechanisms for reliable transmission. In this work, ARQ protocols are preferred to error correction schemes because they require simpler decoding operations than does error correction [5]. The performance of the Stop-and-Wait (SW-ARQ) protocol has been investigated in our previous work [6]. Here, the other two commonly used ARQ schemes, Go-Back-N (GBN-ARQ) and Selective-Repeat (SR-ARQ), have been presented and developed in addition for comparison purposes. Furthermore, since the bio-entities are constrained by information transmission rate, channel capacity and energy limitations [7], the optimal frame lengths for different ARQ schemes are investigated and analysed to obtain better throughput and energy performances. Although there have been several studies on packet size optimisation in traditional wireless and wired networks [8], these are not directly applicable to the bacterial scenario.

The engineering aspects of molecular communication have a research background that has become established in recent years. Generally, molecular communication research can be categorized based on five partially overlapping topics, including modulation techniques [9][10], channel modelling and noise analysis [11][12], coding techniques [13], protocols [14] and simulation tools [15]. **Different propagation schemes exist for molecular communication channels, thus, for each propagation scheme and modulation technique, there exist different channel models and noise models [16].** For example, **for diffusion-based communication,** Atakan and Akan [17] consider the time slotted, On-Off Keyed (OOK) modulation based channel which has been represented as a binary symmetric channel (BSC), where ligand receptors are applied for detection at the receiver. Einolghozati et al. [11][12] analyse and model the molecular communication channel with consideration of the ISI between consecutive transmission symbols. **In [18], the most relevant diffusion-based noise sources, including particle sampling noise and particle counting noise, are considered. Moreover, regarding the diffusion models with flow, the authors in [19] represent the communication with an additive inverse Gaussian noise channel model.** In addition, energy and throughput efficiency issues when utilizing ARQ schemes have been studied in sensor networks [20].

The contributions of this paper are as follows. To our knowledge, although ARQ protocols are well-known concepts in networking and coding theory, this is the first time that the application of known genetic logic gates [21] has been proposed for mapping existing ARQ protocols to QS based molecular communications. Instead of applying the previous presented energy models in [13][22] which take the energy consumption of the NAND gate as equivalent to that obtained from the hydrolysis of one molecule of adenosine triphosphate (ATP), this paper proposes a new energy model, which utilises the recent experimentally validated synthetic biological logic gates in [21]. This provides a promising method of applying synthetic biological components for the development of molecular communication techniques. Also, the paper investigates the optimal frame size for data transmission in channel capacity and energy constrained bacterial quorum communication systems to enhance throughput and energy efficiency. Specifically, this research shows how different protocols and frame lengths can be chosen to optimize different modes of bacterial communication. The results presented show that the throughput and energy efficiency depend on both channel conditions and characteristics of transceivers. Also, when a certain number of molecules are released at the start of each transmission, there is an optimal frame length to achieve the best throughput and energy efficiency. This work could be used in improving the sensitivity of bacterial biosensors and drug delivery systems. The rest of this paper is organized as follows: In Section 2, the basic bacterial communication scheme and the transceiver models are introduced, followed by the establishment of the channel in Section 3. In Section 4, use of the ARQ protocols and error detection codes to enhance the system performance is presented, and the throughput and energy efficiency analysis are introduced. This is followed by the numerical results and discussions in Section 5, where the optimal packet size is determined by maximizing the throughput and energy efficiency. Finally, Section 6 gives the concluding remarks and possible future work.

2. BACTERIAL QUORUM COMMUNICATION SETUP

The high degree of randomness and limited capabilities of a particular bacterium makes communication between two individual bacteria unreliable [2]. In addition, the communication delays can be fairly large due to biological actions such as transcription and translation. So an individual bacterium is very primitive and unreliable and hence incapable of transferring information by itself. Hence, to achieve communication system reliability, we here utilise the communication model between two populations of bacteria proposed in [2] where a cluster of bacteria large enough to collectively achieve reliable transmission and reception of molecular information is trapped in a chamber to form a node. The model consists of the transmitter node and the receiver node containing genetically modified bacteria able to sense specific types of signals and respond accordingly [2], and the communication channel. Molecular communication comprises production of the signalling molecules by the transmitter, their propagation through the medium undergoing Brownian motion and concentration sensing by the receiver node to produce luminescence whose steady-state measurement produces a concentration estimation and hence decodes the transmitted information [2]. Encoding is via the variation of the concentration

Trans. Emerging Tel. Tech. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

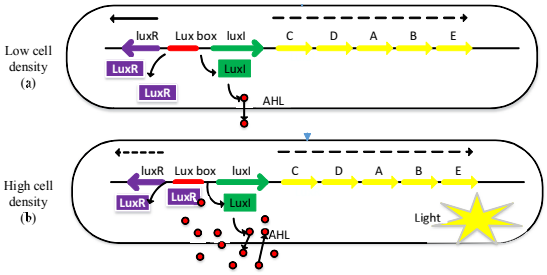

Figure 1. Luminescence production in *V. fischeri*. (a) The system is not active and there is basal transcription of luxR, luxI and luxCDABE. With the bacterial population increase, the autoinducers accumulate until the concentration exceeds a threshold which allows the binding between AHL and LuxR. (b) LuxR is bound to AHL, activating the transcription of the luminescence genes.

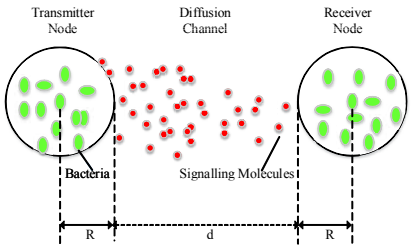

Figure 2. Bacterial communication scheme.

of signalling molecules and its transmission relies on diffusion in a theoretically infinite space.

In this proposed model, both the transmitter and receiver nodes contain m instances of the bacterium *V. fischeri*. **Two QS systems of the bacteria [23], *ain* and *lux*, using Acyl Homoserine Lactone (AHL) as signalling molecules, are considered in this work.** These bacteria are motile rods, 0.8-1.3μm in diameter and 1.8-2.4μm **in length [24]. The former QS system controls normal motility and is based on N-octanoyl-homoserine lactone (C8-HSL), denoted as Type I autoinducer, which is synthesized by AinS (regulated by gene ains) and detected by a hybrid sensor kinase AinR [23].** Bioluminescence in this bacterium is controlled by the ***lux*** system, which is composed of two regulatory genes, luxI and luxR, coding for proteins LuxI and LuxR, respectively. **The process of luminescence production is as follows:** at low cell densities when only a small number of bacteria are present, the signal N-(3-oxo-hexanoyl)-L-homoserine lactone (3-oxo-C6-HSL), which **is denoted as Type II autoinducer**, is synthesized by the protein LuxI, and is produced by the bacteria at a low level. Then the molecules diffuse out of the bacteria cells and propagate into the surrounding environment. When the bacterial population increases, Type II concentration around the node will grow interacting with its ligand receptor, the LuxR protein, if the concentration of the signal reaches a critical threshold. The LuxR/Type II complex binds to a region of DNA called the *lux* box, activating the transcription of the bioluminescence operon, which is made up of the luxCDABE genes. In addition, the LuxR/Type II complex also causes Type II autoinducer (via LuxI) to be produced at a higher level. Thus, Type II autoinducer auto-induces its own synthesis. This process is illustrated in Fig. 1.

In this work, the number of bacteria in each node is assumed to be constant. For example, the chamber could be a chemostat, in which a bacterial population could be maintained at a constant density; this situation has considerable similarity with bacterial growth in natural environments [25]. After certain initial adjustments, the overflow rate could be regulated and the inflow rate of nutrients would be capable of regulating and maintaining a constant population in the growth chamber, through the genetic regulated bacterial growth, division and death [25]. **It is assumed that each bacterium can sense and produce two different types of AHL molecules, specifically Type I and Type II autoinducers mentioned above. Hence, each bacterium must be equipped in general with two distinct receptor types (for Type I and Type II molecules) to perform its functionality as a transmitter or receiver. However, depending on the different functionalities, as a transmitter or receiver, only one type of receptor is activated. For the bacteria in the transmitter node, only Type I receptors (AinR) are activated and the gene expression of ains is repressed, while for bacteria in the receiver node, only Type II receptors (LuxR) are enabled and the gene expression of luxI is repressed, a process which can be controlled by proper enzymes.**

3. CHANNEL MODEL

In the channel model shown in Fig. 2, AHLs propagate through the channel via a diffusion process in a three dimensional medium, which is assumed to be extremely large compared to the size of the information molecules. Furthermore, collisions between these molecules are neglected and their motion is inspired by the forces produced by the constant random thermal molecular fluctuations within the fluid medium. The bacteria inside the transmitter node can produce various concentrations of Type II molecules to be transmitted through the channel by the stimulation of different levels of concentration of **specific stimulants**. The emitted signalling molecules then diffuse through the channel to the receiver which is at a distance d from the transmitter. At the receiver, each bacterium senses the concentration of Type II molecules through the corresponding Type II receptors **(LuxR)**, followed by the production of GFP by the bacteria, which is used to decode the input signal concentration.

Using QS, the bacteria cells in the receiver can synchronously respond to the molecules as they arrive. At the receiver, although the molecules pass through the bacteria cells in the node, the concentration of signalling molecules and the luminescence output will not be affected since the Type I receptors are not activated enabling the channel can be modelled as a Communication via Diffusion (CvD) channel as follows.

The proposed channel is a binary asymmetric channel (BAC) with binary input and binary output, and a probability of error. To effectively represent the transmitted symbols, the propagation time is divided into time slots of equal length, denoted by t_p, in each of which one symbol propagates. Information is encoded by concentration with binary representation. Specifically, if the number of information molecules arriving at the receiver in a certain time slot exceeds a threshold τ, the symbol is interpreted as "1"; otherwise, it will be taken as "0". Moreover, employing OOK modulation, the release of molecules in a time slot represents a binary one while their absence for the same duration represents a binary zero. Errors may be caused by ISI where a signal in which one symbol interferes with subsequent symbols. It should be noticed that the received signals tend to spread to subsequent symbols and smear into each other when a sequence of symbols is transmitted. ISI is related to the medium properties, the symbol propagation distance and the threshold value selected. Here, some information molecules may arrive at the receiver after the current time slot according to the diffusion dynamics leading to the incorrect decoding of the next received symbol.

So the information molecules propagate through the fluid medium undergoing Brownian motion which introduces randomness and a probability that the molecule will hit the receiver at a time slot. At a certain time duration t, the capture probability in a three dimensional environment is given by [22]:

$$P(r,t) = \frac{R}{R+d}\text{erfc}\left\{\frac{d}{2\sqrt{Dt}}\right\} \tag{1}$$

where d is the distance between the information molecule and a receiver with radius R, the value of which is related to the number of bacteria m in the receiver node. The diffusion coefficient D is $4.9\times10^{-6}\text{cm}^2\text{s}^{-1}$, which is settled as a conservative value for AHL in water at 25°C [26].

The communication channel is a Binomial one, where each molecule either arrives at the receiver or does not. ISI means that the number of molecules received in a time slot, denoted by N_{hit} is made up of the molecules sent at the start of the current time slot (N_c) and the sum of those sent at the start of the previous symbol durations (N_p). It is assumed that n information molecules are released at the start of each symbol. Also, the transmitted information includes k bits. Due to ISI, the i^{th} $(i\in[2,k])$ symbol can be affected by the symbols from $(i-1)$ previous time slots. Also, a_j $(j\in[2,i])$ is used to represent the transmitted information symbol from the $(j-1)^{\text{th}}$ previous time slot and a_1 represents the transmitted symbol in the current time slot. For the i^{th} symbol, the number of molecules received within the current time duration is a random variable and follows a binomial representation, which can be approximated with a normal distribution [27]:

$$N_c \sim a_1\mathcal{B}(n,P_1) \sim a_1\mathcal{N}(nP_1, nP_1(1-P_1)) \tag{2}$$

where P_1 represents the capture probability with receiver radius R, transmission distance d and symbol duration t_p, which can be calculated from equation (1).

It has been stated that for the i^{th} symbol, there exist $(i-1)$ previous time slots. Thus, there are 2^i different cases in binary representation, the decimal representations of which are 0 to (2^i-1). For example, there are two previous time slots for the third symbol, the 8 corresponding cases of which can be represented by (000,001,010,011,100,101,110,111), where the last bit represents the current symbol and the other two bits represent the previous symbols. In other words, there exist 2^{i-1} different error patterns for a fixed transmitted symbol in the current time slot. For the $q^{\text{th}}(q\in[1,2^{i-1}])$ error pattern, the number of left over molecules $N_{p(q)}$ belonging to all of the previous time slots can be found from [27]:

$$\begin{aligned} N_{p(q)} &\sim \sum_{j=2}^{i} a_j\left(\mathcal{B}\left(n,p_{j-1}^{j}\right)\right) \\ &\sim \sum_{j=2}^{i} a_j\left(\mathcal{N}\left(np_{j-1}^{j}, np_{j-1}^{j}\left(1-p_{j-1}^{j}\right)\right)\right) \\ &\sim \mathcal{N}\left(\mu_{p(q)},\sigma_{p(q)}^2\right) \end{aligned} \tag{3}$$

where $p_{j-1}^{j} = p_j - p_{j-1}$, and p_j is the capture probability with receiver radius R, transmission distance d and time duration of jt_p; $\mu_{p(q)},\sigma_{p(q)}^2$ are the mean and variance of the distribution of the number of left over molecules from all the previous time slots, respectively. The total number of molecules $N_{\text{hit}(q)}$ received in the i^{th} symbol duration is the sum of N_c and $N_{p(q)}$.

If the transmitted symbol in the current time slot is a "1" $(a_1=1)$, the number of received molecules needs to be larger than τ for correct decoding. Thus for the q^{th} error pattern, the total number of molecules that have arrived in the current time slot is represented by:

$$N_{\text{hit}(q)} = N_{p(q)} + N_c \sim \mathcal{N}\left(\mu_{\text{hit}(q)},\sigma_{\text{hit}(q)}^2\right) \tag{4}$$

where $\mu_{\text{hit}(q)}$ and $\sigma_{\text{hit}(q)}^2$ are the expectation and variance of the distribution of $N_{\text{hit}(q)}$. Thus, the error probability for this case is:

$$p_{e1(q)} = \frac{1}{2^i}P\left(N_{\text{hit}(q)}<\tau\right) = \frac{1}{2^i}\left(1-Q\left(\frac{\tau-\mu_{\text{hit}(q)}}{\sqrt{\sigma_{\text{hit}(q)}^2}}\right)\right) \tag{5}$$

where the Q-function is the tail probability of the standard normal distribution. Similarly, for the case when the symbol in the current time slot is 0, the total number of molecules received in the i^{th} symbol duration is:

Trans. Emerging Tel. Tech. 0000; 00:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

$$N_{\text{hit}(q)} = N_{p(q)} \sim \mathcal{N}\left(\mu_{p(q)}, \sigma^2_{p(q)}\right) \quad (6)$$

Hence, the error probability for the case when the transmitted symbol in the current time slot is 0 is given by:

$$p_{e0(q)} = \frac{1}{2^i} P\left(N_{\text{hit}(q)} > \tau\right) = \frac{1}{2^i} Q\left(\frac{\tau - \mu_{p(q)}}{\sqrt{\sigma^2_{p(q)}}}\right) \quad (7)$$

It is assumed that the transmitter sends "0"s and "1"s with equal frequency. Due to the fact that there exist 2^i different error patterns, that the capture probability is a function of R, d, t_p, and that the receiver radius R is determined by the number of bacteria m in the node, the error probability for given values of n, τ, m, d, t_p can be calculated using:

$$p_{er} = \sum_{q=1}^{2^{i-1}} \left(p_{e1(q)} + p_{e0(q)}\right) \quad (8)$$

The bit error rate (BER) is a key parameter employed to assess the performance of communication systems and here BER refers to the probability of one bit error when information symbols are transmitted in the diffusion based communication channel. According to [7], most molecules arrive at the receiver in a relatively short time while only a few molecules arrive after a very long period of time, which will lead to an unsatisfactorily increasing average hitting time. In this model, the symbol duration t_p is chosen as the time at which 60% of the molecules have arrived at the receiver, i.e. $P(r, t)$ in equation (1) is taken as 0.6. It has been stated that the error probability is a function of parameters n, τ, m, d, t_p. Thus in this paper, for given values of n, m, d, t_p, the BER is taken as the value when the error probability calculated by equation (8) reaches its minimum for $\tau \in [1, n]$. The corresponding values of threshold are applied in the following research.

4. ARQ PROTOCOLS

Channels with a relatively high BER level cause frequent packet corruptions and out-of-sequence delivery, needing error check codes and ARQ mechanisms for effective error detection and recovery, respectively [5]. ARQ forms the basis of peer-to-peer protocols that provide for the reliable transfer of information and is an error control technique to ensure that a data stream is delivered accurately to the user despite transmission errors. The basic elements of ARQ protocols are: information transferring frames, acknowledgement frames (ACKs) and negative acknowledgements (NAKs) to signify the receipt of a given frame; the timeout mechanism required to maintain the flow of frames.

Advances in synthetic biology, particularly the foundation of the BioBricks database [28], enable many capabilities based on genetically engineered bacteria, including timing, counting, clocking, logic gates, pattern detection and intercellular communication [29]. Moreover, because ARQ mechanisms and error control operations can be implemented through circuits [5], these may thus be implemented in bacterial communication systems as well. The SW-ARQ protocol has been investigated in our previous work [6] which shows that it suffers from inefficiency due to the fact that the channel is idle between the transmission of the message and the reception of its ACK or NAK. Theoretically, the GBN-ARQ and SR-ARQ protocols offer a better performance but have a higher requirement for buffers at the receiver. This paper analyses the three basic ARQ schemes to determine the optimal frame size for different modes of the channel parameters to enhance throughput and energy efficiency in the bacterial quorum communication system.

Here, the transmitter generates a sequence of information frames for transmission, each of which contains a header with sequence numbers essential for in-sequence delivery, information bits and error detection bits. Cyclic redundancy check (CRC) codes are used to effect error detection, represented by polynomials, having good error sensing performance, fast encoding and decoding capabilities, and applicability to varying message lengths [5]; they are appended to the frame to determine if error occurs during transmission. The three commonly used versions of generator polynomial in industry are:

- CRC8 $= x^8 + x^7 + x^6 + x^4 + x^2 + 1$
- CRC12 $= x^{12} + x^{11} + x^3 + x^2 + 1$
- CRC16 $= x^{16} + x^{15} + x^2 + 1$

It is assumed that the information flows only in one direction, from the transmitter to the receiver. The reverse communication channel is used only for the transmission of ACKs/NAKs. **As has been stated** in Section 2, for the transmitter node, only Type I receptors are activated, while for the receiver node, only Type II receptors are enabled. In addition, because the signalling molecules used to encode the acknowledgement frames in the reverse channel (Type I molecules) are different from autoinducers used to encode the information frames in the forward channel (Type II molecules), they will not interfere with each other. Also, at the receiver node, the inhibition of **the expression of gene luxI** makes it impossible to generate **extra** Type II molecules, providing a more accurate decoding of the transmitted information at the receiver. **Similarly, at the transmitter node, the repression of gene expression of ains represses the synthesis of extra Type I molecules, making it more accurate to decode the acknowledgement information at the transmitter**. Therefore, both the transmitter and receiver are able to generate and receive different types of signalling molecules to avoid adjacent channel interference. Moreover, genetic marking techniques, such as fluorescent labelling technology, could be applied to distinguish between **different types of molecules used for sending messages and that for acknowledgement messages for better observation** [30].

Generally, shorter frames are more reliable, while longer frame sizes are associated with greater loss rates which may result in reductions of throughput and energy efficiency despite their increased information block size. Hence, we intuitively expect an optimal frame size that balances these conflicting interests. Traditionally, error control is decoupled from data link layer packet size optimisation [8]. However, error control parities consume valuable transceiver time and energy which must be taken into account in bacterial nodes. The encoding/decoding time and energy also need to be incorporated, which will be discussed later. Our approach to packet length optimization is unique in this regard. In the following subsections, we briefly examine the time and energy consumption characteristics and define a suitable optimization metric.

4.1. Energy model

In a conventional wireless communication system, the energy budgets of the transmitters and receivers result in a constrained system performance. Energy is usually provided by limited battery power and should be used efficiently to maximize the system performance while attaining acceptable operational device lifetime values [7]. For bacterial communications, due to the size of bio-entities and the limited energy storage functionalities of the transceivers, energy efficiency is also a design parameter.

In this paper, both the transmitter and receiver nodes are considered able to store energy. The energy required during the communication process is provided by the culture media which are designed to provide all the essential nutrients in solution for bacterial growth. During the communication process, protein and mRNA molecules have to be constantly synthesised to counter their degradation, consuming power provided by the hydrolysis of ATP, which is the universal currency of energy transfer between cells in living organisms [31]. Breaking off phosphate groups enables various processes in transcription and translation to move forward in a nearly irreversible way. The hydrolysis of one molecule of ATP provides nearly $20k_BT$ of energy, where k_B is Boltzmann's constant and T is the absolute temperature [22]. The energy cost of synthesizing an mRNA molecule in ATP is $E_{\text{mRNA}} \approx 30L_{\text{mRNA}}$, where L_{mRNA} is the length of the mRNA molecule in nucleotides [31]. Similarly, the energy cost of synthesizing a protein molecule in ATP is $E_{\text{prot}} \approx 50L_{\text{prot}}$ where L_{prot} is the length of the protein molecule in amino acids. These estimates are speculative since the energy consumption in bacterial communications has yet to be fully investigated. To implement the necessary CRC codes, logical operations in traditional communications are used [5], such as the logical representation of the systematic encoder and decoder of CRC-8 in Fig. 3.

It is clear in Fig. 3 that the logical operations may be reduced to an arrangement of XOR gates, NOR gates and shift registers. Thus, we need an insight into the ATP cost of implementing the logical operations required for the

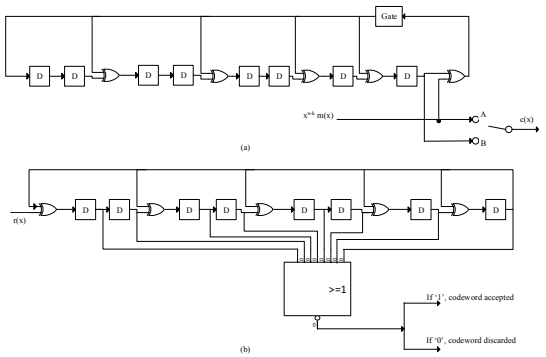

Figure 3. (a) Systematic encoder for CRC-8, where $m(x)$ is the information bits and $c(x)$ is the code word; (b) Systematic decoder for CRC-8, where $r(x)$ is the received code word.

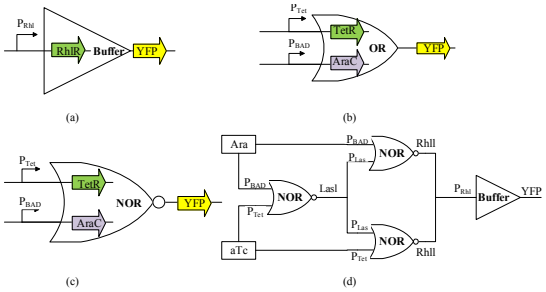

Figure 4. (a) Buffer gate with P_{Rhl} (downstream gene rhlR and corresponding protein RhlR) as the input promoter and yellow fluorescent protein (YFP) as the output; (b) Genetic OR gate with P_{BAD} (downstream gene araC and corresponding protein AraC) and P_{Tet} (downstream gene tetR and corresponding protein TetR) as the input promoters and YFP as the output; (c) Genetic NOR gate with P_{BAD} and P_{Tet} as the input promoters and YFP as the output; (d) XOR gate which is built with three NOR gates and a buffer gate. P_{BAD} and P_{Tet} are activated in the presence of arabinose (Ara) and anhydrotetracycline (aTc), respectively. The promoter P_{las} (downstream gene lasR and corresponding protein LasR) is activated by signal LasI [21].

encoder and decoder. In this paper, the synthetic genetic logic gates constructed in [21] are considered to design the energy model. The four logic gates which will be used to design the energy model are displayed in Fig. 4.

NOR gates are functionally complete, which means that any computational operation can be implemented by layering them. In Fig. 4(d), the XOR gate is made up of three NOR gates and one buffer gate, which are connected in series. The output of the first gate is the expression of the AHL synthase (LasI or RhlI). AHL diffuses through the cell membrane and binds to its cognate transcription factor, leading to the activation of the promoter, which acts as the input to the next logic gate. In other words, AHLs are used as signal-carrying 'wires' to connect the logic gates [21]. However, it should be noted that since multiple gates can be layered to build more complex computations, it remains difficult to predict how a combination of circuits

Trans. Emerging Tel. Tech. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

will behave on the basis of the functions of the individuals [21].

Due to the fact that not all the concentrations of the elements (e.g., promoters, mRNAs and proteins) involved in the logic gates have been given in [21], the energy consumption computations of logic gates are simplified by taking the number of all the inputs as 1, which means that we do not consider the actual required concentration of the inputs of the logic gate implementations. **Also, for bacteria, the distribution of individual proteins may be described a linear relationship [protein]$=\gamma\times$[mRNA] [32]. The constant of proportionality, γ, is a random variable that ranges between approximately 300 and 600 with a mean of 540 proteins/mRNA. Thus in this work, we take the ratio of protein to mRNA to be the mean value $\gamma = 540$. Although this simplification is inevitably an approximation and may be improved by future experimental discoveries, this work is considered to be useful as it provides a method to combine experimentally validated synthetic biological logic gates with the energy consideration for molecular communications**. We introduce the nomenclature as follows: L_{xyz} is the length in amino acids of the protein or promoter xyz; E_{xyz} is the energy in ATP units required to synthesise or transcript the protein or promoter xyz. Moreover, we assume equally probable zeros and ones in the data stream so that the mean energy cost is the mean of the energy values for each of the possible input-output states of the logic gates considered.

For the buffer, the output is the same as the input and so the energy cost when the input is binary "1" is:

$$\begin{aligned} E_{\text{buffer1}} &= E_{\text{rhlR}} + E_{\text{RhlR}} + E_{\text{YFP}} \\ &= 30L_{\text{rhlR}} + 50\gamma L_{\text{RhlR}} + 50\gamma L_{\text{YFP}} \end{aligned} \tag{9}$$

Since a binary "0" input produces a binary "0" output, the energy consumption is zero so:

$$E_{\text{buffer}} = E_{\text{buffer1}}/2 = 15L_{\text{rhlR}} + 25\gamma L_{\text{RhlR}} + 25\gamma L_{\text{YFP}} \tag{10}$$

For an OR gate, the output is "0" only when both the inputs are "0". For the case "00", there is no energy cost. For case "01", only the promoter P_{BAD} is activated, and the protein YFP is synthesised. Thus in this case, the energy consumption is:

$$\begin{aligned} E_{\text{or}_{01}} &= E_{\text{araC}} + E_{\text{AraC}} + E_{\text{YFP}} \\ &= 30L_{\text{araC}} + 50\gamma L_{\text{AraC}} + 50\gamma L_{\text{YFP}} \end{aligned} \tag{11}$$

For the case "10", only the promoter P_{Tet} is activated, and the output is YFP. Thus in this situation, the energy cost is:

$$\begin{aligned} E_{\text{or}_{10}} &= E_{\text{tetR}} + E_{\text{TetR}} + E_{\text{YFP}} \\ &= 30L_{\text{tetR}} + 50\gamma L_{\text{TetR}} + 50\gamma L_{\text{YFP}} \end{aligned} \tag{12}$$

For the case "11", both promoters are activated and the output is YFP. Thus the energy consumption for this case is:

$$\begin{aligned} E_{\text{or}_{11}} &= E_{\text{araC}} + E_{\text{AraC}} + E_{\text{tetR}} + E_{\text{TetR}} + E_{\text{YFP}} \\ &= 30L_{\text{araC}} + 50\gamma L_{\text{AraC}} + 30L_{\text{tetR}} + 50\gamma L_{\text{TetR}} + 50\gamma L_{\text{YFP}} \end{aligned} \tag{13}$$

Hence, the energy cost for the OR gate is given by:

$$E_{\text{or}} = \frac{1}{4}\left(E_{\text{or}_{00}} + E_{\text{or}_{01}} + E_{\text{or}_{10}} + E_{\text{or}_{11}}\right) \tag{14}$$

The output of the NOR gate shown in Fig. 4(c), is binary "1" only if both inputs are binary "0". For the case "00", neither of the input promoters is activated and expressed, while the output is YFP. Thus the energy cost for this case is the energy required to synthesise YFP, which is denoted by:

$$E_{\text{nor}_{00}} = E_{\text{YFP}} = 50\gamma L_{\text{YFP}} \tag{15}$$

Similarly, for case "01", only promoter P_{BAD} is activated and its downstream gene AraC is expressed, while YFP is not synthesised. Hence, the energy cost for this case is:

$$E_{\text{nor}_{01}} = E_{\text{araC}} + E_{\text{AraC}} = 30L_{\text{araC}} + 50\gamma L_{\text{AraC}} \tag{16}$$

For case "10", only promoter P_{Tet} is activated and the downstream gene tetR is expressed so the energy consumption is:

$$E_{\text{nor}_{10}} = E_{\text{tetR}} + E_{\text{TetR}} = 30L_{\text{tetR}} + 50\gamma L_{\text{TetR}} \tag{17}$$

For case "11", both the input promoters are activated and the corresponding downstream genes are expressed. Thus, the energy cost for this case is:

$$\begin{aligned} E_{\text{nor}_{11}} &= E_{\text{araC}} + E_{\text{AraC}} + E_{\text{tetR}} + E_{\text{TetR}} \\ &= 30L_{\text{araC}} + 50\gamma L_{\text{AraC}} + 30L_{\text{tetR}} + 50\gamma L_{\text{TetR}} \end{aligned} \tag{18}$$

Then, the NOR gate energy consumption is found from:

$$E_{\text{nor}} = \frac{1}{4}\left(E_{\text{nor}_{00}} + E_{\text{nor}_{01}} + E_{\text{nor}_{10}} + E_{\text{nor}_{11}}\right) \tag{19}$$

The XOR gate only produces a binary "1" output when the inputs are different. For the case "00", both the inputs P_{BAD} and P_{Tet} are inactivated. However, the promoter P_{las} is activated by the quorum signal LasI. Thus for this case, the energy cost can be denoted by:

$$\begin{aligned} E_{\text{xor}_{00}} &= E_{\text{LasI}} + 2\left(E_{\text{lasR}} + E_{\text{LasR}}\right) \\ &= 50\gamma L_{\text{LasI}} + 60L_{\text{lasR}} + 100\gamma L_{\text{LasR}} \end{aligned} \tag{20}$$

Similarly, for case "01", among the three NOR gates of the XOR gate, the promoter P_{Tet} of two NOR gates are activated. Also, the protein YFP and RhlI are synthesised. Moreover, in this case, the promoter P_{Rhl} is expressed, with the expression of gene rhlR and synthesis of protein RhlR. Thus the energy cost can be given as:

$$\begin{aligned} E_{\text{xor}_{01}} &= 2\left(E_{\text{tetR}} + E_{\text{TetR}}\right) + E_{\text{YFP}} + E_{\text{RhlI}} + E_{\text{rhlR}} + E_{\text{RhlR}} \\ &= 60L_{\text{tetR}} + 100\gamma L_{\text{TetR}} + 50\gamma L_{\text{YFP}} + 50\gamma L_{\text{RhlI}} \\ &\quad + 30L_{\text{rhlR}} + 50\gamma L_{\text{RhlR}} \end{aligned} \tag{21}$$

In the case "10", among the three NOR gates of the XOR gate, the promoter P_{BAD} of two NOR gates are activated. Also, the protein YFP and RhlI are synthesised. Moreover, in this case, the promoter P_{Rhl} is expressed, with the expression of gene rhlR and synthesis of protein RhlR. Thus the energy cost can be given as:

$$\begin{aligned} E_{\text{xor}_{10}} &= 2\left(E_{\text{araC}} + E_{\text{AraC}}\right) + E_{\text{YFP}} + E_{\text{RhlI}} + E_{\text{rhlR}} + E_{\text{RhlR}} \\ &= 60L_{\text{araC}} + 100\gamma L_{\text{AraC}} + 50\gamma L_{\text{YFP}} + 50\gamma L_{\text{RhlI}} \\ &\quad + 30L_{\text{rhlR}} + 50\gamma L_{\text{RhlR}} \end{aligned} \tag{22}$$

For the case "11", among the three NOR gates of the XOR gate, both the promoters P_{Tet} and P_{BAD} of two NOR gates are activated. Hence, the energy cost of this case is displayed as:

$$\begin{aligned} E_{\text{xor}_{11}} &= 2\left(E_{\text{tetR}} + E_{\text{TetR}}\right) + 2\left(E_{\text{araC}} + E_{\text{AraC}}\right) \\ &= 60L_{\text{tetR}} + 100\gamma L_{\text{TetR}} + 60L_{\text{araC}} + 100\gamma L_{\text{AraC}} \end{aligned} \tag{23}$$

In a similar fashion to that used for the NOR gate, the energy cost of XOR gate can be described by:

$$E_{\text{xor}} = \frac{1}{4}\left(E_{\text{xor}_{00}} + E_{\text{xor}_{01}} + E_{\text{xor}_{10}} + E_{\text{xor}_{11}}\right) \tag{24}$$

4.2. SW-ARQ

In SW-ARQ, the transmitter waits for a receiver ACK/NAK after transmitting a frame before transmitting the next frame. The receiver checks a received frame for errors and sends an ACK or NAK through the feedback path depending on whether an error was found. If an ACK is received by the transmitter, the transmitter transmits the next frame, if not it will retransmit the previously sent frame. However, if the ACK/NAK is lost, the transmitter waits for a timeout period until the next transmission. So the retransmission continues until a frame is received correctly and is positively acknowledged, or the number of retransmissions reaches a certain threshold. To avoid ambiguities a sequence number can be added to the frames. The protocol continues in this manner until all the frames are transmitted successfully.

The throughput efficiency is defined as the number of information bits correctly transmitted divided by the total number of bits transmitted by the transmitter node. It is assumed an (n_c, k) code is used in the protocol, where k is the number of information bits and n_c is the total number of bits in the information frame including the information bits, sequence numbers and error checking bits. The probability that a code word is received incorrectly at the receiver is represented by:

$$P = 1 - (1 - p_e)^{n_c} \tag{25}$$

where p_e is the raw channel BER, which can be calculated from Section 3. It should be noted here that we only consider independent bit errors, other than burst error conditions with no interleaving. One successful transmission of a code word requires a total time T_t, which is:

$$T_t = 2t_p + 2t_{ps} + t_t + t_{tk} = 2t_p + 2t_{ps} + \frac{n_c}{R_b} + \frac{l}{R_b} \tag{26}$$

where t_p is the propagation time between the transmitter and the receiver for one bit, which can be calculated according to Section 3. The time from when the first bit of a frame arrives at the receiver to when the last bit arrives is denoted by t_t (for an information frame) or t_{tk} (for an ACK/NAK frame). The logical operations needed for CRC codes can be implemented using genetic circuits with a processing time, t_{ps}, which is set as four hours according to [21]. In this time interval, an equivalent logical operation behaviour can be produced [21]. The variable l is the number of bits for the acknowledgement frame which is of the same value as the number of sequence bits and R_b is the bit rate of the transmission channel. The boundary condition for the choice of R_b is represented by:

$$(n_c \cdot t_p + t_{ps}) \cdot R_b \leq n_c \tag{27}$$

The boundary condition above indicates that in the time period to transmit one code word which contains n_c bits of information, no more than n_c bits are sent by the transmitter. Here, R_b is chosen as the maximum value according to equation (27). In addition, the timeout period is set to be exactly equal to the sum of round trip propagation delay and the CRC processing time. However, not all error patterns are able to be detected in CRC implementations [33]. Assuming that P_u is the probability of undetected frame error, the probability that a transmission for a given frame is the last transmission can be calculated from $(1 - P + P_u)$. Generally, CRC can detect all single, double and odd bits in error [34].

In the case of an error, for the SW-ARQ scheme, the average number of attempts N_{r_sw} required to transmit a code word successfully is given by:

$$N_{r\_\text{sw}} = \sum_{i=1}^{\infty} iP^{i-1}(1 - P + P_u) = \frac{1 - P + P_u}{(1 - P)^2} \tag{28}$$

Therefore, the average time required to successfully transmit the code is: $T_{r\_\text{sw}} = N_{r\_\text{sw}} T_t$. For transmission in an uncoded system, the transmission time of the code would have been: $T_{tu} = k/R_b$. Thus, the throughput efficiency of the SW-ARQ scheme is calculated by:

$$\eta(\text{sw\_th}) = \frac{T_{tu}}{T_{r\_\text{sw}}} = \frac{k/R_b}{\left(2t_p + 2t_{ps} + \frac{n_c}{R_b} + \frac{l}{R_b}\right)\left(\frac{1-P+P_u}{(1-P)^2}\right)} \tag{29}$$

Since CRC implementations can be represented by combinations of logic gates, their energy costs are needed as the first step in calculating the energy efficiency of the ARQ protocols. The systematic encoder in Fig. 3 is made up of 8 buffer gates and 5 XOR gates, while the systematic decoder constitutes 8 buffer gates, 5 XOR gates and one 8-input NOR gate, which can be constructed with 6 OR

Trans. Emerging Tel. Tech. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

gates and 1 NOR gate due to the 2-input limitation of the proposed logic gates in Fig. 4. So the energy cost of a CRC-8 encoder can be described by:

$$E_{\text{encode8}} = 8E_{\text{buffer}} + 5E_{\text{xor}} \tag{30}$$

The energy cost of CRC-8 decoder can be denoted by:

$$E_{\text{decode8}} = 8E_{\text{buffer}} + 5E_{\text{xor}} + 6E_{\text{or}} + E_{\text{nor}} \tag{31}$$

Similarly, according to [5], the energy cost for CRC-12 encoder and decoder can be represented by:

$$E_{\text{encode12}} = 12E_{\text{buffer}} + 5E_{\text{xor}} \tag{32}$$

$$E_{\text{decode12}} = 12E_{\text{buffer}} + 5E_{\text{xor}} + 10E_{\text{or}} + E_{\text{nor}} \tag{33}$$

In addition, the energy consumption for a CRC-16 encoder and decoder can be represented by:

$$E_{\text{encode16}} = 16E_{\text{buffer}} + 3E_{\text{xor}} \tag{34}$$

$$E_{\text{decode16}} = 16E_{\text{buffer}} + 3E_{\text{xor}} + 14E_{\text{or}} + E_{\text{nor}} \tag{35}$$

As stated in Section 3, the number of molecules emitted at the start of each time slot is denoted by n. Hence for an (n_c,k) code, the number of molecules required to construct an information frame is denoted by $E_{\text{info}} = 50n \cdot n_c \cdot L_{\text{AI2}}$, where L_{AI2} is the length of the Type II signalling molecule in amino acids. At the receiver, the energy required to construct the LuxR receptors for Type II is given by $E_{\text{re}} = 50n \cdot n_c \cdot L_{\text{LuxR}}$, where L_{LuxR} is the length of LuxR molecule. In the feedback channel, the number of molecules required to construct an ACK/NAK frame is represented by $E_{\text{ack}} = 50n \cdot l \cdot L_{\text{AI1}}$, where L_{AI1} is the length of the Type I signalling molecule in amino acids. In addition, the energy required for buffers in both the transmitter and receiver is represented by $E_{\text{transb}} = (L_{\text{send}} + L_{\text{rec}}) \cdot E_{\text{buffer}}$, where L_{send} and L_{rec} represent the sending window size and receiving window size, respectively. So the energy consumption of a frame in one hop is $E = E_{\text{info}} + E_{\text{re}} + E_{\text{ack}} + E_{\text{encode}} + E_{\text{decode}} + E_{\text{transb}}$. where E_{encode} and E_{decode} are the energy consumption for CRC encoder and decoder, respectively, for different CRC polynomials; and can be obtained from equation (30 $\sim$ 35). Hence, the average energy consumption for transmitting a code word is:

$$E_{\text{sw}} = E \cdot N_{r\_\text{sw}} \tag{36}$$

In the case of transmission via an uncoded system, with no CRC and ARQ mechanisms applied, the energy required to transmit a code word is: $E_0 = 50n \cdot k \cdot L_{\text{AI2}} + 50n \cdot k \cdot L_{\text{LuxR}}$.

In this paper, energy efficiency is expressed by the ratio of the energy consumption to transmit a frame without ARQ to the energy cost to transmit a frame with retransmission schemes. Thus, for SW-ARQ protocol, the energy efficiency can be described as: $\eta(\text{sw\_en}) = E_0/E_{\text{sw}}$. Using the GenBank database [35], the length of promoters and proteins mentioned above are given in Table. I.

Table I. Promoters and proteins lengths

Promoters	length in nucleotides	Proteins	Length in amino acids
araC	879bp	LasI	201aa
tetR	627bp	RhlI	201aa
lasR	720bp	YFP	243aa
rhlR	726bp	AI1	216aa
		AI2	160aa
		LuxR	250aa
		TetR	228aa
		AraC	292aa
		LasR	80aa
		RhlR	241aa

4.3. GBN-ARQ

In the GBN-ARQ technique, the transmitter continuously transmits a block of W (often known as the window size) frames without waiting for the acknowledgement for the individual frame. Each frame must be buffered (stored) until a valid ACK arrives, in case retransmission is needed. The length W represents the maximum number of frames that may be outstanding simultaneously. The receiver keeps track of the sequence number of the next frame it expects to receive, and sends that number with every ACK it sends and only sends the NAK if errors are detected. When the transmitter receives a NAK for the first time, it stops transmission and resends all the frames which were transmitted prior to stopping of transmission but starting from the frame for which NAK is received, and discards the frames transmitted prior to the frame in error from the memory. In short, the receiver will discard any frame that does not have the exact sequence number it expects (either a duplicate frame it already acknowledged or an out-of-order frame it expects to receive later) and will resend an ACK for the last correct in-order frame. Once the sender has sent all of the frames in its window, it will detect that all of the frames since the first lost frame are outstanding, and will go back to sequence number of the last ACK it received from the receiver process and fill its window starting with that frame and continue the process over again. For example, if the fourth frame of the block is the first negatively acknowledged from when up to Wth frame has been transmitted, the transmitter will then discard first to third frames from its buffer and retransmit all the frames from fourth to Wth. It is the best situation when none of the frames in one block is negatively acknowledged and successful transmission of W packets can be achieved within a minimum of one timeout period. Compared with the best situation in SW-ARQ where one frame is involved with one timeout period, GBN-ARQ provides throughput and energy improvement in theory.

There are two different types of GBN-ARQ: continuous and non-continuous. In the continuous scheme, after transmission of a block of W packets, the transmitter

does not need to wait for the acknowledgements of these packets before starting the transmission of next block. For continuous GBN-ARQ scheme, there are W packets in each block. The boundary condition is represented by $W > T_t/t_t$, where T_t and t_t have the same meanings as in equation (26). It can be derived that:

$$W > \frac{2\left(t_p + t_{ps}\right) R_b + n_c + l}{n_c} \tag{37}$$

In this case, the average time for successful transmission of a frame is given by:

$$\begin{aligned} T_{r\_\text{con\_gbn}} &= t_t + \sum i P^i \left(1 - P + P_u\right) T_t \\ &= \frac{n_c}{R_b} + \frac{P\left(1 - P + P_u\right)}{(1-P)^2} T_t \end{aligned} \tag{38}$$

In the case of transmission using an uncoded system, the transmission time of the code T_{tu} has already been given above. Thus, the throughput efficiency of GBN-ARQ scheme is calculated by: $\eta(\text{con\_gbn\_th}) = T_{tu}/T_{r\_\text{con\_gbn}}$. For this continuous GBN-ARQ scheme, the average number of attempts required per successful transmission of a frame can be calculated by: $N_{r(\text{con-gbn})} = T_{r\_\text{con\_gbn}}/T_t$. Hence, the energy efficiency for this case is given by: $\eta(\text{con\_gbn\_en}) = E_0/(E \cdot N_{r(\text{con-gbn})})$.

In contrast with the case of continuous mode, in the non-continuous mode, before starting the transmission of the next block, the transmitter has to wait for the previous block's packet acknowledgements. In this case, the average number of transmission times for the successful transmission of a frame is given by [33]:

$$\begin{aligned} N_{r(\text{unc-gbn})} &= \frac{1}{W} \sum \left(iW + 1\right) P^i \left(1 - P + P_u\right) \\ &= \frac{\left(1 - P + P_u\right)\left(1 - P + WP\right)}{W\left(1 - P\right)^2} \end{aligned} \tag{39}$$

where $i = 0, 1, 2, \ldots$. Therefore, the average time required to successfully transmit the code is: $T_{r\_\text{unc\_gbn}} = N_{r(\text{unc-gbn})} T_t$. Hence, it can be concluded that the throughput efficiency of non-continuous GBN-ARQ scheme is calculated by: $\eta(\text{unc\_gbn\_th}) = T_{\text{tu}}/T_{r\_\text{unc\_gbn}}$. Also, the energy efficiency for this case is given by: $\eta(\text{unc\_gbn\_en}) = E_0/\left(E \cdot N_{r(\text{unc-gbn})}\right)$.

4.4. SR-ARQ

SR-ARQ operates in a similar way to GBN-ARQ, sending a number of frames specified by a window size, but only retransmits the frame for which a NAK is received or timed out. The receiver accepts and buffers out-of-order frames, requiring more receiver buffer space.

As for GBN-ARQ, for the SR technique, there are also similar two cases.

Case I: continuous scheme when $W > T_t/t_t$:

Here the average transmission time for successful transmission of a frame is straightforward as below [33]:

$$T_{r\_\text{con\_sr}} = \frac{T_{t\_\text{con\_sr}}}{1 - P + P_u} \tag{40}$$

where $T_{t\_\text{con\_sr}}$ is the time required for one successful transmission of a frame in continuous SR-ARQ scheme, which is different from that of SW-ARQ and GBN-ARQ stated above. It can be calculated using [33]:

$$T_{t\_\text{con\_sr}} = \frac{W t_t + k_0 \cdot 2t_p}{W} \tag{41}$$

where k_0 is the number of errors out of W, which can be represented by [33]:

$$k_0 = \sum_{i=1}^{W} i p^i \left(1 - P + P_u\right)^{W-i} \tag{42}$$

For an uncoded system, the transmission time remains that used in the previous sections. Thus, the throughput efficiency of SR-ARQ scheme is calculated by: $\eta(\text{con\_sr\_th}) = T_{tu}/T_{r\_\text{con\_sr}}$. For this continuous SR-ARQ scheme, the average number of attempts required per successful transmission of a frame can be calculated by: $N_{r(\text{con-sr})} = T_{r\_\text{con\_sr}}/T_{t\_\text{con\_sr}}$. So the energy efficiency for this case is given by: $\eta(\text{con\_sr\_en}) = E_0/(E \cdot N_{r(\text{con-sr})})$.

Case II: when $W < T_t/t_t$:

In this case, the average number of transmission required for successful transmission of a frame is given by [33]:

$$N_{r(\text{unc-sr})} = \frac{1}{W} \sum i P^{i-1} \left(1 - P + P_u\right) = \frac{1 - P + P_u}{W(1-P)^2} \tag{43}$$

where $i = 0, 1, 2, \ldots$. Therefore, the average time required to successfully transmit the code is: $T_{r\_\text{unc\_sr}} = N_{r(\text{unc-sr})} T_t$. It can be concluded that the throughput efficiency of non-continuous SR-ARQ scheme is calculated by: $\eta(\text{unc\_sr\_th}) = T_{\text{tu}}/T_{r\_\text{unc\_sr}}$. Also, energy efficiency for this case is given by: $\eta(\text{unc\_sr\_en}) = E_0/\left(E \cdot N_{r(\text{unc-sr})}\right)$.

5. RESULTS

In this section, we compare the performance of the throughput and energy efficiency of ARQ schemes, respectively. Also, considering the limited functionalities of bacterial nodes, optimal frame lengths for minimizing channel throughput and energy efficiency are being analysed.

5.1. Parameter setup

Based on Section 4, the channel throughput and energy efficiency performance for SW-ARQ, GBN-ARQ and SR-ARQ protocols are discussed in terms of different channel parameters, particularly the information frame length. In addition, optimal frame lengths for various channel conditions are investigated. We assume equally probable "0"s and "1"s. The total data length is set to between 1 and 1500 bits, broken into frames for transmission. CRC check bits and a 3-bit sequence number are appended to each information frame. The primary focus is the optimization

Trans. Emerging Tel. Tech. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

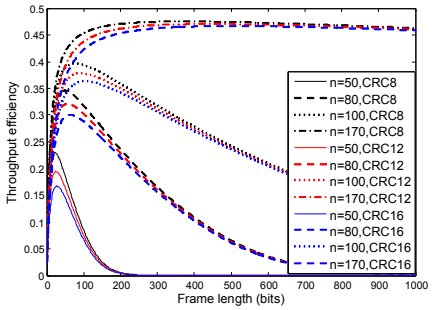

Figure 5. Throughput efficiency versus frame length for SW-ARQ.

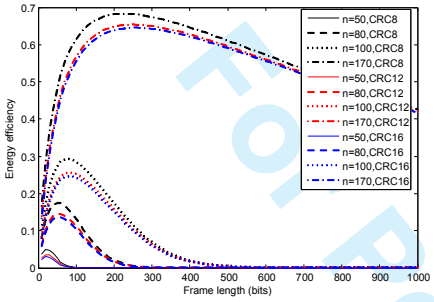

Figure 6. Energy efficiency versus frame length for SW-ARQ.

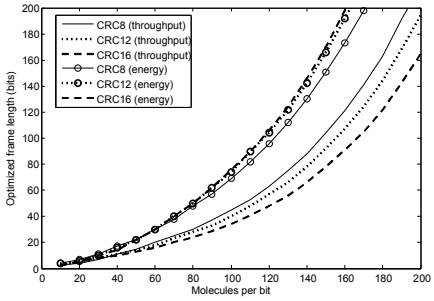

Figure 7. SW-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different CRC polynomials.

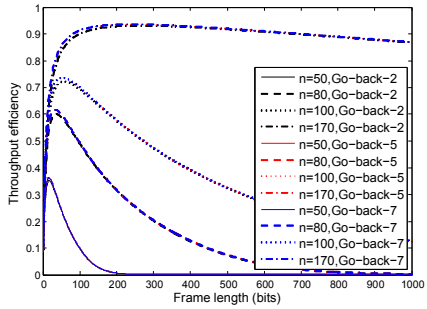

Figure 8. Throughput efficiency versus frame length for GBN-ARQ.

of the frame length. All information frames are also set to the same length and only one bit is allowed to transmit in one time slot. The transmission distance is set to be $4\mu m$ and there are 100 bacteria in the receiver node. The lengths of the promoters and proteins are given in Table I.

5.2. SW-ARQ

The throughput and energy efficiency for varying system frame lengths were simulated, with different CRC polynomials and different number of molecules per bit. The results are given in Fig. 5 and Fig. 6, showing the throughput and energy efficiency for various frame sizes, CRC polynomials and number of molecules emitted at the start of each time slot, respectively. They both indicate that when a certain CRC polynomial is used, more molecules per bit result in better throughput and energy performances. Also, with the same number of molecules per bit, CRC-8 outperforms CRC-12 and CRC-16 because of the complex logical operations with more bits appended. Both the throughput and energy efficiency are upper bounded, with maximum attainable values as low as 0.45 and 0.7, respectively, for the condition $n = 170$, CRC-8. In addition, it is clear that in each case, there exist a unique maximum value for either throughput or energy efficiency, with a corresponding optimal frame size.

Fig. 7 shows the optimal frame lengths for maximizing the throughput and energy efficiency of the SW-ARQ scheme. The optimal frame length grows with the number of molecules per bit. Also, when using a certain CRC, the optimal frame length required to reach the maximum energy efficiency is larger than that to reach the largest throughput efficiency.

5.3. GBN-ARQ

For a communication system with GBN-ARQ applied, the transmitter has a limit on the number of frames that can be outstanding. The appending of an x-bit sequence number to each information frame means that the sequence of frames carries sequence numbers with decimal representation $[0, (2^x - 1)]$. In general, the window size W for GBN-ARQ needs to be $(2^x - 1)$ or less to avoid ambiguities [36]. Thus, in this model, the window size could be chosen between 1 and 7. In addition, considering the fact that the operation is much more complex for substantially increased numbers of CRC check bits [5], CRC-8 is applied in the following investigations due to its relatively lower operation time and energy consumption.

Fig. 8 and Fig. 9 show the throughput and energy efficiency of the GBN-ARQ protocol for various values of frame size, window size and number of molecules emitted at the start of each time slot, respectively. Both figures indicate that the performances are almost the same for window sizes of 5 and 7. Also, both the throughput and energy efficiency are higher with a larger window size and more molecules per bit. Hence, under reliable

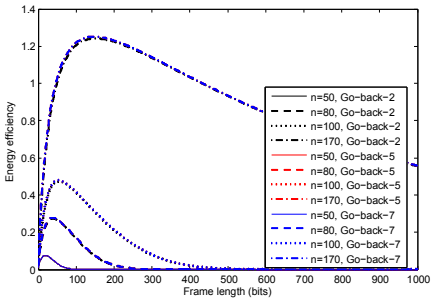

Figure 9. Energy efficiency versus frame length for GBN-ARQ.

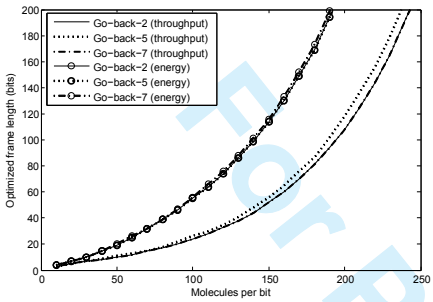

Figure 10. GBN-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.

channel conditions when the number of molecules per bit is large, one can operate at significantly higher packet lengths and still achieve near-optimal throughput and energy efficiency, while the margin for error is much smaller under harsh channel conditions. In a similar way to SW-ARQ there is in each case a certain value of frame size to maximize the throughput and energy efficiency.

Fig. 10 shows the optimal frame length of GBN-ARQ for maximizing the throughput and energy efficiency, respectively. The optimal frame length grows with the number of molecules per bit. Also, for fixed numbers of molecules per bit, the corresponding optimal frame length for different window sizes to achieve the maximum throughput is smaller than that to reach the maximum energy efficiency; and both values are slightly smaller than for SW-ARQ in Fig. 7.

5.4. SR-ARQ

Here there is also a limit on the maximum send window size, since the receiver needs a window size greater than one to store out-of-order positively acknowledged frames. Usually, the window sizes of the transmitter and the receiver are identical and when an x-bit sequence number is appended to each information frame, the window size W needs to be 2^{x-1} or less to avoid duplicate transmission [36]. Thus, in this model, the window size of the transmitter and the receiver should be equal and within the range 1 to 4 inclusive. CRC-8 will be applied.

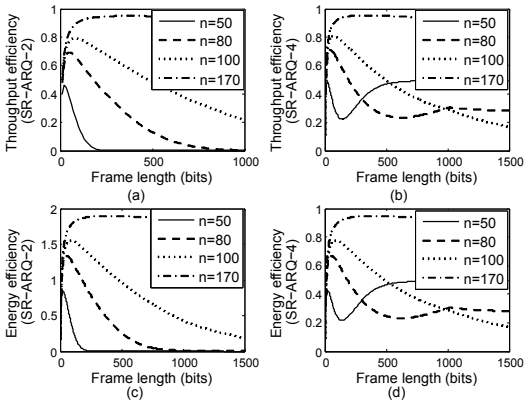

Figure 11. SR-ARQ performance: (a) throughput efficiency versus frame length for window size of 2; (b) throughput efficiency versus frame length for window size of 4; (c) energy efficiency versus frame length for window size of 2; (d) energy efficiency versus frame length for window size of 4.

Fig. 11 shows results obtained for the throughput and energy efficiency of SR-ARQ scheme using different frame lengths, numbers of molecules per bit and window sizes. It is noted that the performance of window size of 3 is almost the same as that of window size of 4, so is not plotted. For a window size of 2, the curve trends of both the throughput and energy efficiency are similar to that of GBN-ARQ, but the values are higher. However, for a window size of 4, a different behaviour is observed when the number of molecules per bit is relatively small. For 50 molecules, the initial trend is actually downwards before a steady-state value is attained as a result of the poor BER producing many retransmissions. For 80 molecules the increasing frame length eventually overwhelms the improvements from the SR-ARQ as the BER is still poor. This behaviour differs from that of SW-ARQ and GBN-ARQ, and disappears when using more molecules and hence achieving a better BER.

Fig. 12 shows the optimal frame lengths for maximizing the throughput and energy efficiency for the SR-ARQ scheme where it may be seen that the optimal frame length grows with the number of molecules per bit. Also, for the same number of molecules per bit, the optimal frame length obtained by maximizing throughput efficiency is smaller, compared with that obtained by maximizing energy efficiency.

6. CONCLUSIONS AND FUTURE WORK

In recent years, bacteria have been considered as one approach for molecular communication using QS. In this paper, a bacterial communication network model through a diffusive channel between transmitter and receiver bacteria populations is proposed. For the first time, the widely used ARQ protocols have been employed, specifically

Trans. Emerging Tel. Tech. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

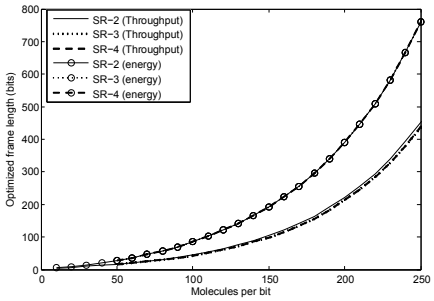

Figure 12. SR-ARQ optimal frame length versus molecules per bit to maximize the throughput and energy efficiency, for different window sizes.

utilising CRC coding, SW-ARQ, GBN-ARQ and SR-ARQ. Existing frame size optimization techniques are not applicable in the case of bacterial communications due to its limited capabilities. Both throughput and energy efficiency were chosen as the optimization metrics. The optimal fixed frame size was then determined for a given set of channel parameters by maximizing the throughput and energy efficiency. For each of these three ARQ schemes, the transmission distance and the bacteria population in the receiver node were set as fixed values. Given the increasing complexity and delay from high order CRC polynomial operation, CRC-8 was selected for GBN-ARQ and SR-ARQ. For each of the three ARQ schemes, both the throughput and energy efficiency are larger with a larger number of molecules per bit and a larger window size (for GBN-ARQ and SR-ARQ). In other words, increasing the window size used enhances performance but the gains saturate meaning that a value of 7 is found to be a highly satisfactory compromise for GBN-ARQ, and a value of 4 is the best for SR-ARQ. Also, it should be noted that for both the throughput and energy efficiency, the maximum value for GBN-ARQ and SR-ARQ is higher than that of SW-ARQ, when the frame length is a fixed value. In addition, for a certain number of molecules per bit used, to achieve the best throughput and energy efficiency, the required frame length for SR-ARQ scheme is slightly larger than SW-ARQ scheme, followed by GBN-ARQ scheme. The simulation results described show how traditional ARQ schemes perform in bacterial communication networks and how the parameters can be optimized to achieve a better channel performance. It must be stressed, however, that the simulations are conceptual and intended to lay the groundwork for ongoing and detailed study. We recognize that there are details to be filled in. These include a mechanism for a cluster of transmitter bacteria to release information molecules at an appropriate rate, progress in connecting biological logic gates based on transcription and translation [21] to create complex coding, windowing and sequencing operations. Nevertheless, we consider that the broad conclusions with respect to the performance of the established ARQ schemes over the bacterial diffusion channel to be valid.

In addition to the topics above, future work may include investigation of the implementation of ARQ schemes in bacterial communication for practical use and multiple-hop research, and take burst errors into consideration. We would then envisage that a possible application of ARQ protocols in bacterial communications would be to improve the sensitivity of bacterial biosensors and drug delivery systems.

REFERENCES

1. B. L. Bassler. 2003. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current Opinion in Microbiology* 1999; 2(6), pp. 582-587, DOI: 10.1016/S1369-5274(99)00025-9.
2. A. Einolghozati, M. Sardari, F. Fekri. Design and Analysis of Wireless Communication Systems Using Diffusion-Based Molecular Communication Among Bacteria. *IEEE Transactions on Wireless Communications* 2013; 12(12), pp. 6096-6105, DOI: 10.1109/TWC.2013.101813.121884.
3. W. B. Miller Jr. 2013. *The Microcosm Within: Evolution and Extinction in the Hologenome*. Universal-Publishers.
4. Y. Tanouchi, D. Tu, J. Kim, L. You. Noise reduction by diffusional dissipation in a minimal quorum sensing motif. *PLOS Computational Biology* 2008; 4(8), p. e1000167, DOI: 10.1371/journal.pcbi.1000167.
5. T. K. Moon. 2005. *Error correction coding: mathematical methods and algorithms*. Wiley-Blackwell.
6. C. Bai, M. S. Leeson, M. D. Higgins. Performance of SW-ARQ in bacterial quorum communications. *Nano Communication Networks* 2015; 6(1), pp. 3-14, DOI: 10.1016/j.nancom.2014.11.001.
7. M. Ş. Kuran, H. B. Yilmaz, T. Tugcu, B. Özerman. Energy model for communication via diffusion in nanonetworks. *Nano Communication Networks* 2010; 1(2), pp. 86-95, DOI: 10.1016/j.nancom.2010.07.002.
8. Y. Sankarasubramaniam, I. F. Akyildiz, S. W. McLaughlin. Energy efficiency based packet size optimization in wireless sensor networks. In: *Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on*, 2003, pp. 1-8.
9. M. U. Mahfuz, D. Makrakis, H. T. Mouftah. On the characterization of binary concentration-encoded molecular communication in nanonetworks. *Nano Communication Networks* 2010; 1(4), pp. 289-300, DOI: 10.1016/j.nancom.2011.01.001.
10. M. S. Kuran, H. B. Yilmaz, T. Tugcu, I. F. Akyildiz. Modulation techniques for communication via diffusion in nanonetworks. In: *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1-5, DOI: 10.1109/icc.2011.5962989.

11. A. Einolghozati, M. Sardari, F. Fekri. Capacity of diffusion-based molecular communication with ligand receptors. In: *Information Theory Workshop (ITW), 2011 IEEE*, 2011, pp. 85-89, DOI: 10.1109/ITW.2011.6089591.
12. A. Einolghozati, M. Sardari, A. Beirami, F. Fekri. Capacity of discrete molecular diffusion channels. In: *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2011, pp. 723-727, DOI: 10.1109/ISIT.2011.6034228.
13. Y. Lu, M. D. Higgins, M. S. Leeson. Comparison of Channel Coding Schemes for Molecular Communications Systems. *IEEE Transactions on Communications* 2015; 63(11), pp. 3991-4001, DOI: 10.1109/TCOMM.2015.2480752.
14. Z. P. Li, J. Zhang, T. C. Zhang. Concentration Aware Routing Protocol in Molecular Communication Nanonetworks. *Applied Mechanics and Materials* 2014; 556, pp. 5024-5027, DOI: 10.4028/www.scientific.net/AMM.556-562.5024.
15. I. Llatser, D. Demiray, A. Cabellos-Aparicio, D. T. Altilar, E. Alarcn. N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks. *Simulation Modelling Practice and Theory* 2014; 42, pp. 210222, DOI: 10.1016/j.simpat.2013.11.004.
16. T. Nakano, A. W. Eckford, T. Haraguchi. 2013. *Molecular Communication*. Cambridge University Press.
17. B. Atakan, O. Akan. An information theoretical approach for molecular communication. In: *2007 Bio-Inspired Models of Network, Information and Computing Systems (Bionetics)*, 2007, pp. 33-40, DOI: 10.1109/BIMNICS.2007.4610077.
18. M. Pierobon, I. F. Akyildiz. Diffusion-Based Noise Analysis for Molecular Communication in Nanonetworks. *IEEE Transactions on Signal Processing* 2011; 59(6), pp. 2532-2547, DOI: 10.1109/TSP.2011.2114656.
19. K. V. Srinivas, A. W. Eckford, R. S. Adve. Molecular Communication in Fluid Media: The Additive Inverse Gaussian Noise Channel. *IEEE Transactions on Information Theory* 2012; 58(7), pp. 4678-4692, DOI: 10.1109/TIT.2012.2193554.
20. A. Ghosh, H. S. Cho. Throughput and energy efficiency of a cooperative hybrid ARQ protocol for underwater acoustic sensor networks. *Sensors* 2013; 13(11), pp. 15385-15408, DOI: 10.3390/s131115385.
21. A. Tamsir, J. J. Tabor, C. A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 2011; 469(7329), pp. 212-215, DOI: 10.1038/nature09565.
22. M. S. Leeson, M. D. Higgins. Error correction coding for molecular communications. In: *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 6172-6176, DOI: 10.1109/ICC.2012.6364980.
23. A. Kuo, N. V. Blough, P. V. Dunlap. Multiple N-acyl-L-homoserine lactone autoinducers of luminescence in the marine symbiotic bacterium Vibrio fischeri. *Journal of bacteriology* 1994; 176(24), pp. 7558-7565.
24. F. Cousteau. 2011. *Ocean: the world's last wilderness revealed*. Dorling Kindersley Ltd.
25. N. G. Ravichandra. 2013. *Fundamentals of Plant Pathology*. Prentice-Hall of India Pvt.Ltd.
26. P. S. Stewart. Diffusion in biofilms. *Journal of Bacteriology* 2003; 185(5), pp. 1485-1491, DOI: 10.1128/JB.
27. H. B. Yilmaz, C. B. Chae. Arrival modelling for molecular communication via diffusion. *Electronics Letters* 2014; 50(23), pp. 1667-1669, DOI: 10.1049/el.2014.2943.
28. R. P. Shetty, D. Endy, T. F. Knight Jr.. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* 2008; 2(1), pp. 1-12, DOI: 10.1186/1754-1611-2-5.
29. Guopeng Wei, P. Bogdan, R. Marculescu. Efficient modeling and simulation of bacteria-based nanonetworks with BNSim. *IEEE Journal on Selected Areas in Communications* 2013; 31(12), pp. 868-878, DOI: 10.1109/JSAC.2013.SUP2.12130019.
30. H. Sahoo. Fluorescent labeling techniques in biomolecules: a flashback. *RSC Advances* 2012; 2(18), pp. 7017-7029, DOI: 10.1039/C2RA20389H.
31. R. Sarpeshkar. Analog synthetic biology. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2014; 372(2012), p. 20130110, DOI: 10.1098/rsta.2013.0110.
32. P. Lu, C. Vogel, R. Wang, X. Yao, E. M. Marcotte. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology* 2007; 25(1), pp. 117-124, DOI: 10.1038/nbt1270.
33. C. T. Bhunia. 2006. *Information Technology Network and Internet*. New Age International.
34. S. P. Pramod, A. Rajagopal, S. K. Akshay. FPGA Implementation of Single Bit Error Correction using CRC. *International Journal of Computer Applications* 2012; 52(10), pp. 15-19, DOI: 10.5120/8238-1471.
35. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler. GenBank. *Nucleic Acids Research* 2005; 33(suppl 1), pp. D33-D38, DOI: 10.1093/nar/gki063.
36. I. Widjaja, A. Leon-Garcia. 2004. *Communication Networks Fundamental Concepts and Key Architectures*. Mc GrawHill.