THE UNIVERSITY OF WARWICK

warwickpublicationswrap

highlight your research

http://wrap.warwick.ac.uk

Analysis of ARQ Protocols for Bacterial Quorum Communications

Chenyao Bai[a], Mark. S. Leeson[a], Matthew D. Higgins[a,*]

[a]*School of Engineering, University of Warwick, Coventry, CV4 7AL, UK*

Abstract

Quorum sensing (QS) is used to describe the communication between bacterial cells, whereby a coordinated population response is controlled through the synthesis, accumulation and subsequent sensing of specific diffusible chemical signals called autoinducers, enabling a cluster of bacteria to regulate gene expression and behavior collectively and synchronously, and assess their own population. As a promising method of molecular communication, bacterial populations can be programmed as bio-transceivers to establish information transmission using molecules. In this work, to investigate the key features for molecular communication, a bacterial QS system is introduced, which contains two clusters of bacteria, specifically *Vibrio fischeri*, as the transmitter node and receiver node, and the diffusive channel. The transmitted information is represented by the concentration of autoinducers with on-off keying (OOK) modulation. In addition, to achieve better reliability, transmission efficiency and channel throughput performance, different Automatic Repeat reQuest (ARQ) protocols are taken into consideration. This configuration is investigated via simulation and the consequent results discussed. The performance of the system is evaluated in terms of transmission time, efficiency, bit error rate (BER) and channel throughput. Results show that Selective-Repeat (SR-ARQ) performs better than Go-Back-N (GBN-ARQ), while the performance of Stop-N-Wait (SW-ARQ) varies for different channel conditions, which is quite different from the performance of ARQ schemes in traditional networking areas.

Keywords: Bacterial Communication, Quorum Sensing, ARQ Protocols, Molecular Communication, Nano Communications

1. Introduction

Instead of language, bacteria communicate with each other through a process called quorum sensing (QS), using signalling molecules, called autoinducers, which are released into their immediate environment [1]. Also, they are able to measure the concentration of the signalling molecules within a population, which means that the accumulation of autoinducers enables a single cell to sense the number of bacteria (cell density) [1]. In particular, if the concentration of molecular signals in the medium exceeds a certain threshold, an individual bacterium in a population releases more molecules into the environment [2], which will in turn increase the density of signalling molecules over time producing a positive feedback process. Thus, the concentration of external autoinducers is correlated with the bacterial cell population density. The output of the QS mechanism can be in various forms and one example is the production of luminescence [2]. The phenomenon of QS was first found in the luminescence marine bacteria *V. fischeri* and *V. harveyi*, which were found to be luminescent when the local bacterial population is high

[1]. The discovery of this phenomenon changed the general perception of many individual, simple organisms existing in the natural world. Bacteria can alter the target gene expressions by changing the signalling molecular concentration, which enables coordinated behaviours, such as competence, virulence, biofilm development, sporulation, light production and attacking suitable hosts, and rapid adaptation to environmental changes, based on the local density of the bacterial population [3]. Different bacterial species use different classes of signalling molecules, which have minor variations such as different length side chains and side-chain decorations, to communicate [4]. In some situations, a single bacterial species can have more than one QS system using more than one signalling molecule, which means that a specific bacterial species may respond to different classes of autoinducers in different ways [5]. In addition, it has also been demonstrated that interspecies communication via QS exists, which is referred to as cross talk [5]. In this paper, a diffusive bacterial communication network between two populations of bacteria has been modelled and analysed, using the well-studied bacteria species, *V. fischeri*, which is most famous for its aforementioned bioluminescence and has been used to study toxicity of aquatic environments. The transmitted information is represented by the concentration of signalling molecules, which will then be encoded into data frames, with the release of molecules represented by a binary '1'

and no release represented by a binary '0'.

Noise, which may result from gene expressions at the intracellular level and the diffusion of autoinducers, presents a major challenge for the robust function and performance of natural and engineered quorum sensing networks [6]. For the bacterial communication system proposed in this paper, the effect of inter symbol interference (ISI) caused by molecular diffusion is considered as a major source of impairment. This may result in data packet corruption and out-of-sequence delivery, making it necessary to apply error detection rules and Automatic Repeat reQuest (ARQ) mechanisms for reliable transmission. In this work, ARQ protocols are used rather than error correction schemes because error detection requires much simpler decoding operations than does error correction [7]. The term ARQ was first introduced by Chang [8], after which three widely used ARQ schemes, including Stop-N-Wait (SW-ARQ), Go-Back-N (GBN-ARQ) and Selective-Repeat (SR-ARQ), have been presented and developed [9]. The performance of the SW-ARQ protocol has been investigated in our previous work in [10]. In this paper, we will compare the performance of the three kinds of ARQ schemes for bacterial communication.

The contributions of this paper are as follows. To our knowledge, although GBN-ARQ and SR-ARQ are well-known concepts in networking and coding theory, this is their first use in a bacterial quorum communication system to enhance reliability. Specifically, this research maps existing protocol concepts to biological QS processes and shows how different protocols and parameters can be fitted to different modes of bacterial communication. Also, transmission delay, transmission efficiency, BER and channel throughput of the system are evaluated when different parameters are applied. Results show that the parameter settings are quite different compared with that of the ARQ techniques used in traditional communication fields. 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 transmitter and receiver models are introduced, followed by the establishment of the channel in Section 3. In Section 4, the ARQ protocols and error detection codes are used to enhance the system performance, followed by the results and discussions in Section 5. Finally, Section 6 gives the conclusions 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 delay in the communication process can be fairly large due to biological actions such as transcription and translation. In other words, an individual bacterium is very primitive and unreliable and hence incapable of transferring information by itself. Hence, to achieve reliability of the communication system, here, the communication model between two populations of bacteria which is proposed in [2] is taken into consideration. In this model, a cluster of bacteria trapped in a chamber, which is collectively capable for reliable transmission and reception of molecular information, is considered as a node [2]. The model consists of the transmitter node, the receiver node and the communication channel. Both the transmitter and receiver nodes are considered to be genetically modified bacteria, which can sense specific types of signals and respond accordingly [2]. Molecular communication between two nodes can be made up of three procedures. The transmitter node produces the signalling molecules by adequate stimulation, then these molecules propagate through the medium undergoing Brownian motion and finally the receiver node senses the concentration of the local signalling molecules and takes appropriate actions. The communication system is assumed to be in a theoretically infinite space. The transmitted information is encoded via the concentration of signalling molecules, i.e. the embedding of the information is by alteration of the concentration of the molecules and its transmission relies on diffusion. The output of the receiver node, in the form of luminescence, is measured in steady-state to estimate the concentration of signalling molecules at the vicinity of the node, and hence decode the transmitted information [2].

In this proposed model, both the transmitter and receiver nodes contain m instances of the bacterium *V. fischeri*, which is the most commonly studied QS system in gram-negative bacteria. These bacteria are motile, gram-negative rods, $0.8 - 1.3\mu m$ in diameter and $1.8 - 2.4\mu m$ in length [11]. Bioluminescence in this bacterium is controlled by the QS system, which is composed by two regulatory genes, *luxI* and *luxR*, coding for proteins *LuxI* and *LuxR*, respectively. At low cell densities when only a small number of bacteria are present, the signal (3-oxo-C6-HSL, an N-acyl homoserine lactone or AHL), which is synthesised by the protein *LuxI*, 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 bacteria population increases, the concentration of AHLs around the node will grow. If the concentration of the signal reaches a critical threshold, it is able to interact with the *LuxR* protein, which acts as the ligand receptor for AHL. The *LuxR*/AHL 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*/AHL complex also triggers the AHL (via *LuxI*) to be produced at a higher level. Thus the AHL is said to auto-induce its own synthesis. The process of bioluminescence is illustrated in Fig. 1.

In this work, the number of bacteria in each node is assumed to be constant. The bacteria inside the node can grow, divide and die to maintain the constant population through the process of gene regulation [12]. It is assumed that each bacterium can sense and produce two
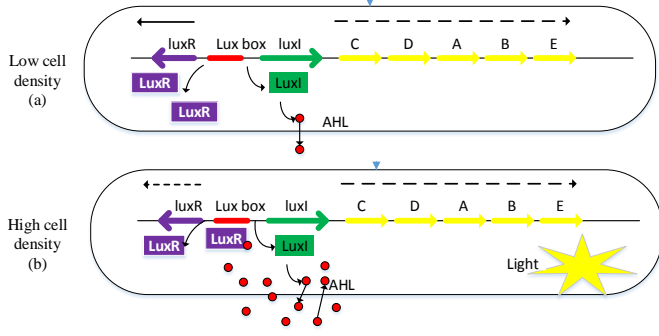
Figure 1: Luminescence production in *V. fischeri*. (a) The system is not active and there is basal transcription of *luxR*, *luxI* and *lux-CDABE*. With the bacterial population increase, the autoinducers accumulate until the concentration exceeds a threshold which allows the binding between AHL and *LuxR*. (b) *LuxR* is bound to AHL, activating the transcription of the luminescence genes.
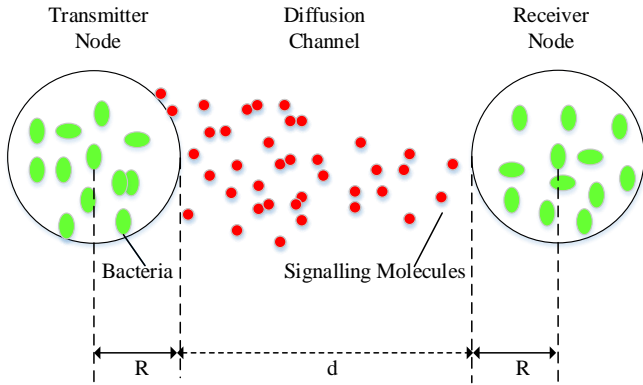

Figure 2: Bacterial communication scheme.

different types of AHL molecules, denoted as type I and type II [13]. 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. Type II molecules are produced with the stimulation of type I molecules, while luminescence is produced when stimulated by type II molecules. However, depending on the different functionalities, as a transmitter or receiver, only one type of receptors is activated. For the bacteria in the transmitter node, only type I receptors are enabled, while for bacteria in the receiver node, only type II receptors are enabled, a process which can be controlled by proper enzymes. 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 type I molecules surrounding them [2]. At the receiver, each bacterium senses the concentration of type II molecules through type II receptors, followed by the production of luminescence by bacteria, which is used to decode the input signal concentration.

3. Channel Model

In the channel model which is 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 motion within the fluid medium. The transmitter encodes its information into the concentration of signals. The emitted signalling molecules then diffuse through the channel to the receiver which is at a distance d from the transmitter. At the receiver, the transmitted information is decoded according to the concentration of type II molecules detected by type II receptors.

Due to the process of QS, the bacteria cells in the receiver can synchronously respond to the molecules as they arrive. In addition, at the receiver, although the molecules can 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 in the node are not activated. Thus 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, which is known as crossover probability. To effectively represent the transmitted symbols, the propagation time is divided into time slots, also called symbol durations, which have the equal length. Only one symbol propagates in single time slot, which is denoted by t_s. The information is encoded by concentration with binary representation. Specifically, if the number of information molecules arriving at the receiver at a certain time slot exceeds a threshold τ, the symbol is represented by '1'. Otherwise, it will be interpreted as '0'. Moreover, with OOK modulation employed, the release of molecules in a time slot represents a binary one while their absence for the same duration represents a binary zero. However, errors may be caused by Inter Symbol Interference (ISI), which is a form of distortion of a signal in which one symbol interferes with subsequent symbols. It is an unavoidable consequence of both wired and wireless communication systems and is known to degrade the error performance of communication systems, particularly when the system is stochastic [14]. It can be noticed that the received signals tended to spread to adjacent symbols and smeared into each other when a sequence of symbols are transmitted [15]. The ISI effect is related to the properties of the medium used, the distance of the symbol propagation and the selection of the threshold value. In the diffusive communication system here, some information molecules may arrive at the receiver after the current time slot according to the diffusion dynamics, which will lead to the incorrectly decoding of the received symbol of the next time slot.

In essence, the information molecules propagate through

3

the fluid medium undergoing Brownian motion which is a random procedure and a probabilistic behaviour, which means that there is a probability that the molecule will hit the receiver at a time slot. At a certain time duration t, the capture probability, which is defined as the probability of success at a specific time period, in a three dimensional environment is given by [16]:

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

where d is the distance away of the information molecule from the 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 $4.9 \times 10^{-6} \text{cm}^2\text{s}^{-1}$, which is settled as a conservative value for AHL in water at $25°C$ [17].

In this work, the communication channel is a binary one, where each molecule arrives at the receiver or does not. Due to ISI, the number of molecules received in a time slot which is 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 sent at the start of each symbol duration. Also, the transmitted information includes k bits. Due to ISI, the ith $(i \in [2, k])$ symbol can be affected by the symbols from $(i-1)$ previous time slots. For the ith 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 [18]:

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

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

The number of left over molecules N_p belonging to all of the previous time slots can be given as:

$$
\begin{aligned}
N_p &\sim \sum_{j=1}^{i-1} (\mathcal{B}(n, P_{j+1}) - \mathcal{B}(n, P_j)) \\
&\sim \sum_{j=1}^{i-1} \left[\mathcal{N}(nP_{j+1}, nP_{j+1}(1 - P_{j+1})) \right. \\
&\quad \left. - \mathcal{N}(nP_j, nP_j(1 - P_j)) \right] \\
&\sim \mathcal{N}(\mu_{p_i}, \sigma_{p_i}^2) \quad (3)
\end{aligned}
$$

where P_j is the capture probability with receiver radius R, transmission distance d and time duration of jt_s. Particularly, for the case of $j = 1$, P_j represents the capture probability in one time slot, which is identical to P_1 in equation (2). $\mu_{p_i}, \sigma_{p_i}^2$ are the expectation 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_{hit} received in the ith symbol duration is the summation of N_c and N_p.

It can be derived from equation (3) that μ_{p_i} and $\sigma_{p_i}^2$ can be described by:

$$\mu_{p_i} = n(P_i - P_1) \quad (4)$$

$$\sigma_{p_i}^2 = nP_1(1 - P_1) + 2n \sum_{j=2}^{i-1} P_j(1 - P_j) + nP_i(1 - P_i) \quad (5)$$

For the ith symbol $(i > 0)$, there exist $(i - 1)$ previous time slots and thus 2^i different cases, i.e. $[0, (2^i - 1)]$, represented as binary numbers. 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 third symbol and the other two bits represent the previous symbols. If the transmitted symbol in the current time slot is 1, the number of received molecules need to be larger than τ for correct decoding. The total number of molecules that have arrived at the ith time slot is represented by:

$$N_{\text{hit}} = N_p + N_c \sim \mathcal{N}(\mu_{\text{hit}(i)}, \sigma_{\text{hit}(i)}^2) \quad (6)$$

where $\mu_{\text{hit}(i)}$ and $\sigma_{\text{hit}(i)}^2$ are the expectation and variance of the distribution of N_{hit} and can be represented by:

$$\mu_{\text{hit}(i)} = \mu_{p_i} + nP_1 = nP_i \quad (7)$$

$$\sigma_{\text{hit}(i)}^2 = 2n \sum_{j=1}^{i-1} P_j(1 - P_j) + nP_i(1 - P_i) \quad (8)$$

Thus, the error probability for this case is:

$$
\begin{aligned}
p_{e(1)} &= P(N_{\text{hit}} < \tau) \\
&= 1 - Q \left(\frac{\tau - nP_i}{\sqrt{n \left(2 \sum_{j=1}^{i-1} P_j(1 - P_j) + P_i(1 - P_i) \right)}} \right)
\end{aligned}
$$
$$\quad (9)$$

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 ith symbol duration is:

$$N_{\text{hit}} = N_p \sim \mathcal{N}(\mu_{p_i}, \sigma_{p_i}^2) \quad (10)$$

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

$$
\begin{aligned}
p_{e(0)} &= P(N_{\text{hit}} > \tau) \\
&= Q \left(\frac{\tau - n(P_i - P_1)}{\sqrt{n \left(P_1(1 - P_1) + 2 \sum_{j=2}^{i-1} P_j(1 - P_j) + P_i(1 - P_i) \right)}} \right)
\end{aligned}
$$
$$\quad (11)$$

4

Due to the fact that the capture probability is a function of R, d, t_s, 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_s can be calculated from $\left(p_{e(0)} + p_{e(1)}\right)/2$.

The bit error rate (BER) is considered as a key parameter which is often employed to assess the performance of communication systems that transmit information from one position to another. Here, the BER refers to the probability of one bit error when information symbols are transmitted in the diffusion based communication channel. According to [18], 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 the unsatisfied increasing average hitting time. In this model, the symbol duration t_s is chosen as the time before 60% of the molecules arrive at the receiver. In this paper, the BER is the value when the error probability reaches the minimum for $\tau \in [1, n]$. The corresponding values of threshold τ is applied in the following research. The results and discussions will be presented in Section 5. Theoretically speaking, the application of ARQ will decrease the BER.

4. ARQ Protocols

A time varying channel with a relatively high BER level causes frequent packet corruptions and out-of-sequence delivery, which needs error check codes and ARQ mechanisms for effective error detection and recovery, respectively [19]. ARQ, which is usually located at the data link layer, is an error control technique to ensure that a data stream is delivered accurately to the user despite errors in transmission. ARQ forms the basis of peer-to-peer protocols that provide for the reliable transfer of information. The basic elements of ARQ protocols include the information frames that transfer the information, the acknowledgement frames (ACKs), the negative acknowledgement (NAK) and the time out mechanism. The ACK/NAK signifies the receipt of a given frame. The time out mechanism is required to maintain the flow of frames.

Generally, a population of bacteria is able to perform distributed information processing, where each bacterium in the colony is capable of information storing, processing, and interpreting information [20]. Also, the advance of synthetic biology, particularly the foundation of the BioBricks database [21], enables many types of capabilities based on genetically engineered bacteria, including timing, counting, clocking, logic gates, pattern detection and intercellular communication [22]. Moreover, because ARQ mechanisms and error control operations can be implemented through circuits [7], it is possible to implement both techniques in bacterial communication systems as well. In addition, genetically engineered bio entities can harvest energy from biological systems and require no external energy sources, which is therefore expected to be energy efficient [23]. Such capabilities greatly improve the bacterial cooperation level during the communication process, making it possible to meet the requirement of high processing and memory intensive ARQ protocols. Generally speaking, the SW-ARQ protocol has been investigated in our previous work [10] 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 [24]. In this paper, the three basic ARQ schemes will be analysed separately and then compared with each other.

In SW-ARQ technique, after transmitting a frame, the transmitter waits for an ACK/NAK from the receiver before transmitting the next frame. On receiving a frame, the receiver checks errors using error detection techniques. If no error is found, the receiver sends an ACK to the sender through the feedback path. If errors are detected, a NAK is transmitted back to the transmitter. If ACK is received by the transmitter, the transmitter transmits the next frame. Otherwise, it will retransmit the previously sent frame. However, if the ACK/NAK is lost, the transmitter waits for a time out period until next transmission. In short, the retransmission continues until a frame is received correctly and it is positively acknowledged, or the number of retransmissions reaches a certain threshold. In addition, in order to avoid ambiguities, sequence number can be added to the frames. The protocol continues in this manner until all the frames are transmitted successfully.

In the GBN-ARQ technique, the transmitter continuously transmits a block of W (W is often known as 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. 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. The receiver 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 dis-
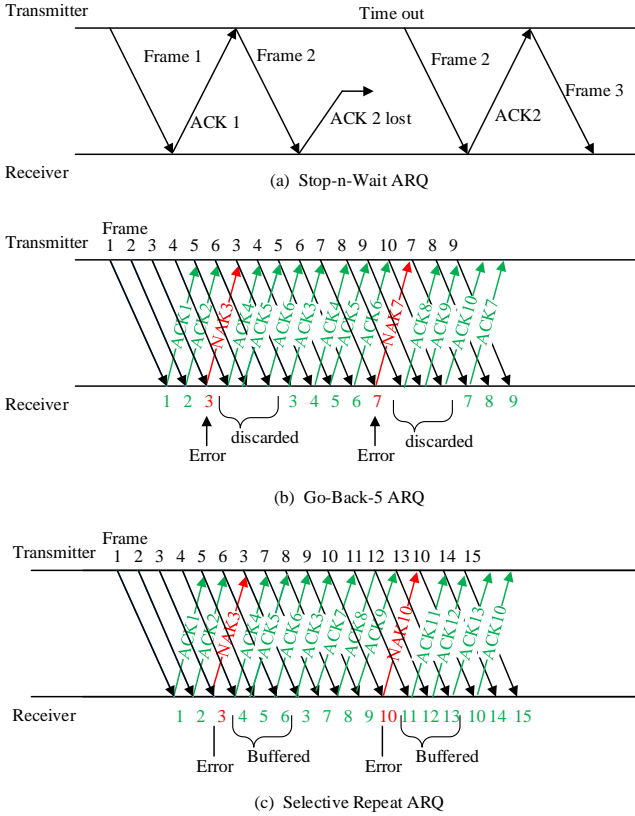
Figure 3: Examples of three ARQ schemes: (a) SW-ARQ. (b) GBN-ARQ with window size of 5. (c) SR-ARQ.

card 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 involves with minimum two propagation delays. Compared with the best situation in SW-ARQ where one frame is involved with two-propagation delays, GBN-ARQ provides throughput improvement in theory.

SR-ARQ operates in a similar way to GBN-ARQ but only retransmits the frame for which a NAK is received. The receiver accepts out-of-order frames and buffers them. This requires more buffer at the receiver to store correct out of sequence frames. Fig. 3 shows examples of the process of transmitting a series of frames with ACKs and NAKs of the three ARQ schemes.

In the channel model here, the transmitter node generates a sequence of information frames for transmission. Each information frame contains a header which contains sequence numbers that are essential for in-sequence delivery, information bits and error detection bits. Here, the error detection is effected via cyclic redundancy check (CRC) codes, which are represented by polynomials, having good error sensing performance, fast encoding and decoding capabilities, and applicability to varying message lengths [25]; these are appended to the frame to determine if error occurs during transmission. 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. In Section 2, it may be observed that for the transmitter node, only type I receptors are activated, while for the receiver node, only type II receptors are enabled. Also, it has been explained in Section 3 that the transmitted information is represented by the concentration of type II molecules. Thus the information frame is made up of certain number of information bits, which are encoded by type II molecules. Similarly, the acknowledgement frames are composed of fixed number of bits encoded by type I molecules. 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 type I receptors makes it impossible to generate extra type II molecules, providing a more accurate decoding of the transmitted information at the receiver. Therefore, both transmission and reception 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 the type I molecules used for sending messages and that for acknowledgement messages [26].

5. Results

In this section the simulation results for the ARQ protocols are discussed in terms of transmission delay, transmission efficiency, BER and channel throughput. Due to our previous works on SW-ARQ [10], we will mainly concentrate on the performance of GBN-ARQ and SR-ARQ.

5.1. Parameter setup

Through the simulation of the information transmission, in this section the channel performance results for the GBN-ARQ and SR-ARQ protocols are discussed in terms of transmission delay, BER, transmission efficiency and channel throughput. Here the total amount of information to be transmitted is set to be 2000 bits, which will be broken into frames for transmission. As has been explained in Section 3 that the transmitter node implements a 1 bit memory, which is either binary '0' or '1', based on the concentration of emitted autoinducers. Similarly, specific biological units could be designed to act as molecule storage, which may simply either be the molecular communication environment where the molecules diffuse and wait for the transceiver and encoder nanomachines to intake them, or a physical component, for instance, a vesicle or liposome embedded in the transceivers to store molecules [27]. In this way, the logical bit sequences of CRC check bits and a 3-bit sequence number can be stored in separate

molecule storages and then appended to each information frame in a controlled manner. The parameters that can be varied in the systems are the transmission distance d, the number of bacteria in the receiver node m, the frame length N, the CRC polynomial and the window size W. Considering the fact there is little difference for the performances of different CRC polynomials [10] and the operation is much more complex for substantially increased numbers of CRC check bits [28], CRC-8, with polynomial representation $x^8 + x^2 + x + 1$, is applied here due to its relatively lower operation time and energy consumption. All information frames are also supposed to be of the same length. In this channel, only one bit is allowed to transmit in one time slot. Thus, the basic delay t_0, in the absence of errors, from the time a frame is transmitted until the corresponding acknowledgement is received is calculated from:

$$t_0 = 2t_{\text{prop}} + 2t_{\text{proc}} + Wt_f + t_{\text{ack}} = 2t_{\text{prop}} + 2t_{\text{proc}} + W\frac{n_c}{R_b} + \frac{n_a}{R_b} \tag{12}$$

In this equation, for a transmitted frame, the first bit that is input into the channel appears at the output of the channel after a propagation time t_{prop}, which can be calculated according to Section 3; the end of the frame is received at the receiver after t_f additional seconds. For GBN-ARQ and SR-ARQ, W frames are allowed to be transmitted continuously without waiting for ACK/NAK for individual frame. The receiver sends an acknowledgement frame that will require t_{ack} seconds of transmission time. After an additional propagation delay, the acknowledgement frame is received at the transmitter. CRC codes can be implemented by logic gate operations [28], which makes it possible to operate CRC coding using genetic circuits. Here, t_{proc} is the processing time for CRC implementation and it is ignored for simplicity in this paper. The term n_c is the total number of bits in the information frame including the information bits, sequence numbers and error checking bits, W is the transmitter window size and n_a is the number of bits for the acknowledgement frame which is of the same value as the number of sequence bits. R_b is the bit rate of the transmission channel. In our proposed model, there is one bit transmitting in each time slot. The boundary condition to choose R_b is represented by:

$$(n_c \cdot t_{\text{prop}} + t_{\text{proc}}) \cdot R_b \leq n_c \tag{13}$$

The boundary condition above indicates that in the time period of transmitting one codeword 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 (13). In addition, the timeout period is set to be exactly equal to the sum of round trip propagation delay and the CRC processing time.

The effective information transmission rate of the system in the absence of errors is given by:

$$R_{\text{eff}}^0 = \frac{\text{bits delivered}}{\text{total time}} = \frac{n_c - n_0}{t_0} \tag{14}$$

where n_0 is the number of overhead bits in a frame.

The effective information transmission rate of the protocol when errors occur is given by:

$$R_{\text{eff}} = \frac{\text{information bits delivered}}{\text{average total time per frame}} = \frac{n_c - n_0}{t_{\text{ave}}} \tag{15}$$

where t_{ave} is the average time to transmit a frame. The transmission efficiency is obtained by $R_{\text{eff}}/R_{\text{eff}}^0$.

In addition, channel throughput is the ratio between the amount of information bits which has been transmitted (2000 bits in this paper), and the total time for transmission.

5.2. Go-Back-N ARQ

For a communication system with GBN-ARQ applied, the transmitter has a limit on the number of frames that can be outstanding. With an x-bit sequence number appended to each information frame, the sequence of frames carry the 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 [29]. Thus, in this model, the window size could be chosen between 1 and 7. Also, it should be noted that when the window size is set to 1, so that only one frame is allowed to be outstanding, this reduces to the SW-ARQ scheme.

The transmission distance is first taken into consideration which is shown in Fig. 4. Here the frame length is set to be 100 bits, the window size is 7 and CRC-8 is employed (these parameters are investigated subsequently). Compared with the performance when no CRC and ARQ are used at all, the channel which uses ARQ performs better with a lower BER. Moreover, the transmission delay, BER, efficiency and channel throughput all improve with smaller transmission distances because over a larger distance, the increased BER results in more transmission times per frame for average. Thus lower transmission efficiency and channel throughput should be observed with larger distances, which perfectly fits the results in Fig. 4(b) and Fig. 4(d). In what follows, a representative transmission distance is taken as $4\mu m$.

Fig. 5 displays the results for different window sizes for a frame length of 100 bits with the number of bacteria in the receiver equal to 100. Here, the efficiency and BER do not vary significantly so are not plotted. Fig. 5(a) shows that when the number of molecules per bit is low (approximately 10-100), the transmission delay is larger using a larger window size. This is because when the number of molecules per bit is at lower end of what is practical, the channel crossover probability is relatively high, leading to greater transmission times for each block of transmission frames. Also, all the frames starting from the first negatively acknowledged frame in the sending window need to be retransmitted. Thus, larger window size
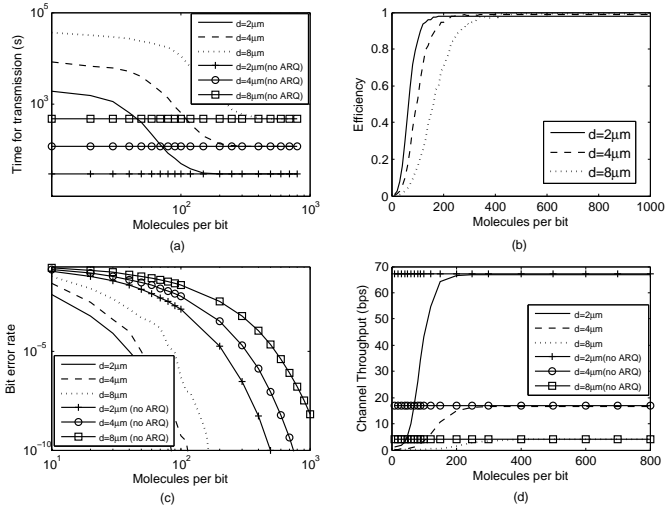
Figure 4: Channel performance of GBN-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.

Figure 6: Channel performance of GBN-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.
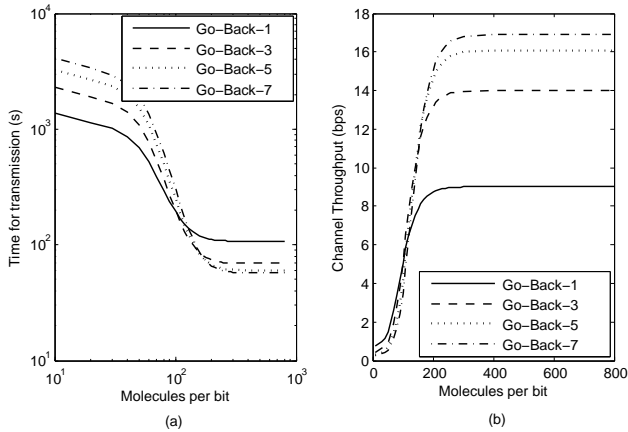

Figure 5: Channel performance of GBN-ARQ for different window sizes: (a) transmission delay; (b) channel throughput.

causes many more retransmissions, resulting in increased transmission time and lower channel throughput, which perfectly matches the result in Fig. 5(b). However, as the number of molecules per bit increases, the error probability of the channel is small enough to sharply decrease the number of retransmissions. So in this situation, the channel performs better with a larger window size. Also, it should be recalled that the condition of Go-Back-1 reduces to SW-ARQ. Hence, compared with SW-ARQ, GBN-ARQ is inefficient when the number of molecules per bit is small. From the results above, a window size of 7 is employed for the subsequent investigations.

In Fig. 6, different numbers of bacteria in each node are considered maintaining the frame length of 100 bits and window size of 7. Results show that less time will be consumed when information bits are transmitted through the channel if there is a larger population of bacteria in
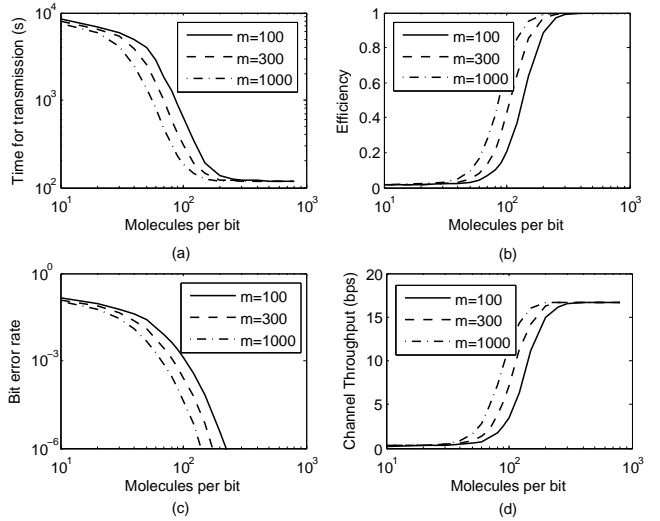
the receiver node. Also, the transmission time converge for all values of m when the number of molecules per bit is relatively high. The reason is that the effect of the bacterial population on the receiver radius R is only in a small degree, resulting in little influence on the capture probability according to equation (1). Also, with the increase in the number of molecules per bit, the error probability of the channel is quite small, which will lead to a sharp decrease of the number of retransmissions. Hence, for relatively larger number of molecules per bit, the differences among the retransmission times per frame for different values of m are quite slight, leading to the convergence for transmission time. Similar conditions happen for the channel throughput performance. Moreover, fewer packet corruptions occur during the transmission process and the transmission efficiency and throughput are higher for larger populations of bacteria.

Fig. 7 shows the system performance when the frame length is different with other parameters taking on the values stated previously. Results show that the error probability is larger when there is a larger frame length. This is because the probability of an error frame is calculated by

$$p_f = 1 - (1 - p)^N \tag{16}$$

where p is the probability of one bit error and N is the frame length. It is clear that when frame length is larger, there is a larger probability that transmission error occurs in the frame. However, there are two crossover points for the transmission time, indicating that when the number of molecules per bit is in the range approximately between 40 and 200, larger frame length performs better. In addition, the channel throughput with smaller frame length is larger when the number of molecules emitted at the start of each time slot is fewer than approximately 200. These
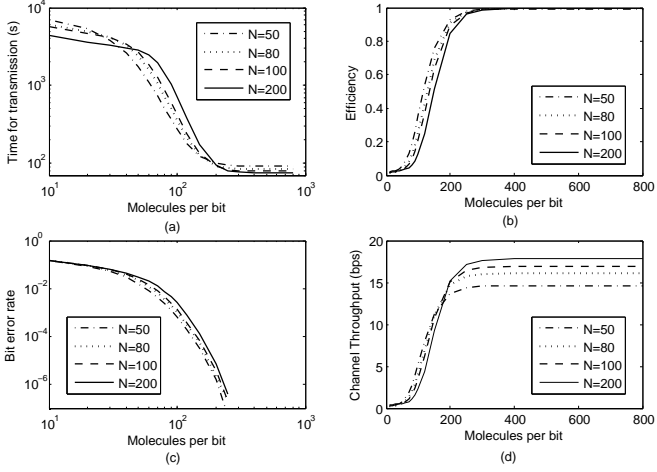
8

Figure 7: Channel performance of GBN-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

phenomena suggest that for a certain value of the number of molecules per bit n, there should exist an optimized frame size to achieve the best channel throughput, which will be further investigated in future work.

5.3. Selective-Repeat ARQ

In a similar way to the GBN-ARQ scheme, there is also a limit on the maximum send window size for SR-ARQ scheme. Also, due to the fact that the receiver needs to store the out-of-order positively acknowledged frame it receives, the receiver's window is required to be larger than 1. Usually, the window sizes of the transmitter and the receiver are identical. In general, when an x-bit sequence number is appended to each information frame, the window size W for SR-ARQ needs to be 2^{x-1} or less to avoid duplicate transmission [29]. Thus, in this model, the window size of the transmitter and the receiver should be equal and within the range 1 to 4 inclusive.

The transmission distance is first taken into consideration which is shown in Fig. 8. Here the frame length is set to be 100 bits, the window size is set to be the maximum value 4 and CRC-8 is employed. The transmission delay per frame, BER, efficiency and channel throughput all improve with smaller transmission distances because over a larger distance, the increased BER results in more transmission times per frame. In the following investigations, the transmission distance is thus taken as $4\mu m$ to make better comparison with the performance of GBN-ARQ.

Fig. 9 displays the results for different window sizes ranging from 1 to 4, where the frame length is 100 bits and bacteria population is 100. The size of the transmitter and receiver windows are the same, denoted by W. Here, the transmission efficiency and BER performances are almost the same so are not plotted. Fig. 9(a) shows that when the number of molecules per bit is low (approximately 10-100), the transmission delay is slightly larger with a larger
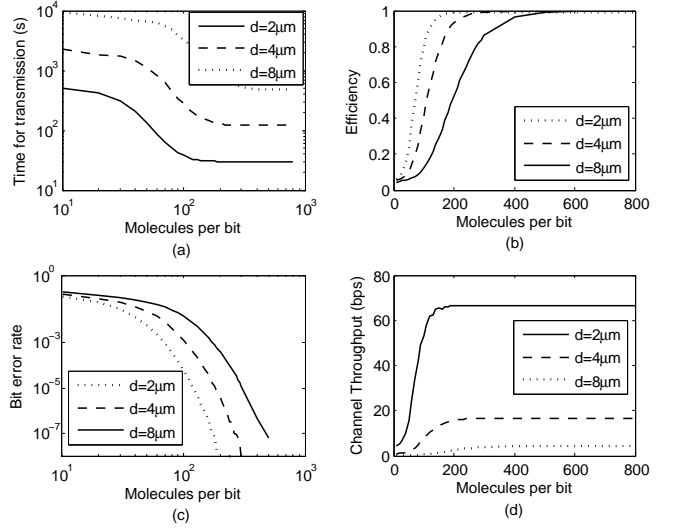

Figure 8: Channel performance of SR-ARQ for different transmission distances: (a) Transmission delay; (b) Efficiency; (c) BER; (d) channel throughput.
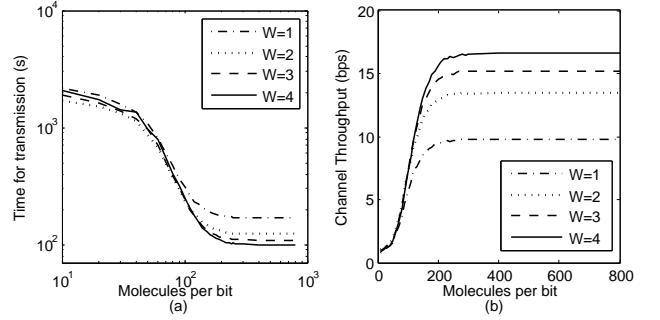

Figure 9: Channel performance of SR-ARQ for different window sizes: (a) transmission delay; (b) channel throughput.

window size. Fig. 9(b) shows that the channel throughput is significantly increased by the use of larger window sizes, without any crossover points, which is quite different from that of GBN-ARQ in Fig. 5(b) because only the negatively acknowledged frames need to be retransmitted in SR-ARQ. The condition of $W = 1$ reduces to SW-ARQ. It is clear that when the number of molecules per bit is relatively small (< 100), there is not much difference in both the transmission time and channel throughput for SR-ARQ ($W > 1$) and SW-ARQ ($W = 1$), while SR-ARQ requires a larger storage capability. Hence, in a similar fashion as GBN-ARQ, SR-ARQ is inefficient when the number of molecules per bit is small. From the results above, a window size of 4 is employed for the subsequent investigations for better throughput.

In Fig. 10, the number of bacteria in each node is taken into consideration, ranging from 100 to 1000, maintaining the frame length of 100 bits and window size of 4. Results show that a smaller amount of time will be consumed when information bits are transmitted through the
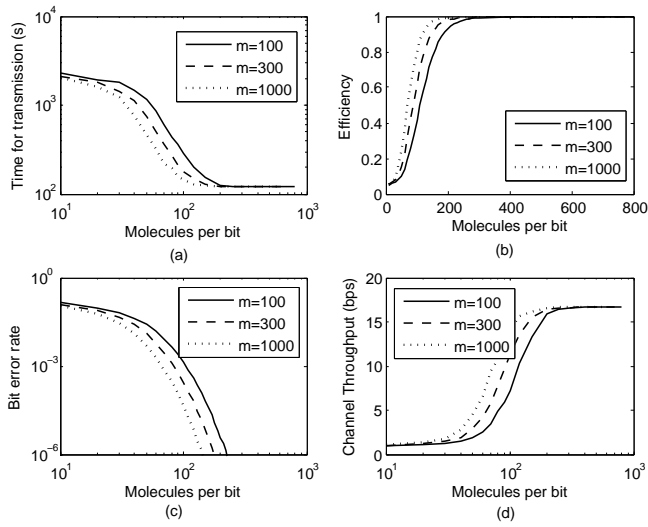
Figure 10: Channel performance of SR-ARQ for different number of bacteria: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.
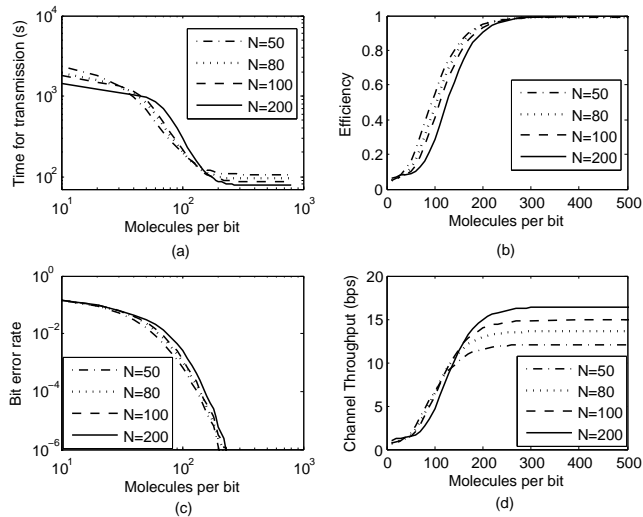

Figure 11: Channel performance of SR-ARQ for different frame lengths: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

channel if there is a larger population of bacteria in the receiver node. Moreover, fewer packet corruptions occur during the transmission process and the transmission efficiency and throughput are higher in this situation.

Fig. 11 shows the system performance for different frame lengths, with other parameters taking on the values stated previously. Results show that the error is larger when there is a larger frame length. This is because that when frame length is larger, there is a greater probability that transmission errors occur in the frame, leading to increased transmission times per frame and hence lower transmission efficiency. However, there are two crossover points for the transmission time, indicating that when the
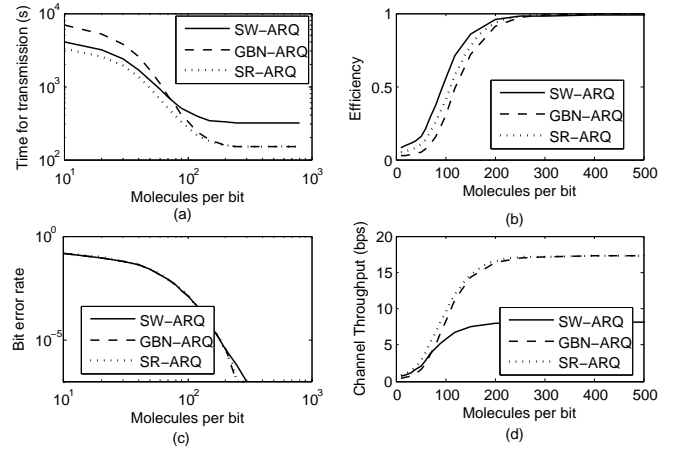

Figure 12: Comparisons of Channel performance for different ARQ schemes: (a) transmission delay; (b) efficiency; (c) BER; (d) channel throughput.

number of molecules per bit is between 40 and 160, larger frame lengths perform better. In addition, the throughput efficiency with smaller frame lengths is greater when the number of molecules emitted at the start of each time slot is in the approximate range zero to 140. In a similar way to the performance of GBN-ARQ, these phenomena show that there should be an optimized frame size to achieve the best channel throughput for each value of the number of molecules per bit and this will bear future investigation.

Finally, to make comparisons between the three ARQ schemes, the parameters are set to be the same, including transmission distance $d = 4\mu m$, the number of bacteria in the receiver $m = 100$, frame length is 100 bits, and window size is 4; the results are shown in Fig. 12.

Fig. 12 shows that when the number of molecules per bit is in the range between 10 to approximately 100, SW-ARQ and SR-ARQ take less time and have a slightly larger channel throughput. Also, SR-ARQ has a slightly better performance than SW-ARQ. According to the non-ARQ BER performances shown in Fig. 4(c), when the number of molecules per bit is in a smaller range, the crossover over probability of the channel is higher, which will lead to greater transmission times for each block of transmission frames. For SW-ARQ and SR-ARQ, only the frame which is negatively acknowledged or timed out is being retransmitted. While for the GBN-ARQ, all the frames starting from the negatively acknowledged frame in the sending window will be retransmitted, leading to longer transmission times and lower channel throughput. However, when the number of molecules per bit is higher (> 120), the crossover probability of the channel will be low (less than $\approx 10^{-5}$), leading to a rapid decrease in the transmission times for each block of transmission frames. Thus, in this situation, the transmission time consumption and channel throughput of GBN-ARQ and SR-ARQ are approximately the same, especially when the number of molecules per bit is larger than 400. Also, there is little difference in error

performance for the three types of ARQ schemes. More-over, GBN-ARQ has the worst transmission efficiency. It is surprising that using ARQ schemes in bacterial communications produces performance that differs markedly from that in traditional areas. In conventional computer networks, the performance of SR-ARQ, GBN-ARQ and SW-ARQ is in a decreasing order. However, in bacterial quorum communications, due to the characteristics including long propagation and processing delays and that only one bit is allowed in the communication channel, SR-ARQ performs better than GBN-ARQ, while the performance of SW-ARQ varies according to the range of the number of emitted molecules at the start of each time slot.

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 diffusion channel between transmitter and receiver bacteria populations is proposed. For the first time, the widely used ARQ protocols have been employed, specifically utilising CRC coding, SW-ARQ, GBN-ARQ and SR-ARQ. For each of these three ARQ protocols, BER and delay performance worsen as the transmission distance increases, leading to the choice of $4\mu m$ as a feasible and representative value. Given the increasing complexity and delay from high order CRC polynomial operation, CRC-8 is selected. 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 better for SR-ARQ. Also, the error performance is better with a larger bacterial population in the receiver node. Moreover, there are two crossover points for the transmission time, indicating that when the number of molecules per bit is in a range of approximately 40 to 200 for GBN-ARQ and 40 to 160 for SR-ARQ, larger frame lengths perform better, with reduced time consumption. In addition, the throughput efficiency with smaller frame lengths is greater when the number of molecules emitted at the start of each time slot is in a range of up to 200 for GBN-ARQ and up to 140 for SR-ARQ. Finally, with identical parameter settings, the performance of the three ARQ schemes has been investigated, indicating that SR-ARQ performs better than GBN-ARQ, while the performance of SW-ARQ varies according to the range of the number of emitted molecules at the start of each time slot; this is quite different from the performance of ARQ schemes in traditional networking areas. The simulation results described show how traditional ARQ schemes perform in bacterial communication networks and how the parameters may be determined 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 biological logic gates based on transcription and translation [30] 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 includes the optimization of the frame length to achieve the best channel throughput and the investigation of the energy consumption. 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.
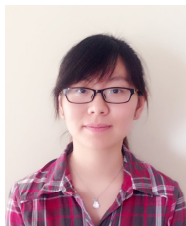
References

[1] B. L. Bassler. How bacteria talk to each other: regulation of gene expression by quorum sensing. Current Opinion in Microbiology 1999;2(6):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):6096-6105. doi:10.1109/TWC.2013.101813.121884.

[3] H. B. Kaplan, E. P. Greenberg. Diffusion of autoinducer is involved in regulation of the vibrio fischeri luminescence system. Journal of Bacteriology 1985;163(3):1210–1214.

[4] S. T. Rutherford, B. L. Bassler. Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harbor Perspectives in Medicine 2012;2(11). doi:10.1101/cshperspect.a012427.

[5] W. B. Miller Jr. The Microcosm Within: Evolution and Extinction in the Hologenome. Universal-Publishers; 2013.

[6] 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.

[7] G. Serban, C. Anton, L. Ionescu, I. Tutanescu, A. Mazare. Implementation of a 64-bit hybrid SR-ARQ algorithm on FPGA. In: 2011 International Conference on Applied Electronics (AE), IEEE 2011, p. 1-4.

[8] S. Chang. Theory of information feedback systems. IRE Transactions on Information Theory 1956;2(3):29-40. doi:10.1109/TIT.1956.1056807.

[9] J. Li, Y. Q. Zhao. Resequencing Analysis of Stop-and-Wait ARQ for Parallel Multichannel Communications. IEEE/ACM Transactions on Networking 2009;17(3):817-830. doi:10.1109/TNET.2009.2020820.

[10] C. Bai, M. S. Leeson, M. D. Higgins. Performance of SW-ARQ in bacterial quorum communications. Nano Communication Networks 2015;6(1):3-14. doi:10.1016/j.nancom.2014.11.001.

[11] F. Cousteau. Ocean: the world's last wilderness revealed. Dorling Kindersley Ltd; 2011.

[12] I. F. Akyildiz, F. Fekri, R. Sivakumar, C. R. Forest, B. K. Hammer. Monaco: fundamentals of molecular nano-communication networks. Wireless Communications, IEEE 2012;19(5):12-18. doi:10.1109/MWC.2012.6339467.

[13] J. G. Cao, E. A. Meighen. Purification and structural identification of an autoinducer for the luminescence system of Vibrio harveyi. Journal of Biological Chemistry 1989;264(36):21670-21676.

[14] M. S. Leeson. Performance analysis of direct detection spectrally sliced receivers using Fabry-Perot filters. Journal of Lightwave Technology 2000;18(1):1325. doi:10.1109/50.818902.

[15] L. Koschel, A. Kortke, F. Undi. Multi-gigabit communication with hybrid pulse shaping and fractional oversampling. In: 2012 IEEE 23rd International Symposium on Personal Indoor

and Mobile Radio Communications (PIMRC) 2012, p.2437-2442. doi:10.1109/PIMRC.2012.6362765.

[16] M. S. Leeson, M. D. Higgins. Error correction coding for molecular communications. In: 2012 IEEE International Conference on Communications (ICC) 2012, p.6172-6176. doi: 10.1109/ICC.2012.6364980.

[17] P. S. Stewart. Diffusion in biofilms. Journal of Bacteriology 2003;185(5):1485–1491. doi:10.1128/JB.

[18] M. Ş. Kuran, H. B. Yilmaz, T. Tugcu, B. Özerman. Energy model for communication via diffusion in nanonetworks. Nano Communication Networks 2010;1(2):86–95. doi: 10.1016/j.nancom.2010.07.002.

[19] E. Ustunel, I. Hokelek, O. Ileri. A cross-layer goodput enhancement considering CRC coding and ARQ dynamics. In: 2012 IEEE Symposium on Computers and Communications (ISCC) 2012, p. 000023–000028. doi:10.1109/ISCC.2012.6249263.

[20] E. B. Jacob, Y. Shapira, A. I. Tauber. Seeking the foundations of cognition in bacteria: From Schr'odinger's negative entropy to latent information. Physica A: Statistical Mechanics and its Applications 2008;359:495–524. doi: 10.1016/j.physa.2005.05.096.

[21] R. P. Shetty, D. Endy, T. F. Knight Jr.. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering 2008;2(1):1–12. doi:10.1186/1754-1611-2-5.

[22] Guopeng Wei, P. Bogdan, R. Marculescu. Efficient modeling and simulation of bacteria-based nanonetworks with BN-Sim. IEEE Journal on Selected Areas in Communications 2013;31(12):868–878. doi:10.1109/JSAC.2013.SUP2.12130019.

[23] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos, J. Shuai. Molecular communication and networking: opportunities and challenges. IEEE Transactions on NanoBioscience 2012;11(2):135–148. doi:10.1109/TNB.2012.2191570.

[24] H. Jianhua, K. R. Subramanian, D. Donghua, W. Wei. Novel methods for the performance analysis of adaptive hybrid selective repeat ARQ. Computer Communications 2000;23(16):1548–1557. doi:10.1016/S0140-3664(00)00215-2.

[25] T. S. Baicheva. Determination of the best CRC codes with up to 10-bit redundancy. IEEE Transactions on Communications 2008;56(8):1214–1220. doi:10.1109/TCOMM.2008.070033.

[26] H. Sahoo. Fluorescent labeling techniques in biomolecules: a flashback. RSC Advances 2012;2(18):7017–7029. doi: 10.1039/C2RA20389H.

[27] T. Nakano, T. Suda, Y. Okaie, M. Moore, A. Vasilakos. Molecular Communication Among Biological Nanomachines: A Layered Architecture and Research Issues. IEEE Transactions on NanoBioscience 2014;13(3):169–197. doi: 10.1109/TCOMM.2008.070033.

[28] Todd K. Moon. Error correction coding: mathematical methods and algorithms. Wiley-Blackwell; 2005.

[29] I. Widjaja, A. Leon-Garcia. Communication Networks Fundamental Concepts and Key Architectures. Mc GrawHill; 2004.

[30] A. Tamsir, J. J. Tabor, C. A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 2011;469(7329):212–215. doi: 10.1038/nature09565.

protocols.

Chenyao Bai received her B.S. degree in Optical Information Science and Technology from Shanxi University, China, in 2011 and her M.Sc. degree in Electronic Systems with Sensor Technology from University of Warwick, UK, in 2012. She is currently pursuing her Ph.D. study in School of Engineering at University of Warwick since 2012. Her research interests include molecular communication, quorum sensing, nanonetworks, mathematical modelling, error correction and detection codes and communication

Mark S. Leeson received the degrees of BSc and BEng with First Class Honors in Electrical and Electronic Engineering from the University of Nottingham, UK, in 1986. He then obtained a PhD in Engineering from the University of Cambridge, UK, in 1990. From 1990 to 1992 he worked as a Network Analyst for National Westminster Bank in London. After holding academic posts in London and Manchester, in 2000 he joined the School of Engineering at Warwick, where he is now a Reader. His major research interests are coding and modulation, nanoscale communications and evolutionary optimization. To date, Dr. Leeson has over 220 publications and has supervised twelve successful research students. He is a Senior Member of the IEEE, a fellow of both the UK Institute of Physics and of the UK Higher Education Academy.

Matthew D. Higgins received his MEng in Electronic and Communications Engineering and PhD in Engineering from the School of Engineering at the University of Warwick in 2005 and 2009 respectively. Remaining at the University of Warwick, he then progressed through several Research Fellow positions with leading defence and telecommunications companies before undertaking two years as a Senior Teaching Fellow. As of July 2012 Dr. Higgins now holds the position of Assistant Professor. His major research interests are the modelling of optical propagation characteristics in underwater, indoor and atmospheric conditions. Dr. Higgins is a Member of the IEEE.